Complexity Dichotomies for Counting Problems

Jin-Yi Cai
Computer Sciences Department
University of Wisconsin
Madison, WI 53706
Email: jyc@cs.wisc.edu

Xi Chen
Department of Computer Science
Columbia University
New York, NY 10027
Email: xichen@cs.columbia.edu

January 24, 2016
6.1 A Dichotomy for a Single Ternary Signature ........................................ 179
6.2 Reductions Between Holant and \#CSP ............................................... 186
6.3 Holant\textsuperscript{c} Problems ............................................................... 190
6.4 A Dichotomy for \#CSP\textsuperscript{d} ......................................................... 199
6.5 Eulerian Orientation ........................................................................... 206

7 Holant Dichotomy for Symmetric Constraints ......................................... 215
  7.1 Vanishing Signatures ........................................................................... 215
    7.1.1 Characterization by Recurrence Relations .................................... 217
    7.1.2 Characterization of Vanishing Signature Sets ............................... 221
    7.1.3 Characterization via Holographic Transformation .......................... 226
  7.2 Theorem Statement and Proof of Tractability ..................................... 227
  7.3 A Sample of Problems ........................................................................ 228
  7.4 Outline of Hardness Proof for Theorem 7.19 ...................................... 232
  7.5 Dichotomy for One Signature of Arity 4 .......................................... 234
  7.6 Vanishing Signatures Revisited .......................................................... 248
    7.6.1 Orientation with a Pair from $V^+$ and $V^-$ .............................. 248
    7.6.2 Combining Vanishing Signatures with Others .............................. 253
  7.7 $A$-transformable and $P$-transformable Signatures ........................... 259
    7.7.1 Characterization of $A$- and $P$-transformable Signatures .......... 259
    7.7.2 Dichotomies when $A$- or $P$-transformable Signatures Appear .... 265
  7.8 Proof of Theorem 7.19 ....................................................................... 270
  7.9 Decidability of the Dichotomy .............................................................. 275

8 Planar \#CSP for Symmetric Constraints ................................................ 279
  8.1 Introduction ......................................................................................... 279
  8.2 Unary Interpolation Revisited ............................................................. 282
  8.3 Planar Pairing ..................................................................................... 285
  8.4 Domain Pairing ................................................................................... 287
  8.5 No-Mixing of Tractable Signatures .................................................... 290
  8.6 Pinning for Planar Graphs ................................................................. 296
8.7 Planar CSP Dichotomy
Chapter 1

Counting Problems

This book is a study of the computational complexity of counting problems, especially those problems which can be expressed as a sum-of-product computation. Our aim is to give a systematic presentation to a body of work, mostly from the past ten years, that gives some comprehensive classifications to these sum-of-product computations from the perspective of computational complexity. All these sum-of-product problems are computable within the level of the complexity class \#P. The classification theorems are stated in the following form, known as a dichotomy: For a class of problems expressible within a framework, every problem in the class is either computable in polynomial time, or it is \#P-hard to compute, i.e., it is as hard as any other problem in the class \#P.

1.1 Counting Problems and Models of Computation

We assume the readers have some basic and preliminary knowledge about computational complexity such as P and NP, for example at a level provided by an undergraduate course on the subject of the Theory of Computing. This knowledge is not crucial however, since we will not use many existing technical results but rather present the framework and develop our own proof techniques gradually. Any reader wishing to be more thoroughly acquainted with the full scope of computational complexity theory can consult a standard textbook, e.g. [?, ?, ?]. On the other hand, a reader with no prior exposure to complexity theory but with a strong mathematical background, for example at a level provided by a solid undergraduate mathematics education, should be able to follow all proofs in the book, provided he or she is willing to accept a handful of results without proof. These results can be found elsewhere, and the particular proofs of which do not impact an understanding of the material in this book. Such a reader, however, would benefit from a wider exposure to complexity theory in order to gain some insight as to why certain problems are investigated and certain questions are asked.

Briefly, in computational complexity theory a problem consists of an infinite set of problem instances, rather than one particular instance. For example, the problem DETERMINANT
is to compute the determinant of an arbitrarily given matrix, not one particular matrix. Solving a problem is to say that there is an algorithm that solves all problem instances. The formal notion of an algorithm is a Turing machine, an idealized computer that carries out step-by-step operations according to a finitary program, valid for all problem instances. The efficiency of the algorithm is measured in terms of the maximum number of steps $T(n)$ the Turing machine may take for all problem instances of size $n$. E.g., for integer matrices, the problem instance of DETERMINANT is an integer square matrix where the size is the sum of the bit length of all the matrix entries. Formally the complexity classes P and NP are defined for decision problems, those problems for which the answer to each instance is either Yes or No. For example, the problem VERTEX COVER is the following decision problem: Given a problem instance consisting of a graph $G = (V, E)$ and an integer parameter $k$, whether there is a subset of vertices $S \subseteq V$ such that $|S| \leq k$ and every edge in $E$ is incident to at least one vertex from $S$. A decision problem $\Pi$ is in the complexity class P if there is an algorithm solving the problem, i.e., gives a correct Yes or No answer to every problem instance, such that $T(n)$ is bounded above by a fixed degree polynomial. The class NP consists of all decision problems $\Sigma$ for which there exists some problem $\Pi$ in P, and some polynomial $p(\cdot)$, such that for all problem instances $x$ of $\Sigma$, $x$ is a Yes instance of $\Sigma$ iff there exists some $y$ with size $|y| \leq p(|x|)$ and the pair $\langle x, y \rangle$ is a Yes instance of $\Pi$. Intuitively, $y$ is a certificate of size polynomially bounded in $x$, such that it can be verified in polynomial time that $y$ certifies that $x$ is a Yes instance of $\Sigma$.

There are historical reasons and also internal logic why the definitions of P and NP are formulated in terms of decision problems. A typical problem, even though it may appear more naturally as a search problem or a numerical problem, can usually be restated as a decision problem such that solving the decision problem is equivalent to solving the search or numerical problem within a polynomial factor in efficiency. E.g., if there is an algorithm that solves the decision problem VERTEX COVER, then by binary search on the value $k$ one can find the minimum size $k_0$ of any vertex cover of $G$. Furthermore, by considering a sequence of at most $O(k_0 n)$ graphs obtained by removing some vertices and their incident edges from $G$ of $n$ vertices, and repeatedly invoking the decision algorithm, one can compute a minimum size vertex cover of $G$. For the DETERMINANT problem one can formulate a decision problem as whether the determinant of an integer matrix is greater than a threshold value. It is easy to compute an a priori bound for a given determinant, then a binary search will find the exact determinant value in polynomial time from the decision algorithm. Such reductions show that there is a theoretical equivalence between a decision problem and its search or numerical version; however in reality a polynomial time algorithm usually directly computes the output value, without going through the decision version.

In this book we mostly study counting problems. For every NP problem $\Sigma$ one can formulate a corresponding counting problem using the problem $\Pi$ in P which defines $\Sigma$ in terms of certificate. For any problem instance $x$ of $\Sigma$, we define the function $f(x)$ as the number of certificates $y$ such that the pair $\langle x, y \rangle$ is a Yes instance of $\Pi$. The class of all such functions is denoted as $\#P$. This complexity class, defined by Valiant [?] in the study of the complexity of the permanent function, will play a central role in this book. Natural counting problems
corresponding to decision problems at the level of NP that have a nonnegative integer solution can all be formulated as a problem in \#P. In order to include more problems at this level we also consider problems whose solutions are not necessarily nonnegative integers, such as those from statistical physics. To capture their computational complexity, we consider the closure of \#P under polynomial time Turing reductions, namely, those problems solvable by a polynomial time algorithm given access to a hypothetical algorithm to some problem in \#P. Typical problems in \#P include counting the number of satisfying assignments to a Boolean formula, or counting the number of vertex covers in a graph. Weighted versions of these problems as well as sum-of-product computations such as partition functions from statistical physics can be easily formulated as problems polynomial time reducible to \#P.

The formal Turing machine model is naturally suited to the study of computation over discrete structures, such as integers or graphs. However in this book it is more natural to consider computation over the real or complex numbers. Doing so causes a technical issue of how one may represent exactly the individual real or complex numbers, and how to account for the complexity of various operations on these numbers. This is a question on the model of computation. One can, for example, take the computational model on real numbers by Blum-Shub-Smale [?]. This is more intuitive; however in a strict logical sense this is not equivalent to the classical Turing machine model, and the results we obtain would not have the same logical meaning. So, formally we still consider the classical Turing machine model, and restrict the objects of computation to algebraic (complex) numbers. Thus, technically, every number \( \alpha \) is specified by a finite irreducible polynomial \( f(X) \in \mathbb{Z}[X] \) with integer coefficients, and a small disk of rational radius containing a unique root \( \alpha \) of \( f(X) \). We will discuss more on this issue of models of computation in Section 1.5. But this is not a central issue for the type of sum-of-product computations we treat in this book; basically the theory can be developed in any reasonable model of computation where sum and product can be efficiently computed.

**Some basic notations.** We denote by \( \mathbb{N} \) the set of natural numbers \( \{0, 1, 2, \ldots\} \), by \( \mathbb{Z} \) the set of integers \( \{\ldots, -2, -1, 0, 1, 2, \ldots\} \), and \( \mathbb{Z}_+ \) the set of positive integers \( \{1, 2, 3, \ldots\} \). We denote by \( \mathbb{Q} \), \( \mathbb{R} \) and \( \mathbb{C} \) the sets of rational numbers, (algebraic) real and complex numbers respectively. We denote \( i = \sqrt{-1} \). All graphs are finite and undirected, unless stated otherwise. Graphs may or may not have loops and parallel edges that should be clear from the context. If \( \alpha \) and \( \beta \) are finite bit strings, then \( \alpha_i \) denotes its \( i \)-th bit, and \( \alpha \oplus \beta \) denotes its bit-wise XOR string. For binary string \( \alpha \), its Hamming weight is denoted by \( \text{wt}(\alpha) \), i.e., the number of 1’s. We use \( \leq^P_T \) or simply \( \leq_T \) to denote polynomial time Turing reducibility, and \( \equiv^P_T \) or simply \( \equiv_T \) to denote polynomial time Turing equivalence.

### 1.2 Three Classes of Counting Problems

We now formally define three frameworks for the types of sum-of-product computations in this book. Fix a finite domain \([q] = \{1, 2, \ldots, q\}\), for a positive integer \( q \). If \( q = 2 \) it is
called the Boolean domain. We consider any set of functions \( F \) on domain \([q]\), where each \( f \in F \) maps from \([q]^m\) to a commutative semiring \( \mathbb{R} \), for some \( m \geq 0 \), called the arity of \( f \). If \( m = 0 \), then \( f \) is a constant. If \( m = 1, 2 \) or \( 3 \), then \( f \) is called a unary, binary or ternary function respectively. We will almost always take \( \mathbb{R} \) to be the (algebraic) complex numbers, but we still denote it as \( \mathbb{C} \). Functions in \( F \) are also called signatures or local constraint functions. A signature \( f \) is symmetric if its value is invariant under permutation of its variables. For the Boolean domain this means that the value of \( f \) depends only on the Hamming weight of its input.

1.2.1 Spin Systems or Graph Homomorphism Problems

A spin system on a graph \( G = (V, E) \) is the following model and it comes from statistical physics. Let \([q] = \{1, \ldots, q\}\) be a finite domain, where \( q \geq 1 \) is an integer. We consider all vertex assignments \( \sigma : V \to [q] \). There is an edge function \( f : [q]^2 \to \mathbb{C} \). For each assignment \( \sigma \) we have an evaluation \( \prod_{(u,v) \in E} f(\sigma(u), \sigma(v)) \), a product over every edge \((u, v) \in E\). Then we define the partition function

\[
Z_f(G) = \sum_{\sigma : V \to [q]} \prod_{(u,v) \in E} f(\sigma(u), \sigma(v)).
\] (1.1)

The partition function represents the total energy or a normalizing factor of a system as one sums over every possible configuration of the particles.

The value \( f(\sigma(u), \sigma(v)) \) is the local contribution, the product \( \prod_{(u,v) \in E} f(\sigma(u), \sigma(v)) \) is the weight of the assignment \( \sigma \), and the partition function is the sum of weights over all assignments. If \( G \) is an undirected graph as is typically the case, we require \( f \) to be a symmetric function, \( f(i,j) = f(j,i) \), for all \( i, j \in [q] \).

If \( q = 2 \) this is called a 2-spin system, and for general \( q \) it is called a \( q \)-spin system.

Well-known examples of a 2-spin system include the Ising model, where \( f(0,0) = f(1,1) = a \) and \( f(0,1) = f(1,0) = b \) for some two constants \( a \) and \( b \). Sometimes there is also a vertex weight function, represented by a unary function normalized to \( u(0) = 1 \) and \( u(1) = \lambda \). Then the partition function of the Ising model is

\[
Z_{a,b,\lambda}(G) = \sum_{\sigma : V \to [2]} q^{|\{(u,v) \in E : \sigma(u) = \sigma(v)\}|} \lambda^{|\{(u,v) \in E : \sigma(u) \neq \sigma(v)\}|} \chi_{\{|v \in V : \sigma(v) = 1\}|}.
\] (1.2)

**Exercise**: Fix any constant \( J \) one can define the Hamiltonian function \( H(\sigma) \) for any assignment \( \sigma : V \to \{-1, +1\} \),

\[
H(\sigma) = -J \sum_{(u,v) \in E} \sigma(u)\sigma(v).
\]

(It is called ferromagnetic if \( J > 0 \) and antiferromagnetic if \( J < 0 \).) Then the partition function of the Ising model can be defined by \( Z(G) = \sum_{\sigma : V \to \{-1, +1\}} e^{-\beta H(\sigma)} \), where \( \beta \geq 0 \).
is called the inverse temperature. Show that this function \( Z(G) \) can be realized by the Sum-of-Product expression \( Z_{a,b,\lambda}(G) \).

A generalization of the Ising model is called the Potts model where for \( q \geq 2 \) and \( i, j \in [q] \), \( f(i, i) = a \) and \( f(i, j) = f(j, i) = b \) for \( i \neq j \). One can normalize \( b = 1 \) and write \( a \) as \( 1 + \gamma \) for a parameter \( \gamma \). In this parameterization the partition function of the Potts model is

\[
Z_{\text{Potts}}(G; q; \gamma) = \sum_{\sigma: V \rightarrow [q]} \prod_{(u,v) \in E} (1 + \gamma \delta(\sigma(u), \sigma(v))),
\]

where \( \delta(i, j) = 1 \) if \( i = j \), and 0 otherwise.

The partition function of the Potts model can be linked to the Tutte polynomial \( T_G(x, y) \) by setting \( \gamma = y - 1 \) and \( q = (x - 1)(y - 1) \). Indeed, one way to define the Tutte polynomial in terms of \( q \) and \( \gamma \) is that \((x - 1)^{\kappa(V,F)}(y - 1)^{|V|}T_G(x, y)\) is equal to

\[
Z_{\text{Tutte}}(G; q, \gamma) = \sum_{F \subseteq E} q^{\kappa(V,F)}\gamma^{|F|},
\]

where \( \kappa(V,F) \) is the number of connected components in the spanning subgraph \((V,F)\).

While the Tutte polynomial is defined for any \( q \), if we restrict to a positive integer \( q \), then

\[
Z_{\text{Tutte}}(G; q, \gamma) = Z_{\text{Potts}}(G; q, \gamma).
\]

To prove this equality, we consider expanding the product \( \prod_{(u,v) \in E} (1 + \gamma \delta(\sigma(u), \sigma(v))) \) in (1.3) as a sum indexed by \( F \subseteq E \), which collects a factor \( \gamma^{|F|} \), namely

\[
\prod_{(u,v) \in E} (1 + \gamma \delta(\sigma(u), \sigma(v))) = \sum_{F \subseteq E} \gamma^{|F|} \prod_{(u,v) \in F} \delta(\sigma(u), \sigma(v)).
\]

The product term indexed by \( F \) is 1 iff \( \sigma(u) = \sigma(v) \) for all \( u \) and \( v \) in a same connected component of the subgraph \((V,F)\), and 0 otherwise. In (1.3) the sum \( \sum_{\sigma: V \rightarrow [q]} \) has exactly \( q^{\kappa(V,F)} \) such terms.

Spin Systems are also called graph homomorphisms, and they come from a different source. Given two graphs \( G \) and \( H \), a graph homomorphism from \( G \) to \( H \) is a map \( \sigma \) from the vertex set of \( G \) to the vertex set of \( H \) such that for every edge \((u,v)\) in \( G \), the image is also an edge in \( H \). In general multigraphs are allowed, thus for example an edge in \( G \) can be mapped to a loop in \( H \). The counting problem is to compute the number of graph homomorphisms from \( G \) to \( H \).

We will fix an \( H \), and consider the computational problem where \( G \) is the input. A special case is when \( H = K_q \), the complete graph on \( q \) vertices, without self loops. In this case, a graph homomorphism from \( G \) to \( H \) is a valid coloring of the vertices of \( G \) using at most \( q \) distinct colors, as any adjacent pair of vertices of \( G \) must be mapped to distinct vertices of
Figure 1.1: Target graphs $H$ and the combinatorial counting problems they define as $\#H$-coloring problems.

Figure 1.2: Target graph $H$ for counting $q$-particle Widom-Rowlinson configurations for $q \in \{2, 3, 4\}$ as an $\#H$-coloring problem.

Figure 1.3: Target graph $H$ for counting $q$-type Beach model configurations for $q \in \{2, 3, 4\}$ as an $\#H$-coloring problem.
Partly due to this special case, counting graph homomorphisms to $H$ in general is also called $\#H$-colorings, and $H$ is called the target graph.

Graph homomorphisms can express a variety of combinatorial problems. In fact its principal purpose is to express and then treat a wide variety of locally defined graph properties in a uniform way. For example, if $H$ is a graph consisting of two vertices $\{T, F\}$ and two edges $\{(T, T), (T, F)\}$, a loop and an edge between the two vertices (see Figure 1.1a), then the $\#H$-coloring problem is the counting problem of VERTEX COVER, i.e., to count the number of vertex covers in graph $G$, denoted as $\#VC$. Indeed a homomorphism from $G$ to this $H$ is a two-coloring of vertices of $G$ such that the subset of vertices of $G$ mapped to $T$ forms a vertex cover. By flipping the intended meaning of $T$ and $F$, this also counts the number of independent sets in $G$. Of course this simply reflects the fact that a subset $S \subseteq V(G)$ is a vertex cover iff its complement $S^c \subseteq V(G)$ is an independent set. As another example, suppose $H$ is the two-vertex directed graph connected by a single directed edge and both vertices have one directed self-loop (see Figure 1.1b). Then the $\#H$-coloring problem takes directed graphs as input. If the input is a directed acyclic graph, then it defines a partial order, and the $\#H$-coloring problem is to compute the number of antichains (or equivalently, the number of lower sets, i.e., downward closed sets in the partial order, or equivalently, the number of upper sets) in this partial order. If the input is not acyclic, then every strongly connected component must be mapped to one vertex in $H$, and the $\#H$-coloring problem is to count the number of antichains in the induced partial order after collapsing each strongly connected component.

**Exercise:** Show that $\#VC$ can be expressed as $\#H$-coloring by defining explicitly the binary constraint function.

**Exercise:** Show that there is a bijection between the set of antichains and the set of lower sets. Conclude that the number of antichains is the same as the number of lower sets. By symmetry the same is proved for upper sets.

More generally, we consider weighted graph homomorphisms. Let $A$ be a $q$-by-$q$ matrix over $\mathbb{C}$. Given a graph or a directed graph $G = (V, E)$, the graph homomorphism problem is to compute

$$Z_A(G) = \sum_{\sigma: V \rightarrow [q]} \prod_{(u, v) \in E} A_{\sigma(u), \sigma(v)}. \quad (1.4)$$

The target graph $H$ is now defined by the matrix $A$, which is the weighted adjacency matrix of $H$. When $A$ is a 0-1 matrix, then this is the unweighted version of graph homomorphism. The matrix $A$ can also be identified as a binary function. In the case of VERTEX COVER the function is the binary Boolean OR function. For INDEPENDENT SET this is the NAND function. For the problem of counting antichains, the function is the binary IMPLICATION function. For $q$-coloring, this is the binary DISEQUALITY function on domain $[q]$.

This is exactly the same notion of a partition function in statistical physics. The Ising
model [?] corresponds to the partition function with matrix \( A = \begin{bmatrix} a & b \\ b & a \end{bmatrix} \). The Potts model [?] corresponds to the partition function with matrix \( A = J_q + \gamma I_q \), were \( J_q \) is the \( q \)-by-\( q \) matrix of all 1’s and \( I_q \) is the \( q \)-by-\( q \) identity matrix, and \( \gamma \) is a parameter.

The \( q \)-particle Widom-Rowlinson model [?] corresponds to the \( \#H \)-coloring problem in which the domain size is \( q + 1 \), \( H \) is the star graph on \( q + 1 \) vertices and all vertices have self loops (see Figure 1.2). The \( q \)-type Beach model [?, ?] corresponds to the \( \#H \)-coloring problem in which the domain size is \( 2q \), \( H \) is the complete graph on \( q \) vertices, each of these \( q \) vertices has a pendant vertex, and all \( 2q \) vertices have a self loop (see Figure 1.3). For more on this connection with statistical physics, see [?, Chapter 4] or [?, Chapter 2] (as well as [?]).

Graph Homomorphism can be viewed as a special case of counting constraint satisfaction problems, where instead of one binary constraint function there can be a set of constraint functions.

### 1.2.2 Constraint Satisfaction Problems

A counting Constraint Satisfaction Problem (\( \#\text{CSP} \)) [?] is parametrized by a set of local constraint functions \( \mathcal{F} \). It is denoted by \( \#\text{CSP}_q(\mathcal{F}) \) when the signatures in \( \mathcal{F} \) are defined over a domain \([q]\) of size \( q \). An instance of \( \#\text{CSP}_q(\mathcal{F}) \) is a finite set of variables \( x_1, x_2, \ldots, x_n \), and a finite set \( C \) of clauses. Each clause is a constraint \( f \in \mathcal{F} \) of some arity \( m \) depending on \( f \) together with a sequence of \( m \) (not necessarily distinct) variables \( x_{i_1}, \ldots, x_{i_m} \in \{x_1, x_2, \ldots, x_n\} \). The output is

\[
\sum_{x_1, \ldots, x_n \in [q]} \prod_{(f(x_{i_1}, \ldots, x_{i_m})) \in C} f(x_{i_1}, \ldots, x_{i_m}).
\]  

When it is the Boolean domain (i.e. \( q = 2 \)), we denote it simply as \( \#\text{CSP}(\mathcal{F}) \). If \( \mathcal{F} \) consists of a single function \( f \) we write \( \#\text{CSP}(f) \) for \( \#\text{CSP}(\{f\}) \). We write similarly \( \#\text{CSP}(f, g) \) if \( \mathcal{F} = \{f, g\} \), and \( \#\text{CSP}(\mathcal{F}, g) \) for \( \#\text{CSP}(\mathcal{F} \cup \{g\}) \), etc.

The canonical example of a \( \#\text{CSP} \) is \( \#\text{SAT} \), or counting Boolean Satisfiability, the problem of counting the number of satisfying assignments to a given Boolean formula. As a constraint satisfaction problem, it is \( \#\text{CSP}(\mathcal{F}) \), with \( \mathcal{F} = \{\text{OR}_m \mid m \geq 1\} \cup \{\neq 2\} \), where \( \text{OR}_m \) is the OR function of arity \( m \) and \( (\neq 2) = [0, 1, 0] \) is the binary DISEQUALITY function. Here are several more well-known examples of \( \#\text{CSP}(\mathcal{F}) \)’s over the Boolean domain and their corresponding constraint sets:
SAT has $\mathcal{F} = \{ \text{OR}_m \mid m \geq 1 \} \cup \{ \neq \}$
3SAT has $\mathcal{F} = \{ \text{OR}_3, \neq \}$
1-in-3SAT has $\mathcal{F} = \{ \text{EXACT-ONE}_3, \neq \}$
NAE-3SAT has $\mathcal{F} = \{ \text{NOT-ALL-EQUAL}_3, \neq \}$

Mon-SAT has $\mathcal{F} = \{ \text{OR}_m \mid m \geq 1 \}$
Mon-3SAT has $\mathcal{F} = \{ \text{OR}_3 \}$
Mon-1-in-3SAT has $\mathcal{F} = \{ \text{EXACT-ONE}_3 \}$
Mon-NAE-3SAT has $\mathcal{F} = \{ \text{NOT-ALL-EQUAL}_3 \}$

By $\#\text{CSP}^d(\mathcal{F})$, we denote the special case of $\#\text{CSP}(\mathcal{F})$ in which every variable appears a multiple of $d$ times. Note that $\#\text{CSP}(\mathcal{F})$ is the same as $\#\text{CSP}^1(\mathcal{F})$, and $\#\text{CSP}^2(\mathcal{F})$ is the same as every variable appearing an even number of times.

### 1.2.3 Holant Problems

A signature grid $\Omega = (G, \pi)$ over $\mathcal{F}$ consists of a graph $G = (V, E)$ and a mapping $\pi$ that assigns to each vertex $v \in V$ an $f_v \in \mathcal{F}$ and a linear order of the incident edges at $v$. The arity of $f$ is equal to the degree at $v$, and the incident edges at $v$ are associated with the input variables of $f_v$. If all signatures in $\mathcal{F}$ are symmetric then there is no need to assign an order for incident edges at any $v$. A Holant problem is parametrized by a set of signatures, or local constraint functions, $\mathcal{F}$.

**Definition 1.1.** For a set $\mathcal{F}$ of signatures over a domain $[q]$, we define Holant$^q(\mathcal{F})$ as:

**Input:** A signature grid $\Omega = (G, \pi)$ over $\mathcal{F}$;

**Output:**

$$\text{Holant}^q(\Omega; \mathcal{F}) = \sum_{\sigma : E \rightarrow [q]} \prod_{v \in V} f_v(\sigma |_{E(v)}),$$

where

- $G = (V, E)$, and $E(v)$ denotes the incident edges of $v$ and
- $\sigma |_{E(v)}$ denotes the restriction of $\sigma$ to $E(v)$, and $f_v(\sigma |_{E(v)})$ is the evaluation of $f_v$ on the ordered input tuple $\sigma |_{E(v)}$.

In Part I of this book we exclusively present the theory over the Boolean domain $q = 2$. We use Holant$(\mathcal{F})$ to denote Holant$^2(\mathcal{F})$. We also denote Holant$(\Omega)$ or Holant$\Omega$ for Holant$(\Omega; \mathcal{F})$, omitting $\mathcal{F}$ in the expression when $\mathcal{F}$ is clear from context. We use $G$ in place of $\Omega$ when $\pi$ is clear from context. We write Holant$^q(\mathcal{F}, g)$ for Holant$^q(\mathcal{F} \cup \{ g \})$. When $\mathcal{F}$ is a finite set of signatures, we sometimes just list the signatures it contains. For example, if $\mathcal{F} = \{ f, g \}$, then instead of writing Holant$^q(\{ f, g \})$, we may also write Holant$^q(f, g)$. This is especially true when $\mathcal{F}$ is a singleton set.
A signature $f$ of arity $m$ over the domain $[q]$ can be denoted by $(f_0, \ldots, f_x, \ldots, f_{q^m-1})$, where $f_x$ is the output of $f$ on $x \in [q]^m$, ordered lexicographically as integers from 0 to $q^m - 1$. This listing of values of $f$ in lexicographical order is the same as in a truth table. It is a vector in $\mathbb{C}^q$, or a tensor in $(\mathbb{C}^q)^\otimes m$ indexed by $[q]^m$. A symmetric signature $f$ of arity $m$ over the Boolean domain can be expressed as $[f_0, f_1, \ldots, f_m]$, where $f_w$ is the value of $f$ on inputs of Hamming weight $w$. An example is the \textsc{Equality} signature $(=_m) = [1, 0, \ldots, 0, 1]$ of arity $m$, where there are $m - 1$ entries of 0.

We give examples of Holant problems using a symmetric signature $f$ of arity $m$. If $f$ is the only available signature and its arity is $m$, then the graph $G$ in a Holant problem using $\{f\}$ must be $m$-regular, i.e., every vertex of the graph $G$ has degree $m$. Here are four examples over the Boolean domain.

$$\text{Holant}(G; f) \text{ counts } \begin{cases} \text{matchings} & \text{in } G \text{ when } f = \text{At-Most-One}_m; \\ \text{perfect matchings} & \text{in } G \text{ when } f = \text{Exact-One}_m; \\ \text{cycle covers} & \text{in } G \text{ when } f = \text{Exact-Two}_m; \\ \text{edge covers} & \text{in } G \text{ when } f = \text{OR}_m. \end{cases}$$

An example for any domain of size $q$ is that $\text{Holant}_q(G; \text{All-Distinct})$ counts proper edge colorings in $G$ using at most $q$ colors, i.e., at every vertex of $G$ the incident edges are colored with distinct colors from $[q]$. Each of these five examples are expressed as Holant problem in a straightforward manner. A less obvious example is that $\text{Holant}(G; \frac{1}{2}[3,0,1,0,3])$ counts Eulerian orientations of a 4-regular graph $G$; this will become clear when we develop the theory further.

As stated above, one can view a signature of arity $m$ as a tensor along with a default choice of basis $\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_m} \mid i_1 i_2 \cdots i_m \in \{0,1\}^m\}$, where $e_0 = (1, 0)$ and $e_1 = (0, 1)$. Readers who are familiar with quantum computation may want to write it as $e_0 = |0\rangle$, $e_1 = |1\rangle$, and $e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_m} = |i_1 i_2 \cdots i_m\rangle$. When doing so, a signature grid is equivalent to a tensor network, and the Holant of the signature grid is equal to the scalar that remains after contracting all edges in the corresponding tensor network.

A \textit{planar signature grid} is a signature grid such that its underlying graph is planar and for some planar embedding, for every vertex $v$, the linear order of the incident edges at $v$ agrees with the cyclic order of the incident edges at $v$ in the embedding starting with some particular edge. We use $\text{Pl-Holant}_q(\mathcal{F})$ to denote the restriction of $\text{Holant}_q(\mathcal{F})$ to planar signature grids. For signature sets $\mathcal{F}$ and $\mathcal{G}$, a \textit{bipartite signature grid} over $(\mathcal{F} \mid \mathcal{G})$ is a signature grid $\Omega = (H, \pi)$ over $\mathcal{F} \cup \mathcal{G}$, where $H = (V, E)$ is a bipartite graph with bipartition $V = (V_1, V_2)$ such that $\pi(V_1) \subseteq \mathcal{F}$ and $\pi(V_2) \subseteq \mathcal{G}$. Signatures in $\mathcal{F}$ are considered as row vectors (or covariant tensors); signatures in $\mathcal{G}$ are considered as column vectors (or contravariant tensors) [?]. We use $\text{Holant}_q(\mathcal{F} \mid \mathcal{G})$ to denote the restriction of $\text{Holant}_q(\mathcal{F} \cup \mathcal{G})$ to bipartite signature grids over $(\mathcal{F} \mid \mathcal{G})$, and $\text{Holant}_q(\Omega; \mathcal{F} | \mathcal{G})$ to denote the value on $\Omega$. A \textit{planar bipartite signature grid} is one that is both planar and bipartite. We use $\text{Pl-Holant}_q(\mathcal{F} \mid \mathcal{G})$ to denote the restriction to these signature grids, and $\text{Pl-Holant}_q(\Omega; \mathcal{F} | \mathcal{G})$ to denote the value on $\Omega$. We use LHS and RHS to stand for left-hand side and right-hand side respectively.
A signature $f$ of arity $m$ is degenerate if there exist unary signatures $u_j \in \mathbb{C}^q$ ($1 \leq j \leq m$) such that $f = u_1 \otimes \cdots \otimes u_m$. Replacing such an $f$ by the $m$ unary signatures $u_j$, one for each associated edge, does not change the Holant value. A symmetric degenerate signature has the form $u^{\otimes m}$, for some unary $u \in \mathbb{C}^q$. Replacing a signature $f \in \mathcal{F}$ by a constant multiple $cf$, where $c \neq 0$, does not change the complexity of Holant$_q(\mathcal{F})$. It merely introduces a nonzero factor $c^n$ to Holant$_q(\Omega; \mathcal{F})$, where the graph has $n$ vertices.

**Exercise:** Prove that the Holant value is unchanged when $f = u_1 \otimes \cdots \otimes u_m$ is replaced by unary signatures $u_j$, with $u_j$ assigned to the $j$-th incident edge of every vertex assigned $f$.

**Exercise:** Prove that if a degenerate signature $f = u_1 \otimes \cdots \otimes u_m$ is symmetric, then it has the form $u^{\otimes m}$, for some unary $u \in \mathbb{C}^q$.

We use the term tractable as a shorthand for polynomial time computable. We allow $\mathcal{F}$ to be an infinite set. For Holant$_q(\mathcal{F})$ to be tractable, the problem must be computable in polynomial time when the input graph $G$ and the description of the signatures appearing in $G$ are included in the input size. In contrast, we say Holant$_q(\mathcal{F})$ is #P-hard if there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard. We say a signature set $\mathcal{F}$ is tractable (resp. #P-hard) in the context of Holant problems if the corresponding counting problem Holant$_q(\mathcal{F})$ is tractable (resp. #P-hard). Similarly for a signature $f$, we say $f$ is tractable (resp. #P-hard) if $\{f\}$ is. We also speak of a signature or signature set as being tractable or #P-hard for various restricted classes of Holant problems, such as those defined over planar, bipartite, or planar and bipartite graphs. The meaning of the restriction should be clear from context.

We can express #CSP$_d^q(\mathcal{F})$ as a Holant problem. An instance of #CSP$_d^q(\mathcal{F})$ has the following bipartite view. Create a node for each variable and each clause. Connect a variable node to a clause node if the variable appears in the clause. This bipartite graph is also known as the constraint graph. To each variable vertex, we assign the Equality signature of the appropriate arity. To each clause vertex, we assign the constraint used in that clause. Under this view, we see that

$$\#\text{CSP}_d^q(\mathcal{F}) \equiv_T \text{Holant}_q(\mathcal{E} \mathcal{Q}_d \mid \mathcal{F}),$$

(1.6)

where $\mathcal{E} \mathcal{Q}_d = \{\omega_k \mid k \geq 1\}$ is the set of Equality signatures of arities equal to a multiple of $d$. We denote by Pl-#CSP$_d^q(\mathcal{F})$ the restriction of #CSP$_d^q(\mathcal{F})$ to inputs with a planar constraint graph. Again we drop $q$ and write Pl-#CSP$_d$ if $q = 2$, and drop $d$ if $d = 1$. The construction above also shows that

$$\text{Pl-#CSP}_d^q(\mathcal{F}) \equiv_T \text{Pl-Holant}_q(\mathcal{E} \mathcal{Q}_d \mid \mathcal{F}).$$

(1.7)

If $d \in \{1, 2\}$, then more is true.

**Lemma 1.2.** Let $\mathcal{F}$ be a set of signatures over a domain of size $q$. If $d \in \{1, 2\}$, then

$$\#\text{CSP}_d^q(\mathcal{F}) \equiv_T \text{Holant}_q(\mathcal{E} \mathcal{Q}_d \cup \mathcal{F}) \quad \text{and} \quad \text{Pl-#CSP}_d^q(\mathcal{F}) \equiv_T \text{Pl-Holant}_q(\mathcal{E} \mathcal{Q}_d \cup \mathcal{F}).$$

17
Proof. By (1.6) and (1.7), it suffices to show

$$\text{Holant}_q(\mathcal{E}Q_d \mid \mathcal{F}) \equiv \text{Holant}_q(\mathcal{E}Q_d \cup \mathcal{F})$$

and correspondingly for Pl-Holant$_q(\mathcal{E}Q_d \mid \mathcal{F})$. It both cases, the reduction from left to right in the equivalence is trivial; just ignore the bipartite restriction. For the other direction, we take a signature grid for the problem on the right and create a bipartite signature grid for the problem on the left such that both signature grids have the same Holant value up to an easily computable factor. If the initial graph is planar, then the final graph will also be planar, so this will prove both equivalences.

If two signatures in $\mathcal{F}$ are assigned to adjacent vertices, then we subdivide all edges between them and assign the binary $\text{EQUALITY}$ signature ($=2$) $\in \mathcal{E}Q_d$ to all new vertices. Suppose $\text{EQUALITY}$ signatures ($=n$), ($=m$) $\in \mathcal{E}Q_d$ are assigned to adjacent vertices connected by $\ell$ edges. If $n = m = \ell$, then these two vertices are disjoint from the rest, and we simply remove these two vertices. The Holant of the resulting signature grid differs from the original by a factor of $q$. Otherwise, we contract all $\ell$ edges, merge the two vertices into one, and assign ($=n+m-2\ell$) $\in \mathcal{E}Q_d$ to the new vertex.

\[\square\]

1.3 Reductions

We now introduce several types of reductions we will use in this book, including gadget constructions, holographic transformations, and interpolations.

1.3.1 Gadget Constructions

Local Gadget Constructions via $\mathcal{F}$-gates

A basic type of reduction is what might be generally known as a gadget construction. In the context of Holant problems, we create “local” gadget constructions in order to realize a signature. Fix a set $\mathcal{F}$ of signature over a domain $[q]$ of size $q$. We say a signature $f$ is realizable or obtainable from $\mathcal{F}$ if there is a gadget with some dangling edges such that each vertex is assigned a signature from $\mathcal{F}$, and the resulting graph, when viewed as a black-box signature with inputs on the dangling edges, is exactly $f$.

Formally, such a notion is defined by an $\mathcal{F}$-gate. An $\mathcal{F}$-gate $F$ is similar to a signature grid $(G, \pi)$ for $\text{Holant}(\mathcal{F})$ except that $G = (V, E, E')$ is a graph with regular edges in $E$ and some dangling edges in $E'$. The dangling edges define external variables for the $\mathcal{F}$-gate. They are ordered, and usually presented pictorially by starting at the edge marked with a diamond and proceeding counterclockwise (see Figure 1.4 for an example.)
An F-gate F with m dangling edges defines the function

\[ \Gamma(y_1, \ldots, y_m) = \sum_{\sigma : E \to [q]} \prod_{v \in V} f_v(\tilde{\sigma} |_{E(v)}), \]

where \((y_1, \ldots, y_m) \in [q]^m\) is an assignment on the dangling edges, \(\tilde{\sigma}\) is the extension of the assignment \(\sigma\) on the internal edges \(E\) by the assignment \((y_1, \ldots, y_m)\) on the dangling edges \(E'\), and \(f_v\) is the signature assigned by \(\pi\) at \(v\). We call this function \(\Gamma\) the signature of the F-gate. We also call an F-gate a gadget. If the signature of an F-gate is invariant under cyclic permutations of inputs, then we omit the diamond since it is unnecessary. We say that such signatures are rotationally symmetric. If the signature is invariant under any permutation of inputs, then we say it is symmetric. These notions are defined in terms of the signature function \(\Gamma\). The signature of an F-gate may be rotationally symmetric or symmetric as a function, even though the graph defining it may not be so.

An F-gate is planar if the underlying graph can be embedded in the plane without edge crossings and the dangling edges are in the outer face in the specified order. Now suppose we have two signature sets \(\mathcal{F}\) and \(\mathcal{G}\) in the context of a bipartite Holant problem \(\text{Holant}_q(\mathcal{F} \mid \mathcal{G})\). Then an \((\mathcal{F} \mid \mathcal{G})\)-gate is an \((\mathcal{F} \cup \mathcal{G})\)-gate such that the underlying graph is bipartite, the vertices in one part are assigned signatures from \(\mathcal{F}\), and the vertices in the other part are assigned signatures from \(\mathcal{G}\). Furthermore, we say that an \((\mathcal{F} \mid \mathcal{G})\)-gate is on the left (resp. on the right) if each vertex incident to a dangling edge is assigned a signature from \(\mathcal{F}\) (resp. \(\mathcal{G}\)).

Using F-gates, we can reduce one Holant problem to another.

**Lemma 1.3.** Let \(\mathcal{F}\) be a set of signatures over a domain of size \(q\). If there exists an F-gate with signature \(f\), then

\[ \text{Holant}_q(\mathcal{F}, f) \leq_T \text{Holant}_q(\mathcal{F}). \]

Similar statements hold for gadgets that are planar, bipartite, or both for Holant problems defined over the same class of graphs.
Proof. Let $F$ be an $F$-gate with signature $f$. Given an instance $\Omega$ of Holant$_q(F,f)$, we replace every occurrence of $f$ by the $F$-gate $F$ to obtain an instance $\Omega'$ of Holant$_q(F)$. Since $f$ is the signature of the $F$-gate $F$, the Holant values for these two signature grids are identical. Furthermore, the size of $F$ is a constant with respect to $\Omega$, so the size $\Omega'$ is only a constant factor larger than $\Omega$.

Even for a very simple signature set $F$, the signatures for all $F$-gates can be quite complicated and expressive. In Chapter 4 we will see that for the simple class of weighted Exact-One functions, which define weighted Perfect Matchings, the resulting signature set can be quite expressive. These are called matchgates.

**Signature Matrix**

It is convenient to write a signature as a matrix. An immediate advantage is that a matrix is more of a pictorial representation than a vector, which aids understanding. However, the more important reason is to simplify the computation of the signature of a gadget.

**Definition 1.4.** Let $f$ be a signature of arity $n$ over a domain of size $q$. The signature matrix of $f$ with parameter $\ell$ is an $q^\ell$-by-$q^{n-\ell}$ matrix for some integer $0 \leq \ell \leq n$ in which the first $\ell$ inputs (in order) are the row index and the remaining $n - \ell$ inputs (in reverse order) are the column index.

If the arity of $f$ is even, then the signature matrix of $f$, without specifying a parameter, is the signature matrix of $f$ with parameter $\ell = \frac{n}{2}$ and is denoted by $M_f$.

The purpose of reversing the order of the column index is so that we can use matrix product in gadget computations. If $f = (f_{00}, f_{01}, f_{10}, f_{11})$ has arity 2 over the Boolean domain, then $M_f = \begin{bmatrix} f_{00} & f_{01} \\ f_{10} & f_{11} \end{bmatrix}$. If $g$ is a signature of arity 4 over the Boolean domain with $g(i,j,k,\ell) = g^{ijk\ell}$, then

$$M_g = \begin{bmatrix} g_{0000} & g_{0010} & g_{0011} & g_{0011} \\ g_{0100} & g_{0110} & g_{0111} & g_{0111} \\ g_{1000} & g_{1010} & g_{1011} & g_{1011} \\ g_{1100} & g_{1110} & g_{1111} & g_{1111} \end{bmatrix}.$$

Notice the reversal of order in the column index $(k,\ell)$, i.e., the columns are listed in the order $k\ell = 00, 10, 01, 11$. Let $F$ be an $F$-gate with signature $f$ of arity $n$. We often depict $F$ with $\ell$ dangling edges protruding to the left and $n - \ell$ dangling edges protruding to the right to aid in the mapping from $F$ to the signature matrix of $f$ with parameter $\ell$. If $F$ and $F'$ are two $F$-gates of arity 4 with signature matrices $M$ and $M'$, then the matrix product $MM'$ is the signature matrix of the $F$-gate linking the two $F$-gates, namely by merging the 4th and 3rd edge (in cyclic order) of $F$ with the 1st and 2nd edge (also in cyclic order) of $F'$ respectively. In particular, if $F$ and $F'$ are planar $F$-gates, then this linking operation produces a planar $F$-gate.
1.3.2 Holographic Transformation

Sometimes two counting problems may appear different, but are really the same problem under a different guise. Holographic transformation is a primary tool to establish this quantitative connection.

It is a simple fact that in a graph, the number of vertex covers is equal to the number of independent sets. The reason is that the complement of one type is the other. At the vertex assignment level, this is accomplished by an exchange of the assignment 0 and 1. This is a type of combinatorial mapping.

In a holographic transformation the 0-1 assignments are no longer considered purely as discrete objects, but as embedded in a vector space as basis elements, and then we consider linear transformations of the ambient space. A holographic transformation puts the assignments of 0 and 1 into a superposition much like the states of a qubit in quantum computing. However, quantum computation is not required—or even any computation at all. Just like the example with counting vertex covers and independent sets, we are merely describing a mathematical proof that two different looking problems are actually the same. By undergoing a change of basis we can connect different problems from this perspective, and prove that they have the same (or closely related) answers. The map between vertex covers and independent sets can also be described by a linear transformation which is simply a basis exchange $e_0 = (1, 0) \leftrightarrow e_1 = (0, 1)$. In general, a holographic transformation will be much more algebraically “mixing”.

To formally introduce the idea of holographic transformations, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value as follows. For each edge in the graph, we replace it by a path of length two. (This operation is called the 2-stretch of the graph and yields the Edge-Vertex incidence graph.) Each new vertex is assigned the binary Equality signature (=2).

For a $q$-by-$q$ matrix $T$ and a signature $f$ of arity $n$, written as a column vector in $\mathbb{C}^q^n$, we write $Tf = T^\otimes nf$ as the transformed signature. For a signature set $\mathcal{F}$, define $\mathcal{F}T = \{Tf \mid f \in \mathcal{F}\}$. For a signature $f$ of arity $n$ written as a row vector, we similarly define $fT = fT^\otimes n$, and $\mathcal{F}T$. Here the tensor product denoted by $\otimes$ is the Kronecker product, namely, if $X_{a \times b}$ and $Y_{c \times d}$ are two matrices, then $X \otimes Y$ is $ac \times bd$, with entry $X_{ij}Y_{k\ell}$ at row indexed by $(i, k) \in [a] \times [c]$ and column indexed by $(j, \ell) \in [b] \times [d]$, both ordered lexicographically. Tensor power is defined inductively $X^\otimes n = X^{\otimes n-1} \otimes X$. Whenever we write $T^\otimes nf$ or $T\mathcal{F}$, we view the signatures as column vectors; similarly for $fT^\otimes n$ or $\mathcal{F}T$ as row vectors.

Let $T$ be an invertible $q$-by-$q$ matrix. The holographic transformation defined by $T$ is the following operation: given a bipartite signature grid $\Omega = (H, \pi)$ of Holant($\mathcal{F} \mid \mathcal{G}$), for the same bipartite graph $H$, we get a new grid $\Omega' = (H, \pi')$ of Holant($\mathcal{F}T \mid T^{-1}\mathcal{G}$) by replacing each signature in $\mathcal{F}$ or $\mathcal{G}$ with the corresponding signature in $\mathcal{F}T$ or $T^{-1}\mathcal{G}$. For this reason, signatures in $\mathcal{F}$ are called covariant and those in $\mathcal{G}$ contravariant. Valiant’s Holant Theorem [?] (see also [?]), Theorem 1.5, states that the Holant value is unchanged,
under a holographic transformation.

**Theorem 1.5.** Let \( F \) and \( G \) be sets of complex-valued signatures over a domain of size \( q \). Suppose \( \Omega \) is a bipartite signature grid over \((F \mid G)\). If \( T \in \text{GL}_q(\mathbb{C}) \), then

\[
\text{Holant}_q(\Omega; F \mid G) = \text{Holant}_q(\Omega'; FT \mid T^{-1}G),
\]

where \( \Omega' \) is the corresponding signature grid over \((FT \mid T^{-1}G)\).

**Proof.** We transform \( \Omega \) to \( \Omega' \) in several steps, preserving the Holant value in each step. The essential idea is that operator application is associative. We illustrate this in Figure 1.5.

Let \( \Omega_0 = \Omega \). Let \( G = (U; V; E) \) be the graph underlying \( \Omega_0 \). Vertices in \( U \) are assigned signatures in \( F \) by \( \pi \) while vertices in \( V \) are assigned signatures in \( G \) by \( \pi \). Let \( e = (u, v) \in E \) be an edge with endpoints \( u \in U \) and \( v \in V \). Initially, Figure 1.5a depicts the neighborhood of \( u \) and \( v \) in \( \Omega_0 = \Omega \). In this example, both \( u \) and \( v \) are incident to three other edges but the vertices incident to the other ends of these edges are not shown. We do the following operations for each edge \( e = (u, v) \in E \): Subdivide \( e \) and assign (=2) to the new vertex \( w \). Let the resulting signature grid be \( \Omega_1 \). See Figure 1.5b. It is clear that

\[
\text{Holant}_q(\Omega_0; F \mid G) = \text{Holant}_q(\Omega_1; FT \mid T^{-1}G).
\]

Then we subdivide \( w \) to get two adjacent vertices \( u' \) and \( v' \) so that we now have a path \((u, u', v', v)\) for every edge \( e = (u, v) \in E \) in \( G \). Assign to \( u' \) and \( v' \) the binary signature (call them \( h_u \) and \( h_v \), respectively) whose signature matrix is \( T \) and \( T^{-1} \) respectively. If \( T \) is not a symmetric matrix, then these signatures are not symmetric and it matters which edge corresponds to which input. The first input for \( h_u \), represented by the row index in \( T \), corresponds to the edge \( \{u, u'\} \). The first input for \( h_v \), represented by the row index in \( T^{-1} \), corresponds to the edge \( \{u', v'\} \), so that it is merged with the second input for \( h_u \), represented by the column index in \( T \). Let the resulting signature grid be \( \Omega_2 \). See Figure 1.5c, which indicates the first inputs of \( h_u \) and \( h_v \) by the rotated placement of \( T \) and \( T^{-1} \), with the first input to their left. If we contract the edge \( \{u', v'\} \) within the dashed box, then we get back \( \Omega_1 \). Thus, the Holant value is unchanged,

\[
\text{Holant}_q(\Omega_0; F \mid G) = \text{Holant}_q(\Omega_1; F \mid G) = \text{Holant}_q(\Omega_2; F \mid G) = \text{Holant}_q(\Omega_3; F \mid G) = \text{Holant}_q(\Omega_4; F \mid G).
\]

Now \( \Omega_3 \) in Figure 1.5d is really the same as \( \Omega_2 \), except we think of the order of contraction operations differently by associating the binary \( T \) and \( T^{-1} \) to the original signatures \( f \) and \( g \) first.

\[
\text{Holant}_q(\Omega_2; F \mid G) = \text{Holant}_q(\Omega_3; F \mid G) = \text{Holant}_q(\Omega_4; F \mid G) = \text{Holant}_q(\Omega_5; F \mid G).
\]

Finally we contract \( \{u, u'\} \) and \( \{v', v\} \). This defines \( \Omega' = \Omega_4 \). After doing so, we once again have the bipartite graph \( G \). What has changed is the assignment to each vertex. If vertex \( u \in U \) of degree \( \deg(u) \) was assigned \( f \in F \) in \( \Omega \) then it is now assigned \( fT^\otimes \deg(u) \),
Figure 1.5: Neighborhood around two adjacent vertices.
Figure 1.5: Neighborhood around two adjacent vertices.
and if vertex $v \in V$ of degree $\deg(v)$ was assigned $g \in G$ in $\Omega$ then it is now assigned $(T^{-1})^\otimes \deg(v)g$. In general, the new assignment to each vertex is the transformed signature in $\mathcal{F}T$ or $T^{-1}G$ respectively, as claimed. Hence,

$$\text{Holant}_q(\Omega_3; \mathcal{F} \cup G \cup \{T, T^{-1}\}) = \text{Holant}_q(\Omega_4; \mathcal{F}T \mid T^{-1}G).$$

Therefore, an invertible holographic transformation does not change the complexity of the Holant problem in the bipartite setting. Furthermore, there is a special kind of holographic transformation, the orthogonal transformation, that preserves the binary equality and thus can be used freely in the standard setting. We denote by $O_q(\mathbb{C})$ the group of $q \times q$ orthogonal matrices over $\mathbb{C}$. Note that $T \in O_q(\mathbb{C})$ iff $T^{-1}T = I_q$ iff the holographic transformation $(=2) \mapsto (=2)T^\otimes 2 = (=2)$. The latter identity can be checked by its matrix form, because for a binary signature $f$, if $M$ is its matrix form, then $T^TM$ is the matrix form of the binary signature $fT^\otimes 2$.

**Exercise:** Verify that for a binary covariant signature (row vector) $f$ with signature matrix $M$, the signature matrix of $fT^\otimes 2$ is $T^TM$. If $f$ is a binary contravariant signature (column vector) with signature matrix $M$, then the signature matrix of $(T^{-1})^\otimes 2f$ is $T^{-1}M(T^{-1})^T$. Generalize this to signatures of higher arity.

**Theorem 1.6.** Let $\mathcal{F}$ be a set of complex-valued signatures over a domain of size $q$. Suppose $\Omega$ is a signature grid over $\mathcal{F}$. If $H \in O_q(\mathbb{C})$, then

$$\text{Holant}_q(\Omega; \mathcal{F}) = \text{Holant}_q(\Omega'; H\mathcal{F}),$$

where $\Omega'$ is the corresponding signature grid over $H\mathcal{F}$.

**Proof.** $\text{Holant}_q(\Omega; \mathcal{F}) = \text{Holant}_q(\Omega; (=2) \mid \mathcal{F})$. \qed

We use Theorems 1.5 and 1.6 to reduce between both tractable problems and both #P-hard problems. Some of our reductions between tractable problems use the following definition, where $\mathcal{C}$ is usually some known tractable class of signatures.

**Definition 1.7.** Let $\mathcal{F}$ be any set of complex-valued signatures over a domain of size $q$. We say $\mathcal{F}$ is $\mathcal{C}$-transformable if there exists a $T \in \text{GL}_q(\mathbb{C})$ such that $(=2)T^\otimes 2 \in \mathcal{C}$ and $\mathcal{F} \subseteq T\mathcal{C}$.

Note that if $\text{Holant}_q(\mathcal{C})$ is tractable over any set of graphs, then $\text{Holant}_q(\mathcal{F})$ is tractable over the same set of graphs for any $\mathcal{C}$-transformable set $\mathcal{F}$.

**Lemma 1.8.** Let $\mathcal{F}$ be any set of complex-valued signatures over a domain of size $q$. If $\mathcal{F}$ is $\mathcal{C}$-transformable, then

$$\text{Holant}_q(\mathcal{F}) \leq_T \text{Holant}_q(\mathcal{C}),$$

where both problems are defined over the same set of graphs.
Another Identity

Holographic transformations between two signature grids give the same Holant value. The following lemma shows a different kind of reason why two Holant problems must always have the same Holant value.

**Lemma 1.9.** Let $x$ and $y$ be any numbers in $\mathbb{C}$. Then for every signature grid $\Omega$ having a $(2,4)$-regular bipartite underlying graph,

$$\text{Holant}(\Omega; [0,1,0] \mid [x,y,1,0,0]) = \text{Holant}(\Omega; [0,1,0] \mid [0,0,1,0,0]).$$

**Proof.** In any signature grid $\Omega$ having a $(2,4)$-regular bipartite underlying graph, because $[0,1,0]$ is the only signature on the left, any nonzero term in the Holant sum must assign 1 to exactly half of the edges in $G$. On the right side, if some copy of $[x,y,1,0,0]$ contributes an $x$ or $y$ in some assignment, then less than half of its incident edges are assigned 1. To compensate, some other copy of $[x,y,1,0,0]$ must have more than half of its incident edges assigned 0, so it contributes a factor of 0.

**Exercise:** Prove that for all $x, y \in \mathbb{C}$ not both 0, there is no holographic transformation between the two Holant problems in Lemma 1.9. Thus the converse of Theorem 1.6 does not hold.

### 1.3.3 Polynomial Interpolation

Polynomial interpolation is a powerful technique to prove $\#P$-hardness for counting problems. We can introduce this technique with an example from [?] by Valiant.

![Figure 1.6: Some graphs obtained from an initial graph $G$ in the proof of Lemma 1.10.](image)

\[ G = G_0 \quad G_1 \quad G_2 \quad G_\ell \]

**Lemma 1.10.** $\#\text{PerfectMatching} \leq_T \#\text{Matching}$

26
Proof. Let $G = (V, E)$ be a graph. We want to determine the number of perfect matchings in $G$ assuming that we have an oracle to count (all, not necessarily perfect) matchings.

For every integer $0 \leq \ell \leq n$, we construct a graph $G_\ell$ from $G$ in the following way. For each vertex $v \in V$ we will “grow a thistle”, namely we add $\ell$ new vertices $v_k$ ($1 \leq k \leq \ell$), each connected to $v$ by one edge. See Figure 1.6 for some examples of these graphs beginning with a specific graph $G$. In $G_\ell$ each $v_k$ has degree 1 and clearly it can only be matched to $v$.

If $v$ is matched to some $v_k$ ($1 \leq k \leq \ell$), then $v$ cannot be matched to any other $v_\ell$ for $\ell \neq k$ nor can it be matched to any other original vertex of $G$.

Let $m_k$ be the number of matchings in $G$ that omit $k$ vertices. Then we can express the number of matchings in $G_\ell$ as

$$\sum_{k=0}^{n} (\ell + 1)^k m_k = \#\text{Matching}(G_\ell).$$

This is because a matching of $G$ that omits exactly $k$ vertices can be extended to exactly $(\ell + 1)^k$ matchings in $G_\ell$, and every matching in $G_\ell$ is obtained uniquely this way from a matching of $G$. To see this, let $M$ be a matching of $G$ that omits $k$ vertices. Then each of the $k$ vertices in $G$ that is not matched by $M$ has $\ell + 1$ possibilities with respect to a matching $M'$ in $G_\ell$ extending $M$, because it can be matched with one of the $\ell$ new neighbors in $G_\ell$ or remain unmatched. Conversely, for every matching $M'$ of $G_\ell$, we can consider its subset consisting of edges in $G$. This defines a matching $M$ in $G$ and $M'$ is one of the extensions from $M$, and such an $M$ is unique.

We collect these equations to form the linear system

$$\begin{bmatrix}
(0+1)^0 & (0+1)^1 & \cdots & (0+1)^n \\
(1+1)^0 & (1+1)^1 & \cdots & (1+1)^n \\
\vdots & \vdots & \ddots & \vdots \\
(n+1)^0 & (n+1)^1 & \cdots & (n+1)^n
\end{bmatrix}
\begin{bmatrix}
m_0 \\
m_1 \\
\vdots \\
m_n
\end{bmatrix}
= \begin{bmatrix}
\#\text{Matching}(G_0) \\
\#\text{Matching}(G_1) \\
\vdots \\
\#\text{Matching}(G_n)
\end{bmatrix}.$$  

Using our oracle, we know the right side. On the left, the coefficient matrix is Vandermonde. It is invertible because the values $\ell + 1$, for $0 \leq \ell \leq n$, are all distinct. Therefore, we can invert this matrix and solve for the unknown $m_k$’s in polynomial time. Then $m_0$, the number of matchings in $G$ that omit no vertices, is the number of perfect matchings $G$ as desired. \qed

The word “polynomial” did not appear in this proof, so what makes it an example of polynomial interpolation? The polynomial is implicit; it is $p(x) = \sum_{k=0}^{n} m_k x^k$. Asking our oracle for the number of matchings in $G_\ell$ is equivalent to evaluating $p(x)$ at $\ell + 1$. Of course this is made possible by the design of the gadget of the “thistle”, a weed like structure, which produced such an expression for $\#\text{Matching}(G_\ell)$. Such designs are where the inventiveness is in an interpolation proof. Polynomial interpolation is the process of converting from points and their evaluations to the coefficients of the polynomial being evaluated (see Figure 1.7), which is what this proof did.
Evaluate
\( x \in \{1, 2, 3, 4\} \)

Interpolate

\[ p(x) = 2x^3 - 3x^2 - 17x + 10 \]

Figure 1.7: Interpolation is the inverse of evaluation.

Since our \( n + 1 \) evaluation points are distinct, we can recover the coefficients of \( p(x) \). In this simple interpolation proof, the coefficient matrix is clearly invertible. For more complicated proofs this can be more demanding mathematically, and of course this works hand-in-hand with the inventive process of designing the right gadget. Given these coefficients, our reduction can proceed by computing any polynomial-time computable function from these coefficients. It is also often the case that we are interested in some evaluation of the interpolated polynomial. In the proof of Lemma 1.10, we evaluated the interpolated polynomial at 0 to obtain \( p(0) = m_0 \).

In fact both problems \#\textsc{Matching} and \#\textsc{PerfectMatching} are \#P-complete. We prove the reverse reduction from the problem of counting all matchings \#\textsc{Matching} to the problem of counting perfect matchings \#\textsc{PerfectMatching}. In contrast the following proof is not by interpolation.

**Lemma 1.11.** \#\textsc{Matching} \( \leq_T \) \#\textsc{PerfectMatching}

**Proof.** Let \( G = (V, E) \) be a graph. We want to determine the number of all, not necessarily perfect, matchings in \( G \) assuming that we have an oracle to count perfect matchings. Suppose \( G \) has \( n \) vertices \( V = \{v_i \mid 1 \leq i \leq n\} \). For every \( k \) such that \( 0 \leq k \leq n/2 \), consider the graph \( G_k = (V', E') \), where \( V' = V \cup \{z_j \mid 1 \leq j \leq n - 2k\} \), a disjoint union of \( V \) and \( n - 2k \) new vertices, and \( E' \) consists of \( E \) and \( n(n - 2k) \) new edges consisting of all pairs \( \{(v_i, z_j) \mid 1 \leq i \leq n, 1 \leq j \leq n - 2k\} \), i.e., a copy of the complete bipartite graph \( K_{n,n-2k} \).
Now it is easy to see that every matching in $G$ of size $k$ has exactly $(n - 2k)!$ extensions to a perfect matching in $G_k$, matching the remaining $n - 2k$ vertices of $V$ to the new vertices in $V'$. Conversely, every perfect matching in $G_k$ is obtained uniquely in this way. Thus by computing $\#\text{PERFECTMATCHING}$ on $G_k$ we obtain the number of matchings in $G$ of size $k$, and summing over all $0 \leq k \leq n/2$ gives $\#\text{MATCHING}$ on $G$.

We note that the proof in Lemma 1.10 is planar in the sense that if the given $G$ is a planar graph then all graphs $G_\ell$ constructed from $G$ are also planar graphs. By contrast, the proof in Lemma 1.11 is not planar in the sense that starting from a planar graph $G$ the graphs $G_k$ constructed from $G$ in the proof are in general not planar. There is an intrinsic reason for this discrepancy in the proof. It turns out that $\#\text{MATCHING}$ is $\#P$-complete even when restricted on planar graphs, while $\#\text{PERFECTMATCHING}$ is only $\#P$-complete for general graphs; we will see in Chapter 4 that $\#\text{PERFECTMATCHING}$ can be computed in polynomial time over planar graphs. Consequently, any proof of Lemma 1.11 that is planar in the above sense would imply a collapse of $\#P$ to $P$. This would easily imply $P = NP$, thus it is a more severe collapse than $P = NP$. This eventuality is considered most unlikely, and philosophically the dichotomy theorems in this book are only meaningful assuming $\#P$ does not collapse to $P$, although the statements and proofs of these theorems do not make this assumption. One of the themes in this book is also to obtain complexity classifications of counting problems which further delineates the boundary between planar and non-planar graphs, and the algorithm for $\#\text{PERFECTMATCHING}$ over planar graphs plays an important role.

The heart of polynomial interpolation as a reduction technique is finding an equation system like (1.9). For the $\ell$-th equation on the left, we have a combination involving some partial sums of the original problem on $G$. On the right, we have the evaluation of the Holant problem on the constructed graph $G_\ell$. The art is how to pick $G_\ell$ so that a relation like (1.8) holds. It is important that the original sum on $G$ is partitioned into at most a polynomial number of partial sums. In this example (1.8) all exponentially many configurations (up to $\binom{n}{k}$) of matchings in $G$ that omit exactly $k$ vertices are combined, with the common property that each of them has exactly $(\ell + 1)^k$ extensions to a matching in $G_\ell$. The construction produces an infinite family $G_\ell$, and size of the graphs in this family should not grow faster than a polynomial in $G$ and $\ell$, so that we can hope to construct them in polynomial time. Finally, they should be indeed constructible in polynomial time and some effective relationship such as (1.8) can be established.

Now we consider an example where the polynomial is quite explicit. The chromatic polynomial, denoted by $\chi(G; \lambda)$, is the unique polynomial in $\lambda$, such that for a natural number $k \in \mathbb{N}$, $\chi(G; k)$ is the number of proper vertex colorings of $G$, i.e., a labeling of the vertices of $G$ such that no two neighboring vertices receive the same color, using at most $k$ colors. A nice exposition of the chromatic polynomial can be found in [?].

The following example of polynomial interpolation is a dichotomy theorem for the chro-
matic polynomial. The reduction we use comes from Linial [?] and the dichotomy was first explicitly stated in [?]. Let \( \chi(\lambda) \) be the problem of evaluating \( \chi(G; \lambda) \) on an input graph \( G \).

**Lemma 1.12.** Let \( \lambda \in \mathbb{C} \). Then \( \chi(\lambda) \) is \#P-hard unless \( \lambda \in \{0, 1, 2\} \), in which case, the problem is computable in polynomial time.

**Proof.** If \( \lambda = 0 \), then \( \chi(G, \lambda) = 0^0 = 1 \) if \( G \) has no vertices and is 0 otherwise. If \( \lambda = 1 \), then \( \chi(G, \lambda) = 1 \) if \( G \) has no edges and is 0 otherwise. If \( \lambda = 2 \), then \( \chi(G, \lambda) = 2^k \) if \( G \) is bipartite with \( k \) connected components and is 0 otherwise.

Now suppose \( \lambda \notin \{0, 1, 2\} \). We reduce from \( \chi(3) \), which is known to be \#P-hard (see, for example [?, Main Theorem, Case (6)] or [?, Proposition 5]). Let \( G \) be a graph with \( n \) vertices, and let \( K_t \) be the complete graph on \( t \) vertices. We use \( G + K_t \) to denote the graph obtained from \( G \) by adding \( K_t \) and all possible edges between the vertices of \( G \) and the vertices of \( K_t \). Then clearly

\[
\chi(G + K_t; \lambda) = \lambda(\lambda - 1) \cdots (\lambda - t + 1)\chi(G; \lambda - t).
\]

when \( \lambda \in \mathbb{N} \). Thus, it must also hold as an equation of polynomials when considering \( \lambda \) as an indeterminate. Thus for \( \lambda \) not an integer between 0 and \( t - 1 \), we have

\[
\chi(G; \lambda - t) = \frac{1}{\lambda(\lambda - 1) \cdots (\lambda - t + 1)}\chi(G + K_t; \lambda).
\]

If \( \lambda \geq 3 \) is an integer, then by setting \( t = \lambda - 3 \), we can directly solve for \( \chi(G; 3) \) via (1.10). Otherwise, \( \lambda \notin \mathbb{N} \), since \( \lambda \neq 0, 1, 2 \). Then choosing \( 0 \leq t \leq n \), using (1.10), we can compute \( \chi(G; \lambda - t) \) by evaluating \( \chi(G + K_t; \lambda) \) at the given \( \lambda \) for a sequence of graphs \( G + K_t \). From these evaluations, we can interpolate the coefficients of \( \chi(G; \lambda) \) and evaluate it at \( \lambda = 3 \). \( \square \)

The proofs of both Lemma 1.10 and Lemma 1.12 involve interpolation of a single variable polynomial. Some reductions between counting problems are accomplished via interpolation of multivariate polynomials. An early example of this occurs in [?, Main Theorem, Case 1], which interpolates a homogeneous polynomial in three variables.

**Lemma 1.13.** \( \#\text{VertexCover} \leq_T \#\text{BipartiteVertexCover} \)

![Diagram](image)

Figure 1.8: Example construction from the proof of Lemma 1.13 with \( \ell = 3 \).
Proof. Given a graph $G$ with $n$ vertices and $m$ edges, we create a graph $G_\ell$ for every $1 \leq \ell \leq N = \binom{m+2}{2}$ in two steps as follows. First we replace every edge of $G$ with $\ell$ parallel edges to obtain a graph $G'_\ell$; this operation is called an $\ell$-thickening on $G$. Then we replace every edge of $G'_\ell$ with a path of length 4; this operation is called a 4-stretching on $G'_\ell$. Since this stretch is by an even length 4, the resulting graph is bipartite. See Figure 1.8 for an example with $\ell = 3$.

We stratify all subsets $S$ of $V$ according to a detailed account on how many edges in $G$ are not covered by $S$, how many edges are covered by exactly one vertex in $S$ and how many are covered by two vertices in $S$. Specifically, let $c_{ijk}$, where $i, j, k \geq 0$ and $i + j + k = m$, be the number of $S \subseteq V$ such that exactly

- $i$ edges in $G$ have neither endpoint in $S$,
- $j$ edges in $G$ have exactly one endpoint in $S$, and
- $k$ edges in $G$ have both endpoints in $S$.

Now define

$$p(x, y, z) = \sum_{\substack{i+j+k=m \\ i, j, k \geq 0}} x^i y^j z^k c_{ijk}.$$  

Then it can be seen that $\#\text{VertexCover}(G) = p(0, 1, 1)$.

Let $v_1, v_2, v_3, v_4$ and $v_5$ be the 5 vertices along a path of length 4. It is a direct verification that there are the following possibilities for vertex covers according to whether exactly zero, or one, or two end points $\{v_1, v_5\}$ are included in a vertex cover of the path. Indeed

1. There are exactly two vertex covers when neither endpoint is in the vertex cover, because then both $v_2$ and $v_4$ must be in the vertex cover, and $v_3$ can be arbitrarily placed;

2. there are exactly three vertex covers when exactly one endpoint, say $v_5$, is in the vertex cover, because then $v_2$ and at least one of $v_3$ and $v_4$ must be in the vertex cover; and

3. there are exactly five vertex covers when both endpoints are in the vertex cover, because then either $v_3$ is in the vertex cover and then $v_2$ and $v_4$ can be arbitrarily placed, or $v_3$ is not in the vertex cover and then both $v_2$ and $v_4$ must be in the vertex cover.

Thus,

$$\#\text{VertexCover}(G_\ell) = p(2^\ell, 3^\ell, 5^\ell) = \sum_{\substack{i+j+k=m \\ i, j, k \geq 0}} (2^i 3^j 5^k) \ell c_{ijk}.$$  

(1.11)

This defines a Vandermonde system that has full rank if $2^i 3^j 5^k \neq 2^{i'} 3^{j'} 5^{k'}$ for all distinct pairs $(i, j, k) \neq (i', j', k')$. This is clearly valid by unique prime factorization. Therefore, we can interpolate $p(x, y, z)$ in polynomial time, and then evaluate it at $p(0, 1, 1)$ to obtain the number of vertex covers of $G$. \qed

31
In most interpolation proofs a crucial point is the following. There is a suitable stratification of the Holant sum on the original instance with at most polynomially many different types (in Lemma 1.13 it is \( N = \binom{m+2}{2} \)). This is usually defined by specifying the number of local configurations induced by a global object (in Lemma 1.13 it is specified by the tuple \( \langle i, j, k \rangle \).) Given a type of local configuration, we can find the factor associated with it in the Holant sum on the constructed instance, and this is independent of the exact distribution of these local configurations. Here the fact that we are dealing with sum-of-product computations is the salient reason.

**Remark.** Let \( P_n \) denote the path graph of length \( n \), with vertices \( v_0, v_1, \ldots, v_n \). We showed in the proof that the vertex covers of \( P_4 \) can be partitioned into three sets represented by the triple \( (2, 3, 5) \). A more general relation for \( P_n \) is the following recurrence relation

\[
\#\text{VertexCover}(P_n) = \#\text{VertexCover}(P_{n-1}) + \#\text{VertexCover}(P_{n-2}).
\]

This is the same recurrence relation as the Fibonacci numbers. The recurrence is to partition the vertex covers of \( P_n \) according to whether it contains the end point \( v_n \) or not. The derivation in the proof of Lemma 1.13 corresponds to the following partition: (1) Both \( v_0 \) and \( v_n \) are in the vertex cover in which case \( v_1 \) and \( v_{n-1} \) must be chosen and then it is the same as vertex covers for \( P_{n-4} \). (2) Exactly one of \( v_0 \) and \( v_n \) is in the vertex cover, say \( v_n \), in which case \( v_1 \) must be chosen and then it is the same as vertex covers for \( P_{n-3} \). However in the account for \( \#\text{VertexCover}(P_n) \) this case should be counted twice. (3) Both \( v_0 \) and \( v_n \) are in the vertex cover in which case it is the same as vertex covers for \( P_{n-2} \). Thus this recurrence is \( x_n = x_{n-4} + 2x_{n-3} + x_{n-2} \) which is consistent with the more succinct Fibonacci recurrence relation. However in the proof of Lemma 1.13 we need to stratify it this way because that is how the common factors \( (2^i 3^j 5^k)^\ell \) can be extracted in the expression \( p(2^\ell, 3^\ell, 5^\ell) \).

For homogeneous polynomials of degree \( d \) in \( n \) variables, the number of monomials is \( \binom{d+n-1}{n-1} \) and thus the minimum number of points that can interpolate such polynomials is \( \binom{d+n-1}{n-1} \), which is polynomial in \( n \) for a fixed \( d \). However often the matrix may not be of full rank. In those case we will have to work harder to prove that interpolation succeeds.

### 1.4 Further Discussion on Models of Computation

In fact, for the computational problems we will consider in this book, every single problem only involves a fixed number of complex numbers as given parameters \( \alpha_1, \alpha_2, \ldots, \alpha_k \), and all other complex numbers involved in the problem belong to the algebraic extension field \( F = \mathbb{Q}(\alpha_1, \alpha_2, \ldots, \alpha_k) \). Thus these parameters \( \alpha_1, \alpha_2, \ldots, \alpha_k \) are part of the specification of the problem, and should not be confused with the computational complexity which is measured by the size of the problem instance.
Because these parameters are part of the problem specification, one may take the following formalistic view. It is known that every finite extension field \( F \) over the rational numbers \( \mathbb{Q} \) has the following structural form \([7]\): \( F \) is a finite algebraic extension \( E(\beta_1, \beta_2, \ldots, \beta_\ell) \) of a certain purely transcendental extension field \( E \) over \( \mathbb{Q} \), having the form \( E = \mathbb{Q}(X_1, \ldots, X_m) \) where \( X_1, \ldots, X_m \) are algebraically independent. \( F \) is said to have a finite transcendence degree \( m \) over \( \mathbb{Q} \). It is known that \( F \) has a unique and well-defined transcendence degree over \( \mathbb{Q} \). Then the formalistic view demands that \( F \) be specified as such (as part of the task of the one who specifies the problem). The \( \beta_j \)'s are specified by irreducible polynomials over \( E \) as usual.

The advantage of this totally general view is that we can then consider the theorems proved in this book logically cover all \( \mathbb{C} \), and not restricted to algebraic numbers. Thus, numbers such as \( e \) or \( \pi \) need not be excluded on the account of models of computation. However, in realistic terms, this approach is not advisable. There is too much unknown about transcendental numbers; e.g., it is still unknown whether \( e + \pi \) or \( e\pi \) are rational, algebraic irrational or transcendental, and it is open whether \( \mathbb{Q}(e, \pi) \) has transcendence degree 2 (or 1) over \( \mathbb{Q} \), i.e., whether \( e \) and \( \pi \) are algebraically independent. By placing the foundation of this theory concerning models of computation on the existence of transcendence degree (which is proved non-constructively), while it may appear to be elegant and sweeping, will obscure the real issue of computational complexity to be addressed in this book. Thus we choose to formally restrict ourselves to algebraic numbers.

However, due to the mathematical inner connections and the proof techniques to be developed in this book, it is inadvisable to restrict ourselves to, say, integers, or \( \{0, 1\} \) entries. By allowing complex numbers we discover structures that would be invisible if we were to artificially restrict ourselves to integers, or even real numbers. Moreover, the scope of algebraic numbers is adequate to develop this theory. This is the reason we choose algebraic numbers for our models of computation.

Because our computations mostly concern with algebraic operations such as sum and product, the consideration of models of computation appears to be not crucial. For brevity and convenience, we will simply use \( \mathbb{C} \) to refer to complex algebraic numbers in \( \mathbb{C} \), and simply call them complex numbers.

### 1.5 An Outline of this Book

Here we give an outline of this book.

The book is divided into two parts. Part I of this book is about the dichotomy theory on Boolean domain problems. A characteristic feature for the theory on Boolean domain problems is that we can prove quite explicit and highly effective complexity dichotomies, in all three frameworks of spin systems, \#CSP, and Holant problems. Another important feature is that we can give quite a complete classification on problems which are \#P-hard in general but polynomial time tractable over planar instances. Part II of the book will deal...
with counting problems over general domains. Two main theorems that will be presented are the complexity dichotomy for graph homomorphisms over $\mathbb{C}$ and for $\#\text{CSP}$ for arbitrary constraint functions taking values in $\mathbb{C}$.

For Part I, after this introductory Chapter 1, we commence the development of our theory in Chapter 2, where we introduce Holant* problems, a restricted class of counting problems in the Holant framework where all unary (or equivalently all degenerate) functions are assumed to be freely available. We will encounter Fibonacci gates, and show that they constitute a main part of the tractable signatures in Holant* problems. In later chapters it will be seen that Fibonacci gates mainly correspond to what is transformable to the so-called product type constraint functions. The main theorem is a complexity dichotomy for Holant* problems for any set of symmetric signatures. One reason we start the presentation of this theory with Holant* problems is that here the theory is relatively simple, very concrete, and yet non-trivial. Also we can illustrate many proof ideas, such as holographic transformations in a simple setting, which will be further developed in later chapters. Another reason is that to be able to obtain all unary functions, either by direct construction or by simulation, is a main theme of proving $\#P$-hardness in later chapters.

In Chapter 3, we consider a second framework, $\#\text{CSP}$ over Boolean domains, and prove a complexity dichotomy for $\#\text{CSP}$ for any set of constraint functions, not necessarily symmetric. Here we introduce affine signatures. Together with signatures of product type, these constitute the tractable constraints for Boolean $\#\text{CSP}$. The important proof technique—interpolation—is employed here. The dichotomy of Chapter 3 is the special case of a much more general dichotomy of $\#\text{CSP}$ (in Part II) restricted to the Boolean domain. Some preparatory theorems can also be derived in a more general setting. However we choose to give a proof in the Boolean domain independent from the general theory because this proof is elementary, and the resulting dichotomy is explicit. The idea of simulating unary signatures in $\#P$-hardness proofs also appears here, which utilizes the dichotomy from Chapter 2 both explicitly as well as in spirit. It is proved that in Boolean $\#\text{CSP}$, one can always obtain two special unary functions, which will define a weaker type of Holant* problems, and will be called Holant* problems.

Chapter 4 starts with the theory of Pfaffian orientations and Kasteleyn’s algorithm (a.k.a. the FKT algorithm) which can count perfect matchings on planar graphs. Then the chapter gives a thorough treatment of the theory of matchgates. It is with matchgates that holographic reductions were first introduced, although in this book we use it already to prove dichotomy theorems in Chapter 2. These matchgates, and together with holographic transformations, are important ingredients in those counting problems that are $\#P$-hard in general but computable in polynomial time for planar graphs. The theory for general matchgates depends on matchgate identities. But for the remainder of Part I all dichotomy theorems concerning matchgates only use symmetric signatures; for these it is logically possible to skip matchgate identities and only use Theorem 4.11. One can derive the form of symmetric matchgate signatures as given in Theorem 4.11, by the necessary parity constraints and the explicit construction in subsection 4.3.3.
Chapter 5 proves a dichotomy for spin systems on regular graphs. A number of proof techniques are introduced in this chapter, which are summarized at the beginning of the chapter. This chapter proves the dichotomy for the case of $k$-regular graphs, for $k = 3$ and $k = 4$. The more general case for all $k$ is also true. However for the purpose of the main dichotomy for Holant problems, we will only need the case $k = 3$ and $4$. This chapter consciously develops the material in a parallel fashion, by introducing technical tools gradually throughout while moving toward the proof of the desired dichotomy theorem. The statement of this dichotomy for regular 3-regular and 4-regular graphs will be a key result for later dichotomies.

Chapter 6 is a transitional chapter. It serves many purposes. It connects Holant problems and #CSP. It also introduces the transformational perspective of looking at dichotomy theorems so that now not only holographic transformations are used as proof techniques, but also they become the language in which the statement of a dichotomy theorem is expressed. In doing so, naturally we introduce group actions. This chapter also proves a dichotomy for Holant problems and for CSP; it also proves Eulerian Orientation is #P-hard.

Chapter 7 proves a main dichotomy theorem for Part I, towards which much of the material in previous chapters are preparatory. The dichotomy is proved for any set of symmetric signatures on Boolean domain without assuming any auxiliary functions. An important new ingredient is the isolation of the vanishing signatures.

Chapter 8 proves a dichotomy for planar #CSP for any set of symmetric signatures on Boolean domain. It shows that for the #CSP framework, a holographic reduction to Kasteleyn’s algorithm, implemented by matchgates, is a universal strategy for those problems which are #P-hard but tractable over planar graphs. It has been proved most recently that for Holant problems this universality is not true—there are additional planar tractable signatures sets.

Most of the dichotomy theorems presented so far are for symmetric function sets, except for #CSP in Chapter 3. In the final chapter of Part I, Chapter 7, we revisit Holant problems of Chapter 2, and prove a Holant* dichotomy for all signatures sets that are not necessarily symmetric, over the Boolean domain. In doing so, we also gain some new perspective on the nature of the tractable signatures sets.

Part II presents the theory on the general domain.
Chapter 2

Fibonacci Gates and Holant* Problems

We start with a special family of signatures called Fibonacci gates. They will be an important building block for tractable Holant problems.

2.1 Fibonacci Gates

Consider the signature $f = [1, 0, 1, 1]$. What is the complexity of Holant($f$)? It is plain that $f$ satisfies $f_{k+2} = f_{k+1} + f_k$, for $0 \leq k \leq 1$. This turns out to be significant.

Definition 2.1. For any $n \geq 1$, a signature $f = [f_0, f_1, \ldots, f_n]$ is a Fibonacci gate if

$$f_{k+2} = f_{k+1} + f_k, \quad 0 \leq k \leq n - 2.$$ 

A set of signatures $\mathcal{F}$ is called Fibonacci if every signature in $\mathcal{F}$ is a Fibonacci gate.

Each Fibonacci gate $f$ has two degrees of freedom, as $f_0$ and $f_1$ are arbitrary; the rest of $f$ are determined by the linear recurrence. In particular any unary signature is a Fibonacci gate. As with the ordinary Fibonacci sequence, the linear recurrence satisfied by $f$ gives the following formula

$$f_k = c_1 \phi^k + c_2 \bar{\phi}^k, \quad 0 \leq k \leq n, \quad (2.1)$$

where $\phi$ is the golden ratio $\frac{1 + \sqrt{5}}{2}$ and $\bar{\phi}$ is its conjugate $\frac{1 - \sqrt{5}}{2}$, both roots of the quadratic polynomial $X^2 - X - 1$. To match the initial values $f_0$ and $f_1$, we have $n = 3$, $c_1 = (-\bar{\phi}/\sqrt{5})f_0 + (1/\sqrt{5})f_1$ and $c_2 = (\phi/\sqrt{5})f_0 + (-1/\sqrt{5})f_1$. For example, for $f = [1, 0, 1, 1]$, we have $c_1 = -\bar{\phi}/\sqrt{5}$ and $c_2 = \phi/\sqrt{5}$.

For $f = [1, 0, 1, 1]$, equation $(2.1)$ can be written as a sum of tensor products

$$f = c_1 \left[ \frac{1}{\phi} \right]^\otimes 3 + c_2 \left[ \frac{1}{\bar{\phi}} \right]^\otimes 3. \quad (2.2)$$
This lists the values of $f$ as a column vector indexed by its input assignments in lexicographic order.

Consider the matrix $M = \begin{bmatrix} 1 & 1 \\ \phi & \phi \end{bmatrix}$. We have $M \otimes 3 \left( c_1 [1] \otimes 3 + c_2 [1] \otimes 3 \right) = f$. This suggests that we take a holographic transformation by the matrix $M$.

Recall that holographic transformations are defined in subsection 1.3.2 of Chapter 1. By Theorem 1.5, for a bipartite Holant problem $\text{Holant}(\mathcal{F} | \mathcal{G})$, the holographic transformation does not change the value

$$\text{Holant}(\Omega; \mathcal{F} | \mathcal{G}) = \text{Holant}(\Omega'; \mathcal{F} M | M^{-1} \mathcal{G}),$$

where $\Omega'$ is the transformed signature grid.

Let’s review how we can treat a problem such as $\text{Holant}(f)$ as a bipartite Holant problem, and in concrete terms what Theorem 1.5 says in this case? We can view $\text{Holant}(f)$ as computing the following tensor contraction operation. Given a 3-regular graph $\Gamma$ where each vertex is labeled by $f$, first we replace each edge $e$ by a path of length two, and label the new vertex by the binary equality function $(\text{=}_2) = [1, 0, 1]$. This transforms the Holant problem $\text{Holant}(f)$ to a bipartite Holant problem $\text{Holant}(\text{=}_2 f)$ on the Edge-Vertex incidence graph without changing its value on $\Gamma$. Next we form the tensor products $\mathcal{R}$ of all left-hand side copies of $(\text{=}_2)$, and $\mathcal{G}$ of all right-hand side copies of $f$, respectively. If $\Gamma$ has $m$ edges, these two tensors $\mathcal{R}$ and $\mathcal{G}$ both have dimension $2^{2m}$, each representing a function of arity $2m$, and there is a one-to-one correspondence between the variables according to the incidence relation in $\Gamma$. The contraction operation $\langle \mathcal{R}, \mathcal{G} \rangle$ is

$$\sum_{x_1, x_2, \ldots, x_{2m} = 0,1} \mathcal{R}(x_1, x_2, \ldots, x_{2m}) \mathcal{G}(x_1, x_2, \ldots, x_{2m}).$$

If we view $\mathcal{R}$ as a (row) vector of dimension $2^{2m}$ and $\mathcal{G}$ as a (column) vector of the same dimension, then the contraction $\langle \mathcal{R}, \mathcal{G} \rangle$ is the dot product of the two vectors.

Now we consider a transformation $(M^{-1}) \otimes 2m$ on $\mathcal{G}$ and $M \otimes 2m$ on $\mathcal{R}$, namely

$$\mathcal{G} \mapsto \mathcal{G}' = (M^{-1}) \otimes 2m \mathcal{G}, \quad \text{and} \quad \mathcal{R} \mapsto \mathcal{R}' = \mathcal{R} M \otimes 2m.$$

(2.3)

It is clear that

$$\langle \mathcal{R}', \mathcal{G}' \rangle = \langle \mathcal{R}, \mathcal{G} \rangle.$$

The transformation in (2.3) can be carried out by distributing $M$ and $M^{-1}$ to each left-hand side copy of $(\text{=}_2)$ and right-hand side copy of $f$. Thus we can consider $(\text{=}_2) \mapsto (\text{=}_2)M \otimes 2$, where $(\text{=}_2)$ is written as a row vector $(1, 0, 0, 1)$, and $f \mapsto (M^{-1}) \otimes 3 f$, where $f$ is given in (2.2). It follows that

$$(M^{-1}) \otimes 3 f = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes 3 + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes 3,$$

and in symmetric signature notation this is $(M^{-1}) \otimes 3 f = [c_1, 0, 0, c_2]$. Note that signatures of the form $[* , 0, \ldots, 0, *]$ force all incident edges to take the same value 0 or 1. They are called Generalized Equalities, or Gen-Eq for short.
Recall that if we write a binary signature \( f \) as a matrix \( F \), with entry \( f(i,j) \) at row \( i \) and column \( j \) (where \( i,j = 0,1 \)), then \( MFM^T \) is the matrix form of \( M \otimes^2 f \), where we write \( f \) as a column vector with entries ordered lexicographically. Similarly \( M^TFM \) is the matrix form of \( fM \otimes^2 \) where we write \( f \) as a row vector. It is clear that the identity matrix \( I \) is the matrix form of \( (=2) \). It follows that for \( M = \begin{bmatrix} 1 & 1 \\ \phi & \phi \end{bmatrix} \), \( (=2)f \) takes the form \( M^Tf = \begin{bmatrix} 1+\phi^2 & 0 \\ 0 & 1+\overline{\phi}^2 \end{bmatrix} \) which is \([1 + \phi^2, 0, 1 + \overline{\phi}^2]\) in symmetric signature notation. Thus both sides of the bipartite Holant problem \( \text{Holant}( (=2) M \otimes^2 | (M^{-1}) \otimes^3 f) \) are GEN-Eq.

This means that the Holant value of \( \text{Holant}( (=2) f) \), which is the same as the Holant value of \( \text{Holant}( (=2) M \otimes^2 | (M^{-1}) \otimes^3 f) \), can be easily computed: On each connected component of the given graph \( \Gamma \), any assignment on a single edge can be uniquely propagated to the entire component. The Holant value on \( \Gamma \) is the product over all connected components.

This can be generalized to any set \( F \) of Fibonacci gates. If \( f = [f_0, f_1, \ldots, f_n] \in F \) is a Fibonacci gate, then equation (2.1) can be written as

\[
\begin{align*}
f &= c_1 \begin{bmatrix} 1 \\ \phi \end{bmatrix}^n + c_2 \begin{bmatrix} 1 \\ \overline{\phi} \end{bmatrix}^n.
\end{align*}
\]

Thus, the same holographic transformation \( f \mapsto (M^{-1}) \otimes^nf \) by \( M \) transforms \( f \) to a GEN-Eq. This shows that \( \text{Holant}(F) \) is tractable.

**Exercise:** Show that there is an orthogonal matrix \( M \in \mathbb{C}^{2 \times 2} \) that transforms every signature of the form (2.4) to a GEN-Eq. Use this fact to prove that \( \text{Holant}(F) \) is tractable for a set of Fibonacci gates \( F \).

There is an alternative proof of the tractability of \( \text{Holant}(F) \) for any finite set of Fibonacci gates \( F \). We turn to this approach next.

**Definition 2.2.** For any \( n \geq 1 \), and a parameter \( \lambda \in \mathbb{C} \), a signature \( f = [f_0, f_1, \ldots, f_n] \) is a generalized Fibonacci gate (with parameter \( \lambda \)) if

\[
f_{k+2} = \lambda f_{k+1} + f_k, \quad 0 \leq k \leq n - 2.
\]

A set of signatures \( F \) is called generalized Fibonacci if for some \( \lambda \in \mathbb{C} \), every signature in \( F \) is a generalized Fibonacci gate with parameter \( \lambda \).

**Theorem 2.3.** For any finite set of generalized Fibonacci gates \( F \), the Holant problem \( \text{Holant}(F) \) is computable in polynomial time.

**Proof.** If \( \Gamma_1, \Gamma_2, \ldots, \Gamma_\kappa \) are the connected components of a graph \( \Gamma \), then

\[
\text{Holant}_\Gamma = \prod_{j=1}^\kappa \text{Holant}_{\Gamma_j}.
\]
So we only need to consider connected graphs as inputs.

Suppose $\Gamma$ has $n$ nodes and $m$ edges. First we cut all the edges in $\Gamma$. A node with degree $d$ can be viewed as an $F$-gate with $d$ dangling edges. Now step by step we merge two dangling edges into one regular edge in the original graph, until we recover $\Gamma$ after $m$ steps. We prove that all the intermediate $F$-gates still have generalized Fibonacci signatures with the same parameter $\lambda$, and at every step we can compute the intermediate signature in polynomial time. After $m$ steps we get $\Gamma$ as an $F$-gate with no dangling edge, the only value of its signature is the Holant value we want. To carry this out, we only need to prove that it is true for one single step. There are two cases, depending on whether the two dangling edges to be merged are in the same component or not. These two operations are illustrated in Fig. 2.1 and Fig. 2.2.

![Figure 2.1: First operation.](image1)

![Figure 2.2: Second operation.](image2)

In the first case, the two dangling edges belong to two components before their merging (Figure 2.1). Let $F$ have dangling edges $y_1, \ldots, y_s, z$ and $G$ have dangling edges $y_{s+1}, \ldots, y_{s+t}, z'$. After merging $z$ with $z'$, we have a new $F$-gate $H$ with dangling edges $y_1, \ldots, y_{s+t}$. Inductively the signatures of $F$ and $G$ are both generalized Fibonacci gates with the same parameter $\lambda$. We show that this remains so for the resulting $F$-gate $H$.

We first prove that $H$ is symmetric. We only need to show that the value of $H$ is unchanged if the values of two inputs are exchanged. Because $F$ and $G$ are symmetric, if both inputs are from $\{y_1, \ldots, y_s\}$ or from $\{y_{s+1}, \ldots, y_{s+t}\}$, the value of $H$ is clearly unchanged. Suppose one input is from $\{y_1, \ldots, y_s\}$ and the other is from $\{y_{s+1}, \ldots, y_{s+t}\}$. By the symmetry of $F$ and $G$ we may assume these two inputs are $y_1$ and $y_{s+1}$. Thus we will fix an arbitrary as-
segment for \(y_2, \ldots, y_s, y_{s+2}, \ldots, y_{s+t}\), and we want to show \(H(0, y_2, \ldots, y_s, 1, y_{s+2}, \ldots, y_{s+t}) = H(1, y_2, \ldots, y_s, 0, y_{s+2}, \ldots, y_{s+t})\).

We will suppress the fixed values \(y_2, \ldots, y_s, y_{s+2}, \ldots, y_{s+t}\) and denote

\[
\begin{align*}
F_{xz} &= F(x, y_2, \ldots, y_s, z), \\
G_{yz} &= G(y, y_{s+2}, \ldots, y_{s+t}, z), \quad \text{and} \\
H_{xy} &= H(x, y_2, \ldots, y_s, y_{s+2}, \ldots, y_{s+t}).
\end{align*}
\]

Then by the definition of Holant, \(H_{ab} = F_{a0}G_{b0} + F_{a1}G_{b1}\), for \(a, b \in \{0, 1\}\).

Because \(F\) and \(G\) are generalized Fibonacci gates with parameter \(\lambda\), we have

\[
H_{01} = F_{00}G_{10} + F_{01}(\lambda G_{01} + G_{00}), \quad \text{and} \quad H_{10} = F_{10}G_{00} + (\lambda F_{01} + F_{00})G_{01}
\]

By the symmetry of \(F\) and \(G\), we have \(H_{01} = H_{10}\).

Now we show that \(H(y_1, \ldots, y_{s+t})\) is also a generalized Fibonacci gate with parameter \(\lambda\). Since we have proved that \(H\) is symmetric, we can choose any two input variables to prove it being Fibonacci. Again, we choose \(y_1\) and \(y_{s+1}\). (This assumes that \(y_1\) and \(y_{s+1}\) exist, i.e., \(F\) and \(G\) are not unary functions. If either one of them is unary, the proof is just as easy.) For any fixed values of all other variables, we have \(H_{00} = F_{00}G_{00} + F_{01}G_{01}\), \(H_{01} = F_{00}G_{10} + F_{01}G_{11}\), and \(H_{11} = F_{10}G_{10} + F_{11}G_{11}\). Now using the fact that both \(F\) and \(G\) are generalized Fibonacci gates with parameter \(\lambda\), it follows that \(H_{11} = \lambda H_{01} + H_{00}\).

If the first two terms of the signatures of \(F\) and \(G\) are \(f_0, f_1\) and \(g_0, g_1\) respectively, then the first two terms of the signature \(H\) are \(h_0 = f_0g_0 + f_1g_1\) and \(h_1 = f_0g_1 + f_1g_0 + \lambda f_1g_1\).

Next we consider the second case, where the two dangling edges to be merged are in the same component (Fig. 2.2). Obviously, the signature for the new gate \(H\) is also symmetric. If \(F = [f_0, f_1, \ldots, f_n]\) is the signature before the merging operation, then the signature after the merging operation is \(H = [f_0 + f_2, f_1 + f_3, \ldots, f_{n-2} + f_n]\). Such an operation preserves the linear recurrence. It follows that \(H\) is also a generalized Fibonacci gate with parameter \(\lambda\), and the signature has already been computed.

We remark that the proof of tractability in Theorem 2.3 is also valid for an infinite set \(\mathcal{F}\) of generalized Fibonacci gates with the same parameter \(\lambda\); however for an infinite set \(\mathcal{F}\) the input size should include the description of the signatures used in the signature grid. See Remark 1 after Theorem 2.12.

### 2.2 Orthogonal Transformation of Fibonacci Gates

Fibonacci gates by Definition 2.1 correspond to the case \(\lambda = 1\) in Definition 2.2. How much generality does an arbitrary parameter \(\lambda\) provide? Can we recapture the tractability of generalized Fibonacci gates by holographic transformation as we did for \(\lambda = 1\)?
The recurrence relation \( f_{k+2} = \lambda f_{k+1} + f_k \) has the characteristic polynomial \( X^2 - \lambda X - 1 \). If \( \lambda \neq \pm 2i \) then there are two distinct eigenvalues \( \alpha, \beta \neq \pm i \), and a general solution to the recurrence has the form \( f_k = a\alpha^k + b\beta^k \), for some constants \( a \) and \( b \) determined by \( f_0 \) and \( f_1 \). Note that \( \alpha\beta = -1 \). More revealing is the following expression in terms of tensor products

\[
f = a \left[ \begin{array}{c} 1 \\ \alpha \end{array} \right] \otimes^n + b \left[ \begin{array}{c} 1 \\ \beta \end{array} \right] \otimes^n.
\]

(2.6)

Note that the two vectors \((1, \alpha)^T\) and \((1, \beta)^T\) are orthogonal, i.e., their dot product is zero, and \(1 + \alpha^2, 1 + \beta^2 \neq 0\). We can scale the entries as follows. Let

\[
T = \begin{bmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{1+\alpha^2}} & \frac{1}{\sqrt{1+\beta^2}} \\ \frac{\alpha}{\sqrt{1+\alpha^2}} & \frac{\beta}{\sqrt{1+\beta^2}} \end{bmatrix}.
\]

Then \( T \) is an orthogonal matrix. Let \( a' = a(1 + \alpha^2)^{n/2} \) and \( b' = b(1 + \beta^2)^{n/2} \), then we have

\[
f = a' \left[ \begin{array}{c} t_{00} \\ t_{10} \end{array} \right] \otimes^n + b' \left[ \begin{array}{c} t_{01} \\ t_{11} \end{array} \right] \otimes^n = T \otimes^n \left( a' \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \otimes^n + b' \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] \otimes^n \right).
\]

(2.7)

Hence the holographic transformation \((T^{-1}) \otimes^n f\) is the GEN-EQ signature \( a' \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \otimes^n + b' \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] \otimes^n = [a', 0, \ldots, 0, b']\). Meanwhile the holographic transformation on the left-hand side keeps the binary EQUALITY \((=2)\) unchanged by \( T \):

\[
(=2) \leftrightarrow (=2)T \otimes^2 = (=2),
\]

(2.8)

because \( T \) is an orthogonal matrix, and in matrix form \((=2)T \otimes^2\) is \(T^2 IT = I\).

Thus for any finite set of generalized Fibonacci gates \( \mathcal{F} \) with parameter \( \lambda \neq \pm 2i \), the Holant problem Holant(\( \mathcal{F} \)), being equivalent to Holant\((=2|\mathcal{F})\), is tractable. This is an alternative proof of Theorem 2.3, when \( \lambda \neq \pm 2i \).

By a limiting argument, one can give a proof of tractability by this holographic approach when \( \lambda = \pm 2i \) as well. However we will refrain from that, because intrinsically there is a difference for the signatures when \( \lambda = \pm 2i \).

Every generalized Fibonacci gate with \( \lambda \neq \pm 2i \) can be transformed to a GEN-EQ by some orthogonal matrix. By composition, for any \( \lambda, \mu \neq \pm 2i \), any generalized Fibonacci gate with parameter \( \lambda \) can also be transformed by some orthogonal matrix to another generalized Fibonacci gate with parameter \( \mu \). We can take the basic Fibonacci gate with \( \lambda = 1 \), as well as GEN-EQ, as a normal form for this class. The orthogonal transformation between these two is the normalized version of the matrix \( M \) from Section 2.1:

\[
\begin{bmatrix} \frac{1}{\sqrt{1+\phi^2}} & \frac{1}{\sqrt{1+\phi^2}} \\ \frac{\phi}{\sqrt{1+\phi^2}} & \frac{1}{\sqrt{1+\phi^2}} \end{bmatrix}.
\]

An invertible holographic transformation maps non-degenerate signatures to non-degenerate signatures. This is because if \( fM \otimes^n = (a_1, b_1) \otimes \cdots \otimes (a_n, b_n) \), then \( f = (a'_1, b'_1) \otimes \cdots \otimes (a'_n, b'_n) \).
where \((a'_k, b'_k) = (a_k, b_k)M^{-1}\). The same argument works for contravariant tensors. Suppose \(f\) is a generalized Fibonacci gate satisfying (2.6). The following are equivalent: (1) \(f\) is non-degenerate, (2) \(ab \neq 0\) in (2.6), (3) its GEN-EQ form \([a', 0, \ldots, 0, b']\) in (2.7) is non-degenerate, and (4) \(a'b' \neq 0\).

Define

\[ \mathcal{F} = \{ f \mid f \text{ satisfies (2.5) for some } \lambda \neq \pm 2i \} \cup \text{GEN-EQ.} \]  

(2.9)

**Theorem 2.4.** For any \(f \in \mathcal{F}\) in (2.9),

1. There exists an orthogonal \(T\) such that \(Tf\) is a GEN-EQ.

2. There exists an orthogonal \(T\) such that \(Tf\) is a Fibonacci gate satisfying Definition 2.1.

3. For all orthogonal \(T\), \(Tf \in \mathcal{F}\).

**Proof.** We only need to prove the last item. By composition and the invertibility of the transformation, we may assume \(f\) is a GEN-EQ function \([a', 0, \ldots, 0, b']\). If \(T = \begin{bmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{bmatrix}\), then

\[ Tf = a' \begin{bmatrix} t_{00} \\ t_{10} \end{bmatrix} \otimes^n + b' \begin{bmatrix} t_{01} \\ t_{11} \end{bmatrix} \otimes^n. \]

If any entry \(t_{ij} = 0\), then by orthogonality \(t_{ij} = 0\) where \(x = 1 - x\), and \(Tf\) is a GEN-EQ. Suppose \(T\) has no zero entries, then \(Tf\) satisfies a recurrence relation with the characteristic roots \(\lambda_1 = t_{10}/t_{00}\) and \(\lambda_2 = t_{11}/t_{01}\). We have \(\lambda_1 \neq \lambda_2\) and \(\lambda_1\lambda_2 = -1\), by the orthogonality of \(T\). Hence the characteristic polynomial of the recurrence is \(X^2 - (\lambda_1 + \lambda_2)X - 1\). Since \(\lambda_1 \neq \lambda_2\) we have \(\lambda_1 + \lambda_2 \neq \pm 2i\). Thus \(Tf \in \mathcal{F}\). \(\square\)

We may consider GEN-EQ gates provide a normal form for generalized Fibonacci gates (with \(\lambda \neq \pm 2i\)) under orthogonal transformations as in (2.7). Curiously, any GEN-EQ of arity at least 3 does not satisfy any recurrence \(f_{k+2} = \lambda f_{k+1} + f_k\). (We can generalize the form of a second order recurrence to \(af_k + bf_{k+1} - af_{k+2} = 0\) for some \((a, b) \neq 0\). Then GEN-EQ does satisfy the recurrence with \((a, b) = (0, 1)\).) The proof of Theorem 2.4 also establishes the following: Other than GEN-EQ itself, any signature \(f\) expressible as (2.7), i.e., an orthogonal transformation of GEN-EQ, is in fact a generalized Fibonacci gate.

**Exercise:** Prove that generalized Fibonacci gates with parameter \(\lambda = \pm 2i\) cannot be transformed to any member of \(\mathcal{F}\) in (2.9).

We now ask what is the right perspective for generalized Fibonacci gates with parameter \(\lambda = \pm 2i\) under holographic transformations?

Let \(\mathcal{F}\) be such a set of signatures. To show Holant(\(\mathcal{F}\)) is tractable, we first get rid of degenerate signatures as follows. Each degenerate signature can be replaced by a set of unary signatures. Then combine a unary signature with any \(f \in \mathcal{F}\) satisfying (2.5) produces
another signature $f'$ with $\text{arity}(f') = \text{arity}(f) - 1$, which also satisfies (2.5) with the same $\lambda$. So we may assume every signature in $F$ is non-degenerate and has arity at least 2.

If every vertex in the given instance graph $H$ has degree at most 2, then the Holant value $\text{Holant}(F)$ can be computed as a matrix product and then taking the trace, per each connected component.

Each $f \in F$ has the form $f_k = c_k \mu^{k-1} + d_k \mu^k$ ($0 \leq k \leq n$ and $\mu = \pm i$), where $X^2 - 2\mu X - 1$ has double root $\mu$. We will suppose $\mu = i$. The case $\mu = -i$ is similar. Consider

$$f' = [c', d', 0, \ldots, 0] = c' \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \otimes^n + d' \sum_{j=1}^n \left\{ \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \otimes^{(j-1)} \otimes \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] \otimes \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \otimes^{(n-j)} \right\},$$

(2.10)

where we set $c' = d - cni/2$ and $d' = ci/2$. Let

$$Z = \left[ \begin{array}{cc} 1 & 1 \\ i & -i \end{array} \right].$$

(2.11)

We apply $Z \otimes^n$ to $f'$ by distributing each $Z$ among the factors of the tensor product in (2.10)

$$Z \otimes^n f' = c' \left[ \begin{array}{c} 1 \\ i \end{array} \right] \otimes^n + d' \sum_{j=1}^n \left\{ \left[ \begin{array}{c} 1 \\ i \end{array} \right] \otimes^{(j-1)} \otimes \left[ \begin{array}{c} 1 \\ i \end{array} \right] \otimes^{(n-j)} \right\}. $$

(2.12)

Then $Z \otimes^n f' = f$. To wit, any entry of $Z \otimes^n f'$ indexed by a bit pattern in $\{0, 1\}^n$ of Hamming weight $k$ is $c' i^k + d' [k(-i)i^{k-1} + (n-k)i^k] = cki^{k-1} + di^k$. Thus the contravariant transformation $(Z^{-1}) \otimes^n f = f'$.

Theorem 2.5. A symmetric signature $[f_0, f_1, \ldots, f_n]$ can be transformed by some invertible holographic transformation to a Fibonacci gate according to Definition 2.1 (equivalently to a
signature in \( \mathcal{F} \) defined in (2.9)) iff there exist three constants \( a, b \) and \( c \), such that \( b^2 - 4ac \neq 0 \), and for all \( 0 \leq k \leq n - 2 \),

\[
a f_k + b f_{k+1} + c f_{k+2} = 0.
\]

(2.13)

Proof. For an invertible \( M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \), we have

\[
M^{\otimes n}[u, 0, \ldots, 0, v] = u \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes n} + v \begin{bmatrix} \gamma \\ \delta \end{bmatrix}^{\otimes n}.
\]

If \( f \) has this form, then \( f \) satisfies a recurrence relation \( a f_k + b f_{k+1} + c f_{k+2} = 0 \) \((0 \leq k \leq n - 2)\), where \( a, b \) and \( c \) satisfy \( b^2 - 4ac \neq 0 \). In particular they are not all \( 0 \). It can be directly verified that the following choice works: \( a = \beta \delta, b = -(\alpha \delta + \beta \gamma) \) and \( c = \alpha \gamma \).

Conversely, from \( a f_k + b f_{k+1} + c f_{k+2} = 0 \), if \( c \neq 0 \) we have a second order linear recurrence with distinct eigenvalues \( \lambda \) and \( \mu \), which gives us the expression

\[
f = u \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes n} + v \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes n} = M[u, 0, \ldots, 0, v].
\]

where \( M = \begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix} \). If \( a \neq 0 \), then we have a reverse linear recurrence. If \( a = c = 0 \), then \( b \neq 0 \), and we have a GEN-EQ.

\[\square\]

Theorem 2.6. Holant(\( \{R_1, R_2, \ldots, R_t\} \mid \{G_1, G_2, \ldots, G_s\} \)) is tractable, where \( G_i \) and \( R_j \) have arities \( n_i \) and \( m_j \) respectively, if there exists an invertible \( M \) such that \( M^{-1}G_i \) and \( R_j M \) are Fibonacci gates. This happens if and only if \( a, b \) and \( c \), such that \( b^2 - 4ac \neq 0 \) and the following two conditions are satisfied:

1. For any \( G_i = [x^{(i)}_1, x^{(i)}_2, \ldots, x^{(i)}_{n_i}] \) and any \( 0 \leq k \leq n_i - 2 \), \( ax^{(i)}_k + bx^{(i)}_{k+1} + cx^{(i)}_{k+2} = 0 \).
2. For any \( R_j = [y^{(j)}_1, y^{(j)}_2, \ldots, y^{(j)}_{m_j}] \) and any \( 0 \leq k \leq m_j - 2 \), \( cy^{(j)}_k - by^{(j)}_{k+1} + ay^{(j)}_{k+2} = 0 \).

Proof. We may take the normal form GEN-EQ, and assume \( (M^{-1})^{\otimes n_i} G_i = [*, 0, \ldots, 0, *] \) and \( R_j M^{\otimes m_j} = [*, 0, \ldots, 0, *] \), for some \( M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \). By the proof of Theorem 2.5, there exist

\[
a = \beta \delta, b = -(\alpha \delta + \beta \gamma) \quad \text{and} \quad c = \alpha \gamma,
\]

such that \( b^2 - 4ac \neq 0 \), and for all \( G_i = [x^{(i)}_1, x^{(i)}_2, \ldots, x^{(i)}_{n_i}] \), \( ax^{(i)}_k + bx^{(i)}_{k+1} + cx^{(i)}_{k+2} = 0 \) \((0 \leq k \leq n_i - 2)\). The characteristic polynomial is \( cX^2 + bX + a \).

If we write \( R_j \) in its transpose, then Theorem 2.5 applies to \( R_j \) with \( (M^{-1})^T = \frac{1}{\det(M)} \begin{bmatrix} \delta & \alpha \\ -\gamma & -\beta \end{bmatrix} \). replacing \( M \). It follows that each \( R_j \) satisfies \( cy^{(j)}_k - by^{(j)}_{k+1} + ay^{(j)}_{k+2} = 0 \) \((0 \leq k \leq m_j - 2)\).

Conversely, suppose all \( G_i \) and \( R_j \) satisfy the linear recurrences. Suppose \( c \neq 0 \). Then for each \( G_i \) the characteristic polynomial of the recurrence is \( X^2 + \frac{b}{c}X + \frac{a}{c} \). Denote by \( \Delta = b^2 - 4ac \). Let \( M = \begin{bmatrix} -\frac{b}{\sqrt{\Delta}} & -\frac{2c}{\sqrt{\Delta}} \\ -\frac{2c}{\sqrt{\Delta}} & -\frac{b + \sqrt{\Delta}}{\sqrt{\Delta}} \end{bmatrix} \). Obviously this is non-singular, and it can be verified that \( G_i = a_i \begin{bmatrix} \frac{2c}{\sqrt{\Delta}} & \frac{b + \sqrt{\Delta}}{\sqrt{\Delta}} \end{bmatrix}^{\otimes n_i} + b_i \begin{bmatrix} \frac{2c}{\sqrt{\Delta}} & \frac{b - \sqrt{\Delta}}{\sqrt{\Delta}} \end{bmatrix}^{\otimes n_i} \), for some constants \( a_i \) and \( b_i \) depending on \( G_i \). Similarly each \( R_j \) satisfies \( R_j = c_j \begin{bmatrix} \frac{2c}{\sqrt{\Delta}} & \frac{b + \sqrt{\Delta}}{\sqrt{\Delta}} \end{bmatrix}^{\otimes m_j} + d_j \begin{bmatrix} \frac{2c}{\sqrt{\Delta}} & \frac{b - \sqrt{\Delta}}{\sqrt{\Delta}} \end{bmatrix}^{\otimes m_j} \), for some \( c_j \) and \( d_j \) depending on \( R_j \). Then both \( (M^{-1})^{\otimes n_i} G_i \) and \( R_j M^{\otimes m_j} \) are GEN-EQ. The proof is similar if \( a \neq 0 \). If \( a = c = 0 \), then \( b \neq 0 \), and both \( G_i \) and \( R_j \) are already GEN-EQ. \[\square\]
2.3 A Dichotomy Theorem for Holant*(\(F\))

Every degenerate signature is a tensor product of unary signatures. If \(F\) consists of degenerate signatures only, each signature grid \(\Omega\) on Holant(\(F\)) is decomposed into a disjoint union of components, one for each vertex, and the Holant value is trivially computable. To isolate more interesting tractable families of signatures, we consider the following Holant problems where all unary functions \(U\) are assumed to be free.

**Definition 2.7.** For any signature set \(F\),

\[
\text{Holant}^*(F) = \text{Holant}(F \cup U).
\]

For bipartite graphs we also denote

\[
\text{Holant}^*(F|G) = \text{Holant}(F | U \cup G \cup U),
\]

for any signature sets \(F\) and \(G\).

We begin with three technical propositions.

**Proposition 2.8.** Let \(n \geq 3\) and let \([x_0, x_1, \ldots, x_n]\) be a non-degenerate symmetric signature. Then for any \(2 \leq m \leq n - 1\), there exists a non-degenerate sub-signature of arity \(m\), unless the signature is of the form \([x_0, 0, \ldots, 0, x_n]\).

**Proof.** Since \(\text{rank} \begin{bmatrix} x_0 & \cdots & x_{n-1} \\ x_1 & \cdots & x_n \end{bmatrix} = 2\), there must be non-zero entries among \(x_0, x_1, \ldots, x_n\).

If all entries are non-zero, then either \(\text{rank} \begin{bmatrix} x_0 & \cdots & x_{n-2} \\ x_1 & \cdots & x_{n-1} \end{bmatrix} = 2\), or \(\text{rank} \begin{bmatrix} x_1 & \cdots & x_{n-1} \\ x_2 & \cdots & x_n \end{bmatrix} = 2\). Otherwise they are both of rank 1, and being non-zero, the second row is a non-zero multiple of the first row in both matrices. Since they share at least one column, the multipliers must be the same, which says that \(\text{rank} \begin{bmatrix} x_0 & \cdots & x_{n-1} \\ x_1 & \cdots & x_n \end{bmatrix} = 1\), a contradiction. Then we use induction to complete the proof.

Now suppose there are zero entries. Assume \([x_0, \ldots, x_n]\) is not of the form \([x_0, 0, \ldots, 0, x_n]\). There exists \(1 \leq i \leq n - 1\) such that \(x_i \neq 0\). Find an \(x_i \neq 0\), for some \(1 \leq i \leq n - 1\), such that \(x_{i-1} = 0\) or \(x_{i+1} = 0\). Now any submatrix containing \(\begin{bmatrix} \cdots & x_{i-1} & x_i & x_{i+1} & \cdots \end{bmatrix}\) has rank 2.

**Definition 2.9.** A signature \([x_0, x_1, \ldots, x_n]\), where \(n \geq 2\), has type I, if there exist \(a\) and \(b\) (not both zero), such that \(ax_k + bx_{k+1} - ax_{k+2} = 0\) \((0 \leq k \leq n - 2)\). To specify the parameters \(a\) and \(b\), we say it is of type \(I(a, b)\). We say it is of type II, if \(x_k + x_{k+2} = 0\) \((0 \leq k \leq n - 2)\).

For arity at least two, type I is the union of generalized Fibonacci gates with \(\text{GEN-Eq}\). For any non-degenerate signature of arity at least three, if it is of type \(I(a, b)\), then \((a, b)\) is unique up to a scalar multiplier, i.e., this \((a, b)\) is uniquely determined as a point on the projective line. Similarly, any non-degenerate signature of arity at least three cannot be both of type I and type II. These can be directly verified.
Proposition 2.10. \([x_0, x_1, x_2, x_3]\) is of type I iff \(\det \begin{bmatrix} x_0 - x_2 & x_1 \\ x_1 - x_3 & x_2 \end{bmatrix} = 0\).

Let \(M = \begin{bmatrix} \alpha & \gamma \\ \beta & \gamma \end{bmatrix}\), and define \([x_0, x_1, x_2, x_3] = M^{\otimes 3}[c, 0, 0, d] = c \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} \gamma \end{bmatrix}^{\otimes 3}\). Then \(\det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} = cd\alpha\gamma (\det M)^2\), \(\det \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} = cd\beta\delta (\det M)^2\), and \(\det \begin{bmatrix} x_0 - x_2 & x_1 \\ x_1 - x_3 & x_2 \end{bmatrix}\) is their sum.

Suppose \([x_0, x_1, x_2, x_3] = M^{\otimes 3}[c, 0, 0, d]\) is non-degenerate. Then, \([x_0, x_1, x_2, x_3]\) is of type I iff the two columns of \(M\) are orthogonal \(\alpha \gamma + \beta \delta = 0\). In this case, it belongs to \(I(a, b)\), where \(a = \alpha \gamma = -\beta \delta\) and \(b = \alpha \delta + \beta \gamma\).

Proof. A straightforward calculation. Note that \([x_0, x_1, x_2]\) is a binary signature having the matrix form \(M \begin{bmatrix} c & 0 \\ 0 & d \end{bmatrix} M^T\). Similarly \([x_0, x_2, x_3] = M^{\otimes 3}[c, 0, 0, d]\) has the form \(M \begin{bmatrix} c \beta & 0 \\ 0 & d \delta \end{bmatrix} M^T\). \(\square\)

Proposition 2.11. Let \([x_0, x_1, x_2, x_3] = M^{\otimes 3}[c, 0, 0, d]\) be non-degenerate. Then, \([x_0, x_1, x_2, x_3]\) is of type II iff the columns of \(M\) are multiples of \((1, i)^T\) and \((1, -i)^T\) respectively.

Proof. By non-degeneracy, \(cd \neq 0\) and \(\det M \neq 0\). The equations \(x_0 + x_2 = 0\) and \(x_1 + x_3 = 0\) are respectively \(ca(\alpha^2 + \beta^2) + d\gamma(\gamma^2 + \delta^2) = 0\) and \(c\beta(\alpha^2 + \beta^2) + d\delta(\gamma^2 + \delta^2) = 0\). Viewed as a linear equation system on \(c\) and \(d\), there are non-zero solutions iff its determinant is zero:

\((\alpha^2 + \beta^2)(\gamma^2 + \delta^2) \det M = 0.\)

It follows that at least one of the factors \(\alpha^2 + \beta^2 = 0\) or \(\gamma^2 + \delta^2 = 0\). If \(\alpha^2 + \beta^2 = 0\), by the linear equation system, since \((\gamma, \delta)\) is a non-zero vector, we get the second factor \(\gamma^2 + \delta^2 = 0\) as well. Similarly starting with \(\gamma^2 + \delta^2 = 0\) we also get \(\alpha^2 + \beta^2 = 0\).

Conversely, if \(\beta = \pm i \alpha\) and \(\delta = \mp i \gamma\), then \(c \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} \gamma \end{bmatrix}^{\otimes 3}\) has the form \([x, y, -x, -y]\). \(\square\)

The essence of Proposition 2.11 is that \(c \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} \gamma \end{bmatrix}^{\otimes 3}\) satisfies the second order linear recurrence with eigenvalues \(\beta/\alpha\) and \(\delta/\gamma\), and it is of type II iff these eigenvalues are \(\pm i\).

Our first major dichotomy theorem in this book is for Holant\(^*(\mathcal{F})\), where \(\mathcal{F}\) is an arbitrary set of symmetric signatures on Boolean variables.

Since all unary signatures can be used for free in Holant\(^*(\mathcal{F})\), we assume the arity of every signature in \(\mathcal{F}\) is greater than one. And since any degenerate signature can be decomposed to unary signatures, we also assume that every signature in \(\mathcal{F}\) is non-degenerate.

Theorem 2.12. Let \(\mathcal{F}\) be a set of non-degenerate symmetric signatures over \(\mathbb{C}\). Then Holant\(^*(\mathcal{F})\) is computable in polynomial time for the following three Classes of \(\mathcal{F}\). In all other cases, Holant\(^*(\mathcal{F})\) is \#P-hard.

A. Every signature in \(\mathcal{F}\) is of arity no more than two;

B. There exist \(a\) and \(b\) (not both zero, depending only on \(\mathcal{F}\)), such that every signature in \(\mathcal{F}\) either (1) has type I\((a, b)\) or (2) has arity two and is of the form \([2a\lambda, b\lambda, -2a\lambda]\).
C. Every signature in $\mathcal{F}$ either (1) has type II or (2) has arity two and is of the form $[\lambda, 0, \lambda]$.

The dichotomy is still valid even if the inputs are restricted to planar graphs.

Remark 1: In order that Holant$^*(\mathcal{F})$ be a finitely specifiable problem with parameter $\mathcal{F}$ we can require that $\mathcal{F}$ be a finite set. However our dichotomy theorem is stronger, and applies to an infinite set $\mathcal{F}$ in the following sense: When $\mathcal{F}$ falls in one of the tractable classes, the Holant problem is computable in polynomial time even when $\mathcal{F}$ is infinite, but the input size of the signature grid includes a description (in symmetric signature notation) of the functions at each node. On the other hand, when $\mathcal{F}$ does not belong to one of the tractable classes, then there is a finite subset $\mathcal{F}' \subseteq \mathcal{F}$ for which Holant$^*(\mathcal{F}')$ is #P-hard. In more detail, this means that there is a finite set of unary functions $\mathcal{U}'$ such that Holant$(\mathcal{F}' \cup \mathcal{U}')$ is #P-hard.

Remark 2: For a more conceptual characterization, the reader is encouraged to compare this dichotomy theorem with Theorem 7.19. However in order to formulate and prove Theorem 7.19 we need to develop the necessary machinery and preparatory results, of which Theorem 2.12 is the first one. The reader is also encouraged to compare Theorem 2.12 with Theorem 3.5 which is a more abstract statement of Theorem 2.12.

This dichotomy theorem essentially says that for Holant$^*(\mathcal{F})$ the only tractable symmetric signature sets $\mathcal{F}$ are those when, after all degenerate signatures are removed, $\mathcal{F}$ consists of functions of arity at most two, or consists of generalized Fibonacci gates under a holographic transformation and some specific binary functions, or one additional class.

Proof of Tractability: Class $\mathfrak{A}$ is when every signature in $\mathcal{F}$ has arity at most two. Then the graph of the signature grid $\Omega$ is a disjoint union of paths and cycles (isolated points contribute a constant; we may assume there are no isolated points.) By matrix multiplication, we can compute the Holant value for a path. The Holant value for a cycle is by taking the trace of a path. The value Holant$_{\Omega}$ is the product over connected components.

Next we consider Class $\mathfrak{B}$. If $a = 0$, all functions in $\mathcal{F}$ are GEN-EQ or binary DIS-EQUALITY, and Holant$^*(\mathcal{F})$ is obviously computable in polynomial time. Suppose $a \neq 0$. All functions in $\mathcal{F}$ having arity greater than 2 are generalized Fibonacci gates. We first assume $X^2 - (b/a)X - 1$ has two distinct eigenvalues $\lambda \neq \mu$. Then $\lambda \mu = -1$ and $\lambda + \mu = b/a$.

All functions in $\mathcal{F}$ have the form $f = u\left[\begin{array}{c} 1 \\ \lambda \end{array}\right] ^ \otimes n + v\left[\begin{array}{c} 1 \\ \mu \end{array}\right] ^ \otimes n$ or multiples of $g = [2, b/a, -2]$.

Under the (inverse) holographic transformation by $M = \left[\begin{array}{cc} 1 & \lambda \\ 1 & \mu \end{array}\right]$, $f \mapsto M \otimes n f$ is a GEN-EQ, $g \mapsto M \otimes 2 g = [0, 4 + (b/a)^2, 0]$, and $(=)_2 = [1, 0, 1]$ is turned into $[1, 0, 1] (M^{-1}) \otimes 2$ which is a multiple of $[1 + \mu^2, 0, 1 + \lambda^2]$, a GEN-EQ. In the meanwhile all unary functions are transformed into unary functions. Connecting any unary function to a GEN-EQ creates another GEN-EQ. Connecting any unary function to any function of arity two creates another unary function. Hence Holant$^*(\mathcal{F})$ is tractable.

Continuing for Class $\mathfrak{B}$, if the eigenvalues $\lambda = \mu$, then $b = \pm 2ai$. Then $a \neq 0$ and
we can normalize to \(a = 1, b = \pm 2i, \lambda = b/2\). The signature \(g = [2, b, -2] = [2, \pm 2i, -2]\) is degenerate. We have proved using the holographic transformation \(Z\) that in this case \(\text{Holant}^*(\mathcal{F})\) is tractable.

For Class \(\mathcal{C}\), we first get rid of all unary functions and binary functions of the form \(\lambda[1,0,1]\). If there is a function \(\lambda[1,0,1]\) applied to two variables (edges) \(e\) and \(e'\), we just merge them into one, remove this function with a global factor \(\lambda\). Combining a unary function with a function of the type \([x,y,-x,-y,\ldots]\) does not change its type. Hence, we may assume all functions satisfy \(x_k + x_{k+2} = 0\). If the input graph \(G(V,E)\) is not bipartite, there is a cycle \(v_1,e_1,v_2,\ldots,v_k,e_k,v_1\) of odd length. We partition all 0-1 assignments for \(E\) into two parts, with a 1-1 correspondence between them. An assignment \(\sigma\) is mapped to \(\sigma'\), which assigns the same values on \(E - \{e_1,\ldots,e_k\}\), but opposite values on \(\{e_1,\ldots,e_k\}\). Under \(\sigma\) and \(\sigma'\), all functions on \(V - \{v_1,\ldots,v_k\}\) evaluate to the same value, and if \(\sigma(e_{j-1}) \neq \sigma(e_j)\) (where the index \(j\) is counted mod \(k\)), the function at \(v_j\) evaluates to the same value under \(\sigma\) and \(\sigma'\), and if \(\sigma(e_{j-1}) = \sigma(e_j)\), the function at \(v_j\) evaluates to opposite values under \(\sigma\) and \(\sigma'\). Consider \(\sigma(e_1),\sigma(e_2),\ldots,\sigma(e_k),\sigma(e_1)\), there must be an even number of times where the value changes in this sequence. Since \(k\) is odd, there are an odd number of \(v_j\)'s, where the functions at \(v_j\) give opposite values under \(\sigma\) and \(\sigma'\). Hence, in the summation, the contributions of \(\sigma\) and \(\sigma'\) are canceled.

If the input graph is bipartite, the problem is \(\text{Holant}(\mathcal{F} | \mathcal{F})\). It is turned into \(\text{Holant}(\mathcal{F'} | \mathcal{F'})\) where \(\mathcal{F'} \subseteq \text{GEN-Eq}\), by the holographic reduction \(Z = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\). This is easily seen by writing each contravariant \(f \in \mathcal{F}\) as \(u \begin{bmatrix} 1 \\ 1 \end{bmatrix} ^{\otimes n} + v \begin{bmatrix} 1 \\ -1 \end{bmatrix} ^{\otimes m} = Z[u,0,\ldots,0,v]\), and each covariant \(g \in \mathcal{F}\) as \(u' \begin{bmatrix} 1 & 1 \end{bmatrix} ^{\otimes m} + v' \begin{bmatrix} 1 & -1 \end{bmatrix} ^{\otimes m}, \) so that \(gZ = u' \begin{bmatrix} 0 & 2 \end{bmatrix} ^{\otimes m} + v' \begin{bmatrix} 2 & 0 \end{bmatrix} ^{\otimes m} = [2^{m}u',0,\ldots,0,2^{m}u']\). Hence \(\text{Holant}(\mathcal{F} | \mathcal{F})\) is tractable.

**Proof Outline for Hardness:** Now for the proof of hardness, we first prove in Lemma 2.13 that the theorem holds if \(\mathcal{F}\) contains a single symmetric signature of arity three. The main technique is holographic reduction. In Lemma 2.14, we prove that if one signature of arity three has the form in Class \(\mathcal{B}\) of Theorem 2.12, and we combine it with another signature of arity two which is not in Class \(\mathcal{B}\), then the Holant* problem is \#P-hard. The idea of the proof of Lemma 2.14 is to reduce it to Lemma 2.13 with holographic reductions. In Lemma 2.15, we prove the same thing for Class \(\mathcal{C}\). In Lemma 2.16 we extend Lemma 2.13 to a single signature of arbitrary arity. Finally we extend the proof to a set of signatures of arbitrary arities.

### 2.3.1 The First Lemma: A Single Ternary Signature

The following lemma says that Theorem 2.12 holds if \(\mathcal{F}\) consists of only one signature of arity three. It serves as the foundation for subsequent lemmas.

**Lemma 2.13.** Let \(f = [x_0,x_1,x_2,x_3]\) be a non-degenerate symmetric signature with arity three, then \(\text{Holant}^*(f)\) is \#P-hard unless \(f\) has type I or type II.
Proof. Assume \( f \) is of neither type I nor type II, we prove that \( \text{Holant}^*(f) \) is \#P-hard. Our starting point is that \( \text{Holant}([0, 1, 1][1, 0, 0, 1]) \) and \( \text{Holant}([1, 0, 1][1, 1, 0, 0]) \) are both \#P-complete [?]. The first problem is to count the number of vertex covers for 3-regular graphs; while the second is to count the number of (not necessarily perfect) matchings for 3-regular graphs. Both of them remain \#P-complete even for planar graphs [?, ?].

Every non-degenerate signature \([x_0, x_1, x_2, x_3]\) belongs to one of the following categories:

- **Category 1.** \( x_k = \alpha_1^{3-k} \alpha_2^k + \beta_1^{3-k} \beta_2^k \), where \( \det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \neq 0 \); or
- **Category 2.** \( x_k = ck \alpha_1^{k-1} + d \alpha_1^k \), where \( c \neq 0 \); or
- **Category 3.** \( x_k = c(3-k) \alpha_2^{3-k} + d \alpha_3^{3-k} \), where \( c \neq 0 \).

This parametrization is obtained by considering the solutions \((a, b, c)^T\) to \( \begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ a & b & c \end{bmatrix} = 0 \), which form a 1-dimensional vector space. Suppose \((a, b, c)^T\) is a nonzero solution. If \( a = c = 0 \) we have \( x_1 = x_2 = 0 \) and a diagonal matrix \( \begin{bmatrix} \sqrt{x_0} & 0 \\ 0 & \sqrt{x_3} \end{bmatrix} \) can be used in Category 1. If \( a \) and \( c \) are not both 0, then we may consider \([x_0, x_1, x_2, x_3]\) satisfies a second order linear recurrence relation (either forward or backward). Depending on whether the characteristic equation has two distinct roots or a double root, we have a case in Category 1 or the other two. Category 3 can be viewed as the reversal of Category 2, so we will omit the proof for Category 3. We take the convention that \( k^{0k-1} = 0, 1, 0, 0 \) for \( k = 0, 1, 2, 3 \) respectively.

For Category 1, we have

\[
 f = [x_0, x_1, x_2, x_3] = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}^{\otimes 3}.
\]

Non-degeneracy of \( f \) implies that \( \alpha_1 \beta_2 - \alpha_2 \beta_1 \neq 0 \). The fact that \( f \) is not of type I or II implies that \( \alpha_1 \beta_1 + \alpha_2 \beta_2 \neq 0 \) by Proposition 2.10, and \( \alpha_1^2 + \alpha_2^2 \neq 0 \) or \( \beta_1^2 + \beta_2^2 \neq 0 \) by Proposition 2.11. By symmetry, we can assume that \( \alpha_1^2 + \alpha_2^2 \neq 0 \).

We apply an orthogonal transformation \( T = \frac{1}{\sqrt{\alpha_1^2 + \alpha_2^2}} \begin{bmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{bmatrix} \) to map the vector \((\alpha_1, \alpha_2)^T\) to \((\alpha_1', 0)^T\), where \( \alpha_1' = \sqrt{\alpha_1^2 + \alpha_2^2} \neq 0 \).

\[
 f' = T^{\otimes 3} f = [x_0', x_1', x_2', x_3'] = \begin{bmatrix} \alpha_1' \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} \beta_1' \\ \beta_2' \end{bmatrix}^{\otimes 3}.
\]

By Theorem 1.6, this transformation does not change the complexity of the Holant problem. So it suffices to prove \#P-hardness for \( f' \). By a scalar multiplication we may assume \( \alpha_1' = 1 \). So, reusing the notation \( f \), we can assume the given signature is

\[
 f = [x_0, x_1, x_2, x_3] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3}.
\] (2.14)
The conditions from Lemma 2.13 become simply both $\alpha, \beta \neq 0$.

Consider $M = \begin{bmatrix} 1 & \alpha \\ 0 & \beta \end{bmatrix}$. We have $f = M^{\otimes 3}[1, 0, 0, 1] = M^{\otimes 3} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 3} \right)$, and

$$(0, 1, 1, 1)(M^{-1})^{\otimes 2} = \frac{1}{\beta^2}(0, \beta, \beta, 1 - 2\alpha).$$

One can check this by computing $M^T \begin{bmatrix} 0 & \beta \\ \beta & 1 - 2\alpha \end{bmatrix} M$. Ignoring $1/\beta^2$, by a holographic reduction the complexity of Holant$([0, \beta, 1 - 2\alpha] | f)$ is the same as the $\#P$-complete vertex cover problem Holant$([0, 1, 1] | [1, 0, 0, 1])$. In particular Holant$([0, \beta, 1 - 2\alpha], f)$ is $\#P$-hard. In order to prove the $\#P$-hardness of Holant$^*(f)$, we only need to show that $[0, \beta, 1 - 2\alpha]$ can be realized by $f$ with some unary signatures.

![Figure 2.3](image)

Figure 2.3: We use this gadget to realize the signature $[0, \beta, 1 - 2\alpha]$. All (three) nodes of degree 3 in this gadget have the signature $f = [x_0, x_1, x_2, x_3]$.

We use the gadget in Figure 2.3 to realize $[0, \beta, 1 - 2\alpha]$, where the two unary signatures $(t_0, t_1)$ and $(s_0, s_1)$ will be determined later. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} \alpha & \beta \\ \beta & 1 - 2\alpha \end{bmatrix} = \begin{bmatrix} \alpha^2 & \alpha \beta \\ \alpha \beta & \beta^2 \end{bmatrix}.$$  

For $f$ in (2.14), if one input is 0, the induced binary signature has its matrix form $A + \alpha B$. If one input is 1, the induced binary signature has matrix form $\beta B$. It follows that the signature in Figure 2.3, as a binary function in matrix form, is

$$(t_0(A + \alpha B) + t_1\beta B)(s_0(A + \alpha B) + s_1\beta B)(t_0(A + \alpha B) + t_1\beta B) = (t_0A + (t_0\alpha + t_1\beta)B)(s_0A + (s_0\alpha + s_1\beta)B)(t_0A + (t_0\alpha + t_1\beta)B).$$

**Exercise:** Show that $t_0(A + \alpha B) + t_1\beta B$ is the signature matrix of the binary gadget obtained by connecting the unary signature $[t_0, t_1]$ to $f$ of (2.14). Deduce that the above matrix is the signature matrix of the binary gadget in Figure 2.3.

Now we use a new set of variables

$$x = t_0, \quad y = t_0\alpha + t_1\beta, \quad z = s_0, \quad w = s_0\alpha + s_1\beta,$$

(2.15)
and write the above matrix as \((xA + yB)(zA + wB)(xA + yB)\). We note that, since \(\beta \neq 0\), for any given \(x, y, z, w\), we can find \(t_0, t_1, s_0, s_1\) to satisfy the relations (2.15). Then, to realize \([0, \beta, 1 - 2\alpha]\), we will choose some \(x, y, z\) and \(w\) such that

\[
(xA + yB)(zA + wB)(xA + yB) = \begin{bmatrix}
0 & \beta \\
\beta & 1 - 2\alpha
\end{bmatrix}.
\]

Substituting \(A\) and \(B\), and denoting \(\alpha^2 + \beta^2\) by \(\gamma\), we have the following:

\[
(xA + yB)(zA + wB)(xA + yB) = w\begin{bmatrix}
\alpha^2(x + y\gamma)^2 & y\alpha\beta\gamma(x + y\gamma) \\
y\alpha\beta\gamma(x + y\gamma) & y^2\beta^2\gamma^2
\end{bmatrix} + z\begin{bmatrix}
(x + y\alpha^2)^2 & y\alpha\beta(x + y\alpha^2) \\
y\alpha\beta(x + y\alpha^2) & y^2\alpha^2\beta^2
\end{bmatrix}
\]

We may choose \(w = (x + y\alpha^2)^2\) and \(z = -\alpha^2(x + y\gamma)^2\) to make the \((1, 1)\) entry zero. The \((1, 2)\) (and \((2, 1)\)) entry is

\[
g_1 = x\alpha\beta^3(x + y\alpha^2)(x + y\gamma);
\]

and the \((2, 2)\) entry is

\[
g_2 = x\gamma(2\alpha^2 + \beta^2 + 2y\alpha^2\gamma).
\]

We want to choose some \(x\) and \(y\) such that \(g_1 = \beta\) and \(g_2 = 1 - 2\alpha\). Recall that \(\alpha\beta \neq 0\). We will choose \(x\) and \(y \neq 0\). As both \(g_1\) and \(g_2\) are homogeneous (of degree 4) in \(x\) and \(y\), we can ignore the common factor \(xy\beta^3\) of \(g_1\) and \(g_2\). It follows that we only have to satisfy that \(g_2/g_1 = (1 - 2\alpha)/\beta\) and \(g_1 \neq 0\), with \(y = 1\). The following equation

\[
\alpha(2\alpha - 1)x^2 + (2\alpha^2 - \alpha + \beta^2)(2\alpha^2 + \beta^2)x + \alpha^2(\alpha^2 + \beta^2)(2\alpha^2 - \alpha + 2\beta^2) = 0
\]

(2.16)

is equivalent to \(\beta g_2 = (1 - 2\alpha)g_1\) after removing \(xy\beta^3\) and setting \(y = 1\). What we have to prove is that the equation in (2.16) has a root \(x \neq 0\) that is not a root of \(g_1 = g_1(x, 1) = 0\). The roots of \(g_1(x, 1) = 0\) are \(x = 0, x = -\alpha^2\) and \(x = -\gamma\). Firstly, when \(x = -\alpha^2\), the expression in (2.16) can be simplified to \(\alpha^2\beta^4 \neq 0\). (Note that when \(g_1 = 0\), the expression in (2.16) is \(\beta g_2\), after removing \(xy\beta^3\), and \(y = 1\).) So \(x = -\alpha^2\) is not a root of (2.16). Secondly when \(x = -\gamma\) the expression in (2.16) can be simplified to \(-\beta^4\gamma\), and if \(x = -\gamma\) is a root, this would force \(\gamma = 0\). So, assuming the expression in (2.16) as a polynomial in \(x\) is indeed of degree 2, and it does not have a double root 0, then we can find a root \(\xi \neq 0\) of (2.16) that is not a root of \(g_1\). Indeed, This \(\xi \neq -\alpha^2\), because \(-\alpha^2\) is not a root of (2.16); \(\xi\) cannot be \(-\gamma\) either, for otherwise \(-\gamma\) would be a root of (2.16) which would force \(\gamma = 0\), and thus \(\xi = -\gamma = 0\), a contradiction. Thus \(\xi\) is a nonzero root of (2.16) but not a root of \(g_1(x, 1)\), as is needed.

Now let us consider the exceptional cases: either \(x = 0\) is a double root of (2.16), or (2.16) has degree less than 2. If \(x = 0\) is a double root of (2.16), we have

\[
(2\alpha^2 - \alpha + \beta^2)(2\alpha^2 + \beta^2) = \alpha^2(\alpha^2 + \beta^2)(2\alpha^2 - \alpha + 2\beta^2) = 0.
\]

52
To satisfy this, since $\alpha \beta \neq 0$, there are only four exceptional cases (A1 to A4): $\alpha = 1, \beta = \pm i$ or $\alpha = -\frac{1}{2}, \beta = \pm \frac{1}{\sqrt{2}}$. On the other hand, if the polynomial in (2.16) has degree less than 2, by $\alpha \neq 0$, we get $\alpha = \frac{1}{2}$. In this case, the polynomial becomes

$$(1/2 + \beta^2)x + (1/4 + \beta^2)/2 = 0.$$ 

This gives us four additional exceptional cases (B1 to B4): $\alpha = \frac{1}{2}, \beta = \pm \frac{1}{\sqrt{2}}$, in which case the polynomial is linear with root $x = 0$; or $\alpha = -\frac{1}{2}, \beta = \pm \frac{1}{\sqrt{2}}$, in which case the polynomial degenerates to a (nonzero) constant. In all other cases, there is a nonzero root of (2.16) that is not a root of $g_1(x, 1)$, completing the #P-hardness proof.

For the cases A1 and A2, we use a new starting problem Holant([1, 1, 0] | [1, 0, 0, 1]), which is the reversal of the previous problem and therefore it is also #P-complete. The problem can also be understood as counting independent sets over 3-regular graphs. Then all previous parts of the proof are still valid, except that the signature of arity two to be realized is

$$(1, 1, 1, 0)(T^{-1})^{\otimes 2} = (1, \frac{1 - \alpha}{\beta}, \frac{1 - \alpha}{\beta}, \frac{\alpha^2 - 2\alpha}{\beta^2}).$$

Substituting $\alpha = 1, \beta = \pm i$, the signature is [1, 0, 1] which is trivially realizable by one edge. So we have proved that it is #P-hard in the cases A1 and A2. Now consider the cases A3 and A4: $\alpha = -\frac{1}{2}, \beta = \pm \frac{1}{\sqrt{2}}$. We will give a different parametrization. For the case A3, we apply an orthogonal transformation $M = \begin{bmatrix} -i & -\sqrt{2} \\ \sqrt{2} & -i \end{bmatrix}$ and a scalar multiplier $(2i)^3$ on the signature $f$ in (2.14), to get $f \mapsto (2i)^3 M^{\otimes 3} f = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 2 \\ 2i \end{bmatrix}^{\otimes 3}$. This is not one of the exceptional cases and we have proved that it is #P-hard. For the case A4, we apply another orthogonal transformation $M' = \begin{bmatrix} i & -\sqrt{2} \\ \sqrt{2} & i \end{bmatrix}$ and a scalar multiplier $(-2i)^3$ on the signature and it becomes $\begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 2 \\ -2\sqrt{2i} \end{bmatrix}^{\otimes 3}$.

The cases B3 and B4 can be shown by the same method as in A4 and A3, using $M'$ and $M$ respectively. The only remaining cases are B1 and B2. Here we will use another gadget similar to the one in Figure 2.3 except we remove the middle edge (including the node labeled $(s_0, s_1)$ and the middle node of degree 3). For B1, $\alpha = \frac{1}{2}, \beta = \frac{1}{2}$, the signature of this gadget is

$$(t_0(A + \alpha B) + t_1 \beta B)^2 = (xA + yB)^2,$$

where the matrices $A$ and $B$ are as before, and with the specific values of $\alpha, \beta, B = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ i & -1 \end{bmatrix}$. By setting $x = i$ and $y = -2i$, we have $(xA + yB)^2 = \begin{bmatrix} 0 & \frac{i}{2} \\ i/2 & 0 \end{bmatrix}$, which is the matrix form of the target signature $[0, \beta, 1 - 2\alpha] = [0, \frac{1}{2}, 0]$. This finishes the case B1. The case B2 can be proved in the same way with $x = 1$ and $y = -2$.

**Exercise:** Prove the cases B3 and B4 using the matrices $M'$ and $M$ in the cases A4 and A3, respectively.
Now we prove for Category 2. In this case \( x_k = c k \alpha^{k-1} + d \alpha^k \), and it satisfies the recurrence \( x_{k+2} = 2 \alpha x_{k+1} - \alpha^2 x_k \), for \( k = 0, 1 \). The characteristic polynomial \( X^2 - 2 \alpha X + \alpha^2 \) has a double root \( \alpha \). If \( \alpha = \pm i \), then \( f \) has type \( I(1, 2\alpha) \). Since we assumed \( f \) is not of type \( I, \alpha \neq \pm i \). Then we can choose some orthogonal matrix to transform the signature to the form \([x, y, 0, 0]\) where \( y \neq 0 \), as follows.

Let \( M = \begin{bmatrix} 1 & \frac{d-1}{c} \\ \alpha & c + \frac{d-1}{c} \end{bmatrix} \) with \( \det M = c \neq 0 \), then the signature \([x_0, x_1, x_2, x_3]\) can be expressed as

\[
(x_0, x_1, x_2, x_1, x_2, x_2, x_3)^T = M^\otimes 3 (1, 1, 1, 0, 1, 0, 0, 0)^T.
\]

In symmetric signature notation \([x_0, x_1, x_2, x_3] = M^\otimes 3 [1, 1, 0, 0] \).

**Exercise:** Verify that \([x_0, x_1, x_2, x_3] = M^\otimes 3 [1, 1, 0, 0] \). Use \( M \) to show that there is another matrix \( M' \) such that \([x_0, x_1, x_2, x_3] = M'^\otimes 3 [0, 1, 0, 0] \).

Let \( M = QR \) be its QR-factorization, i.e., where \( Q \) is orthogonal and \( R \) is upper triangular. In fact for \( M = \begin{bmatrix} 1 & * \\ \alpha & * \end{bmatrix} \), recalling that \( 1 + \alpha^2 \neq 0 \), we can choose our \( Q \) as the (complex) orthogonal matrix \( Q = Q^T = \frac{1}{\sqrt{1 + \alpha^2}} \begin{bmatrix} 1 & \alpha \\ -\alpha & 1 \end{bmatrix} \). Then \( QM = R = \begin{bmatrix} u & w \\ 0 & v \end{bmatrix} \) is upper triangular, where \( u = \sqrt{1 + \alpha^2} \). Because \( \det Q = -1 \) and \( \det R = -\det M = -c \neq 0 \), we have \( uv \neq 0 \).

This \( Q \) is our choice of the orthogonal transformation. It follows that

\[
Q^\otimes 3 [x_0, x_1, x_2, x_3] = (QM)^\otimes 3 [1, 1, 0, 0]
\]

\[
= R^\otimes 3 \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [u]^\otimes 3 + [u] \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}
\]

\[
= [u] \otimes [u] \otimes [w] + [u] \otimes [w] \otimes [u] + [w] \otimes [u] \otimes [u] \quad (2.17)
\]

In the notation for symmetric signatures, this is \([u^3 + 3u^2w, u^2v, 0, 0]\), with \( u^2v \neq 0 \).

By a scalar multiplication, we can normalize the entry \( u^2v \) to 1. So we only have to deal with a signature of the form \([v, 1, 0, 0]\) for an arbitrary given \( v \).

For this signature, we can apply a holographic transformation defined by the matrix \( M' = \begin{bmatrix} 1 & -\frac{v-1}{3} \\ 0 & \frac{1}{3} \end{bmatrix} \) with inverse \( M'^{-1} = \begin{bmatrix} \frac{1}{1} & \frac{v-1}{3} \\ 0 & \frac{1}{1} \end{bmatrix} \). To prove \#P-hardness, we will reduce from the MATCHING problem Holant([1, 0, 1] | [1, 1, 0, 0]). Under a contravariant transformation \((M'^{-1})^\otimes 3 [1, 1, 0, 0] = [v, 1, 0, 0] \), the signature \([1, 1, 0, 0]\) becomes \([v, 1, 0, 0]\). By the corresponding covariant transformation, \([1, 0, 1]\) becomes \([1, \frac{1-v}{3}, 1 + \frac{(1-v)^2}{9}] \):

\[
(1, 0, 0, 1)M'^\otimes 2 = ((1, 0)^\otimes 2 + (0, 1)^\otimes 2)M'^\otimes 2 = (1, \frac{1-v}{3}, \frac{1-v}{3}, 1 + \frac{(1-v)^2}{9}).
\]

We complete this proof by using the same gadget in Figure 2.3 to realize this binary signature, using unary signatures and \([v, 1, 0, 0]\).
We will rename the values $x = t_0$, $y = t_1$, $z = s_0$ and $w = s_1$ in Figure 2.3. The signature of this gadget in matrix form is $(xA + yB)(zA + wB)(xA + yB)$, where $A = \begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

We set $x = 1$. Then $(A + yB)(zA + wB)(A + yB)$ is

$$\begin{bmatrix}
    w \cdot (v + y)^2 + z \cdot (v(v + y) + 2(v + y) & w \cdot (v + y) + z \cdot (v(v + y) + 1) \\
    w \cdot (v + y) + z \cdot (v(v + y) + 1) & w + z \cdot v
\end{bmatrix}.$$

Our goal is to choose $y$, $z$ and $w$ such that it is equal to $\begin{bmatrix} 1 & (1 - v)/3 \\ (1 - v)/3 & 1 + (1 - v)^2/9 \end{bmatrix}$. We can write this requirement as a system of three linear equations in $z$ and $w$, whose coefficient matrix depends on $y$. Then we can complete the proof, if we can choose $y$ such that the following matrix has determinant 0, yet the first two columns have rank 2.

$$\begin{bmatrix}
    (v + y)^2 & v(v + y) + 2(v + y) & 1 \\
    (v + y) & v(v + y) + 1 & (1 - v)/3 \\
    1 & v & 1 + (1 - v)^2/9
\end{bmatrix}.$$  

After some elementary column operations on the first two columns it becomes

$$\begin{bmatrix}
    -(v + y)^2 & 2(v + y) & 1 \\
    0 & 1 & (1 - v)/3 \\
    1 & 0 & 1 + (1 - v)^2/9
\end{bmatrix},$$

whose first two columns obviously have rank 2. Its determinant is

$$-(v + y)^2(1 + (1 - v)^2/9) + 2(v + y)(1 - v)/3 - 1.$$  

Clearly $1 + (1 - v)^2/9$ and $(1 - v)/3$ cannot be both 0. Thus we can set $y$ so that this determinant is 0. This completes the proof of Lemma 2.13. $\square$

### 2.3.2 A Pair of Signatures

Lemma 2.13 shows what happens when there is a single non-degenerate symmetric signature $f$ of arity three. It explicitly isolates two tractable cases: If $\text{Holant}^*(f)$ is not $\#P$-hard, then $f$ must be of type I or type II. The next lemma addresses what happens when we have more than one signature, where one signature happens to be of type I but some other signature does not go along with it.

**Lemma 2.14.** Let $f = [x_0, x_1, x_2, x_3]$ and $g = [y_0, y_1, y_2]$ be non-degenerate symmetric signatures with arity three and two respectively. Suppose $f$ is of type I$(a, b)$ for some $a, b$ (not both zero), but $g$ is not of type I$(a, b)$, i.e., $ay_0 + by_1 - ay_2 \neq 0$, and $g$ is not of the form $[2a\lambda, b\lambda, -2a\lambda]$. Then $\text{Holant}^*(g \mid f)$ is $\#P$-hard. It follows that $\text{Holant}^*(\{f, g\})$ is also $\#P$-hard.
Proof. Since \( f \) is non-degenerate, the pair \((a,b)\) is unique up to a scalar factor. Our proof plan is as follows: We will show that Holant\(^*\)(\( g \mid f \)) is \#P-hard by a holographic reduction where \( g \) is transformed to the binary \textsc{Equality} (=2). Hence we want an invertible matrix \( M \) such that \( f \mapsto M^{\otimes 3} f = f' \) and \( g \mapsto g(M^{-1})^{\otimes 2} = [1,0,1] \). Note that \([1,0,1]\) can be replaced by an edge, and the unary signatures are transformed to unary signatures. Thus the complexity of the problem Holant\(^*\)(\( g \mid f \)) is the same as Holant\(^*\)(\( f' \)). We then apply Lemma 2.13 to \( f' \). Note that \( f' \) is non-degenerate. Otherwise, \( f' \) would be a tensor product of unary signatures, and then \( f = (M^{-1})^{\otimes 3} f' \) would also be a tensor product of unary signatures, and thus degenerate.

The requirement that \( g(M^{-1})^{\otimes 2} = [1,0,1] \) is the same as \( G = M^T M \) for some matrix \( M = \begin{bmatrix} a & \gamma \\ \beta & \delta \end{bmatrix} \), where \( G = \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} \) is the matrix form of \( g \). This is

\[
\begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} = M^T M = \begin{bmatrix} \alpha^2 + \beta^2 & \alpha \gamma + \beta \delta \\ \alpha \gamma + \beta \delta & \gamma^2 + \delta^2 \end{bmatrix}
\]

(2.18)

Such a factorization for a symmetric complex matrix \( G \) exists. If \( y_0 \neq 0 \), we take \( M = \frac{1}{\sqrt{y_0}} \begin{bmatrix} y_0 & y_1 \\ 0 & \Delta \end{bmatrix} \), where \( \Delta = \sqrt{\det G} \). Note that \( \det G \neq 0 \) since \( g \) is non-degenerate. If \( y_0 = 0 \) but \( y_2 \neq 0 \), we can first exchange \( y_0 \) and \( y_2 \) by \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \), and take \( M = \frac{1}{\sqrt{y_2}} \begin{bmatrix} \Delta & 0 \\ y_1 & y_2 \end{bmatrix} \). If \( y_0 = y_2 = 0 \), then we take \( M = \frac{1}{\sqrt{y_1}} Z \) where \( Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \).

Let \( f' = [u_0, u_1, u_2, u_3] = M^{\otimes 3} f \). We want to show that \( f' \) is not of type I or II, the two tractable cases of Lemma 2.13.

First we assume for a contradiction that \( f' \) is of type II. Then it satisfies the recurrence \( u_{k+2} + u_k = 0 \), for \( k = 0,1 \), with characteristic polynomial \( X^2 + 1 \). Therefore \( f' = c \begin{bmatrix} 1 \\ 1 \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} 1 \\ -i \end{bmatrix}^{\otimes 3} \), for some \( c \) and \( d \). We have \( M^{-1} = \Delta^{-1} \begin{bmatrix} \delta & \gamma \\ -\beta & \alpha \end{bmatrix} \), where \( \Delta = \det M \neq 0 \). Then it follows that

\[
f = (M^{-1})^{\otimes 3} f' = c\Delta^{-3} \begin{bmatrix} \delta - \gamma i \\ -\beta + \alpha i \end{bmatrix}^{\otimes 3} + d\Delta^{-3} \begin{bmatrix} \delta + \gamma i \\ -\beta - \alpha i \end{bmatrix}^{\otimes 3}.
\]

Since \( f \) is non-degenerate, \( cd \neq 0 \). By the assumption that \( f \) has type I\((a,b)\), and by Proposition 2.10, we have

\[
\alpha^2 + \beta^2 + \gamma^2 + \delta^2 = 0,
\]

and \( a = \gamma^2 + \delta^2 = -\alpha^2 - \beta^2 \), and \( b = -2(\alpha \gamma + \beta \delta) \) after a scaling. By (2.18), \([y_0, y_1, y_2] = [-a, -b/2, a] \) is of the form \([2a\lambda, b\lambda, -2a\lambda] \), a contradiction.

This completes our first step. Our second step is to show that \( f' \) cannot be of type I either. For a contradiction, suppose \( f' \) is of type I.

We are given that \( f \) is of type I\((a,b)\). If \( a = 0 \) then \( b \neq 0 \) and \( x_1 = x_2 = 0 \). We can write

\[
f = x_0 \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + x_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 3},
\]
and
\[ f' = [u_0, u_1, u_2, u_3] = M^{\otimes 3} f = x_0 \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + x_3 \begin{bmatrix} \gamma \\ \delta \end{bmatrix}^{\otimes 3}. \]

By Proposition 2.10 the dot product of \((\alpha, \beta)^T\) and \((\gamma, \delta)^T\) is zero, i.e.,
\[ y_1 = \alpha \gamma + \beta \delta = 0. \]
This implies that \(ay_0 + by_1 - ay_2 = 0\), a contradiction.

Finally we consider the case \(a \neq 0\). By a scaling we can assume \(a = 1\). Assume temporarily that \(b \neq \pm 2i\), then the recurrence \(x_{k+2} = bx_{k+1} + x_k\) has two distinct eigenvalues \(\lambda\) and \(\mu\) with \(\lambda + \mu = b\) and \(\lambda \mu = -1\). So we can write \([x_0, x_1, x_2, x_3] = c \begin{bmatrix} 1 \\ \lambda \\ 1 \\ \mu \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}^{\otimes 3} = \left[ \frac{1}{\lambda} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] \otimes 3 [c, 0, 0, d]. \]
Then \([u_0, u_1, u_2, u_3] = \left( M \left[ \frac{1}{\lambda} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] \right)^{\otimes 3} [c, 0, 0, d]. \] By Proposition 2.10 we have
\[ \det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = \left( \det M \cdot \det \left[ \frac{1}{\lambda} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] \right)^2 cd \left[ 1 \begin{bmatrix} \lambda \\ 1 \end{bmatrix} M^T M \begin{bmatrix} 1 \\ \mu \end{bmatrix} \right] \]
(2.19)
Note that the dot product of the two columns of \(M \left[ \frac{1}{\lambda} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] \) is \((1 \lambda)M^T M \left[ \frac{1}{\mu} \right] = (1 \lambda) \left[ \frac{y_0}{y_1} \frac{y_1}{y_2} \right] \), which is exactly \(y_0 + by_1 - y_2 \neq 0\). Also \(cd(\lambda - \mu)^2 = \det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} \). Indeed the binary signature \([x_0, x_1, x_2] = c \begin{bmatrix} 1 \\ \lambda \\ 1 \end{bmatrix}^{\otimes 2} + d \begin{bmatrix} 1 \\ \mu \end{bmatrix}^{\otimes 2} \) has matrix form \( \left[ \frac{1}{\lambda} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] \begin{bmatrix} 1 & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} 1 \\ \mu \end{bmatrix} \) T, with determinant \(cd(\lambda - \mu)^2\). Hence (2.19) is equivalent to
\[ \det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = (\det M)^2 \det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} (y_0 + by_1 - y_2) \]
(2.20)
Equation (2.20) is valid for \(b \neq \pm 2i\). However we show that it is valid for \(b = \pm 2i\) by a limiting argument. We take a sequence \(b^{(n)} \to b\), where \(b^{(n)} \neq 0, \pm 2i\). For the given \(g = [y_0, y_1, y_2]\) with \(y_0 + by_1 - y_2 \neq 0\) we may define a sequence \(g^{(n)} = [y_0, \frac{b}{b^{(n)}} y_1, y_2] \to g\), satisfying \(y_0 + b^{(n)}(\frac{b}{b^{(n)}} y_1) - y_2 \neq 0\). From \(g^{(n)}\) we may define \(M^{(n)}\) with \((M^{(n)})^T M^{(n)}\) being the matrix form of \(g^{(n)}\), and satisfying \(M^{(n)} \to M\). Then we can define \(x^{(n)} = [x_0, x_1, x_2^{(n)}, x_3^{(n)}]\) by the same recurrence with parameter \(b^{(n)}\) and then define \(u^{(n)}\) by \(M^{(n)}\) from \(x^{(n)}\) accordingly. By taking limit in (2.20) for this sequence we get (2.20) valid for \(b = \pm 2i\) as well.

The condition that \(f'\) is of type I is equivalent to \(\det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = 0\). We know that \(\det M \neq 0, y_0 + by_1 - y_2 \neq 0\), hence \(\det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} = 0\). As the third column in \(\begin{bmatrix} x_0 & x_1 & x_2 \end{bmatrix}\) is linearly dependent on the first two columns by the recurrence, this says that \(f\) is degenerate, a contradiction.

Lemma 2.15 does the same thing as Lemma 2.14 for tractable cases of type II in Lemma 2.13.

**Lemma 2.15.** Let \(f = [x, y, -x, -y]\) be of type II and \(g = [y_0, y_1, y_2]\), both non-degenerate. If \(g\) is not of type II, i.e., \(y_0 + y_2 \neq 0\), and \(g\) is not of the form \([\lambda, 0, \lambda]\), then Holant\(^*(g | f)\) is \#P-hard. It follows that Holant\(^*(\{f, g\})\) is also \#P-hard.

57
Proof. The proof idea is the same as for Lemma 2.14.

Let $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$ be the matrix such that $g(M^{-1})^\otimes 2 = [1,0,1]$, then $G = \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} = MTM$. Let $f' = M^\otimes 3 f$. Then $f'$ is non-degenerate since $f$ is. Note that $f$ has the form 

$$
[x,y,-x,-y] = c \begin{bmatrix} 1 \\ i \end{bmatrix}^\otimes 3 + d \begin{bmatrix} 1 \\ -i \end{bmatrix}^\otimes 3 ,
$$

where $cd \neq 0$ as $f$ is non-degenerate. We get

$$
f' = c \begin{bmatrix} \alpha + \gamma i \\ \beta + \delta i \end{bmatrix}^\otimes 3 + d \begin{bmatrix} \alpha - \gamma i \\ \beta - \delta i \end{bmatrix}^\otimes 3 .
$$

We want to show that Holant$^*(f')$ is #P-hard. By Lemma 2.13, we have to consider two cases. If $f'$ is of type II, then by Proposition 2.11

$$
0 = (\alpha + \gamma i)^2 + (\beta + \delta i)^2 = (\alpha - \gamma i)^2 + (\beta - \delta i)^2 .
$$

This gives $\alpha \gamma + \beta \delta = 0$ and $\alpha^2 + \beta^2 = \gamma^2 + \delta^2$. By $G = MTM$, this is precisely $y_1 = 0$ and $y_0 = y_2$. Then $[y_0, y_1, y_2]$ has the form $[\lambda, 0, \lambda]$, a contradiction.

Suppose $f'$ has type I. By Proposition 2.10 we have

$$
0 = (\alpha + \gamma i)(\alpha - \gamma i) + (\beta + \delta i)(\beta + \delta i) = \alpha^2 + \beta^2 + \gamma^2 + \delta^2 .
$$

This is precisely $y_0 + y_2 = 0$, a contradiction. \hfill \Box

2.3.3 A Single Signature of Arity $n$

The following lemma extends Lemma 2.13 to a signature with an arbitrary arity.

Lemma 2.16. Let $f = [x_0, x_1, \ldots, x_n]$ be a non-degenerate symmetric signature with arity $n > 3$, then Holant$^*(f)$ is #P-hard unless $f$ has type I or type II.

Proof. Assume that Holant$^*(f)$ is not #P-hard, we prove that $f$ must be of type I or type II. Using unary signatures $\Delta_0 = [1,0]$ and $\Delta_1 = [0,1]$, for any subset $S$ of sub-signatures of $f$, Holant$^*(S) \leq^p T$ Holant$^*(f)$. Hence Holant$^*(S)$ is not #P-hard.

Case A: For every $0 \leq k \leq n - 2$, the sub-signature $[x_k, x_{k+1}, x_{k+2}]$ is non-degenerate.

By Lemma 2.13, there are two tractable types to be considered for $[x_0, x_1, x_2, x_3]$.

(i) $[x_0, x_1, x_2, x_3]$ has type $I(a,b)$ for some non-zero pair $(a,b)$. Such a non-zero pair $(a,b)$ must be unique up to a scalar factor, by the non-degeneracy of $[x_0, x_1, x_2, x_3]$. Then by Lemma 2.14, for every $0 \leq k \leq n - 2$, either $ax_k + bx_{k+1} - ax_{k+2} = 0$ or $[x_k, x_{k+1}, x_{k+2}]$ has the form $[2a\lambda, b\lambda, -2a\lambda]$. 

58
If $a = 0$, then $b \neq 0$, and $x_1 = x_2 = 0$ from the linear equations. Then $[x_0, x_1, x_2]$ is degenerate. A contradiction. So $a \neq 0$ and we can normalize it and assume $a = 1$. Now we show that the form $[2\lambda, b\lambda, -2\lambda]$ cannot appear among $[x_k, x_{k+1}, x_{k+2}]$ ($0 \leq k \leq n-2$). This will conclude that $f$ has type I$(a, b)$.

Suppose the form $[2\lambda, b\lambda, -2\lambda]$ does occur. Such a $\lambda$ certainly is non-zero, otherwise it is degenerate. Also $b \neq \pm 2i$, since $[2, \pm 2i, -2]$ is degenerate. If it occurs as $[x_k, x_{k+1}, x_{k+2}]$ for $k = 0$ or $1$, then together with $ax_0 + bx_1 - ax_2 = 0$ and $ax_1 + bx_2 - ax_3 = 0$, and $a = 1$, we get $b = \pm 2i$. A contradiction. Thus let $k$ be the minimum index such that $[x_k, x_{k+1}, x_{k+2}]$ has the form $[2, b, -2]$, then $k \geq 2$, and there is a sub-signature $[x_{k-2}, x_{k-1}, x_k, x_{k+1}, x_{k+2}] = \lambda[x, y, 2, b, -2]$ with arity 4, where $ax + by - 2a = 0$ and $ay + 2b - ab = 0$. So the sub-signature is $\lambda[2^2 + 2, -b, 2, b, -2]$, of which there is a sub-signature $\lambda[-b, 2, b, -2]$. Applying Lemma 2.15 to the non-degenerate signature pair $[-b, 2, b, -2]$ and $[b^2 + 2, -b, 2, b, -2]$, we have $b^2 + 2 = -2$ or $b^2 + 2 = 2$. The first case would imply that $b = \pm 2i$. So we have $b^2 + 2 = 2$ and thus $b = 0$. In this case, the arity 4 sub-signature is $[2, 0, 2, 0, -2]$. Using a unary signature $[1, 1]$, we can get $[2, 2, 2, -2]$. By Lemma 2.13, Holant$^*$([2, 2, 2, -2]) is #P-hard.

(ii) If $[x_0, x_1, x_2, x_3]$ is of the form $[x, y, -x, -y]$, then by Lemma 2.15, for every $0 \leq k \leq n-2$, we have $x_k + x_{k+2} = 0$ or $[x_k, x_{k+1}, x_{k+2}]$ is of the form $[\lambda, 0, \lambda]$. We prove that $[\lambda, 0, \lambda]$ can not appear. This will conclude that $f$ has type II.

Suppose the form $[\lambda, 0, \lambda]$ does appear among $[x_k, x_{k+1}, x_{k+2}], \lambda \neq 0$ by non-degeneracy. It is easy to see that if it occurred at $k = 0$ or $1$, then $[x_0, x_1, x_2, x_3] = [x, y, -x, -y]$ is degenerate. Let $k$ be the minimum index such that $[x_k, x_{k+1}, x_{k+2}] = [\lambda, 0, \lambda]$, then $k \geq 2$. To its left we have $x_\ell + x_{\ell+2} = 0$, for $0 \leq \ell < k$. It follows that there must be a sub-signature (after a scaling) of the form $[1, 0, -1, 0, -1]$. Again we use a unary signature $[1, 1]$ and get $[1, -1, -1, -1]$. By Lemma 2.13 Holant$^*$([1, -1, -1, -1]) is #P-hard.

Case B: There exists some $0 \leq k \leq n-2$, such that $[x_k, x_{k+1}, x_{k+2}]$ is degenerate.

If $f$ is of the form $[x_0, 0, \ldots, 0, x_n]$, then $f$ has type I$(0, 1)$ and Lemma 2.16 holds. In the following we assume $f$ is not of this form.

By Proposition 2.8, there exists $0 \leq s \leq n-3$, such that the arity three sub-signature $f' = [x_s, x_{s+1}, x_{s+2}, x_{s+3}]$ is non-degenerate. We want to find a non-degenerate sub-signature of arity three which contains a degenerate sub-signature of arity two. Starting from a degenerate sub-signature $g = [x_k, x_{k+1}, x_{k+2}]$ of arity two, if $s \leq k \leq s + 1$, then $g$ is already a sub-signature of $f'$. Otherwise, either $k < s$ or $k > s + 1$. Suppose $k < s$. Consider the sub-signature $[x_{k+1}, x_{k+2}, x_{k+3}]$. If it is degenerate, we can substitute it for $g$, and continue. If it is non-degenerate, then we can substitute $f'$ by $[x_k, x_{k+1}, x_{k+2}, x_{k+3}]$, and it will also be non-degenerate. The proof for $k > s + 1$ is similar.

Thus we can find $f'$ and $g$ as specified. This $f'$ must be of the form $[s^2, sr, r^2, x]$ or $[y, s^2, sr, r^2]$. By symmetry, we only consider $f' = [s^2, sr, r^2, x]$. By Lemma 2.13, we have two cases. For the first case, $f'$ has type I. Then det $[s^2 - r^2, sr, x]$ = 0, by Proposition 2.10. This implies that $sr^2 = r^4$. If $sr \neq 0$, $sr^2 = r^4$ implies that $[s^2, sr, r^2, x]$ is degenerate,
a contradiction. If $sr = 0$, then $r = 0$ by $sr x = r^4$. Since $[s^2, sr, r^2, x] = [s^2, 0, 0, x]$ is non-degenerate, we must have $s^2 \neq 0$ and $x \neq 0$. Because $n > 3$, there must be a sub-signature of $f$ of the form $[s^2, 0, 0, x, z]$ or $[z', s^2, 0, 0, x]$. Suppose we have $[s^2, 0, 0, x, z]$; the other case is similar. Then consider the pair $[s^2, 0, 0, x]$ and $[0, x, z]$. If $z \neq 0$, then Holant$^*(\{[s^2, 0, 0, x], [0, x, z]\})$ is #P-hard by Lemma 2.14. If $z = 0$, then we have a sub-signature $[0, 0, x, 0]$. Holant$^*(\{0, 0, x, 0}\})$ is also #P-hard by Lemma 2.13. Finally for the second case, $f'$ has type II: $s^2 + r^2 = 0$ and $sr + x = 0$. Then the signature must be $[s^2, s^2i, -s^2, -s^2i]$ or $[s^2, -s^2i, -s^2, s^2i]$. Both are degenerate, a contradiction. □

2.3.4 Putting Things Together

Finally we extend these lemmas to handle a set of signatures and finish the proof of Theorem 2.12. The main idea is to find a non-degenerate signature of arity 3, and anchor the proof on this arity 3 signature, namely to show that everything else must conform to it, or else the problem is #P-hard.

Proof of Theorem 2.12: The tractability part has already been proved. We prove the hardness part. We assume every signature in $\mathcal{F}$ is non-degenerate and has arity at least 2.

Suppose Holant$^*(\mathcal{F})$ is not #P-hard. We want to show that $\mathcal{F}$ must fall in one of the three Classes.

If no signature of $\mathcal{F}$ has arity $\geq 3$, then we are done, with Class $\mathcal{A}$. Suppose there are some signatures of arity $\geq 3$. Each such signature is of type I or II by Lemma 2.13 and Lemma 2.16. First suppose all signatures in $\mathcal{F}$ of arity $\geq 3$ are of the same type, namely either for some $(a, b)$ they are all of type I($a, b$) or they are all of type II. If they are all of type I($a, b$), then we may obtain a non-degenerate signature of arity three having the same type by either taking a non-degenerate sub-signature (Proposition 2.8), or connecting it with the unary $[1, 1]$ if it is of type I(0, 1). Lemma 2.14 finishes the proof in this case and we are in Class $\mathcal{B}$. If all signatures in $\mathcal{F}$ of arity $\geq 3$ are of type II, again we can obtain a non-degenerate sub-signature of arity three. Lemma 2.15 then finishes the proof and we are in Class $\mathcal{C}$.

In the following we assume there are more than one types of signatures in $\mathcal{F}$ having arity $\geq 3$. Each such signature declares a type I or II.

- Case 1. Suppose there is a non-degenerate type I(0, 1) signature in $\mathcal{F}$ of arity $\geq 3$. By connecting it with the unary $[1, 1]$ we may assume we have a non-degenerate type I(0, 1) signature $f \in \mathcal{F}$ of arity three. By assumption, there exists another non-degenerate signature $g$ of arity $\geq 3$ and of a different type. By Proposition 2.8, there exists a non-degenerate sub-signature $[x, y, z]$ of arity two. By being not #P-hard, we apply Lemma 2.14 to the pair $f$ and $[x, y, z]$ and conclude that there are two cases: $[x, y, z] = [x, 0, z]$, or $[0, y, 0]$. Being non-degenerate, $xz \neq 0$ in the first case, and $y \neq 0$ in the second case. Within $g$ there is a sub-signature containing $[x, y, z]$. In the first case, it has the form $[w, x, 0, z]$ (or $[x, 0, z, w]$),
then \([w, x, 0]\) (or \([0, z, w]\)) is non-degenerate. We apply Lemma 2.14 once again to \(f\) and \([w, x, 0]\) (or \([0, z, w]\)), and get \(w = 0\). In the second case, it has the form \([u, 0, y, 0]\) (or \([0, y, 0, u]\)). If \(u = 0\), Holant\(^*\)\(([0, 0, y, 0])\) (or Holant\(^*\)\(([0, y, 0, 0])\)) is \#P-hard by Lemma 2.13. Hence in all cases we get a signature of the form \([u, 0, v, 0]\) (or its reversal \([0, v, 0, u]\)) where \(uv \neq 0\). By symmetry we consider \([u, 0, v, 0]\), as \([0, v, 0, u]\) is changed to \([u, 0, v, 0]\) by a 0-1 exchange, and the type I(0,1) signature \(f\) retains its type. We connect a unary \([s, t]\) to \([u, 0, v, 0]\), and get \([su, tv, sv]\). As long as \(st \neq 0\) and \(s^2/t^2 \neq v/u\) this is non-degenerate and Lemma 2.14 on the pair \(f\) and \([su, tv, sv]\) leads to \#P-hardness.

- **Case 2.** Now we assume Case 1 does not apply, but there exists a non-degenerate type I(a,b) signature in \(\mathcal{F}\) of arity \(\geq 3\), where \(a \neq 0\) and \(b \neq \pm 2ai\). By Proposition 2.8, we may assume there is a non-degenerate type I(a,b) signature \(f\) having arity three. We may normalize \((a, b)\) to \(a = 1\), and \(b \neq \pm 2i\).

Then the characteristic equation \(X^2 - bX - 1\) has two distinct roots \(\lambda \neq \mu\). It follows that \(\lambda \mu = -1\), and \(\lambda, \mu \neq \pm i\). Define \(T = \begin{bmatrix} \frac{1}{\sqrt{1+\lambda^2}} & \frac{1}{\sqrt{1+\mu^2}} \\ \frac{\lambda}{\sqrt{1+\lambda^2}} & \frac{\mu}{\sqrt{1+\mu^2}} \end{bmatrix}\), then \(T\) is orthogonal. It is easy to verify that \(f = T^{\circ 3} f'\) for some \(f' = [x_0, 0, 0, x_3]\), a non-degenerate signature of type I(0,1). Performing a holographic transformation by \(T\) on all signatures in \(\mathcal{F}\), we have reduced Case 2 to Case 1.

- **Case 3.** We assume Cases 1 and 2 do not apply, but there exists a non-degenerate type I(1,b) signature in \(\mathcal{F}\) of arity \(\geq 3\), where \(b = \pm 2i\). By Proposition 2.8, we may assume such a signature \(f\) of arity three. We still suppose there exists a non-degenerate signature \(g\) of arity \(\geq 3\) and of a different type; for otherwise we are done. \(g\) could be of type I(1,b') or II, where \(b' = \mp 2i\) (but \(b' \neq b\), and hence \(b' = -b\)). By Proposition 2.8, there exists a non-degenerate sub-signature \([x, y, z]\) of \(g\).

If \(g\) has type I(1,b'), then \(x + b'y - z = 0\). Note that for \(b = \pm 2i\), the binary signature \([2, b, -2]\) is degenerate. Since \([x, y, z]\) is non-degenerate, it cannot be \(\lambda[2, b, -2]\). Hence, by applying Lemma 2.14 to \(f\) and \([x, y, z]\), we get \(x + by - z = 0\). This implies that \(y = 0\) since \(b' \neq b\), and \(x = z \neq 0\) by the non-degeneracy of \([x, y, z]\). But within \(g\) there is a sub-signature \([w, x, 0, x]\) or \([x, 0, x, w]\). As \(x \neq 0\), \([w, x, 0]\) and \([0, x, w]\) are non-degenerate. Applying Lemma 2.14 once again we get a contradiction. If \(g\) is of type II, then we still have \(x + by - z = 0\) as well as \(z = -x\). Since \(b = \pm 2i\), this implies that \([x, y, z]\) is degenerate, also a contradiction.

If Cases 1, 2 and 3 all do not apply, then all non-degenerate signatures in \(\mathcal{F}\) of arity \(\geq 3\) are of type II. This case has been proved before; we are in Class \(\mathfrak{C}\). Theorem 2.12 is proved. \(\square\)
2.4 The Road Ahead

Theorem 2.12 is our first major dichotomy theorem. While it covers a fairly broad class of problems, we should also recognize its limitations. There are at least three aspects in which Theorem 2.12 is restricted.

The first is the restriction to a symmetric set of constraint functions $\mathcal{F}$. One can ask what happens when we remove this restriction of symmetry. The answer will be given in Chapter ?? where we present a dichotomy theorem, Theorem ??, for Holant* problems applicable to a set of constraint functions $\mathcal{F}$ that are not necessarily symmetric. In so doing we will further reveal an alternative perspective on the classification achieved here. However to arrive at Theorem ?? we will have to develop the theory much further; in particular, the proof of Theorem ?? will use the symmetric case Theorem 2.12.

The second is the assumption of the presence of all unary functions $\mathcal{U}$ for free in the definition of Holant* problems. One can ask what happens if this assumption is removed. In Chapter 7 we will present a dichotomy for Holant problems on symmetric signatures without assuming any freely available auxiliary functions, in particular unary functions. The proof will utilize a substantial part of the theory to be developed in the next few chapters. We develop this theory starting with free unary functions. The theory will also consider other freely available auxiliary functions. For example when $\text{EQUALITY}$ functions of all arities are assumed to be free, namely Boolean $\#\text{CSP}$ problems, which we turn to next in Chapter 3. We will also consider the case where only two special unary functions $\Delta_0$ and $\Delta_1$ are free. This is called Holant$^c$ problems. It is an interesting open problem when we remove both the symmetry restriction and the presence of unary functions. It is open even for Holant$^c$ problems when the symmetry restriction is removed. However, the study of Holant* problems by assuming unary functions $\mathcal{U}$ are free is more than an arbitrary choice. As we develop the theory, aside from explicitly using Theorem 2.12, the idea of simulating the free presence of unary functions is a major theme.

Finally the third limitation is that we restrict our study to the Boolean domain in Part I of this book. In Part II we will develop the theory on a general finite domain. However the theory on general domains is not as well-developed as on the Boolean domain. In particular, the analogue of Theorem 2.12 is still open for the general domain.
Chapter 3

Boolean \#CSP

In this chapter we address the following type of counting problems, called Boolean \#CSP. Let $F$ be a set of functions, where each $F \in F$ is a function $F : \{0, 1\}^k \rightarrow \mathbb{C}$, for some $k$ depending on $F$, mapping Boolean variables to the complex numbers. The functions in $F$ are not assumed to be symmetric in general in this Chapter. The Boolean counting constraint satisfaction problem \#CSP$(F)$ is defined as follows: The input $I$ is a finite sequence of constraints on Boolean variables $x_1, x_2, \ldots, x_n$ of the form $F(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$, where $F \in F$. The output is called the partition function

$$Z(I) = \sum_{x_1, x_2, \ldots, x_n \in \{0, 1\}} \prod F(x_{i_1}, x_{i_2}, \ldots, x_{i_k}),$$

where the product is over all constraints occurring in $I$. When $F = \{F\}$ consisting of a single constraint function, we denote \#CSP$(\{F\})$ simply as \#CSP$(F)$.

3.1 The 0-1 Case and Non-Negative Boolean \#CSP

In general, functions $F \in F$ can take arbitrary complex values. If each $F$ takes values 0, 1, then it is identified with a Boolean valued constraint. In that case, $Z(I)$ counts the number of assignments satisfying all the Boolean constraints of $I$.

Typically $F$ is a finite set and considered fixed. The complexity is measured in terms of the input size of $I$. For a finite and fixed $F$ this input size is equivalent to $n$, up to a polynomial factor. Occasionally we will also allow $F$ to be infinite, such as those sets containing all unary functions $U$, in which case the input size must include the description of the constraints used in the input.

If each $F \in F$ takes values 0, 1, then we have the following theorem.

**Theorem 3.1** (Creignou-Hermann). For every finite $F$ consisting of 0-1 valued functions on Boolean variables, \#CSP$(F)$ is \#P-complete, unless each $F \in F$ is the 0-1 indicator function of some affine linear subspace over $\mathbb{Z}_2$, in which case it is in P.
A 0-1 indicator function of some affine linear subspace is just a Boolean conjunction of XOR functions on a (possibly empty) subset of variables and possibly a constant 1 ∈ Z₂. This includes the constant 0, 1 functions, the unary functions Δ₀ = [1, 0] and Δ₁ = [0, 1], the binary Disequality [0, 1, 0] and Equality of any arities (=k). When F has the form specified in the theorem, one can solve #CSP(F) in polynomial time by Gaussian elimination over the field Z₂.

**Definition 3.2.** A function has product type if it can be expressed as a product of unary functions, binary Equality functions ((=₂) = [1, 0, 1]) and binary Disequality functions ((/=₂) = [0, 1, 0]), on not necessarily disjoint subsets of variables. We denote by P the set of all functions of product type.

**Exercise:** Show that a symmetric signature is in P iff it is a Gen-Eq, or a (scalar multiple of) binary Dis-Equality, or a constant function.

**Definition 3.3.** A function is pure affine if it is a constant multiple of a 0-1 indicator function of some affine linear subspace over Z₂.

If all functions in F take non-negative values, then Theorem 3.1 can be generalized to the following theorem.

**Theorem 3.4** (Dyer-Goldberg-Jerrum). For every finite F consisting of non-negative valued functions, #CSP(F) is solvable in polynomial time in the following two cases, and is #P-hard otherwise:

1. Every function in F is of a product type; or
2. Every function in F is a pure affine function.

For either tractable cases, the polynomial time algorithm is quite simple. The main claim is that everything else is #P-hard. Theorem 3.4 is concerned with non-negative valued functions. However the tractability still holds without this restriction. In particular functions in P can be complex-valued, i.e., the unary functions in the definition of a function in P can be complex-valued. Any Gen-Eq is in P. Similarly the constant multiple in the definition of a pure affine function can be from C.

It is also clear that Theorem 3.1 is a special case of Theorem 3.4.

However, allowing real or complex-valued product type or pure affine functions in Theorem 3.4 turns out to be not sufficient to encompass all tractable cases when functions in F are no longer restricted to those taking non-negative values. When the functions F ∈ F take possibly negative or complex values, there are more non-trivial tractable #CSP(F) problems. This is made possible by algebraic cancelations, and they lead to novel efficient algorithms. The goal of this Chapter is to prove an explicit dichotomy theorem for #CSP(F), Theorem 3.7, for any complex-valued function set F. Theorem 3.4, and a fortiori Theorem 3.1, are special cases of this dichotomy theorem.
Before we move on to investigate Boolean #CSP in detail, we can restate Theorem 2.12 in terms of what can be transformed to product type $P$. Note that a symmetric signature is in $P$ if it is a GEN-EQ, or a binary DISEQUALITY, or a degenerate function.

**Theorem 3.5.** Let $F$ be a set of non-degenerate symmetric signatures over $C$. Then Holant$^*$($F$) is #P-hard, unless $F$ satisfies the following conditions, in which case it is computable in polynomial time.

1. All signatures in $F$ have arity at most 2.
2. There exists some $M \in \text{GL}_2(C)$ such that $(=)_2M^{\otimes 2} \in P$ and $F \subseteq M_\Delta$.
3. There exists $\lambda \in \{2i, -2i\}$, such that every signature $f \in F$ of arity $n$ satisfies the recurrence
   \[ f_{k+2} = \lambda f_{k+1} + f_k, \quad \text{for} \quad 0 \leq k \leq n - 2. \]

**Remark:** Case 1 here corresponds to Class $A$ of Theorem 2.12. Case 2 corresponds to all of Class $B$ (except type I$(1, \pm 2i)$) as well as Class $C$. Case 3 corresponds to type I$(1, \pm 2i)$ of Class $B$. As shown in Section 2.2, type I$(1, \pm 2i)$ is also equivalent to the following: under the $Z$ transformation, every $f \in F$ has the form $f = Z[c, d, 0, \ldots, 0]$, or every $f \in F$ has the form $f = Z[0, \ldots, 0, d, c]$, where $Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ \end{bmatrix}$. Recall also that under the covariant transformation by $Z$, the binary EQUALITY is transformed to the binary DISEQUALITY: $(=)_2Z^{\otimes 2} = (\neq)_2$.

**Exercise:** Show that there exists some $M \in \text{GL}_2(C)$ such that $(=)_2M^{\otimes 2}$ is a GEN-EQ and $F \subseteq M_\Delta$, iff there exists an orthogonal $T$ that does this transformation.

**Exercise:** Show that there exists some $M \in \text{GL}_2(C)$ such that $(=)_2M^{\otimes 2}$ is $(\neq)_2$ and $F \subseteq M_\Delta$, iff we may take $M = Z$ to do this transformation.

**Exercise:** Prove that Theorem 3.5 is an equivalent form of Theorem 2.12.

### 3.2 Affine Functions $A$ and $F_1 \cup F_2 \cup F_3$

There is a simple relation between #CSP and Holant problems. We can represent an instance of a #CSP problem by a bipartite graph $G$ where the left-hand side is labeled by variables and the right-hand side is labeled by constraints. We define a signature grid $\Omega$ on $G$ by assigning an EQUALITY function to every variable node on the left-hand side, and every constraint node on the right-hand side has the given constraint function. Then Holant$_\Omega$ is exactly the same as the #CSP problem. In effect, the EQUALITY function on each variable node forces the incident edges to take the same value; this effectively reduces edge assignments in Holant$_\Omega$ to vertex assignments on the left-hand side in the #CSP problem. Thus #CSP problems
are precisely the special case of Holant problems on bipartite graphs where every vertex on the left-hand side is assigned an equality function. The restriction on bipartiteness is not essential; given any constraint function set $\mathcal{F}$, it is easy to show that \#CSP($\mathcal{F}$) is exactly the same as Holant($\mathcal{F} \cup \mathcal{E}$), where $\mathcal{E} = \{=_{k} \mid k \geq 1\}$ is the set of all equality functions of any arity $k$. In other words, \#CSP problems are Holant problems with the implicit assumption that all equality functions are freely available. This is similar to Holant* problems where we assume all unary functions are freely available. In fact we will see that the presence of all equality functions already imply the presence of two special unary functions $\Delta_0$ and $\Delta_1$. Applying $\Delta_0$ or $\Delta_1$ on a variable is called pinning.

On the other hand, Holant problems can be considered as \#CSP problems where every variable appears twice. We will show that the following three families of functions are tractable. Below we write a function as a vector listing its values in lexicographic order of the assignments, which we will also call its truth table, and where $i = \sqrt{-1}$.

$$\mathcal{F}_1 = \{\lambda([1, 0]^\otimes k + i^r[0, \ 1]^\otimes k) \mid \lambda \in \mathbb{C}, k = 1, 2, \ldots, \text{ and } r = 0, 1, 2, 3\};$$

$$\mathcal{F}_2 = \{\lambda([1, 1]^\otimes k + i^r[1, -1]^\otimes k) \mid \lambda \in \mathbb{C}, k = 1, 2, \ldots, \text{ and } r = 0, 1, 2, 3\};$$

$$\mathcal{F}_3 = \{\lambda([1, i]^\otimes k + i^r[1, -i]^\otimes k) \mid \lambda \in \mathbb{C}, k = 1, 2, \ldots, \text{ and } r = 0, 1, 2, 3\}.$$

We explicitly list all the signatures in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ up to an arbitrary constant multiple from $\mathbb{C}$:

1. $[1, 0, \ldots, 0, \pm 1]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_1, r = 0, 2)$
2. $[1, 0, \ldots, 0, \pm i]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_1, r = 1, 3)$
3. $[1, 0, 1, 0, \ldots, 0 \text{ or } 1]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_2, r = 0)$
4. $[1, -i, 1, -i, \ldots, (-i) \text{ or } 1]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_2, r = 1)$
5. $[0, 1, 0, 1, \ldots, 0 \text{ or } 1]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_2, r = 2)$
6. $[i, i, 1, i, \ldots, i \text{ or } 1]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_2, r = 3)$
7. $[1, 0, -1, 0, 1, 0, -1, 0, \ldots, 0 \text{ or } 1 \text{ or } (-1)]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_3, r = 0)$
8. $[1, 1, -1, -1, 1, 1, -1, -1, \ldots, 1 \text{ or } (-1)]$; \hspace{1cm} \hspace{1cm} \hspace{1cm} $(\mathcal{F}_3, r = 1)$
9. $[0, 1, 0, -1, 0, 1, 0, -1, \ldots, 0 \text{ or } 1 \text{ or } (-1)]$; \hspace{1cm} \hspace{1cm} $(\mathcal{F}_3, r = 2)$
10. $[1, -1, -1, 1, 1, -1, -1, 1, \ldots, 1 \text{ or } (-1)]$. \hspace{1cm} \hspace{1cm} $(\mathcal{F}_3, r = 3)$

We can show that Holant$_0$ for any $\Omega = (G, \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$ is computable in $\mathbb{P}$. These functions are all related to each other by holographic reductions. We will prove this tractability by a more general tractability theorem which covers non-symmetric functions as well.

We note that expressions in complex numbers appear naturally, even for real-valued functions. The special case where $r = 1$, $k = 2$ and $\lambda = (1 + i)^{-1}$ in $\mathcal{F}_3$ is noteworthy. In this case we get a real-valued binary symmetric function $H = [1, 1, -1]$. In other words, $H(0, 0) = H(0, 1) = H(1, 0) = 1$ and $H(1, 1) = -1$. The matrix form of this function is the Hadamard matrix $H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Furthermore, if we take $r = 0$, any $k$ and $\lambda = 1$ in $\mathcal{F}_1$ we get the equality function on $k$ variables. If $\Omega = (G, \mathcal{F})$ with $\mathcal{F}$ consisting of exactly the
function $H$ and all \textsc{Equality} functions, then Holant$_0$ is computing the partition function of $\#\text{CSP}(H)$. This special case where the constraint function set in $\#\text{CSP}$ consists of a single binary (symmetric) function is also known as graph homomorphism $Z_H(G)$, or a spin system, on (undirected) graph $G$. What is the value of $Z_H(G)$? If $X$ (respectively $Y$) is the number of (vertex subset-) induced subgraphs of $G$ with an even (respectively odd) number of edges, then $Z_H(G) = X - Y$. Since trivially $X + Y = 2^n$, we have $X = \frac{2^n + Z_H(G)}{2}$ and $Y = \frac{2^n - Z_H(G)}{2}$, thus the problem $Z_H$ essentially computes the number of induced subgraphs with an even (respectively odd) number of edges.

The main result in this chapter is a complexity dichotomy theorem, Theorem 3.7, for complex valued Boolean $\#\text{CSP}$. The family $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ can be naturally generalized to include unsymmetric functions, called the affine family $\mathcal{A}$. In fact the family $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ consists of precisely the unary functions of $\mathcal{A}$ and symmetric non-denegerate functions of $\mathcal{A}$ of arity $\geq 2$. Of course unary functions are also symmetric (but are considered denegerate).

The dichotomy theorem says that a Boolean $\#\text{CSP}$ problem $\#\text{CSP}(\mathcal{F})$ is tractable if \textit{either} all its constraint functions $\mathcal{F}$ are of product type $\mathcal{D}$, \textit{or} all are from this affine family $\mathcal{A}$; otherwise it is $\#\text{P}$-hard.

Suppose $F$ is a function on input variables $x_1, x_2, \ldots, x_k$. For $1 \leq s \leq k$ and $c = 0, 1$, $F^{x_s=c}$ denotes the function $F^{x_s=c}(x_1, \ldots, x_{s-1}, x_{s+1}, \ldots, x_k) = F(x_1, \ldots, x_{s-1}, c, x_{s+1}, \ldots, x_k)$, and $F^{x_s=s}$ denotes the function $F^{x_s=s}(x_1, \ldots, x_{s-1}, x_{s+1}, \ldots, x_k) = \sum_{x_s=0,1} F(x_1, \ldots, x_k)$.

A function of arity $k$ can be expressed by its truth table of length $2^k$. Define

$$\mathcal{D} = \{ F \mid F = [a_1, b_1] \otimes [a_2, b_2] \otimes \cdots \otimes [a_k, b_k], \text{ for some } a_i, b_i \in \mathbb{C} \}$$

to be the set of functions that can be expressed as a tensor product of unary functions, that is, a function in $\mathcal{D}$ on $k$ variables is the product of $k$ unary functions applied to its $k$ variables separately. Functions in $\mathcal{D}$ are called \textit{degenerate}. This notion of degeneracy agrees with the definition of degeneracy for symmetric functions. A binary function is in $\mathcal{D}$ iff its corresponding matrix is singular. Note that $\mathcal{D}$ is a subset of $\mathcal{D}$, the signatures of product type.

For a constraint function $F$ we define its underlying relation, also called its support, by

$$R_F = \{ X \in \{0,1\}^k \mid F(X) \neq 0 \}.$$

We say a relation $R \subseteq \{0,1\}^k$ is affine if it is an affine subspace over $\mathbb{Z}_2$ (including possibly the empty space). It is composed of solutions to a system of linear equations. Equivalently, it satisfies the property that if $\alpha, \beta, \gamma \in R$, then the bit-wise XOR string $\alpha \oplus \beta \oplus \gamma \in R$. To see this, first suppose $R$ is the set of solutions to $AX = r$ for some matrix $A$ and vector $r$ over $\mathbb{Z}_2$. Then clearly for any $\alpha, \beta, \gamma \in R$, $A(\alpha \oplus \beta \oplus \gamma) = r \oplus r \oplus r = r$, and $\alpha \oplus \beta \oplus \gamma \in R$. Now suppose $R$ satisfies this property. Suppose $R \neq \emptyset$. Define $L_R = \{ \alpha \oplus \beta \mid \alpha, \beta \in R \}$. Then $L_R$ is a linear subspace of $\mathbb{Z}_2^2$. Take any $\alpha_0 \in R$. Then $R \subseteq L_R + \alpha_0 = \{ X \oplus \alpha_0 \mid X \in L_R \}$, since for any $\alpha \in R$, $\alpha = \alpha \oplus \alpha_0 \oplus \alpha_0$. Also $L_R + \alpha_0 \subseteq R$ by the fact that $R$ satisfies the property.
If \( R_F \) is affine, we say \( F \) has affine support. We also view relations as functions from \( \{0, 1\}^k \) to \( \{0, 1\} \).

Now we define the affine family of functions \( \mathcal{A} \). Let \( X \) denote the \( k + 1 \) dimensional column vector \( (x_1, x_2, \ldots, x_k, 1)^T \) over the field \( \mathbb{Z}_2 \). Suppose \( A \) is a matrix over \( \mathbb{Z}_2 \). \( \chi_{AX} \) denotes the affine relation on inputs \( x_1, x_2, \ldots, x_k \), whose value is 1 if \( AX \) is the zero vector, and 0 otherwise.

**Definition 3.6.** A function is of affine type if it can be expressed as

\[
\lambda \cdot \chi_{AX} \cdot L_1(X) + L_2(X) + \cdots + L_n(X),
\]

where \( \lambda \in \mathbb{C} \), \( i = \sqrt{-1} \), each \( L_j \) is an integer 0-1 indicator function of the form \( \langle \alpha_j, X \rangle \), where \( \alpha_j \) is a \( k + 1 \) dimensional vector over \( \mathbb{Z}_2 \) and the dot product \( \langle \cdot, \cdot \rangle \) is computed over \( \mathbb{Z}_2 \). The set of all functions of affine type is denoted by \( \mathcal{A} \).

We may compute the dot product as an ordinary integer dot product, and then take the value mod 2, producing an integer value 0 or 1. The additions among \( L_j(X) \) are the usual addition in \( \mathbb{Z} \). It can be computed mod 4, but not mod 2. We usually will omit any non-zero constant factor \( \lambda \), since they do not affect complexity, and can be ignored.

### 3.3 A Dichotomy for Boolean \#CSP

**Theorem 3.7.** Suppose \( \mathcal{F} \) is a set of functions mapping Boolean inputs to complex numbers. If \( \mathcal{F} \subseteq \mathcal{A} \) or \( \mathcal{F} \subseteq \mathcal{P} \), then \#CSP(\( \mathcal{F} \)) is computable in polynomial time. Otherwise, \#CSP(\( \mathcal{F} \)) is \#P-hard.

**Proof Outline:** The polynomial time algorithm for \#CSP(\( \mathcal{P} \)) is easy. Section 3.4 gives a polynomial time algorithm for \#CSP(\( \mathcal{A} \)). Cancelations when summing for functions in \( \mathcal{A} \) play an essential role. The starting point of the hardness result is Theorem 3.14, which says that if \( \mathcal{F} \) contains only one binary symmetric function and is not in \( \mathcal{A} \cup \mathcal{P} \), then the \#CSP problem is \#P-hard. In Lemma 3.18, we prove that \#CSP(\( \mathcal{F} \)) is \#P-hard unless \( \mathcal{F} \) has affine support. This structure is essential in the proof of Lemma 3.19 and Lemma 3.20, the two key lemmas of the hardness reduction. The common strategy of Lemma 3.19 and Lemma 3.20 is to reduce the arity of a given function. In Lemma 3.19, we prove that given a function \( \mathcal{F} \) not in \( \mathcal{A} \), we can simulate (in polynomial time) a unary function \( \mathcal{F}' \notin \mathcal{A} \); In Lemma 3.20, we prove that given a function \( \mathcal{G} \) not in \( \mathcal{P} \), we can simulate (in polynomial time) a binary or a ternary function \( \mathcal{G}' \notin \mathcal{P} \). Then we prove that \#CSP(\( \{\mathcal{F}', \mathcal{G}'\} \)) is \#P-hard. To prove this, we show that we can always combine functions \( \mathcal{F}' \) and \( \mathcal{G}' \) to realize a binary symmetric function which is not in \( \mathcal{A} \cup \mathcal{P} \).

**Exercise:** The tractability criterion in Theorem 3.7 should not be confused with the statement \( \mathcal{F} \subseteq \mathcal{A} \cup \mathcal{P} \). Show that there are \( \mathcal{F} \) satisfying the above statement and yet \#CSP(\( \mathcal{F} \)) is \#P-hard.
There is an alternative expression of a function in \( \mathcal{A} \). We claim that the expressibility as

\[ \lambda \chi_{\mathcal{A}X} i^{L_1(X)+L_2(X)+\ldots+L_n(X)} \]

is equivalent to an expression of the form \( \lambda \chi_{\mathcal{A}X} i^{Q(X)} \) where \( Q \) is a homogeneous quadratic polynomial in \( \mathbb{Z}[X] \) with the additional requirement that every cross term \( x_s x_t \) has an even coefficient, where \( s \neq t \). To see this we observe that each \( L_j(X) \) as an integer sum mod 2 can be replaced by \( (L_j(X))^2 \) as an integer sum mod 4, since \( L_j(X) = 0, 1 \) (mod 2) iff \( (L_j(X))^2 = 0, 1 \) (mod 4). After this, all cross terms have an even coefficient, and \( i^{x_s} \) can be replaced by \( i^{x_s^2} \), for \( x_s = 0, 1 \). Conversely, we can express \( Q \) mod 4 as a sum of squares of affine forms of \( X \), using the extra condition that all cross terms have even coefficients.

**Exercise:** Prove that the class of function \( \mathcal{A} \) is not changed by the following modifications in its definition:

- \( Q \) is a polynomial of degree at most 2 (with possible linear and constant terms, but cross terms are still required have even coefficients).
- \( Q \) is a multilinear polynomial of degree at most 2, where cross terms have even coefficients.

**Exercise:** Prove that a function \( \mathcal{A} \) is not changed by the following modifications in its definition:

- \( Q \) is a polynomial of degree at most 2 (with possible linear and constant terms, but cross terms are still required have even coefficients).
- \( Q \) is a multilinear polynomial of degree at most 2, where cross terms have even coefficients.

**Exercise:** Prove that a function \( \{0,1\}^k \rightarrow \mathbb{Z}_4 \) has a unique multilinear polynomial expression. In particular, suppose \( f = \lambda \chi_{\mathcal{A}X} i^{Q(X)} \), where \( Q(X) \) is a multilinear polynomial on free variables \( X \) of its support. If \( Q \) has degree greater than 2, or has a cross term with an odd coefficient, then \( f \notin \mathcal{A} \).

**Exercise:** Derive a normal form for any real valued affine function \( f \in \mathcal{A} \).

The class \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \) is the restriction of \( \mathcal{A} \) to symmetric functions. More precisely, \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \) consists of scalar multiples of all unary or non-degenerate symmetric functions in \( \mathcal{A} \). We first show that \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \subseteq \mathcal{A} \). Omitting the non-zero constants, the functions in \( \mathcal{F}_1 \) can be expressed as \( \chi_{x_1=x_2=\ldots=x_k} i^{|x|^2/2} \). The functions in \( \mathcal{F}_2 \) can be expressed as \( \chi_{\sum_j x_j=0} \), or \( \chi_{\sum_j x_j=1} \), or \( i^{(\sum_j x_j)^2} \), or \( (-i)^{(\sum_j x_j)^2} = i^3(\sum_j x_j)^2 \), respectively. The functions in \( \mathcal{F}_3 \) can be expressed as \( \chi_{\sum_j x_j=0} i^{\sum_j x_j^2/2} \), or \( \chi_{\sum_j x_j=1} i^{\sum_j x_j^2-1} = -i\chi_{\sum_j x_j=1} i^{\sum_j x_j^2} \), or \( (-1)^{\sum_{s<t} x_s x_t} = i^{2\sum_{s<t} x_s x_t} \), or \( (-1)^{\sum_{s<t} x_s x_t + \sum_j x_j} = i^{2\sum_j x_j^2 + 2\sum_{s<t} x_s x_t} \), respectively.

Conversely suppose \( f(x_1, \ldots, x_k) \in \mathcal{A} \) is symmetric. Assume \( f \) is not identically zero, otherwise it is trivial. Let \( S = \{ x \in \mathbb{Z}_2^k \mid f(x) \neq 0 \} \) be its affine support. Consider all affine linear equations over \( \mathbb{Z}_2 \) that are valid on \( S \), which must be closed under permutations of variables. If the set of equations is empty, then \( S = \mathbb{Z}_2^k \). Assume the set of equations is non-empty. By symmetry, if there is an equation involving only one variable \( x_i = \epsilon \), for some \( 1 \leq i \leq k \), then for all \( 1 \leq j \leq k \), \( x_j = \epsilon \). Thus \( S = \{(0,\ldots,0)\} \) or \( \{(1,\ldots,1)\} \). Assume no equation has only one variable, but there is an equation of two variables \( x_i + x_j = \epsilon \), for some
1 ≤ i < j ≤ k, then by symmetry, this holds for all i ≠ j. If k > 2 and ϵ = 1, we would have $x_1 + x_2 = 1$ and $x_1 + x_2 = (x_1 + x_3) + (x_2 + x_3) = 0 \mod 2$, and S would be empty. Thus for $k > 2$, we must have ϵ = 0 and we have $x_1 = x_2 = \ldots = x_k$. Thus $S = \{(0, \ldots , 0), (1, \ldots , 1)\}$. If k = 2, additionally we also have possibly $S = \{(0, 1), (1, 0)\}$. Suppose all equations involve at least 3 variables. Then k ≥ 3, and every equation must involve all k variables. Otherwise, we have $x_1 + \ldots + x_j = ϵ$ and $x_1 + \ldots + x_{j-1} + x_{j+1} = ϵ$, for some j < k, whose sum gives $x_j + x_{j+1} = 0$ having only two variables. Thus $S = \{x \in \mathbb{Z}_2^k \mid \sum_j x_j = ϵ \mod 2\}$.

Let $f = \chi_S i^{Q(x_1, \ldots , x_k)} \in \mathcal{A}$ be symmetric, where $Q(x_1, \ldots , x_k) = \sum_j a_j x_j^2 + 2 \sum_{s < t} b_{s,t} x_s x_t$. Assume f is either unary or not degenerate if it has arity at least two. The case when the support S has at most two points is easy to prove. Suppose $S = \mathbb{Z}_2^k$. Then it is easy to see that all $a_j$ are the same, and all $b_{s,t}$ are the same, by setting one or two $x_j = 1$ and the other $x_j$’s to 0. Then $Q(x_1, \ldots , x_k)$ takes the form $a \sum_j x_j^2 + 2b \sum_{s < t} x_s x_t = a(\sum_j x_j)^2 + 2(b - a) \sum_{s < t} x_s x_t$. If we denote by $\Sigma = \sum_j x_j$ as an integer sum, and we denote by $(\Sigma \mod 2)$ the integer value 0 or 1, then $i^Q = i^{a(\Sigma \mod 2)}(-1)^{(b-a)((\Sigma^2) \mod 2)}$. By listing the values according to $\Sigma$, we find that either the function f is degenerate, or $f \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$.

Now suppose S is specified by $\sum_j x_j = ϵ \mod 2$, where ϵ = 0, 1. We first substitute $x_k = ϵ + \sum_{j=1}^{k-1} x_j \mod 2$ in the linear forms $L_j$ in the expression $i^{L_1(X) + L_2(X) + \ldots + L_n(X)}$, which can then be turned into the form $i^{Q'(x_1, \ldots , x_{k-1})}$, where $Q' = \sum_{j=1}^{k-1} a'_j x_j^2 + 2 \sum_{1 \leq s < t \leq k-1} b'_{s,t} x_s x_t$ is a quadratic form as above. The remaining proof is similar.

Suppose f is a symmetric degenerate function. Then f has the form $[a, b]^{\otimes n}$ for some unary function $[a, b]$ and n ≥ 1. If either $a = 0$ or $b = 0$, then clearly $f \in \mathcal{A}$, and also $[a, b] \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. Suppose $ab \neq 0$. Then f is a non-zero scalar multiple of $[1, b/a]^{\otimes n}$. Note that the ratio of any two non-zero entries of a function in $\mathcal{A}$ is a power of i, it follows that f belongs to $\mathcal{A}$ iff $b/a$ is a power of i, and also iff $[1, b/a] \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. We conclude that for any $f = [a, b]^{\otimes n}$, $f \in \mathcal{A}$ iff $[a, b] \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$.

We summarize the above discussion as the following theorem.

**Theorem 3.8.** A function f belongs to $\mathcal{A}$ iff it can be expressed as $\lambda \chi_{AX} i^{Q(x_1, \ldots , x_k)}$ where Q is a homogeneous quadratic polynomial over $\mathbb{Z}$ with the additional requirement that every cross term $x_s x_t$ has an even coefficient, where s ≠ t. We may also use all, not necessarily homogeneous, polynomials over $\mathbb{Z}$ of degree at most 2, with the same requirement on cross terms.

The class $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ consists of scalar multiples of all unary and all non-degenerate symmetric functions in $\mathcal{A}$. For any degenerate $f = [a, b]^{\otimes n}$, $f \in \mathcal{A}$ iff $[a, b] \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$.

### 3.4 Tractable Cases

We first show that #CSP($\mathcal{P}$) is tractable. Each constraint function in an instance I of #CSP($\mathcal{P}$) is a product of unary functions, binary EQUALITY functions (=2) and binary...
DISEQUALITY functions ($\neq 2$). We may replace each function by its factors as separate constraints. For the new instance of the \#CSP, define a graph $G$: The vertices are variables, and the edges are ($=2$). On each connected component we may assume no ($\neq 2$) is applied to two variables; otherwise $Z(I) = 0$. Now define another graph $G'$: The vertices are the connected components of $G$, and there is an edge between two components $C_1$ and $C_2$ iff there is a constraint ($\neq 2$)$(x,y)$ applied on some $x$ in $C_1$ and $y$ in $C_2$. Then either $G'$ is bipartite or $Z(I) = 0$. Now suppose $G'$ is bipartite. Take the union of components of $G$ belonging to the same connected component of $G'$. Then there are exactly two truth assignments to all the variables in this union, propagated by ($=2$) or ($\neq 2$). We can easily calculate these two values by multiplying the functional evaluations, taking into account of the unary functions. From these we can calculate $Z(I)$. Hence, \#CSP($\mathcal{P}$) is computable in polynomial time.

Now we analyze \#CSP($\mathcal{A}$). We will ignore the global constant factor $\lambda$ for each $F \in \mathcal{A}$. Firstly, we show how to get rid of the factor $\chi_{AX}$.

**Lemma 3.9.** Let $F(x_1, x_2, \ldots, x_k) = \chi_{AX}^{L_1(X)+L_2(X)+\cdots+L_n(X)} \in \mathcal{A}$. If $AX = 0$ is infeasible over $\mathbb{Z}_2$, then $\sum_{x_1,x_2,\ldots,x_k} F = 0$. Suppose $AX = 0$ is feasible. Then in polynomial time, we can construct another function $H(y_1, y_2, \ldots, y_s) = i^{L_1(Y)+L_2(Y)+\cdots+L_n(Y)} \in \mathcal{A}$, such that $0 \leq s \leq k$, and $\sum_{x_1,x_2,\ldots,x_k} F = \sum_{y_1,y_2,\ldots,y_s} H$.

**Proof.** In polynomial time we can solve the linear system $AX = 0$ over $\mathbb{Z}_2$, and decide if it is feasible. Suppose $AX = 0$ is feasible. Without loss of generality, we can assume that \{y_1, y_2, \ldots, y_s\} $\subseteq$ \{x_1, x_2, \ldots, x_k\} is a set of independent variables over $\mathbb{Z}_2$ and the others are dependent variables, where $0 \leq s \leq k$. Each dependent variable can be expressed by an affine linear sum of $y_1, y_2, \ldots, y_s$. For each $L_j(X)$, we can substitute all the dependent variables and get a 0-1 indicator function of $y_1, y_2, \ldots, y_s$, which we denote by $L'_j(Y)$. So we have

$$\sum_{x_1,x_2,\ldots,x_k} \chi_{AX}^{L_1(X)+L_2(X)+\cdots+L_n(X)} = \sum_{y_1,y_2,\ldots,y_s} i^{L'_1(Y)+L'_2(Y)+\cdots+L'_n(Y)}.$$

The following lemma gives a key property of the function $i^{L_1(X)+L_2(X)+\cdots+L_n(X)}$. This property plays an important role in both the tractability proof and the hardness proof.

**Lemma 3.10.** Let $F(x_1, x_2, \ldots, x_k) = i^{L_1(X)+L_2(X)+\cdots+L_n(X)}$. Let $G(x_2, \ldots, x_k) = F_{x_1=1}/F_{x_1=0}$. Then exactly one of the following two statements hold:

1. (Congruity) There exists a constant $c \in \{1, -1, i, -i\}$ such that for all $x_2, x_3, \ldots, x_k \in \{0, 1\}$ we have $G(x_2, x_3, \ldots, x_k) = c$;

2. (Semi-congruity) There exists a constant $c \in \{1, i\}$ and an affine subspace $S$ of $\mathbb{Z}_2^{k-1}$ with $\dim S = k - 2$, such that

$$G(x_2, x_3, \ldots, x_k) = c \text{ on } S, \quad \text{and } G(x_2, x_3, \ldots, x_k) = -c \text{ on } \mathbb{Z}_2^{k-1} - S.$$
Proof. If for every $1 \leq j \leq n$, the coefficient of $x_1$ is zero in $L_j(X)$, then $G$ is a constant 1. Otherwise, without loss of generality, suppose the coefficients for $x_1$ is nonzero in exactly the first $m$ affine linear forms $L_j(X)$. Obviously, the other $L_j(X)$’s cancel in the ratio $G = F^{x_1=1}/F^{x_1=0}$.

For any assignment to $x_2, x_3, \ldots, x_k$, consider the two assignments obtained by extending to $x_1 = 0$ and $x_1 = 1$, namely $(0, x_2, x_3, \ldots, x_k)$ and $(1, x_2, x_3, \ldots, x_k)$. For each $1 \leq j \leq m$, $L_j(1, x_2, x_3, \ldots, x_k) = 1 - L_j(0, x_2, x_3, \ldots, x_k)$. Therefore the ratio

$$G = F^{x_1=1}/F^{x_1=0} = \prod_{j=1}^{m} i^{1-2L_j(0, x_2, x_3, \ldots, x_k)} = i^m(-1)^{\sum_{j=1}^{m} L_j(0, x_2, x_3, \ldots, x_k)}.$$  

Here $m$ is independent of the assignment on $x_2, x_3, \ldots, x_k$. Since the base is $-1$ now, the sum can be evaluated as a sum mod 2. Therefore there is an affine linear form $\alpha(X) = \sum_{\ell=2}^{k} \alpha_\ell x_\ell + \alpha_{k+1} \pmod{2}$, such that $G = i^m(-1)^{\alpha(X)}$.

If all $\alpha_\ell = 0$, for $2 \leq \ell \leq k$, then this ratio is a constant and we are in the case of Congruity. If $\alpha_\ell = 1$, for some $2 \leq \ell \leq k$, then we have Semi-congruity.

Exercise: Prove Lemma 3.10 using the expression $F(X) = i^{Q(X)}$, where $Q(X)$ is a multilinear polynomial of degree at most 2 and all cross terms have even coefficients.

**Theorem 3.11.** $\#\text{CSP}(\mathcal{A})$ is polynomial time computable.

**Proof.** We first observe that $\mathcal{A}$ is closed under multiplication. Therefore given an instance of $\#\text{CSP}(\mathcal{A})$, the value of the output can be expressed as the summation on a single function $F = \chi_{AX}i^{L_1(X)+L_2(X)+\ldots+L_n(X)} \in \mathcal{A}$. We also note that if $F \in \mathcal{A}$, so is $F^{x_1=c}$, for $c = 0, 1$.

In each step of our algorithm, we reduce the number of variables by at least one and still get a summation of this form.

If the linear system $AX = 0$ over $\mathbb{Z}_2$ is infeasible, the function is a totally zero function and we just output 0. Otherwise by Lemma 3.9 we can remove the factor $\chi_{AX}$ and possibly decrease the number of variables at the same time, and assume it has the form $F = i^{L_1(X)+L_2(X)+\ldots+L_n(X)}$. We apply Lemma 3.10 to remove $x_1$. There are three cases.

Case 1: We have Congruity in Lemma 3.10. Then $F^{x_1=1}/F^{x_1=0}$ is a constant $c$, and

$$\sum_{x_1, x_2, \ldots, x_k} F = (1 + c) \cdot \sum_{x_2, x_3, \ldots, x_k} F^{x_1=0}.$$  

So we get a new summation $\sum_{x_2, x_3, \ldots, x_k} F^{x_1=0}$ and have removed a variable $x_1$.

Case 2: We have Semi-congruity in Lemma 3.10, and $c = 1$. Then on the affine subspace $S$, the ratio $F^{x_1=1}/F^{x_1=0} = 1$, and on the complementary subspace $\mathbb{Z}_2^{k-1} - S$ the ratio
\[ F^{x_1=1}/F^{x_1=0} = -1. \] For all \((x_2, x_3, \ldots, x_k) \in \mathbb{Z}_2^{k-1} - S\), the terms cancel, \(F^{x_1=1}(x_2, x_3, \ldots, x_k) + F^{x_1=0}(x_2, x_3, \ldots, x_k) = 0\). On \(S\), the terms are equal. It follows that

\[
\sum_{x_1, x_2, \ldots, x_k} F = 2 \sum_{x_2, x_3, \ldots, x_k} \chi_S F^{x_1=0}.
\]

Note that \(\chi_S F^{x_1=0}\) is also a function in \(\mathcal{A}\), so we get a new summation of this form and have removed a variable \(x_1\).

Case 3: We have Semi-congruity in Lemma 3.10, and \(c = i\). Then for all \((x_2, x_3, \ldots, x_k)\) in the affine subspace \(S\), we have \(F^{x_1=1}/F^{x_1=0} = i\), and in \(\mathbb{Z}_2^{k-1} - S\), we have \(F^{x_1=1}/F^{x_1=0} = -i\). It follows that

\[
\sum_{x_1, x_2, \ldots, x_k} F = \sum_S (1 + i) F^{x_1=0} + \sum_{\mathbb{Z}_2^{k-1} - S} (1 - i) F^{x_1=0}.
\]

Now we make a simple but crucial observation. The ratio of \(1 + i\) and \(1 - i\) is exactly \(i\), i.e., \(i = \frac{1+i}{1-i}\). As a result we can rewrite the two sums as follows:

\[
\sum_{x_1, x_2, \ldots, x_k} F = \sum_S (1 - i) \cdot F^{x_1=0} \cdot i^{L(X')} + \sum_{\mathbb{Z}_2^{k-1} - S} (1 - i) \cdot F^{x_1=0} \cdot i^{L(X')},
\]

where \(L(X')\), on \(X' = (x_2, x_3, \ldots, x_k, 1)\), is a 0-1 indicator function which takes the value 1 on \(S\) and 0 on \(\mathbb{Z}_2^{k-1} - S\). Thus we can combine the two sums and get

\[
\sum_{x_1, x_2, \ldots, x_k} F = (1 - i) \cdot \sum_{x_2, x_3, \ldots, x_k} \left( F^{x_1=0} \cdot i^{L(X')} \right).
\]

Note that \(F^{x_1=0} \cdot i^{L(X')}\) is also a function in \(\mathcal{A}\). So we get a new summation of this form and have removed a variable \(x_1\).

After at most \(k\) steps we can eliminate all the variables and obtain the value of the initial summation. Both \(k\) and \(n\) are bounded by input size. In each iteration, we either resolve a linear system \(AX = 0\) or compute a linear equation by Lemma 3.10 representing the affine subspace \(S\), both of which can be done in polynomial time. After each iteration, the formula inside the summation gets at most one more factor \(i^{L(X')}\) or \(\chi_S\), so the whole algorithm is in polynomial time.

\[ \square \]

### 3.5 Hardness

In this section, we prove the hardness part of Theorem 3.7. We start with two useful lemmas which provide tools for constructing reductions.

In \(#\text{CSP}\) problems, if we have a constraint function \(F\), and we let a variable \(x_j\) of \(F\) not occur in any other place, then we get \(F^{x_j=*}\). This observation proves the following lemma.
Lemma 3.12. For any $F$ and $F$, and for any variable $x_j$ of $F$,

$$\#\text{CSP}(F \cup \{F^{x_j=x}\}) \leq_T \#\text{CSP}(F \cup \{F\}).$$

The presence of all \textsc{Equality} gates in \#\text{CSP} already provides two important unary functions $\Delta_0$ and $\Delta_1$. Applying $\Delta_0$ or $\Delta_1$ to a variable fixes that variable to the constant 0 or 1 respectively, and thus it is called \textit{pinning}. We can simulate $\Delta_0$ and $\Delta_1$ by the following pinning lemma.

Lemma 3.13 (Dyer-Goldberg-Jerrum). For any $F$, $\#\text{CSP}(F \cup \{\Delta_0, \Delta_1\}) \leq_T \#\text{CSP}(F)$. In particular, for any $F$ and $c \in \{0,1\}$,

$$\#\text{CSP}(F \cup \{F^{x=c}\}) \leq_T \#\text{CSP}(F \cup \{F\}).$$

Proof. Given an instance $I$ of $\#\text{CSP}(F \cup \{\Delta_0, \Delta_1\})$, let $V_0$ and $V_1$ denote the set of variables to which $\Delta_0$ and $\Delta_1$ are applied, respectively. If $V_0 \cap V_1 \neq \emptyset$, then there are no satisfying assignments and the partition function $Z(I) = 0$. We assume $V_0 \cap V_1 = \emptyset$. Let $V_2 = V - (V_0 \cup V_1)$ be the set of other variables. We define an instance $I'$ of $\#\text{CSP}(F)$ which has the variable set $V_2 \cup \{y_0, y_1\}$, where $y_0, y_1$ are two new variables. The constraints of $I'$ are exactly the same as those of $I$, except (1) every variable in $V_0$ (respectively $V_1$), is replaced by $y_0$ (respectively $y_1$), and (2) the constraints of $\Delta_0$ and $\Delta_1$ are not applied. Then $Z(I')$ can be expressed as the sum of four quantities,

$$Z(I') = \sum_{(b_0,b_1) \in \{0,1\}^2} Z(I' \mid b_0 b_1)$$

where $Z(I' \mid b_0 b_1)$ is the sum in $Z(I')$ restricted to assigning $(y_0, y_1)$ to $(b_0, b_1) \in \{0,1\}^2$.

Now we define a second instance $I''$ of $\#\text{CSP}(F)$ which has the variable set $V_2 \cup \{y\}$, where $y$ is a new variable. It is the same as $I'$ except we identify $y_0$ and $y_1$ by $y$. Then clearly

$$Z(I'') = \sum_{b_0=b_1} Z(I' \mid b_0 b_1) = Z(I' \mid 00) + Z(I' \mid 11).$$

Hence

$$Z(I') - Z(I'') = Z(I' \mid 01) + Z(I' \mid 10). \quad (3.1)$$

Obviously $Z(I) = Z(I' \mid 01)$. If every function $F \in F$ has \textit{complementary invariance}, i.e., if $F$ has arity $k$, then for every input $(b_1, \ldots, b_k) \in \{0,1\}^k$, $F(b_1, \ldots, b_k) = F(\overline{b_1}, \ldots, \overline{b_k})$, where $\overline{b_i} = 1 - b_i$, then clearly $Z(I') - Z(I'') = 2Z(I)$, and the lemma is proved.

Suppose for some $F \in F$ of arity $k$, and some $(b_1, \ldots, b_k) \in \{0,1\}^k$, $F(b_1, \ldots, b_k) \neq F(\overline{b_1}, \ldots, \overline{b_k})$, we will define two more instances of $\#\text{CSP}(F)$. The first is $I'_0$ which is the same as $I'$, except it has one more constraint $F$ applied to the sequence of variables $(y_0, \ldots, y_k)$. Note that for any assignment in $I'_0$ if $(y_0, y_1)$ is assigned to $(0,1)$, then $F(y_0, \ldots, y_k)$ will evaluate to $F(b_1, \ldots, b_k)$, and if $(y_0, y_1)$ is assigned to $(1,0)$, then $F(y_0, \ldots, y_k)$ will evaluate
to $F(\overline{b_1}, \ldots, \overline{b_k})$. Of course if $(y_0, y_1)$ is assigned to $(0, 0)$ or $(1, 1)$, then $F(y_{b_1}, \ldots, y_{b_k})$ will evaluate to $F(0, \ldots, 0)$ and $F(1, \ldots, 1)$ respectively.

The second instance we define is $I'_b$ which is the same as $I''$, except it has one more constraint $F$ applied to the sequence of variables $(y, \ldots, y)$. It is now clear that

$$Z(I'_b) - Z(I''_b) = Z(I' | 01)F(b_1, \ldots, b_k) + Z(I' | 10)F(\overline{b_1}, \ldots, \overline{b_k}).$$

From equations (3.1) and (3.2) we can solve $Z(I) = Z(I' | 01)$.

### 3.5.1 One Symmetric Binary Function

For hardness, we first establish the following theorem, which says that Theorem 3.7 holds for a single symmetric binary function.

**Theorem 3.14.** If $[a, b, c] \notin \mathcal{S} \cup \mathcal{P}$, then $\#CSP([a, b, c])$ is $\#P$-hard. Explicitly, for any binary function $[a, b, c]$, where $a, b, c \in \mathbb{C}$, if $\#CSP([a, b, c])$ is not $\#P$-hard, then $[a, b, c]$ takes the form: $[x, 0, y]$, $[0, x, 0]$, $[x^2, xy, y^2]$, $x[1, \pm i, 1]$ or $x[1, \pm 1, -1]$.

Before proving Theorem 3.14, we first note that every one of the five listed exceptional cases are in $\mathcal{S} \cup \mathcal{P}$, and it can be checked directly that all binary symmetric functions in $\mathcal{S} \cup \mathcal{P}$ take one of these five forms.

In this proof we use an important reduction method called polynomial interpolation, first introduced by Valiant [3]. The method involves setting up and then solving a system of linear equations. The solvability of these linear systems is usually by the fact that it is a Vandermonde system.

Here is a simple example. Suppose $\mathcal{F}$ contains some function $F = [1, a, 1]$, where $a \neq 0$ and is not a root of unity. Suppose we want to simulate a function $H = [1, b, 1]$, that is, we want to reduce $\#CSP(\mathcal{F} \cup \{H\})$ to $\#CSP(\mathcal{F})$. Given an instance $I$ of $\#CSP(\mathcal{F} \cup \{H\})$, where there are $n$ occurrences of the constraint $H$, we construct instances $I_j$ of $\#CSP(\mathcal{F})$, for each $j \geq 1$, by replacing each occurrence of the constraint $H(x_{i_1}, x_{i_2})$ in $I$ by $j$ many constraints $F(x_{i_1}, x_{i_2})$. Each occurrence of $H$ takes an input of the form $(0, 0), (0, 1), (1, 0), (1, 1)$. We can write the sum defining $Z(I)$ as a sum over all assignments stratified according to the number of $(1, 0)$ or $(0, 1)$ assigned at the $n$ occurrences of $H$. If exactly $i$ of $n$ occurrences of $H$ are assigned $(1, 0)$ or $(0, 1)$, then the other $n - i$ are assigned $(0, 0)$ or $(1, 1)$. The value $Z(I)$ can be written as the summation

$$Z(I) = \sum_{i=0}^{n} w_i b^i,$$

where $w_i$ is the summation of products of all constraints other than these $n$ occurrences of $H$, over all assignments such that exactly $i$ of these $n$ occurrences of $H$ receive inputs $(1, 0)$ or $(0, 1)$. Note that for assignments $(1, 0)$ or $(0, 1)$, $H$ evaluates to $b$; the factor $b^i$ in (3.3)
comes from this. The other $n - i$ occurrences of $H$ receiving inputs $(0, 0)$ or $(1, 1)$ contribute a factor 1. Meanwhile, we have

\[ Z(I_j) = \sum_{i=0}^{n} w_i a^{ij}. \]  

(3.4)

Note that the same set of values $w_i$ occur here. We let $j = 1, \ldots, n + 1$, then (3.4) becomes a system of linear equations about $w_i$, whose coefficient matrix is a Vandermonde matrix $(a^{ij})$. The values $Z(I_j)$ can be obtained by oracle calls to $\#\text{CSP}(\mathcal{F})$. Since $a$ is non-zero and not a root of unity, $(a^{ij})$ is a non-singular matrix, and we can solve for all $w_i$ in polynomial time from (3.4), which gives us $Z(I)$ by (3.3). This is essentially how every reduction by polynomial interpolation in this chapter will be done.

To get acquainted with the technique, we start with the following theorem, a special case of Theorem 3.14. We illustrate two features. The first is the use of interpolation in a particularly simple setting, where non-negativity avoids many complications. The second is to show that the ability to get unary functions is a powerful step to prove $\#P$-hardness. Once we can interpolate all unary functions, we can appeal to the dichotomy theorem for Holant* problems. This theorem is also a very special case of much more general theorems on graph homomorphisms, in particular the Bulatov-Grohe Theorem (see Chapter ??).

**Theorem 3.15.** Let $[a, b, c]$ be a symmetric binary function, where $a, b, c$ are non-negative real numbers. Then $\#\text{CSP}([a, b, c])$ is $\#P$-hard unless $[a, b, c]$ is of one of the following three forms: $[a, 0, c]$, $[0, b, 0]$ or $[x^2, xy, y^2]$.

**Proof.** We note that $\#\text{CSP}([0, 1, 1])$ counts the number of vertex covers in a graph. Given a graph $G = (V, E)$, we may consider it defines a $\#\text{CSP}([0, 1, 1])$ instance $I$ where each vertex is a Boolean variable and each edge is given the constraint $[0, 1, 1]$ representing the Boolean Or function. Then the product over all edges evaluates to 1 for an assignment $\sigma : V \to \{0, 1\}$ iff $\sigma^{-1}(1)$ is a vertex cover of $G$. Hence $Z(I)$ is exactly the number vertex covers in $G$. This problem is $\#P$-complete.

If we used the constraint $[1, 1, 0]$, then $\#\text{CSP}([1, 1, 0])$ counts the number of independent sets in a graph, since $\sigma^{-1}(1)$ is an independent set iff the product for $\sigma$ evaluates to 1. In fact, by exchanging 0 and 1, we see that these two problems always have the same solution.

We will reduce one of these $\#P$-hard problems to $\#\text{CSP}([a, b, c])$, if $[a, b, c]$ does not have one of the three forms. We assume $b \neq 0$, and not both $a$ and $c$ are 0. Since $b \neq 0$, we may normalize $b = 1$ by replacing $[a, b, c]$ with $[a/b, 1, c/b]$. This only causes $Z(I)$ to change by a factor $b^m$ where $m$ is the number of occurrences of the constraint $[a, b, c]$ in the input instance $I$. Thus we may assume the given constraint is $[x, 1, y]$, for some non-negative $x$ and $y$, where $(x, y) \neq (0, 0)$, and $xy \neq 1$.

First suppose $y \neq 0, 1$. Using $\Delta_1$ we can get the unary function $[1, y]$, namely

\[ \#\text{CSP}([x, 1, y], [1, y]) \leq T \ #\text{CSP}([x, 1, y]). \]
With a positive $y \neq 1$, we can interpolate any unary function using $[1, y]$ as follows. $\#\text{CSP}(\mathcal{F})$ is precisely Holant$(\mathcal{E} \mathcal{Q} \mid \mathcal{F})$ where $\mathcal{E} \mathcal{Q} = \{(\_k) \mid k \geq 1\}$ is the set of all equalities, and $\mathcal{F}$ is any set of constraints. Given any unary function $[u, v]$ and any $\mathcal{F}$, we reduce $\#\text{CSP}(\mathcal{F} \cup \{[u, v]\})$ to $\#\text{CSP}(\mathcal{F} \cup \{[1, y]\})$. Suppose $I$ is an instance for $\#\text{CSP}(\mathcal{F} \cup \{[u, v]\})$ where $[u, v]$ occurs $n$ times. Then $Z(I) = \sum_{i=0}^{n} w_i u^{n-i} v^i$, where $w_i$ is the summation of products of all constraints other than the $n$ occurrences of $[u, v]$, over all assignments such that exactly $i$ of these $n$ occurrences of $[u, v]$ receive input 1. Define an instance $I_j$ for $\#\text{CSP}(\mathcal{F} \cup \{[1, y]\})$ by replacing each occurrence of $[u, v]$ in $I$ by $j$ occurrences of $[1, y]$. Then $Z(I_j) = \sum_{i=0}^{n} w_i y^{j-i}$, for all $j \geq 1$. Take $1 \leq j \leq n + 1$ we obtain a non-singular Vandermonde system, as the positive $y \neq 1$. We can solve for all $w_i$ in polynomial time, and compute $Z(I)$.

Similarly if $x \neq 0, 1$, then using $\Delta_0$ we can get the unary function $[x, 1]$, and then interpolate any other unary function. If $x, y \in \{0, 1\}$, since $(x, y) \neq (0, 0)$ and $xy \neq 1$, the only cases left are $(x, y) = (0, 1)$ or $(1, 0)$. These are the $\#P$-hard problems VERTEX COVER and INDEPENDENT SET respectively.

Now suppose we can interpolate any unary function $[u, v]$. By a finite sequence of reductions, we can interpolate any finite set of unary functions. By the dichotomy theorem for Holant* problems, namely Theorem 2.12 (essentially Lemma 2.14), Holant*($(\_3) \mid [x, 1, y]$) is $\#P$-hard. This means that for some finite subset of unary functions $\mathcal{U}' \subset \mathcal{U}$, Holant($(\_3) \cup \mathcal{U}' \mid \{[x, 1, y]\} \cup \mathcal{U}'$) is $\#P$-hard. Hence Holant*($(\_3) \mid \{[x, 1, y]\} \cup \mathcal{U}'$) is $\#P$-hard for some finite $\mathcal{U}' \subset \mathcal{U}$. The latter is reducible to $\#\text{CSP}([x, 1, y])$, hence it is also $\#P$-hard. This argument will be used often, and for simplicity in what follows we will just say:

$$\text{Holant}^*((\_3) \mid [x, 1, y]) \leq_r \#\text{CSP}([x, 1, y]),$$

and since Holant*($(\_3) \mid [x, 1, y]$) is $\#P$-hard by Theorem 2.12, we conclude that $\#\text{CSP}([x, 1, y])$ is also $\#P$-hard. \hfill $\square$

Now we consider complex valued functions. First we prove two simple lemmas.

**Lemma 3.16.** For any symmetric binary function $[0, b, c]$, where $b, c \in \mathbb{C}$ and $bc \neq 0$, the problem $\#\text{CSP}([0, b, c])$ is $\#P$-hard.

*Proof.* Since $b \neq 0$, we can normalize it and assume $b = 1$. So we rename the constraint function $[0, 1, c]$. First suppose $c$ is a root of unity. Let $c^k = 1$. We can realize $[0, 1^k, c^k] = [0, 1, 1]$ by $k$ repeated copies of $[0, 1, c]$. The problem $\#\text{CSP}([0, 1, 1])$ is counting vertex covers, hence $\#P$-hard. Now suppose $c$ is not a root of unity. Also by assumption $c \neq 0$. We can interpolate $[0, 1, x]$, for any $x$, by polynomial interpolation. In particular, we can interpolate $[0, 1, 1]$ and $\#\text{CSP}([0, 1, 1])$ is $\#P$-hard. \hfill $\square$

**Lemma 3.17.** Let $[1, b, c]$ be a symmetric binary function, where $bc \neq 0$ and $c \neq b^2$. Suppose either $b$ or $c$ is not a root of unity. Then $\#\text{CSP}([1, b, c])$ is $\#P$-hard.
Proof. Suppose $b$ is not a root of unity. Using $\Delta_0$ we can get the unary function $[1, b]$. Then we can interpolate any finitely many unary functions. Then

$$\text{Holant}^*([=_3] | [1, b, c]) \leq_T \#\text{CSP}([1, b, c]).$$

By Theorem 2.12 for Holant* problems, $\#\text{CSP}([1, b, c])$ is $\#\text{P}$-hard.

Next suppose $c$ is not a root of unity. Connect two inputs of $([=_3])$ by $[1, b, c]$, we can get the function $[1, c]$, and interpolate any unary function by polynomial interpolation, as in Theorem 3.15. Hence again by Theorem 2.12, $\#\text{CSP}([1, b, c])$ is $\#\text{P}$-hard. \hfill \qed

Now we prove Theorem 3.14: If $[a, b, c] \notin \mathcal{A} \cup \mathcal{P}$, then $\#\text{CSP}([a, b, c])$ is $\#\text{P}$-hard. The basic idea is to interpolate unary functions and then appeal to the Holant* dichotomy. However in order to interpolate all unary functions we must handle some special cases. For this purpose we also use a gadget which links two copies of the binary function $F = [a, b, c]$. It is realizable in $\#\text{CSP}$ as $G(x, y) = \sum_z F(x, z)F(z, y)$, using an extra variable $z$ not appearing elsewhere.

Proof. of Theorem 3.14. If $a = 0$, we know $bc \neq 0$, otherwise it is in the second or the third exceptional case of Theorem 3.14. So by Lemma 3.16, $\#\text{CSP}([a, b, c])$ is $\#\text{P}$-hard. The case $c = 0$ is symmetric. Since $[a, b, c] \notin \mathcal{A} \cup \mathcal{P}$, we know $b \neq 0$. Therefore we will assume in the following that $abc \neq 0$, and by normalizing, we can assume $a = 1$.

Now we consider the complexity of $\#\text{CSP}([1, b, c])$, with $bc \neq 0$. We assume $c \neq b^2$; otherwise $[1, b, c]$ is degenerate, the third exceptional case of Theorem 3.14. If either $b$ or $c$ is not a root of unity, then $\#\text{CSP}([1, b, c])$ is $\#\text{P}$-hard by Lemma 3.17.

Suppose both $b$ and $c$ are roots of unity. We can realize $G = \left[\begin{array}{cc} 1 & b \\ b & c \end{array}\right]^2 = [1+b^2, b+bc, b^2+c^2]$, which is the signature of two consecutive copies of $[1, b, c]$. $G$ is also non-degenerate. If $\#\text{CSP}(G)$ is $\#\text{P}$-hard so is $\#\text{CSP}([1, b, c])$. We consider the following cases.


   Since $[1, b, c] = [1, -1, c] \notin \mathcal{A} \cup \mathcal{P}$, we know $c \neq \pm 1$. If $c = \pm i$, we get $G = [2, -1 \mp i, 0]$, which is $\#\text{P}$-hard by Lemma 3.16 (or rather a version of Lemma 3.16 on a reversed signature, by exchanging the inputs 0 and 1). If $c \notin \{\pm 1 \pm i\}$, then there are no zero entries in $G$. Since $c$ is a root of unity, and $c \neq \pm 1$, we have $|1 + c^2| \neq 2$. In particular $\frac{1+c^2}{2}$ is not a root of unity. Normalizing we get $[1, \frac{-1-c}{2}, \frac{1+c^2}{2}]$. Hence $\#\text{CSP}([1, b, c])$ is $\#\text{P}$-hard by Lemma 3.17.


   Since $[1, b, c] = [1, -c, c] \notin \mathcal{A} \cup \mathcal{P}$, we know $c \neq \pm 1$. If $c = \pm i$, we get $G = [0, 1 \mp i, -2]$, which is $\#\text{P}$-hard by Lemma 3.16. If $c \notin \{\pm 1 \pm i\}$, then there are no zero entries in $G$. Normalizing we get $[1, \frac{-c-c^2}{1+c^2}, \frac{2c^2}{1+c^2}]$. For $c$ a root of unity but $c \neq \pm 1$, we have $|1 + c^2| \neq |2c^2|$. In particular $\frac{2c^2}{1+c^2}$ is not a root of unity. It follows from Lemma 3.17 that $\#\text{CSP}(G)$ is $\#\text{P}$-hard.
3. \(c = 1\). \(G = [1 + b^2, 2b, 1 + b^2]\).

Since \([1, b, c] = [1, b, 1] \not\in \mathcal{A} \cup \mathcal{P}\), we know \(b \not\in \{\pm 1 \pm i\}\). So \(1 + b^2 \neq 0\), and \(\frac{2b}{1+b^2} \neq 1\). Hence \(\frac{2b}{1+b^2}\) is not a root of unity. Therefore the problem is \#P-hard by Lemma 3.17.

4. \(b \neq -1\), \(b \neq -c\), and \(c \neq 1\). Recall that we are given \(c \neq b^2\) since \([1, b, c] \not\in \mathcal{P}\).

If \(F(x_1, x_2) = [1, b, c]\), then \(F^{x_2=x} = [1 + b, b + c]\) with both entries non-zero. By \(|b| = |c| = 1\), we claim that, \(1 + b = |b + c|\) if and only if \(c = 1\) or \(c = b^2\). For now assume the claim holds. Since we assumed \(c \neq 1\) and \(c \neq b^2\), we have \(\frac{b + c}{1+b} \neq 1\) and thus we can interpolate all unary functions with \([1+b, b+c]\). Therefore \#CSP([1, b, c]) is \#P-hard by Theorem 2.12 for Holant* problems.

To prove the claim one direction is obvious. For the reverse direction, suppose \(1 + b = |b + c|\), then \((1 + b)(1 + b) = (b + c)(b + c)\). Expanding this equation, we get \(b + \bar{b} = cb + \bar{c}b\), i.e., \(Rb = Re\ c/b\). If \(b = e^{i\alpha}\) and \(c = e^{i\beta}\), then \(\cos \alpha = \cos(\beta - \alpha)\), hence \(\beta - \alpha = \pm \alpha \mod 2\pi\). It follows that \(c = 1\) or \(b^2\).

\[\square\]

Theorem 3.14 is the special case of graph homomorphism on the Boolean domain. We will discuss graph homomorphism on a general domain in much more detail in Part II of the book.

### 3.5.2 Affine Support

The following lemma gives a necessary condition for \#CSP(F) being not \#P-hard. This affine support criterion will be much generalized later when we discuss dichotomies for graph homomorphisms and \#CSP on general, i.e., not necessarily Boolean, domains. The proof given here is elementary.

**Lemma 3.18.** If the relation \(R_F\) defined by \(F\) is not affine, then \#CSP(F) is \#P-hard.

**Proof.** We prove by induction on the arity of the function \(F\).

All functions of arity one have affine support. So we assume the arity of \(F\) is at least two. We first consider a function \(F\) of arity two. Suppose \(F\) does not have affine support. This implies that exactly one of its four values is 0. The matrix form of \(F\) is

\[
\begin{bmatrix}
a & b \\
c & d
\end{bmatrix} = \begin{bmatrix} F(0, 0) & F(0, 1) \\ F(1, 0) & F(1, 1) \end{bmatrix},
\]

then in particular \(\det(F) \neq 0\). By taking two copies of \(F\) sharing a free variable \(z\) in the appropriate order \((x, z)\) and \((z, y)\), we can realize the binary function \(H(x, y) = \sum_z F(x, z)F(z, y)\),
whose matrix form is \( H = FF^\top = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix} \). This \( H \) is a symmetric binary function, which can also be denoted by \([a^2 + b^2, ac + bd, c^2 + d^2]\). We can apply Theorem 3.14 to \( H \). Because \( F \) is nonsingular, so is the matrix for \( H \). Because exactly one entry of \( F \) is 0, \( ac + bd \neq 0 \) and \( H \) is not of the form \([x, 0, y]\). Because either \( a^2 + b^2 \neq 0 \) or \( c^2 + d^2 \neq 0 \), \( H \) is not of the form \([0, x, 0]\). So the only remaining possibilities for \( H \in \mathcal{A} \cup \mathcal{P} \) is that \( H \) is of the form \([x, 0, y] \) or \([1, 0, 1] \) or \([x, 1, 1] \). By symmetry, we only need to consider the cases \( a = 0 \) and \( bcd \neq 0 \) or \( b = 0 \) and \( acd \neq 0 \). If \( a = 0 \), we can assume \( b = 1 \) by normalization, and then the function \( H \) is \([1, d, c^2 + d^2] \). If \( H \) is of the form \([x, 0, y] \), we have \( d = \pm 1 \) and \( c = \pm \sqrt{2} \). Then we can realize another symmetric binary function by \( G(x, y) = F(x, y)F(y, x) \). So \( G = [a^2, bc, d^2] = [0, \pm \sqrt{2}, -1] \). \( \text{CSP}(G) \) is \( \#P \)-hard by Theorem 3.14. If \( H \) is of the form \([1, 0, 1] \), we have \( d = \pm 1 \) and \( c = \pm \sqrt{2} \). Then \( G = [a^2, bc, d^2] = [0, \pm \sqrt{2}i, 1] \) and \( \text{CSP}(G) \) is \( \#P \)-hard by Theorem 3.14 again. So we have completed for the \( a = 0 \) case. If \( b = 0 \), we can assume \( a = 1 \) and the function \( H \) is \([1, c, c^2 + d^2] \). If \( H \) is of the form \([x, 1, 1] \), we have \( c = \pm 1 \) and \( d = \pm \sqrt{2}i \). Then we can realize another binary function \( F' \) by \( F'(x, y) = F(x, y)F(y, x) \). In matrix notation \( F' = \begin{bmatrix} a^2 & b^2 \\ c^2 & d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} \). Next we can simulate \( H' = F'F^\top = \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix} \) from \( F' \). In symmetric notation \( H' = [1, -1, 5] \). By Theorem 3.14 \( \text{CSP}(H') \) is \( \#P \)-hard. Finally if \( H \) is of the form \([x, 1, 1, 1] \), we have \( c = \pm 1 \) and \( d = \pm \sqrt{2}i \). Then by the same construction, \( F' = \begin{bmatrix} 1 & 0 \\ 1 & -2 \end{bmatrix} \) and \( H' = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix} \), which in symmetric notation is \( H' = [1, 1, 5] \). So again \( \text{CSP}(H') \) is \( \#P \)-hard. We have completed the proof for the case where the function \( F \) of the form with an affinity less than \( k \), for some \( k \geq 3 \), and now assume the function \( F \) has an affinity \( k \). Since \( R_F \) is not affine, there exist \( \alpha, \beta, \gamma \in R_F \), where \( \alpha = (a_1, \ldots, a_k), \beta = (b_1, \ldots, b_k), \gamma = (c_1, \ldots, c_k) \), such that \( \delta = \alpha \oplus \beta \oplus \gamma = (d_1, \ldots, d_k) \notin R_F \). We only need to prove that we can use \( F \) to simulate a function of smaller arity that does not have affine support.

Divide the index set \([k] \) of input variables of \( F \) into four subsets according to the values of \( \alpha, \beta, \gamma \) as follows:

\[
I = \{ j \mid a_j = b_j \neq c_j \}, \quad J = \{ j \mid a_j = c_j \neq b_j \}, \quad K = \{ j \mid b_j = c_j \neq a_j \}, \quad L = \{ j \mid a_j = b_j = c_j \}.
\]

Since \( a_j, b_j \) and \( c_j \) take Boolean values, \( \{I, J, K, L\} \) is a partition of \([k] \). We also remark that, if \( j, l \in I \), then either \((a_l, b_l, c_l) = (a_j, b_j, c_j) \) or \((a_l, \overline{b_j}, \overline{c_j}) \). The same is true for \( J, K \) and \( L \). Now we have the following four cases, and for each case, we prove our result.

- \( L \) is not empty. There exists \( j \) such that \( a_j = b_j = c_j \).

  We fix the \( j \)th input of \( F \) to be \( a_j \), and get a function \( F^{x_j = a_j} \). Recall that \( \delta = \alpha \oplus \beta \oplus \gamma \), then \( d_j = a_j = b_j = c_j \). This function \( F^{x_j = a_j} \) does not have affine support.

  Now we may assume \( L = \emptyset \) and \([k] = I \cup J \cup K \).

- There are indices \( l \neq j \), such that \((a_l, b_l, c_l) = (a_j, b_j, c_j) \).
Without loss of generality we assume \( l = 1 \) and \( j = 2 \). Define a function of arity \( k - 1 \) by \( H(x_1, x_3, \ldots, x_k) = F(x_1, x_1, x_3, \ldots, x_k) \). \( H \) can be simulated by \( F \), and by the property that \( \alpha, \beta, \gamma \in R_F \) and yet \( \delta \not\in R_F \), \( H \) does not have affine support.

- There are indices \( l \neq j \), such that \((a_l, b_l, c_l) = (\overline{x_l}, \overline{y_l}, \overline{z_l})\).

Clearly both \( l \) and \( j \) belong to the same set \( I \) or \( J \) or \( K \). Without loss of generality we assume \( l = 1 \in I \) and \( j = 2 \in I \). The proof for \( J \) and \( K \) are the same. Then we have

\[
\alpha = (a, \overline{a}, \alpha'), \quad \beta = (a, \overline{a}, \beta'), \quad \gamma = (\overline{a}, a, \gamma'), \quad \text{and} \quad \delta = (\overline{a}, a, \delta') \tag{3.5}
\]

where \( a \in \mathbb{Z}_2 \), \( \alpha', \beta', \gamma' \in \mathbb{Z}_2^{k-2} \), and \( \delta' = \alpha' \oplus \beta' \oplus \gamma' \in \mathbb{Z}_2^{k-2} \). Assume for a contradiction that all functions of the forms \( F_{x_1=b} \) and \( F_{x_1=\pi} \) have affine support (1 \( \leq i \leq k \), \( b = 0, 1 \)).

Consider \( F_{x_1=a} \), whose underlying relation \( R_{F_{x_1=a}} \) is affine. Because \( \alpha, \beta \in R_F \), \( (\overline{a}, \alpha'), (\overline{a}, \beta') \in R_{F_{x_1=a}} \). The summation of \( (\overline{a}, \alpha'), (\overline{a}, \beta'), (a, \gamma'), (a, \delta') \) is the zero vector in \( \mathbb{Z}_2^{k-1} \), so \((a, \gamma') \in R_{F_{x_1=a}} \) iff \((a, \delta') \in R_{F_{x_1=a}} \). This implies that \((a, a, \gamma') \in R_F \) iff \((a, a, \delta') \in R_F \).

Next consider \( F_{x_2=a} \). Because \( \gamma \in R_F \) and \( \delta \not\in R_F \), we have \((\overline{a}, \gamma') \in R_{F_{x_2=a}} \), \((a, \gamma') \not\in R_{F_{x_2=a}} \). As we just proved, \( R_F(a, a, \gamma') = R_F(a, a, \delta') \). If this value is 1, then \((a, \gamma'), (a, \delta') \in R_{F_{x_2=a}} \), but this is impossible for the affine relation \( R_{F_{x_2=a}} \), in view of the four inputs \((\overline{a}, \gamma'), (\overline{a}, \delta'), (a, \gamma'), \) and \((a, \delta') \). So we must have \((a, a, \gamma') \not\in R_F \) and \((a, a, \delta') \not\in R_F \).

Similarly, we can prove that both \((\overline{a}, \overline{a}, \alpha') \) and \((\overline{a}, \overline{a}, \beta') \not\in R_F \). More explicitly, first consider \( F_{x_2=\pi} \). Because \( \alpha, \beta \in R_F \), \((a, \alpha'), (\overline{a}, \beta') \in R_{F_{x_2=\pi}} \). Having an affine support, \( R_{F_{x_2=\pi}} \) evaluates to the same value on \((\overline{a}, \alpha') \) and \((\overline{a}, \beta') \). Thus \( R_F \) evaluates to the same value on \((\overline{a}, \overline{a}, \alpha') \) and \((\overline{a}, \overline{a}, \beta') \). Next consider \( F_{x_1=\pi} \). It also has an affine support. Since \( \gamma \in R_F \) and \( \delta \not\in R_F \), we have \((a, \gamma') \in R_{F_{x_1=\pi}} \) and \((a, \delta') \not\in R_{F_{x_1=\pi}} \). If \((\overline{a}, \overline{a}, \alpha'), (\overline{a}, \overline{a}, \beta') \in R_F \), then \((\overline{a}, \alpha'), (\overline{a}, \beta') \in R_{F_{x_1=\pi}} \). This is impossible for an affine relation \( R_{F_{x_1=\pi}} \), in view of the four inputs \((a, \gamma'), (a, \delta'), (\overline{a}, \alpha'), \) and \((\overline{a}, \beta') \). Thus it follows that \((\overline{a}, \overline{a}, \alpha') \not\in R_F \) and \((\overline{a}, \overline{a}, \beta') \not\in R_F \).

To summarize we have all \((a, a, \gamma'), (a, a, \delta'), (\overline{a}, \overline{a}, \alpha'), (\overline{a}, \overline{a}, \beta') \not\in R_F \).

Finally we consider \( F_{x_1=\pi} \), and calculate as follows:

\[
F_{x_1=\pi}(\overline{a}, \alpha') = F(\overline{a}) + F(\overline{a}, \alpha') = F(\overline{a}) \neq 0,
\]
\[
F_{x_1=\pi}(\overline{a}, \beta') = F(\overline{a}) + F(\overline{a}, \beta') = F(\overline{a}) \neq 0,
\]
\[
F_{x_1=\pi}(a, \gamma') = F(\gamma) + F(a, a, \gamma') = F(\gamma) \neq 0,
\]
\[
F_{x_1=\pi}(a, \delta') = F(\delta) + F(a, a, \delta') = F(\delta) = 0.
\]

This is a contradiction with the assumption that \( R_{F_{x_1=\pi}} \) is affine.

- If there are more than one element in the set \( I \), or in the set \( J \), or in the set \( K \), it is included in the previous two cases. The remaining case is that the sizes of \( I, J, K \) are
all at most one and $L$ is empty. Because $k > 2$, the sizes of $I, J, K$ are exactly one, and $k = 3$. Without loss of generality, let $I = \{1\}$, $J = \{2\}$ and $K = \{3\}$.

A moment’s reflection shows that we can write $\alpha, \beta, \gamma$ and $\delta$ as follows:

$$
\begin{align*}
\alpha &= (p, q, r), \\
\beta &= (p, \overline{q}, r), \\
\gamma &= (\overline{p}, q, r), \\
\delta &= (\overline{p}, \overline{q}, r),
\end{align*}
$$

where $p, q, r \in \mathbb{Z}_2$.

First we consider $F^{x_1 = p}$, which has an affine support by our induction hypothesis on arity. Let $u = (p, q, r)$, and suppose $u \in R_F$. Then $(q, r) \in R_{F^{x_1 = p}}$. Because $\alpha, \beta \in R_F$, then $(q, \overline{r})$ and $(\overline{q}, r)$ both belong to $R_{F^{x_1 = p}}$. Then being affine, $(\overline{q}, \overline{r}) \in R_{F^{x_1 = p}}$. Let $v = (p, \overline{q}, \overline{r})$, then $v \in R_F$.

Next we consider $R_{F^{x_2 = q}}$. By $\alpha, \gamma \in R_F$, we get $(p, \overline{r}), (p, q) \in R_{F^{x_2 = q}}$. By assumption $u \in R_F$, then $(p, r) \in R_{F^{x_2 = q}}$. By $R_{F^{x_2 = q}}$ being affine, we get $(p, \overline{r}) \in R_{F^{x_2 = q}}$. Let $w = (p, q, \overline{r})$, then $w \in R_F$.

Now $\alpha, v, w \in R_F$. This gives us $(p, q), (p, \overline{q}), (\overline{p}, q) \in R_{F^{x_3 = \tau}}$. Since $R_{F^{x_3 = \tau}}$ is affine, $(p, \overline{q}) \in R_{F^{x_3 = \tau}}$. This means that $\delta = (p, q, \overline{r}) \in R_F$, which is a contradiction.

We conclude that in fact $u \notin R_F$.

By tracing the above steps, under the new condition $u \notin R_F$, we get $v \notin R_F$, and also $w \notin R_F$.

Finally we consider $F^{x_3 = r}$. By $\beta, \gamma \in R_F$, we get $(p, \overline{q}), (\overline{q}, q) \in R_{F^{x_3 = \tau}}$. By $u \notin R_F$, we have $(p, q) \notin R_{F^{x_3 = r}}$. By $R_{F^{x_3 = r}}$ being affine, we get $(p, \overline{q}) \notin R_{F^{x_3 = r}}$. i.e., $(p, \overline{q}, r) \notin R_F$.

We have now accounted for all eight points of $\mathbb{Z}_2^3$. Exactly three of them $\alpha, \beta, \gamma$ belong to $R_F$ and the other five do not. It can be directly verified that $R_{F^{x_1 = \tau}}$ has exactly three members $(q, r), (q, \overline{r}), (\overline{q}, r)$, but not $(\overline{q}, \overline{r})$, which is a contradiction to $R_{F^{x_1 = \tau}}$ being affine. This contradiction completes our proof.

\[\square\]

### 3.5.3 Arity Reduction

Now we come to the two key lemmas for the hardness proof. Both proofs inductively reduce the arity of a function. Suppose $F \not\subset \mathcal{A}$ and $F \not\subset \mathcal{P}$. Let $F \not\subset \mathcal{A}$ and $G \not\subset \mathcal{P}$, where $F, G \in \mathcal{F}$. (It is possible that $F = G$). From $F$ and from $G$, we recursively simulate functions with smaller arities, keeping the property of being not in $\mathcal{A}$ and not in $\mathcal{P}$ respectively. After the two lemmas we handle the base case of the induction.
Lemma 3.19. If \( F \not\in \mathcal{A} \), then either \(#\text{CSP}(F)\) is \#P-hard, or we can simulate a unary function \( H \not\in \mathcal{A} \), that is, there is a reduction from \(#\text{CSP}\{F,H\})\) to \(#\text{CSP}(F)\).

Proof. We prove by induction on the arity of the function \( F \). If \( F \) has arity one, then we may let \( H = F \).

Inductively we assume the lemma has been proved for functions with arity less than \( k \), for some \( k \geq 2 \). Now let \( F \) have arity \( k \). In the following proof, for each case, we construct some functions that can be simulated in \(#\text{CSP}(F)\), but have arity less than \( k \), and then assume they are in \( \mathcal{A} \) (otherwise, it is proved by induction). Finally we prove that either the problem is \#P-hard, or we get a unary function \( H \not\in \mathcal{A} \) or reach a contradiction.

Since the constant function 0 is in \( \mathcal{A} \), \( F \) has a non-empty support \( R_F \). We first assume \( R_F \) is not the whole space \( \mathbb{Z}_2^k \). By Lemma 3.18, either \(#\text{CSP}(F)\) is \#P-hard, or \( R_F \) is affine. Suppose \( R_F = \chi_{AX} \), and \( x_1, x_2, \ldots, x_s \) (\( 0 \leq s < k \)) are free variables of \( AX = 0 \). The function \( F_{x_1=1} : x_{s+2} = \cdots = x_k = * \) can be simulated by \( F \) and has an arity less than \( k \). Thus by our assumption \( F_{x_1=1} : x_{s+2} = \cdots = x_k = * \in \mathcal{A} \). Then obviously \( F \) is equal to \( \chi_{AX} F_{x_1=1} : x_{s+2} = \cdots = x_k = * \) and thus \( F \in \mathcal{A} \). This is a contradiction to the assumption that \( F \not\in \mathcal{A} \).

So we may assume \( R_F = \mathbb{Z}_2^k \). We normalize \( F \) by dividing \( F \) by the nonzero constant \( F(0,0,\ldots,0) \) so that we may assume \( F(0,0,\ldots,0) = 1 \). By our assumption \( F_{x_1=0} \in \mathcal{A} \) so all values of \( F(0,x_2,\ldots,x_k) \) are powers of \( i \). Also \( F_{x_2=0}, F_{x_2=1} \in \mathcal{A} \), and since all values of \( F_{x_1=0, x_2=0} \) and \( F_{x_1=0, x_2=1} \) are powers of \( i \), so are all values of \( F_{x_2=0} \) and \( F_{x_2=1} \). In particular so are all values of \( F_{x_1=1, x_2=1} \).

Now we apply Lemma 3.10 to \( F_{x_2=0} \in \mathcal{A} \) and \( F_{x_2=1} \in \mathcal{A} \). Accordingly we have the following three cases.

1. Both \( F_{x_2=0} \) and \( F_{x_2=1} \) have Congruity. We will denote the function \( F_{x_1=a, x_2=b} \) by \( F_{ab} \). Let \( c \) and \( c' \in \{1, -1, i, -i\} \) be the two constants in the definition of Congruity for \( F_{x_2=0} \) and \( F_{x_2=1} \) respectively, in Lemma 3.10. Thus \( F^{10}/F^{00}(x_3,\ldots,x_k) = c \) and \( F^{11}/F^{01}(x_3,\ldots,x_k) = c' \).

   (a) \( c = c' \).

   This means \( F_{x_1=1}/F_{x_1=0} \) is a constant \( c \in \{1, -1, i, -i\} \). Suppose \( c = i^r \). Then \( F = (i^{x_1}) \cdot F_{x_1=0} = i^{x_1^2} F_{x_1=0} \). Since \( F_{x_1=0} \) is in \( \mathcal{A} \) by arity, this shows that \( F \) is also in \( \mathcal{A} \). A contradiction.

   (b) \( c = -c' \).

   We will use the notation \( [\alpha(X)] \) to denote the 0-1 indicator function for an affine linear function \( \alpha(X) \) over \( \mathbb{Z}_2 \). For any input \( X \), \( [\alpha(X)] \) takes value 0 if \( \alpha(X) = 0 \) in \( \mathbb{Z}_2 \), and it takes value 1 if \( \alpha(X) = 1 \) in \( \mathbb{Z}_2 \). Since \( c \in \{1, -1, i, -i\} \), there exists an \( r \) such that \( i^r = c/i \). Then we claim that

\[
F = (i^{x_1})^r \cdot i^{x_1 \otimes x_2 + [x_2]+[x_2]+[x_2]} \cdot F_{x_1=0}.
\]
To verify this, first suppose \( x_1 = 0 \), then the right-hand side is \( 1^{i[x_2]}F^{x_1=0} = F^{x_1=0} \).

Now let \( x_1 = 1 \), then the right-hand side is \( i^r \cdot 1^{1-[x_2]+3|x_2]}F^{x_1=0} = c(-1)^{x_2}F^{x_1=0} \).

This is \( cF^{00} = F^{10} \), if \( x_2 = 0 \). For \( x_2 = 1 \), the expression is \(-cF^{01} = cF^{01} = F^{11} \).

The claim is proved. Now since \( F^{x_1=0} \) has arity less than \( k \), \( F^{x_1=0} \in A \). But then the claim implies that \( F \in A \) as well. A contradiction.

(c) \( c = ic' \) or \( c = -ic' \).

Assign an arbitrary assignment for \( x_3, \ldots, x_k \). Let \( P \) be the resulting function on \( x_1, x_2 \). In matrix form, where the rows are indexed by \( x_1 = 0, 1 \) and columns are indexed by \( x_2 = 0, 1 \), we have \( P = \begin{bmatrix} u & v \\ \pm ic'u & \pm ic'v \end{bmatrix} \), for some \( u, v \) taking values in \( \{\pm 1, \pm i\} \). Let \( Q(x_1, x_2) = P^3(0, x_2)P(x_1, x_2) \), a product of functions, realizable by pinning \( \Delta_0 \). In matrix form, this is \( Q = \begin{bmatrix} u^4 & v^4 \\ \pm ic'u^4 & \pm ic'v^4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \pm ic' & c' \end{bmatrix} \). Here we used the fact that the values of \( P \) are powers of \( i \). Now \( Q^{22\times2} \) is a unary function \( [2, (1 \pm i)c'] \) which has unequal and non-zero norms \( 2 \neq |(1 \pm i)c'| = \sqrt{2} \) and hence not in \( A \).

2. One of \( F^{x_1=0} \) and \( F^{x_1=1} \) has Congruity and the other has Semi-congruity. Without loss of generality, assume \( F^{x_1=0} \) has Congruity and \( F^{x_1=1} \) has Semi-congruity. The other case is similar.

By Congruity, there is a constant \( c \in \{1, -1, i, -i\} \), such that \( F^{10}/F^{00}(x_3, \ldots, x_k) = c \) for all \( x_3, \ldots, x_k \in \mathbb{Z}_2^{k-2} \). By Semi-congruity there is a constant \( c' \in \{1, -1, i, -i\} \), and a \((k - 3)\)-dimensional affine linear subspace \( S \subset \mathbb{Z}_2^{k-2} \), represented by an affine linear function \( \alpha(x_3, \ldots, x_k) = 0 \), such that on \( S \), \( F^{11}/F^{01}(x_3, \ldots, x_k) = c' \) and on \( \mathbb{Z}_2^{k-2} - S \), \( F^{11}/F^{01}(x_3, \ldots, x_k) = -c' \). We note that to have Semi-congruity, it must be the case that \( k \geq 3 \), and one of the coefficients of \( x_3, \ldots, x_k \) in \( \alpha(x_3, \ldots, x_k) \) must be non-zero. Without loss of generality, let it be the coefficient of \( x_3 \).

Fix an arbitrary assignment to \( x_4, \ldots, x_k \) (if \( k = 3 \) this step is vacuous), this gives a function \( P(x_1, x_2, x_3) \). By changing the constant term in \( \alpha \) and \( c' \) to \(-c' \) if necessary we may assume \( x_3 = 0 \) gives a point with \( \alpha(x_3, \ldots, x_k) = 0 \).

Now we will use a special notation to represent \( P(x_1, x_2, x_3) \).

\[
P = \begin{bmatrix} z & x & c_x \\ y & c_y & c_z \\ w & -c'_w \end{bmatrix}.
\]

This symbol is to suggest a cube and is to be read as follows: The left (right) 4 entries are function values with \( x_1 = 0 \) \((x_1 = 1)\); the top (bottom) 4 entries are function values with \( x_2 = 0 \) \((x_2 = 1)\); finally the inner (outer) 4 entries are values with \( x_3 = 0 \) \((x_3 = 1)\).

Let \( Q(x_1, x_2, x_3) = P(x_1, x_2, x_3)(P(0, x_2, x_3))^3 \). This corresponds to taking the 3rd power of each of left 4 nodes \((x, y, z, w)\) and multiplying to itself and the node to its right. We get \( Q = \begin{bmatrix} \frac{1}{1}c & c \\ \frac{1}{1}c' & -c' \end{bmatrix} \), since \( x^4 = y^4 = z^4 = w^4 = 1 \). Next let \( R(x_1, x_2, x_3) = Q(x_1, x_2, x_3)(Q(x_1, x_2, 0))^3 \). This gives \( R = \begin{bmatrix} \frac{1}{1}1 & \frac{1}{1}1 \\ \frac{1}{1}1 & \frac{1}{1}1 \end{bmatrix} \), since \( c^4 = c'^4 = 1 \). Then \( R^{x_1=0} = \begin{bmatrix} \frac{1}{1}1 & \frac{1}{1}1 \\ \frac{1}{1}1 & \frac{1}{1}1 \end{bmatrix} \).
Both $F_{x_2=0}$ and $F_{x_2=1}$ have Semi-congruity. Let $F^{10}/F^{00} = c$ on $\alpha(x_3, \ldots, x_k) = 0$ and $-c$ on $\alpha(x_3, \ldots, x_k) = 1$. Similarly $F^{11}/F^{01} = c'$ on $\beta(x_3, \ldots, x_k) = 0$ and $-c'$ on $\beta(x_3, \ldots, x_k) = 1$. Here $c, c' \in \{1, -1, i, -i\}$, and $\alpha, \beta$ are two non-trivial affine linear functions.

(a) $c \neq \pm c'$. Since $\beta$ is non-trivial, we may assume the coefficient of $x_3$ in $\beta$ is non-zero. Fix any assignment to $x_4, \ldots, x_k$, we may assume without loss of generality $x_3 = 0$ satisfies $\beta = 0$. We have the following function $P(x_1, x_2, x_3)$, which in our symbolic notation is $P = \frac{x_1^{\delta} x_2^{\epsilon} x_3^{c}}{y^{c'} x^{w}}$, where $\epsilon, \delta \in \{\pm 1\}$ depending on the assignment of $x_4, \ldots, x_k$. If $\epsilon = \delta$, the two entries $\pm cz$ and $\pm cx$ both take the same $+c$ or $-c$ multiplier, then we have obtained a ternary function of the same form in Case 2 and therefore we can continue in exactly the same way. So assume $\epsilon = -\delta$. By renaming $c$ as $-c$, we may assume the two entries are in fact $-cz$ and $+cx$ respectively. Now we take $Q(x_1, x_2, x_3) = P(0, x_2, x_3)^3 P(x_1, x_2, x_3)$. Finally let $R(x_1, x_2, x_3) = Q(x_1, 0, x_3)^3 Q(x_1, x_2, x_3)$. Then we have

$$Q = \frac{1}{1} c \frac{-c}{1} \frac{1}{1} c' \frac{1}{1} = \frac{1}{1} c' \frac{1}{1} \frac{-c}{c'}. \frac{1}{1}.$$

Note that $c^3 c' = c'/c$, since $c^4 = 1$. It is easy to see that $R_{x_1=\ast, x_3=0} = \frac{1}{1} + \frac{1}{1} c'/c$. Since $c'/c \neq \pm 1$ we have $c'/c = \pm i$. Then this unary function $[1, 1 \pm i] \not\in \mathcal{A}$ since it has unequal non-zero norms $2 \neq [1 \pm i]$.

(b) $c = \pm c' \in \{1, -1\}$. In this case $F_{x_1=1}/F_{x_1=0}$ only takes values $\pm 1$. Then $R_{F_{x_1=\ast}}$ is precisely where $F_{x_1=1}/F_{x_1=0} = 1$. If $R_{F_{x_1=\ast}}$ is not affine, we have #P-hardness by Lemma 3.18. So let $R_{F_{x_1=\ast}}$ be defined by an affine linear $\gamma(x_2, \ldots, x_k) = 0$. It can be directly verified that

$$F = F_{x_1=0} \cdot \{[x_1]+[x_1]+[x_1]+[x_1]\} = \{1, x_3, \ldots, x_k \mid \alpha(x_3, \ldots, x_k) = 0\} \cup \{(1, x_3, \ldots, x_k) \mid \beta(x_3, \ldots, x_k) = 0\}.$$
\[\sum_{i=3}^{k} \alpha_i x_i + a = 0\] over \(\mathbb{Z}_2\) defines the set \(S\). Denote this affine linear function by \(\gamma\), then it can be verified that
\[F = F^{x_1=0} \cdot i[x_1+\gamma]+[\gamma]+[\gamma].\]

We have \(F^{x_1=0} \in \mathcal{A}\) by arity, which implies that \(F \in \mathcal{A}\). A contradiction.

Now suppose some coefficients of \(x_3,\ldots,x_k\) in \(\alpha\) and \(\beta\) differ. Without loss of generality, suppose the coefficient of \(x_3\) is 0 in \(\alpha\), and is 1 in \(\beta\) respectively. Fix any assignment to \(x_4,\ldots,x_k\), then the value of \(\alpha\) is fixed, and yet by setting \(x_3\) to 0 or 1, the value of \(\beta\) flips. Then we get a function \(P(x_1,x_2,x_3) = \frac{z}{w}\frac{x+\epsilon y}{by-\delta w}\) for some \(\epsilon,\delta = \pm i\). From here the proof is completed as in Case 2.

\[\square\]

We observe that for any \(a,b \neq 0\), the ternary functions \(H = [a,0,b,0]\) and \([0,a,0,b] \notin \mathcal{P}\). In fact if \(H\) were to be in \(\mathcal{P}\), any expression as a member of \(\mathcal{P}\) must not use any binary Disequality (\(\neq\)), since there is an assignment having a non-zero value \(H \neq 0\), with \(x_i = x_j = 0\), for any \(i \neq j \in \{1,2,3\}\), and the third variable set accordingly. Similarly it cannot use any binary Equality (\(=\)) since there is also an assignment having a non-zero value \(H \neq 0\), with \(x_i \neq x_j\). But unary functions alone cannot define these functions either.

**Lemma 3.20.** For any function \(F \notin \mathcal{P}\), either \(\#CSP(F)\) is \(\#P\)-hard, or we can simulate, using \(F\), a function \([a,0,1,0]\) (or \([0,1,0,a]\)), where \(a \neq 0\), or a binary function \(H \notin \mathcal{P}\) having no zero values.

**Proof.** Suppose \(F\) has arity \(k\). Then \(k \geq 2\) since \(F \notin \mathcal{P}\) and \(\mathcal{P}\) contains all unary functions. Define a \(\{0,1\}\)-matrix by listing all elements of \(R_F\) in some order. The rows are indexed by \(R_F\) and the columns are indexed by \(1,\ldots,k\) corresponding to \(x_1,\ldots,x_k\).

We first remove any column which is all-0 or all-1. If we remove an all-0 column corresponding to \(x_i\), then any \(X \in R_F\) has \(x_i = 0\). The updated matrix corresponds to \(R_F^{x_i=0}\). Similarly if we remove an all-1 column corresponding to \(x_i\), then any \(X \in R_F\) has \(x_i = 1\). The updated matrix corresponds to \(R_F^{x_i=1}\). If two columns are identical or are complementary in every bit, we remove one of them. If the columns at \(x_i\) and \(x_j\) are identical, then any \(X \in R_F\) has \(x_i = x_j\). Similarly for a pair of complementary columns at \(x_i\) and \(x_j\), any \(X \in R_F\) implies that \(x_i \oplus x_j = 1\). In both cases, the updated matrix by removing column \(j\) corresponds to \(R_F^{x_j=\pm}\).

We remove columns as long as possible. We claim that this removal process maintains the property of not belonging to \(\mathcal{P}\). Suppose we removed an all-0 column at \(x_i\), to get \(G = F^{x_i=0}\). Since any \(X \in R_F\) has \(x_i = 0\), we have \(F = \Delta_0(x_i) \cdot G\), where \(\Delta_0(x_i)\) is the unary function \([1,0]\) applied to \(x_i\). This expression shows that if \(G \in \mathcal{P}\) then \(F \in \mathcal{P}\), a contradiction. The case with removing an all-1 column is similar, where we use the unary function \(\Delta_1(x_i)\) instead. If we removed the column at \(x_j\) identical to the column at \(x_i\), then
\( G = F^{x_j=\ast} \) and \( F = \chi_{x_i=x_j} \cdot G \). Finally for the removal of a complementary column at \( x_j \) we have \( G = F^{x_j=\ast} \) and \( F = \chi_{x_i=0} \cdot G \). In every step, we maintain \( G \notin \mathcal{P} \).

Now we suppose there is some \( G \notin \mathcal{P} \) where no more columns can be removed by the above process. There must be some columns left in the matrix, otherwise the function just before the last column removal is a unary function, hence in \( \mathcal{P} \). In fact since \( G \notin \mathcal{P} \), the arity of \( G \) is at least two. For simplicity we still denote it by \( k \). Thus \( k \geq 2 \). We have two cases:

Case 1: \( |R_G| < 2^k \). By Lemma 3.18, we may assume \( R_G \) is affine, given by a linear system on \( x_1, \ldots, x_k \) over \( \mathbb{Z}_2 \). We have that \( |R_G| = |R_F| \neq 0 \). This is because we never deleted any rows, and distinct rows produce distinct rows in the column removal process; so the number of rows remains the same. Clearly \( |R_G| > 1 \) since otherwise every column (of length one) is all-0 or all-1. Without loss of generality, assume \( x_1, \ldots, x_s \) are free variables in the linear system defining \( R_G \), and \( x_{s+1}, \ldots, x_k \) are dependent variables. \( |R_G| = 2^k \) is a power of 2. We have shown that \( s \geq 1 \). By \( |R_G| < 2^k \), we have \( s < k \). We claim \( s \geq 2 \). If instead \( s = 1 \), then every \( x_2, \ldots, x_k \) is dependent on \( x_1 \) on \( R_G \), so the column at \( x_2 \) must be an all-0 or all-1 column, or be identical or complementary to \( x_1 \), and the removal process would have continued. The expression of \( x_k \) in terms of \( x_1, \ldots, x_s \) must involve at least two non-zero coefficients; otherwise the column at \( x_k \) must be an all-0 or all-1 column, or be identical or complementary to another column. Without loss of generality, say the coefficients of \( x_1, x_2 \) are non-zero.

Let \( P(x_1, x_2, x_k) = G^{x_1=0, \ldots, x_i=0, x_{s+1}=\ast, \ldots, x_k=\ast} \) (the two sets of variables \( \{x_3, \ldots, x_s\} \) and \( \{x_{s+1}, \ldots, x_k\} \) could be empty). Note that the sum over \( x_{s+1}, \ldots, x_k \) do not introduce any cancelations because for any assignment \( x_1, \ldots, x_s \) there is a unique nonzero value for each variable \( x_{s+1}, \ldots, x_k \). It follows that \( R_P = \chi_{x_1=0} x_2 x_k = c \) for some \( c \in \mathbb{Z}_2 \).

The affine linear equation \( x_1 \oplus x_2 \oplus x_k = c \) is symmetric. Now we define a symmetrized function \( H(x_1, x_2, x_k) = \prod_{\sigma \in S_3} P(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(k)}) \), where \( S_3 \) is the symmetry group on three letters \( \{1, 2, k\} \). This \( H \) is a symmetric function on \( \{x_1, x_2, x_k\} \) and has support \( R_H = R_P \). Thus, after normalizing, \( H = [a, 0, 1, 0] \) or \( [0, 1, 0, a] \) where \( a \neq 0 \). This ternary function \( H \notin \mathcal{P} \).

Case 2: \( |R_G| = 2^k \). If for all \( 1 \leq j \leq k \), the ratio \( G^{x_j=1}/G^{x_j=0} \) is a constant function \( c_j \), (since \( |R_G| = 2^k \) there are no divisions by zeros), then \( G = c_0 \cdot [1, c_1] \otimes \cdots \otimes [1, c_k] \), where the constant \( c_0 = G^{x_1=0, \ldots, x_k=0} \), and the unary function \( [1, c_j] \) is on \( x_j \). This gives \( G \in \mathcal{P} \), a contradiction.

Now suppose for some \( j \), \( G^{x_j=1}/G^{x_j=0} \) is not a constant function. Without loss of generality, we assume \( j = 1 \). The Boolean hypercube on \( \{x_2, \ldots, x_k\} \in \{0, 1\}^{k-1} \) is connected by edges which flip just one bit. Without loss of generality, along some flip of \( x_2 \) we have two different values of \( G^{x_1=1}/G^{x_1=0} \). Suppose \( G^{x_1=1}/G^{x_1=0}(0, a_3, \ldots, a_k) \neq G^{x_1=1}/G^{x_1=0}(1, a_3, \ldots, a_k) \), for some \( a_3, \ldots, a_k \). Set \( x_3 = a_3, \ldots, x_k = a_k \), we get a binary function \( H(x_1, x_2) = G(1, 0, a_3, \ldots, a_k) \). We have \( H(1, 0)/H(0, 0) \neq H(1, 1)/H(0, 1) \), hence the rank of \( H = \begin{bmatrix} H(0, 0) & H(0, 1) \\ H(1, 0) & H(1, 1) \end{bmatrix} \) is two.
Suppose $H \in \mathcal{R}$. If a defining expression of $H$ uses binary \textsc{Equality} or \textsc{Disequality} functions, then we can set $x_1, x_2$ so that $H = 0$. But $H$ is never zero. So $H$ is defined by unary functions alone. But such a function has a matrix form of rank at most one. This contradiction proves that $H \not\in \mathcal{R}$ and completes our proof. □

### 3.5.4 Proof of Boolean \#CSP Dichotomy

Now we are ready to complete the proof for the main theorem of this chapter: Theorem 3.7.

**Proof of Theorem 3.7:** In Section 3.4 we have shown that \#CSP($\mathcal{A}$) and \#CSP($\mathcal{P}$) are computable in polynomial time.

If $\mathcal{F} \subsetneq \mathcal{A}$, by Lemma 3.19, either \#CSP($\mathcal{F}$) is \#P-hard, or we can simulate a unary function not in $\mathcal{A}$. After dividing by a non-zero factor, such a function takes the form $F = [1, \lambda]$, where $\lambda \not\in \{0, \pm1, \pm i\}$. If $\mathcal{F} \subsetneq \mathcal{P}$, by Lemma 3.20, either \#CSP($\mathcal{F}$) is \#P-hard, or we can simulate a function $P = [a, 0, 1, 0]$, or $P' = [0, 1, 0, a]$, where $a \neq 0$, or a binary function $H \not\in \mathcal{P}$ having no zero values.

Firstly, we prove \#CSP($\{F, P\}$) is \#P-hard. Clearly $P^{x_1=\ast} = [a, 1, 1]$. If $a \not\in \{1, -1\}$, \#CSP($[a, 1, 1]$) is \#P-hard by Theorem 3.14. If $a \in \{1, -1\}$, we can construct $Q(x_1, x_2) = \sum_{x_3} P(x_1, x_2, x_3)F(x_3) = [a, \lambda, 1]$, which is $[\pm1, \lambda, 1]$. Both problems \#CSP($[\pm1, \lambda, 1]$) are \#P-hard by Theorem 3.14. The proof for \#CSP($\{F, P'\}$) is the same.

Secondly, we prove \#CSP($\{F, H\}$) is \#P-hard. After normalizing, we may suppose $H = [1 \ x \ z]$, where $xyz \neq 0$. Since $\mathcal{P}$ is a superset of $\mathcal{R}$, $H$ is non-degenerate. Hence $z \neq xy$.

There are two cases, depending on whether $z = -xy$. For the case $z \neq -xy$, we construct a symmetric function $H(x_1, x_2)H(x_2, x_1) = [1, xy, z^2]$. By the conditions $xyz \neq 0, z \neq \pm xy$, it is impossible to be the first three tractable cases in Theorem 3.14. If it is in the last two tractable cases, then $xy$ is a power of $i$. Now we can form the function $H(x_1, x_2)H(x_2, x_1)F(x_1)F(x_2)$, which is $[1, \lambda xy, \lambda^2 z^2]$. This function has no zero entry and its matrix form has rank 2, so it is not of the first three tractable cases in Theorem 3.14. If it were in the last two tractable cases, then $\lambda xy$ is a power of $i$, which implies that $\lambda = (\lambda xy)/(xy)$ itself is a power of $i$. However since $[1, \lambda] \not\in \mathcal{A}$, $\lambda$ is not a power of $i$.

For the case $z = -xy$, We construct some binary functions with a non-negative integer parameter $s$ as follows:

\[
\sum_{x_3} H(x_1, x_3)H(x_2, x_3)(F(x_3))^s = [1 + \lambda^s x^2, (y + \lambda^s xz), (y^2 + \lambda^s z^2)] = [1 + \lambda^s x^2, y(1 - \lambda^s x^2), y^2(1 + \lambda^s x^2)].
\]

As $\lambda$ is not a power of $i$, at most one of the two values $x^2$ and $\lambda x^2$ can be a power of $i$. Now we choose $s = 0$ or $s = 1$ so that $\lambda^s x^2 \not\in \{\pm1, \pm i\}$. 88
After normalizing, we may write the function \([1 + \lambda^*x^2, y(1 - \lambda^*x^2), y^2(1 + \lambda^*x^2)]\) as \([1, y(1 - \lambda^*x^2)/(1 + \lambda^*x^2), y^2]\), noticing that \(1 + \lambda^*x^2 \neq 0\). We claim that this function is not one of the five tractable cases from Theorem 3.14. Since there are no zero entries, clearly it is not the first two cases. Its matrix form has rank 2, therefore it is not the third case. If it were the fourth tractable case \([1, \pm i, 1]\), then \(y = \pm 1\), and \((1 - \lambda^*x^2)/(1 + \lambda^*x^2) = \pm i\). This implies that \(\lambda^*x^2 = \pm i\), which is impossible. If \([1, y(1 - \lambda^*x^2)/(1 + \lambda^*x^2), y^2] = [1, \pm 1, -1]\), the fifth tractable case, then \(y = \pm 1\), and again \((1 - \lambda^*x^2)/(1 + \lambda^*x^2) = \pm i\), also impossible.

The proof of Theorem 3.7 is complete.

If we restrict to real (algebraic) valued constraint functions, we have the following theorem due to Bulatov, Dyer, Goldberg, Jalsenius and Richerby.

Theorem 3.21 (Bulatov et al.). Suppose \(F\) is a set of functions mapping Boolean inputs to (algebraic) real numbers. \(#\text{CSP}(F)\) is \(\#P\)-hard, unless

- Either every function in \(F\) takes the form \(\lambda \cdot \chi_{AX} \cdot (-1)^Q(X)\), where \(\lambda \in \mathbb{R}\), \(\chi_{AX}\) is a 0-1 indicator function with a matrix \(A\) over \(\mathbb{Z}_2\), and \(Q(X)\) is a homogeneous quadratic polynomial over \(\mathbb{Z}_2\);

- Or every function in \(F\) is a product of \((=_2), \neq_2)\) and real valued unary functions.

In these two cases, \(#\text{CSP}(F)\) is computable in polynomial time.

Exercise: Prove that Theorem 3.21 follows from Theorem 3.7. Similarly show that Theorem 3.1 and Theorem 3.4 can be derived from Theorem 3.21 or Theorem 3.7.
Chapter 4

Matchgates and Holographic Algorithms

Some counting problems are \#P-hard on general instances, but are tractable over planar structures. In fact, one of the most remarkable algorithms for counting problems is Kasteleyn’s algorithm (a.k.a. the FKT algorithm) to count the number of perfect matchings for any planar graph. Starting from this chapter, we will gradually build a theory that aims to show that this is not only a beautiful polynomial time counting algorithm, but also a universal strategy in a very strong technical sense, for a broad class of counting problems. This will be one of the highlights in Part I of this book, and will be built on the theory of matchgates and holographic algorithms. We note that a corresponding theory on problems over general domains has not yet been developed, and will be a major challenge for the future. To fully develop this topic we start with Pfaffian orientations on planar graphs.

4.1 Pfaffian Orientations

An \( n \times n \) matrix \( A \) is called skew-symmetric if \( A_{i,j} = -A_{j,i} \), for \( 1 \leq i, j \leq n \). We assume the entries of \( A \) are from a field of characteristic \( \neq 2 \). Then \( A_{i,i} = 0 \).

Every permutation \( \pi = \left( \begin{array}{c} 1 \ 2 \ \cdots \ 2k \\ i_1 \ i_2 \ \cdots \ i_{2k} \end{array} \right) \) defines a partition of \([2k]\) into disjoint pairs \( \bar{\pi} = \{\{i_1, i_2\}, \{i_3, i_4\}, \ldots, \{i_{2k-1}, i_{2k}\}\} \). The Pfaffian of an \( n \times n \) skew-symmetric matrix \( A \) is defined as follows. Suppose \( n = 2k \geq 2 \) is even, then

\[
\text{Pf}(A) = \sum_\pi \epsilon_\pi A_{i_1,i_2}A_{i_3,i_4}\cdots A_{i_{2k-1},i_{2k}} \tag{4.1}
\]

where the sum is over all permutations \( \pi = \left( \begin{array}{c} 1 \ 2 \ \cdots \ 2k \\ i_1 \ i_2 \ \cdots \ i_{2k} \end{array} \right) \) such that,

\[
i_1 < i_2, \ i_3 < i_4, \ldots, \ i_{2k-1} < i_{2k} \quad \text{and} \quad i_1 < i_3 < \ldots < i_{2k-1}. \tag{4.2}
\]

The symbol \( \epsilon_\pi \) in (4.1) denotes the sign of the permutation \( \pi \); it is +1 or −1 depending on whether the parity of \( \pi \) is even or odd, respectively. We note that there is a natural 1-1
We make a simple but important observation: If \( n \) is odd, then we define \( \text{Pf}(A) = 0 \). By convention if \( n = 0 \) we define \( \text{Pf}(A) = 1 \).

For any permutation \( \pi = \left( \begin{array}{ccc} 1 & 2 & \cdots & 2k \\ i_1 & i_2 & \cdots & i_{2k} \end{array} \right) \), not necessarily satisfying the stipulation (4.2), define

\[
a_\pi = \epsilon\pi A_{i_1, i_2} A_{i_3, i_4} \cdots A_{i_{2k-1}, i_{2k}}.
\]

We make a simple but important observation: If \( \tilde{\pi} = \pi' \), then \( a_\pi = a_{\pi'} \), i.e., the expression \( a_\pi \) has the same value if we list the partition \( \tilde{\pi} = \{(i_1, i_2), (i_3, i_4), \ldots, (i_{2k-1}, i_{2k})\} \) in any order of the pairs, as well as in any order of the two labels of each pair. This follows from the fact that \( A \) is skew-symmetric. This invariance will be used repeatedly.

We say two pairs of labels \( i_{2j-1} < i_{2j} \) and \( i_{2\ell-1} < i_{2\ell} \) form a crossover, or an overlapping pair, if \( i_{2j-1} < i_{2\ell-1} < i_{2j} < i_{2\ell} \) or \( i_{2\ell-1} < i_{2j-1} < i_{2\ell} < i_{2j} \). Suppose \( \pi \) is a permutation satisfying the stipulation (4.2). Let \( c(\pi) \) be the number of crossovers among the pairs in the partition \( \tilde{\pi} \). Then the sign \( \epsilon_\pi \) is also \((-1)^{c(\pi)}\). Hence it can be computed as the parity of the number of crossovers. To see this, consider the permutation \( \pi \) and consider a sequence of adjacent transpositions which moves the sequence \((1, 2, \ldots, 2k - 1, 2k)\) to \((i_1, i_2, \ldots, i_{2k-1}, i_{2k})\). We have \( i_1 = 1 \) by (4.2). The number of transpositions that will bring \( i_2 \) to the position right after 1 is the number of labels strictly between \( i_1 = 1 \) and the number \( i_2 \), and has the same parity as the number of crossovers the pair \( \{i_1, i_2\} \) forms with all other pairs of labels \( \{(i_3, i_4), \ldots, (i_{2k-1}, i_{2k})\} \) in \( \tilde{\pi} \). After \( i_1, i_2 \) are placed in the first two positions, if \( n > 2 \), then \( i_3 \) is the minimum among all other labels by (4.2), and is currently located right after the first two elements. Then we move \( i_4 \) to the position right after \( i_3 \). The proof is completed by induction. We note that the definition of Pfaffian and this notion of crossovers apply to any totally ordered label set, in particular to any index set of a principal submatrix obtained from \( A \) by selecting a common subset of rows and columns.

The Pfaffian can be computed in polynomial time [?, ?>. A key relation to determinant is the following theorem due to Caylay [?, ?].

**Theorem 4.1.** For any \( n \times n \) skew-symmetric matrix \( A \),

\[
\det(A) = \left[\text{Pf}(A)\right]^2.
\]

Note that for odd \( n \), a skew-symmetric matrix \( A \) has \( \det(A) = 0 \). For even \( n \), an algorithm similar to Gaussian elimination can be derived using the identity \( \text{Pf}(BAB^T) = \det(B)\text{Pf}(A) \), for any \( n \times n \) matrix \( B \). It can be even computed in NC [?].

Let \( G = (V, E) \) be a simple undirected graph without self loops and parallel edges. Suppose the vertices of \( G \) are labeled by a totally ordered set, e.g, \( V = \{1, 2, \ldots, n\} \). Assign an indeterminate \( x_e \) for every edge \( e = \{u, v\} \in E \). Then we define the skew-symmetric
adjacency matrix $A = A(G)$ of the graph $G$ to be
\[
A_{u,v} = \begin{cases} 
  x_e & \text{if } e = \{u, v\} \in E \text{ and } u < v \\
  -x_e & \text{if } e = \{u, v\} \in E \text{ and } u > v \\
  0 & \text{if } \{u, v\} \notin E
\end{cases} \tag{4.3}
\]

Note that for any permutation $\pi$, the partition $\tilde{\pi}$ is a perfect matching if all pairs are edges. Denote by $\mathcal{M}(G)$ the set of all perfect matchings of $G$. There is a 1-1 correspondence between non-zero terms in $\text{Pf}(A)$ and $\mathcal{M}(G)$. For any $M \in \mathcal{M}(G)$ there are $2^{k!}$ permutations of the form $\pi = (1 \ 2 \ \cdots \ 2k)$ that can represent $M$, i.e., $\tilde{\pi} = M$, of which the unique permutation that satisfies the stipulation (4.2) is called a canonical expression. All other permutation expressions for $M$ are obtained by permuting the order of the pairs $\{i_{2k-1}, i_{2k}\}$ and the order within each pair.

We define the weight of a perfect matching $M$ in $G$ to be the product of edge weights $x_e$ for all $e \in M$, namely $\Gamma(M) = \Gamma_G(M) = \prod_{e \in M} x_e$. It follows that $a_\pi = \pm \Gamma(M)$, if $\tilde{\pi} = M$.

An orientation of a graph $G$ assigns one direction to each edge of $G$. We denote by $u \rightarrow v$ if the edge $\{u, v\}$ is oriented from $u$ to $v$. In this case $u$ is its tail, and $v$ is its head. For an oriented graph $\vec{G}$ we modify the skew-symmetric matrix $A$ in (4.3) to be $B = B(\vec{G})$:
\[
B_{u,v} = \begin{cases} 
  x_e & \text{if } e = \{u, v\} \in E \text{ and } u \rightarrow v \\
  -x_e & \text{if } e = \{u, v\} \in E \text{ and } v \rightarrow u \\
  0 & \text{if } \{u, v\} \notin E
\end{cases} \tag{4.4}
\]

In other words, we change the sign at both entries $A_{u,v}$ and $A_{v,u}$ if $u < v$ and $\{u, v\}$ is oriented $v \rightarrow u$.

Given an orientation, we will consider the Pfaffian of $B$
\[
\text{Pf}(B) = \sum_{\pi} \epsilon_\pi B_{i_1,i_2} B_{i_3,i_4} \cdots B_{i_{2k-1},i_{2k},} \tag{4.5}
\]
where the sum is over permutations $\pi$ satisfying (4.2). For a perfect matching $M$ in an oriented graph $\vec{G}$, suppose $M = \tilde{\pi}$, define the Pfaffian term
\[
\text{Pf}_G(M) = \epsilon_\pi B_{i_1,i_2} B_{i_3,i_4} \cdots B_{i_{2k-1},i_{2k}}. \tag{4.6}
\]

It is a term in $\text{Pf}(B)$ when $\pi$ is the canonical expression for $M$ and it is equal to either the weight of the perfect matching $\Gamma_G(M)$ or its negation $-\Gamma_G(M)$. However we can choose any permutation $\pi$ representing $M$, and the value $\text{Pf}_G(M)$ is defined, which does not depend on the particular partition $\pi$ corresponding to $M$, due to the invariance observed earlier.

**Definition 4.2.** For any perfect matching $M$ in an oriented graph $\vec{G}$, the sign of the perfect matching $M$ with respect to this orientation is
\[
\text{sgn}_\vec{G}(M) = \frac{\text{Pf}_G(M)}{\Gamma_G(M)} \in \{-1, 1\}. \tag{4.7}
\]

When $\vec{G}$ is clear from the context, we write $\text{sgn}(M)$ instead.
We note that \( \text{sgn}(M) \) can be computed with any partition \( \pi \) corresponding to \( M \), i.e., \( \overline{\pi} = M \), due to the invariance of \( \text{Pf}_{\overline{G}}(M) \). In particular, \( \text{sgn}(M) \) can be computed simply as the sign \( \epsilon_\pi \) of the permutation \( \pi = \left( \begin{array}{cccc} 1 & 2 & \ldots & 2k \\ i_1 & i_2 & \ldots & i_{2k} \end{array} \right) \) where each matching edge \( \{i_{2\ell-1}, i_{2\ell}\} \in M \) is listed by its orientation \( i_{2\ell-1} \rightarrow i_{2\ell} \). This permutation \( \pi \) is in general not a canonical expression.

If \( M \) and \( M' \) are two perfect matchings, the symmetric difference \( M \oplus M' \) consists of a set of cycles of even length that are alternating between edges from \( M \) and \( M' \). For any orientation on a graph \( G \), if \( C \) is a cycle of even length, then we say it is evenly oriented if there are an even number of edges oriented in one direction, and oddly oriented otherwise. Clearly this notion does not depend on the direction since \( C \) has an even length.

**Lemma 4.3.** For any orientation on a graph \( G \), and perfect matchings \( M \) and \( M' \), if \( k \) is the number of evenly oriented cycles in \( M \oplus M' \) with respect to the orientation, then

\[
\text{sgn}(M) \cdot \text{sgn}(M') = (-1)^k.
\]

**Proof.** We first show that if this equality holds for one orientation then it holds for all orientations. Let \( e \) be an edge in \( \overline{G} \) and suppose we reverse its orientation. If \( e \notin M \oplus M' \) then obviously the reversal has no effect on the equality. Suppose \( e \in M \oplus M' \). Then \( k \) is changed by one, and exactly one of \( \text{sgn}(M) \) and \( \text{sgn}(M') \) is changed. Hence we may choose any orientation convenient for \( M \) and \( M' \) to prove our equality. So we choose one which gives a cyclic orientation on each cycle of \( M \oplus M' \) (either direction of the cyclic orientation for each cycle is acceptable). We may orient all other edges arbitrarily. In this orientation, \( k \) is the number of cycles in \( M \oplus M' \).

Next we will relabel the vertices of \( \overline{G} \). If we apply a permutation \( \sigma \) on the vertices of \( \overline{G} \), the Pfaffian term for a perfect matching \( M \) is changed as follows. Suppose \( M \) corresponds to the permutation \( \pi = \left( \begin{array}{cccc} 1 & 2 & \ldots & 2k \\ i_1 & i_2 & \ldots & i_{2k} \end{array} \right) \) where each matching edge \( \{i_{2\ell-1}, i_{2\ell}\} \in M \) is listed by its orientation \( i_{2\ell-1} \rightarrow i_{2\ell} \). We write its corresponding Pfaffian term \( b_{\pi} = \epsilon_\pi B_{i_1,i_2}B_{i_3,i_4} \cdots B_{i_{2k-1},i_{2k}} \). If \( \sigma = \left( \begin{array}{cccc} i_1 & i_2 & \ldots & i_{2k} \\ j_1 & j_2 & \ldots & j_{2k} \end{array} \right) \), then each matching edge is also oriented as \( j_{2\ell-1} \rightarrow j_{2\ell} \) in the new labeling. The skew-symmetric matrix \( B' \) defined by the new labeling has \( B'_{j_1,j_2}B'_{j_3,j_4} \cdots B'_{j_{2k-1},j_{2k}} = B_{i_1,i_2}B_{i_3,i_4} \cdots B_{i_{2k-1},i_{2k}} \). But the sign \( \epsilon_\sigma \) becomes \( \epsilon_{\sigma \pi} = \epsilon_\sigma \epsilon_\pi \). Hence the change to \( \text{sgn}(M) \) is to multiply it by \( \epsilon_\sigma \). The same is true for the change to \( \text{sgn}(M') \).

Note that we wrote \( \sigma \) as \( \left( \begin{array}{cccc} i_1 & i_2 & \ldots & i_{2k} \\ j_1 & j_2 & \ldots & j_{2k} \end{array} \right) \) in reference to \( \pi \) that corresponds to \( M \). However the conclusion that we change \( \text{sgn}(M) \) by multiplying it with \( \epsilon_\sigma \) is independent of how we express \( \sigma \). In particular, for \( M' \), we can write \( \sigma \) in reference to \( \pi' \) corresponding to \( M' \), and the conclusion is the same. Hence the product \( \text{sgn}(M) \cdot \text{sgn}(M') \) remains unchanged. Meanwhile the value \( k \) is clearly the same. So we may choose to label all vertices as follows. Pick any cycle of \( M \oplus M' \) and start at any edge of this cycle in \( M \). Label its tail as 1 and its head as 2. Label the rest of the cycle along the chosen orientation using the next unused integers. Thus if the cycle has length \( 2\ell \), then its vertices are labeled from 1 to \( 2\ell \).
consecutively along the orientation starting at the tail of one edge of \( M \). Then starting with \( 2\ell + 1 \) we label the next cycle of \( M \oplus M' \), if there is any, in exactly the same way. After all cycles of \( M \oplus M' \) are dealt with, we label all remaining vertices so that each edge in \( M \cap M' \) is labeled consecutively with the next unused integers, and increasing from tail to head.

Now the Pfaffian term \( \operatorname{Pf}_G(M) \) for \( M \) is just the product of all edge weights \( \Gamma_G(M) \). This is seen easily if we write \( b_\pi \) in the way where the permutation \( \pi \) is the identity, and we list all matched edges in the oriented order. For the Pfaffian term \( b_{\pi'} \) corresponding to \( M' \), we still list the product part \( B_{i_1,i_2}B_{i_3,i_4} \cdots B_{i_{2k-1},i_{2k}} \) in the oriented order for each matched edge. The sign \( \epsilon_{\pi'} \) for the permutation is as follows. The part for the first cycle of length \( 2\ell \) in \( M \oplus M' \) has the form \( (1 \ 2 \ \ldots \ 2\ell - 1 \ 2\ell) \), which is an even cycle as a permutation, and has an odd parity. The permutation \( \pi' \) is simply a product of these cycles in the permutation group, disjoint and one for each cycle in the graph \( M \oplus M' \). Hence it has parity \( \epsilon_{\pi'} = (-1)^k \).

**Definition 4.4.** An orientation of a connected plane graph is called a Pfaffian orientation if along the boundary of every non-outer face, there are an odd number of clockwise oriented edges.

**Lemma 4.5.** Any Pfaffian orientation in a connected plane graph \( G \) satisfies the following property: For every cycle \( C \), the number of clockwise oriented edges of \( C \) is of the opposite parity to the number of vertices contained within \( C \). (This number does not include the vertices on the cycle \( C \)).

**Proof.** Let \( V \) and \( E \) be the number of vertices and edges contained within \( C \), respectively, and let \( \ell \) be the number of edges on \( C \), which is also the number of vertices on \( C \). The vertices contained within \( C \) are those in the interior of the region bounded by \( C \); they do not include those on the cycle \( C \). Similarly the edges within \( C \) do not include those on \( C \). An edge is a bridge edge iff it has the same face on both sides. Then in a traversal of the boundary of a face \( \Delta \), if there is a bridge edge \( e \), it is traversed twice, once in each direction. In particular in either orientation of \( e \), it contributes one to the count of clockwise oriented edges along the boundary of \( \Delta \). For a non-bridge edge \( e \), depending on its orientation, it contributes one to the count of clockwise oriented edges to exactly one face it borders on but not the other.

Suppose there are \( F \) faces bounded by \( C \), and let \( c_i \) be the number of clockwise oriented edges on the boundary of the \( i \)-th face \( (1 \leq i \leq F) \). Each \( c_i \) is odd by assumption, therefore \( F \equiv \sum_{i=1}^{F} c_i \) (mod 2). By Euler’s formula, counting the face formed by the exterior of \( C \), we have \((V + \ell) - (E + \ell) + (F + 1) = 2\). It follows that \( E = V + F - 1 \).

If we add up all the clockwise oriented edges among all boundary edges in \( F \) faces, each interior edge within \( C \) regardless orientation contributes one and each clockwise oriented edge on \( C \) contributes one. Hence \( \sum_{i=1}^{F} c_i = E + c \), where \( c \) is the number of clockwise oriented edges on \( C \).
It follows that
\[ F \equiv \sum_{i=1}^{F} c_i = E + c = V + F - 1 + c \quad (\text{mod } 2), \]
and hence \( V + c \equiv 1 \quad (\text{mod } 2). \)

If \( G \) has a Pfaffian orientation, then with respect to this orientation every two perfect matchings \( M \) and \( M' \) must have the same sign: \( \text{sgn}(M) = \text{sgn}(M') \). In fact since every cycle of \( M \oplus M' \) contains a perfect matching (possibly empty) in its interior, the number of vertices within the cycle must be even. Each cycle of \( M \oplus M' \) has an even length, consisting of alternatingly edges from \( M \) and \( M' \). By Lemma 4.5 the cycle is oddly oriented. By Lemma 4.3, \( \text{sgn}(M) \cdot \text{sgn}(M') = 1 \).

Hence for a Pfaffian orientation, every Pfaffian term has the same sign.

**Definition 4.6.** The perfect matching polynomial, \( \text{PerfMatch}(G) \), is the following:
\[
\text{PerfMatch}(G) = \sum_{M \in \mathcal{M}(G)} \prod_{e \in M} x_e \quad (4.8)
\]
where \( \mathcal{M}(G) \) is the set of all perfect matchings in \( G \) and \( x_e \) corresponds to the edge \( e \) in \( G \).

We will usually assign a value \( w(e) \) in a field to each \( x_e \) as the edge weight for \( e \). The default choice of the field is \( \mathbb{C} \); but one can replace \( \mathbb{C} \) by any commutative ring. To discuss the computation of the perfect matching polynomial \( \text{PerfMatch}(G) \), we usually fix an assignment of weights \( w(e) \in \mathbb{C} \) for \( x_e \). The case \( w(e) = 1 \) for all edges \( e \) in \( G \) is called the unweighted case; in this case \( \text{PerfMatch}(G) \) is the number of perfect matchings in \( G \). For a graph \( G \) with at least one vertex, if it has no perfect matchings, then \( \text{PerfMatch}(G) = 0 \). By convention if \( G \) is the empty graph with no vertices, we define \( \text{PerfMatch}(G) = 1 \). Pfaffian orientation can be used to compute \( \text{PerfMatch}(G) \) for any planar graph \( G \).

**Theorem 4.7** (Kasteleyn). Every connected planar graph has a Pfaffian orientation. Such an orientation can be constructed in polynomial time, leading to a polynomial time algorithm to compute \( \text{PerfMatch}(G) \) for any weighted planar graph \( G \).

**Proof.** A graph can tested for planarity, and if so a planar embedding of the graph can be computed in polynomial time \( ?? \). Assume \( G \) is a connected planar graph, and given a planar embedding. If \( G \) is a tree, then any orientation is acceptable. If \( G \) is not a tree, then choose any edge on the boundary of the outer face which belongs to a cycle. (Such an edge exists by a simple induction: Take any edge \( e \) on the boundary of the outer face. If \( e \) is a bridge edge, then its removal disconnects \( G \) into two parts. Since \( G \) is not a tree, at least one part is not a tree. The traversal of the boundary of the outer face of \( G \) consists of the concatenation of two traversals of the boundaries of the outer faces of the two parts, in sequence. Then by induction, one edge of the traversal belongs to a cycle.) Let \( F \) be the non-outter face containing this edge \( e \). By induction we can construct a Pfaffian orientation for \( G - \{e\} \), the graph with the same vertex set as \( G \) but with edge \( e \) removed. By definition, in this Pfaffian
orientation every non-outer face of $G - \{e\}$ contains an odd number of clockwise oriented edges. Now add $e$ back, and orient $e$ appropriately we can guarantee that $F$ also has an odd number of clockwise oriented edges. This completes the proof by induction.

For a Pfaffian orientation, let $B = B(G)$ be the skew-symmetric matrix. Then either $\text{Pf}(B) = \text{PerfMatch}(G)$ or $\text{Pf}(B) = -\text{PerfMatch}(G)$. The equality is a polynomial equality: for the given Pfaffian orientation, either $+$ holds for all weight values, or $-$ holds for all weight values. Setting all weight values to 1, we can decide which sign is valid for the particular orientation (unless there is no perfect matching and $G$ is non-empty, in which we can safely output $\text{PerfMatch}(G) = 0$). Then we can compute $\text{PerfMatch}(G)$ for the actual weight values. If $G$ has connected components $G_1, G_2, \ldots, G_m$, then $\text{PerfMatch}(G) = \prod_{i=1}^{m} \text{PerfMatch}(G_i)$.  

**Exercise:** Show that $\text{PerfMatch}(G)$ can be computed in polynomial time for weighted planar (multi)graphs $G$ with self-loops and multiple edges.

### 4.2 Matchgates

Valiant initially introduced *matchgates* as a way to show that a non-trivial, though restricted, fragment of quantum computation can be simulated in classical polynomial time [?]. At the heart of this simulation technique is a way to encode certain quantum states by a classical computation of perfect matchings, and to simulate certain quantum gates by the so-called matchgates. Subsequently he introduced holographic algorithms using matchgates [?]. It is in this context that holographic transformations were introduced. This idea turns out to have a wider applicability as a general reduction method. We have already encountered its use for the dichotomy of Holant* problems in Chapter 2. In the context of holographic algorithms based on matchgates, this is essentially a reduction from a problem to the planar Perfect Matching problem. Theorem 4.7 shows that there is a remarkable polynomial time algorithm—Kasteleyn’s algorithm—to compute the weighted sum of perfect matchings for a planar graph. We introduce matchgates in this section.

**Definition 4.8.** A matchgate is an undirected weighted plane graph $G$ with a subset of distinguished nodes on its outer face, called the external nodes, ordered in a clockwise order.

A plane graph is a planar graph given with a particular planar embedding. Without loss of generality, we assume all edge weights are non-zero; zero weighted edges can be deleted.

Let $G$ be a matchgate with $k$ external nodes. For each length-$k$ bitstring $\alpha$, $G$ defines a subgraph $G^\alpha$ obtained from $G$ by the following operation: For all $1 \leq i \leq k$, if the $i$-th bit of $\alpha$ is 1, then we remove the $i$-th external node and all its incident edges. Thus, $G^{\alpha_0 \ldots \alpha_0} = G$, and $G^{\alpha_1 \ldots \alpha_1}$ is obtained from $G$ by removing all external nodes and their incident edges.
Definition 4.9. We define the signature of a matchgate $G$ as the vector $\Gamma_G = (\Gamma_G^\alpha)$, indexed by $\alpha \in \{0, 1\}^k$ in lexicographic order, as follows:

$$\Gamma_G^\alpha = \text{PerfMatch}(G^\alpha) = \sum_{M \in \mathcal{M}(G^\alpha)} \prod_{e \in M} w(e).$$ (4.9)

For a perfect matching $M \in \mathcal{M}(G^\alpha)$ we define $\Gamma_G^\alpha(M) = \prod_{e \in M} w(e)$ as the perfect matching term, equal to the product of the edge weights for the matching $M$. Where $G$ is clear, we omit the subscript $G$, and write $\Gamma$ for $\Gamma_G$, $\Gamma^\alpha$ for $\Gamma_G^\alpha$, and $\Gamma^\alpha(M)$ for $\Gamma_G^\alpha(M)$.

We denote by $\mathcal{M}$ the set of all matchgate signatures.

Definition 4.10. A matchgate signature $\Gamma$ is symmetric if, for all $\alpha$ and $\beta$ of equal Hamming weight, $\Gamma^\alpha = \Gamma^\beta$.

In other words, the value of a symmetric signature entry is only a function of how many 1’s there are in its index, not their particular pattern. These signatures are important because they have a clear combinatorial meaning.

Regarding symmetric signatures that are realizable as matchgate signatures, we will prove

Theorem 4.11. A symmetric signature is the signature of a matchgate iff it has the following form, for some $a, b \in \mathbb{C}$ and integer $k$ (we take the convention that $0^0 = 1$):

1. $[a^kb^0, 0, a^{k-1}b, 0, a^{k-2}b^2, 0, \ldots, a^0b^k]$ (arity $2k \geq 2$)
2. $[a^kb^0, 0, a^{k-1}b, 0, a^{k-2}b^2, 0, \ldots, a^0b^k, 0]$ (arity $2k + 1 \geq 1$)
3. $[0, a^kb^0, 0, a^{k-1}b, 0, a^{k-2}b^2, 0, \ldots, a^0b^k]$ (arity $2k + 1 \geq 1$)
4. $[0, a^kb^0, 0, a^{k-1}b, 0, a^{k-2}b^2, 0, \ldots, a^0b^k, 0]$ (arity $2k + 2 \geq 2$).

We will postpone the proof of this theorem until we have properly developed the theory of matchgates.

Some Problems

We list some problems that can be solved in polynomial time using holographic algorithms with matchgates. The first several problems are all from the initial paper by Valiant [?].

We start with a problem motivated by statistical physics. An orientation of an undirected graph $G$ is an assignment of a direction to each of its edges. An “ice problem” involves counting the number of orientations of a graph such that certain local constraints are satisfied. Pauling [?] initially proposed such a model for planar square lattices, where the constraint was that an orientation assigned exactly two incoming and two outgoing edges at every node. The question of determining how the number of such orientations grows for various planar repeating structures has been extensively studied [?, ?, ?, ?, ?, ?].
**#PL-3-NAE-ICE**

**INPUT:** A planar graph \( G = (V, E) \) of maximum degree 3.

**OUTPUT:** The number of orientations such that no node has all incident edges directed toward it or all incident edges directed away from it.

Hence #PL-3-NAE-ICE counts the number of no-sink-no-source orientations. A node of degree one will preclude such an orientation. We assume every node has degree 2 or 3. To solve this problem by a holographic algorithm with matchgates, we design a signature grid based on \( G \) as follows: We attach to each node of degree 3 a signature \([0; 1; 1; 0]\) (as a contravariant tensor). This represents a Not-All-Equal or Nae gate of arity 3. For any node of degree 2 we use a binary Nae signature, i.e., a binary Disequality signature \((\neq) = [0, 1, 0]\) (also a contravariant tensor). For each edge in \( E \) we use a binary Disequality signature \((\neq)\) (as a covariant tensor), which stands for an orientation from one node to the other. (To express such a problem, it is completely arbitrary to label one side as contravariant and the other side as covariant tensors, cf. Section ?? of Chapter 2. They are called generators and recognizers, respectively.) From the given planar graph \( G \) we obtain a signature grid \( \Omega \), where the underlying graph \( G' \) is the Edge-Vertex incidence graph of \( G \).

By definition, \( \text{Holant}_\Omega \) is an exponential sum where each term is a product of appropriate entries of the signatures. Each term is indexed by a 0-1 assignment on all edges of \( G' \); it has value 0 or 1, and it has value 1 iif it corresponds to an orientation of \( G \) such that at every vertex of \( G \) the local Nae constraint is satisfied. Therefore \( \text{Holant}_\Omega \) is precisely the number of valid orientations required by #PL-3-NAE-ICE.

Note that the signature \([0, 1, 1, 0]\) is not the signature of any matchgate according to Theorem 4.11. A simple reason for this is that a matchgate signature, being defined in terms of perfect matchings, cannot have non-zero values for inputs of both odd and even Hamming weight.

However, under a holographic transformation using \( H = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \) with \( H^{-1} = 2H \),

\[
(H^{-1})^\otimes_3 [0, 1, 1, 0] = (H^{-1})^\otimes_3 \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}^\otimes_3 - \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\otimes_3 - \begin{bmatrix} 0 \\ 1 \end{bmatrix}^\otimes_3 \right\} = [6, 0, -2, 0],
\]

\[
(H^{-1})^\otimes_2 [0, 1, 0] = H^\otimes_2 \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}^\otimes_2 - \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\otimes_2 - \begin{bmatrix} 0 \\ 1 \end{bmatrix}^\otimes_2 \right\} = [2, 0, -2],
\]

and

\[
[0, 1, 0] H^\otimes_2 = \frac{1}{2} [1, 0, -1],
\]

as the matrix form of this binary signature is \[\frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\]. These signatures are all realizable as matchgate signatures, by Theorem 4.11. More concretely, we can exhibit the requisite matchgates in Figures 4.1a and 4.1b.

Hence #PL-3-NAE-ICE is precisely the following Holant problem on planar graphs:

\[
\text{Holant}([0, 1, 0] | [0, 1, 0], [0, 1, 1, 0]) \equiv_T \text{Holant}(\frac{1}{2}[1, 0, -1] | [2, 0, -2], [6, 0, -2, 0]).
\]
An often useful perspective on holographic transformations is to think of a matrix $M = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in GL_2(\mathbb{C})$ as representing two generalized truth values of $[b_0 = (a \ b), b_1 = (c \ d)]$.

We consider a covering problem next. A cycle is a sequence of edges through distinct nodes that starts and ends at the same vertex. A chain is a sequence of edges through distinct nodes that starts and ends at distinct vertices. A cycle-chain cover in $G = (V,E)$ is a set of cycles and chains that are vertex disjoint and the union of their vertex sets is all of $V$. For real values $(x, y)$, the $(x, y)$ cycle-chain sum of $G$ is the sum of $x^i y^j$ over all cycle-chain covers $C$, where $i$ and $j$ are the numbers of cycles and chains in $C$ respectively. Note that the $(1,1)$ cycle-chain sum is the total number of cycle-chain covers $G$. Note also that the $(2,0)$ cycle-chain sum is a weighted sum of cycle-chain covers $C$ of $G$ with only cycles, and the weight is $2^i$ if $C$ has $i$ cycles. In particular $1/2$ of the $(2,0)$ cycle-chain sum mod 2 is the parity of the number of Hamiltonian cycles of $G$. The number of Hamiltonian cycles is #P-hard [?], and the proof can be adapted to show that if nodes of both degrees 2 and 3 are allowed, then the parity problem is $\oplus$P-complete [?], which is NP-hard under randomized reduction [?]. Similarly, the $(2,4)$ cycle-chain sum is also $\oplus$P-hard.

**#PL-3-CYCLE-CHAIN**

**INPUT:** A planar 3-regular graph $G = (V,E)$.

**OUTPUT:** The $(1,1)$ cycle-chain sum of $G$.

To solve this problem by a holographic algorithm with matchgates, we design a signature grid as follows: We represent each node of degree 3 in $V$ by a generator with signature $[0, 1, 1, 0]$. For each edge in $E$ we use a recognizer with a binary EQUALITY signature.
\(=2\) = \([1, 0, 1]\), which enforces a 0-1 assignment of the edges, and is interpreted as either taking or not taking the edge in the cycle-chain cover. The signature \([0, 1, 1, 0]\) at each node of degree 3 enforces that the subgraph formed by the edges taken is a spanning subgraph (i.e., a subgraph containing all vertices) and has degree either 1 or 2 at every vertex, hence a cycle-chain cover.

By a holographic transformation using \(H\), both \([6, 0, -2, 0] = (H^{-1}) \otimes 3 [0, 1, 1, 0]\) and \(\frac{1}{2}[1, 0, 1] = [1, 0, 1] H \otimes 2\) are realizable as matchgate signatures (Figures 4.1a and 4.1c). Hence \#PL-3-CYCLE-CHAIN is computable in polynomial time.

The next problem is a Satisfiability problem.

**#PL-3-NAE-SAT**

**Input:** A planar formula \(\Phi\) consisting of a conjunction of NAE clauses each of size 3.

**Output:** The number of satisfying assignments of \(\Phi\).

This is a variant of 3SAT. Lichtenstein [?] defined the notion of planar formulae. A Boolean formula is planar if it can be represented by a planar graph where vertices represent variables and clauses, and there is an edge iff the variable or its negation appears in that clause. The SAT problem is when the gate for each clause is the Boolean Or. When SAT is restricted to planar formulae it is still NP-complete, and its corresponding counting problem is \#P-complete. Moreover, for many connectives other than NAE (e.g., EXACTLY-ONE) the unrestricted or the planar decision problems are still NP-complete, and the corresponding counting problems are \#P-complete [?].

We design a signature grid as follows: To each NAE clause we assign a generator with signature \([0, 1, 1, 0]\). To each Boolean variable we assign a generator with signature \((=k)\) where \(k\) is the number of clauses the variable appears, either negated or unnegated. Further, if a variable occurrence is negated we have a recognizer \([0, 1, 0]\) along the edge that joins the variable generator and the NAE generator, and if the variable occurrence is unnegated then we use a recognizer \([1, 0, 1]\) instead. Under a holographic transformation using \(H\), \((=k)\) is transformed to

\[
(H^{-1}) \otimes k \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes k + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \otimes k \right\} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes k + \begin{bmatrix} 1 \\ -1 \end{bmatrix} \otimes k = 2[1, 0, 1, 0, \ldots].
\]

Thus, by Theorem 4.11, all the signatures used can be realized by matchgates under a holographic transformation.

The next problem is essentially a decision problem, rather than a counting problem. It is stated as an optimization problem minimizing the cardinality of a subset to be removed. With an additional threshold parameter \(k\) as input, one can easily define a decision problem on whether a subset of cardinality at most \(k\) can be removed.

**PL-NODE-BIPARTITION**

**Input:** A planar graph \(G = (V, E)\) of maximum degree 3.
OUTPUT: The minimum cardinality of a subset $V' \subset V$ such that the deletion of $V'$ and its incident edges results in a bipartite graph.

The decision problem is known to be NP-complete for maximum degree 6 [?]. See [?] for a general approach to such “node deletion” problems. Note that numerous other planar NP-complete problems, such as Hamiltonian cycles and minimum vertex covers are NP-complete already for degree 3 [?, ?].

To solve this problem, we may assume all vertices are of degree 2 or 3, as retaining any vertex of degree 1 does not violate bipartiteness. We will use generator signatures $[x, 1, x]$ and $[x, 1, 1, x]$ for a vertex of degree 2 and 3 respectively. These are realizable by matchgates under $H$. For each edge we use a recognizer signature $[0, 1, 0]$ which results in an orientation. For any orientation, each source or sink node contributes a factor $x$, and the other nodes contribute a factor 1. The Holant sum is a polynomial in $x$, where the largest $i$ for which $x^i$ has a non-zero coefficient is exactly the maximum number of nodes that a bipartite graph can have that is obtained by deleting nodes and incident edges from $G$.

The maximum possible degree this polynomial can have is $|V|$. Thus by evaluating the Holant at $|V| + 1$ distinct values of $x$, we can interpolate the polynomial and reconstruct it. Consequently we can find its actual degree $d$, which gives $|V| - d$ as the answer to the instance of PL-NODE-BIPARTITION.

We note that if instead of considering node deletion we consider edge deletion, this is just another way of defining the well known problem of MAX-CUT, which is NP-hard (and even NP-hard to approximate [?, ?, ?, ?]). Planar MAX-CUT is known to be in P [?]. The following problem is a joint generalization of both. It can be solved similarly in polynomial time by a holographic algorithm with matchgates.

Exercise: Show that, for any $x$, the signatures $[x, 1, x]$ and $[x, 1, 1, x]$ can be realized by matchgates under $H$.

Exercise: Show that $[x, y, y, x]$ can be realized by matchgates under $H$, for all $x$ and $y$. Then show that the following problem PL-NODE-EDGE-BIPARTITION is solvable in polynomial time.

**PL-NODE-EDGE-BIPARTITION**

**Input:** A planar graph $G = (V, E)$ of maximum degree 3. A non-negative integer $k \leq |V|$.

**Output:** The minimum $l$ such that deletion of at most $k$ nodes (including all of their incident edges) and $l$ more edges results in a bipartite graph.

We now consider a matching problem. Jerrum [?] showed that counting the number of (not necessarily perfect) matchings in a planar graph is \#P-complete, and Vadhan [?] subsequently proved that it remains \#P-complete even for planar bipartite graphs of degree six. For degree two the problem can be solved easily and one might have conjectured that all
other nontrivial cases are \( \#P \)-complete. However, the following problem \cite{?} can be solved in polynomial time.

**\#X-MATCHINGS**

**INPUT:** A planar weighted bipartite graph \( G = (V, E, W) \) where \( V \) has bipartition \( V_1, V_2 \) and the nodes in \( V_1 \) have degree 2.

**OUTPUT:** The sum of the masses of all matchings of all sizes where the mass of a matching is the product of (i) the weights of all the edges present in the matching, and (ii) the quantity \(- (w_1 + \ldots + w_k)\) for all the \( V_2 \) nodes that are not matched, where \( w_1, \ldots, w_k \) are the weights of the edges incident to that (unmatched) node.

Note that if every \( V_2 \) node has degree 4 and every edge has weight one, then computing \#X-MATCHINGS gives the number of matchings, but each weighted by \((-4)^k\), where \( k \) is the number of unmatched \( V_2 \) nodes. Computing this mod 5 gives the number of matchings mod 5.

![Figure 4.2: A matchgate with an underlying star graph](image)

To solve this problem, we use a different matrix \( M = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \). Consider the matchgate pictured in Figure 4.2. This star graph has a central vertex \( v_0 \) and \( k \) external nodes labeled \( v_1, \ldots, v_k \), and the edge \((v_0, v_i)\) has weight \( w_i \) (1 \( \leq \) \( i \) \( \leq \) \( k \)). Its signature \( \Gamma \) has non-zero entries only at weight \( k-1 \), and the value \( \Gamma^\alpha = w_i \), where \( \alpha = 1^{i-1}01^{k-i} \) has a single 0 at bit position \( i \). The covariant transformation is \( \Gamma M^{\otimes k} = \sum_{i=1}^{k} w_i (1, 0)^{\otimes (i-1)} \otimes (-1, 1) \otimes (1, 0)^{\otimes (k-i)} \). This signature has non-zero values \(- (w_1 + \ldots + w_k)\) at \( 0^k \), and \( w_i \) at \( 0^i10^{k-i} \). All other entries of \( \Gamma M^{\otimes k} \) are zero.

To each vertex in \( V_2 \) we attach the recognizer signature \( \Gamma M^{\otimes k} \) above. To each vertex in \( V_1 \) we attach the At-Most-One signature \([1, 1, 0]\) as a generator. Under a contravariant transformation \((M^{-1})^{\otimes 2}[-1, 0, 1] = [1, 1, 0] \). We verify that \( M^{\otimes 2}[1, 1, 0] = [-1, 0, 1] \) by its matrix form

\[
\begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}
\]

and \([1, 0, 1]\) can be realized by a matchgate consisting of a path of length 3, where the two end vertices are external nodes, and the weights of the three edges are respectively 1, 1 and \(-1\) along the path.
This problem is motivated by its proximity to counting the number of (not necessarily perfect) matchings in a planar graph, which is \#P-complete \cite{?}. Still, the quantity \(- (w_1 + \ldots + w_k)\) seems a little artificial. If one were to be able to replace \(- (w_1 + \ldots + w_k)\) by 1, then one would be able to count all (not necessarily perfect) matchings in such planar bipartite graphs. Later in this Chapter, we will develop the theory of matchgates to address the problem of when a desired set of signatures can be realized by a holographic algorithm with matchgates. This theory will show that one cannot replace \(- (w_1 + \ldots + w_k)\) by 1.

Now we consider a curious counting Satisfiability problem.

**\#_7PL-RTW-MON-3CNF**

**INPUT:** A planar 3CNF Boolean formula where each variable appears positively and in exactly two clauses (Planar, Read-Twice, Monotone, 3CNF.)

**OUTPUT:** Count the number of satisfying assignments modulo 7.

Let us first consider simply the counting problem for this restricted class of Boolean formulae, which is denoted as \#PL-RTW-MON-3CNF. This problem is known to be \#P-complete. Its graphic representation is simply a 2-3 regular graph, i.e., a bipartite graph where every vertex on the left has degree 2 representing a Boolean variable, and every vertex on the right has degree 3 representing an OR clause. For a given planar, read-twice, monotone 3CNF formula \( \Phi \), we design a signature grid by assigning the binary \textsc{Equality} signature \([1, 0, 1]\) as a generator for each variable, and the Boolean \textsc{Or} signature on three variables \([0, 1, 1, 1]\) as a recognizer on each clause. We will consider these signatures over the finite field \( \mathbb{Z}_7 \). Let \( M = \begin{bmatrix} 4 & 1 \\ 5 & 1 \end{bmatrix} \), with inverse \( M^{-1} = \begin{bmatrix} -1 & 1 \\ 5 & -4 \end{bmatrix} \). We verify that, over \( \mathbb{Z}_7 \),

\[
[0, 1, 1, 1] (M^{-1})^3 = [(1, 1)^3 - (1, 0)^3] (M^{-1})^3 = (4, -3)^3 - (-1, 1)^3 = [2, 0, 2, 0].
\]

Meanwhile \( M^2 [1, 0, 1] \) has its matrix form

\[
MI_2 M^T = \begin{bmatrix} 4 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix} = [3, 0, 5].
\]

Both \([2, 0, 2, 0]\) and \([3, 0, 5]\) can be realized as matchgate signatures, by Theorem 4.11.

We remark that \#_2PL-RTW-MON-3CNF, the parity version of the counting problem \#PL-RTW-MON-3CNF, is \( \oplus \)P-complete \cite{?}, thus NP-hard under randomized reductions \cite{?}. The polynomial time solvability for mod 7 is a surprise, especially in view of its hardness mod 2. In this Chapter we will develop the theory to give an explanation of this mystery.

Finally we consider a geometric problem.

**2-COLOR-COUNTING**

**INPUT:** A set \( S \) of \( n \) points on a plane is given, where no three points are colinear. Also given is a set of line segments between some pairs of points in \( S \). We assume no 3 line segments intersect at a point not in \( S \), and every point of \( S \) is incident to either 2 or 3 line segments.
The number of 2-colorings for the line segments which satisfy the following conditions: (1) for every point in $S$, the incident line segments are not monochromatic; (2) when two line segments cross each other, they have different colors.

![Figure 4.3: A matchgate with signature $\Gamma$ indexed by $\{0,1\}^4$. Edges without explicitly labeled weight all have weight 1. Listed lexicographically $\Gamma = (1,0,0,-1,0,1,-1,0,0,-1,1,0,-1,0,0,1)^T$. Also see Fig 4.5 in Section 4.3.2. Under the holographic transformation $H$ this matchgate has signature $G = (H^{-1})^{\otimes 4}\Gamma$, up to a nonzero factor, $G^{0101} = G^{1010} = 1$ and $G^\alpha = 0$ for other $\alpha \in \{0,1\}^4$.](image)

We use the NAE generator signatures $[0,1,0]$ or $[0,1,1,0]$ for each point with either two or three incident line segments respectively. For each intersection of two line segments we use the generator signature $G = (H^{-1})^{\otimes 4}\Gamma$, where $\Gamma$ is the signature of the matchgate given in Figure 4.3, and $H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. $G$ has exactly two non-zero values $G^{0101} = G^{1010} = 1$. This ensures that the coloring on the line segment is properly propagated past an intersection, and two intersecting line segments are colored differently. For each line segment between two consecutive points which are from $S$ or are intersection points, we add a binary EQUALITY recognizer signature $(=)$. Under a holographic transformation by $H$ (which is symmetric and, up to a scalar, orthogonal and thus its own inverse), $[0,1,0]$, $[0,1,1,0]$ and $(=) = [1,0,1]$ are all realizable. It follows that the problem 2-COLOR-COUNTING can be solved in polynomial time.

The fact that the matchgate signature $\Gamma$ is as stated can be directly verified. However after Subsection 4.3.2 it will become clear why this construction works.

**Exercise:** Show that we may compute the signature $G = (H^{-1})^{\otimes 4}\Gamma$ in problem 2-COLOR-COUNTING as follows: the 4 by 4 matrix form of $G$ is the product $(H^{-1})^{\otimes 2}M(\Gamma)(H^{-1})^{\otimes 2} \Gamma^T$, where $M(\Gamma)$ is the 4 by 4 matrix form of $\Gamma$ which has rows and columns indexed by $\{0,1\}^2$, in lexicographic order, and has entry $\Gamma^{ijkl}$ in entry at row $(ij)$ and column $(kl)$. Conclude that the 4 by 4 matrix form of $G$ is, up to a nonzero scalar, $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. The two nonzero entries are indexed by 0101 and 1010 respectively. Written with the column listed in reverse
order it is, up to a nonzero scalar,

\[
\begin{bmatrix}
G^{0000} & G^{0010} & G^{0001} & G^{0011} \\
G^{0100} & G^{0110} & G^{0101} & G^{0111} \\
G^{1000} & G^{1010} & G^{1001} & G^{1011} \\
G^{1100} & G^{1110} & G^{1101} & G^{1111}
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]

For most holographic algorithms that have been studied, the signatures needed are symmetric signatures. For this problem we used an asymmetric signature \( G \). Holographic algorithms that employ asymmetric signatures have not been extensively studied, and will be an interesting topic for the future. Another notable feature about 2-COLOR-COUNTING is that the problem is not \textit{a priori} stated for a planar graph. It is in the process of forming the signature grid that we obtain a planar graph.

### 4.3 The Theory of Matchgates

In this Section we develop the theory of matchgates. The main result is that there is a set of identities, called Matchgate Identities, which characterize matchgate signatures.

#### 4.3.1 Matchgate Identities

Without loss of generality we may assume the underlying graph of any matchgate is connected. If it is not connected and there are several connected components \( G_i \). Consider a clockwise traversal of all the external nodes. We may consider the planar embedding is on the sphere with one fixed point in the outer face designated as \( \infty \). We temporarily connect each external node to \( \infty \) by non-intersecting paths. As we clockwise-traverse from one external node \( u \) to the next \( v \), if they belong to different components \( G_i \) and \( G_j \), we can connect \( u \) to \( v \) by a path of length 2: \( u, e = \{u, w\}, w, e' = \{w, v\}, v \), adding a new node \( w \), and one extra node \( w' \) with an edge \( \{w, w'\} \). This gadget can be made disjoint from all the temporary paths to \( \infty \), and also disjoint from each other. All new edges (there are three edges on each such gadget) have weight 1. In any perfect matching, \( w \) is matched to \( w' \) and therefore this gadget has no effect on the signature. Then we remove the temporary paths to \( \infty \). Now the matchgate graph is 1) connected, and 2) its outer face uniquely well-defined for the given planar embedding, with a connected boundary.

By default we assume the edge weights are complex numbers in Definition 4.8 and 4.9. Thus \( \Gamma_G \in \mathbb{C}^2 \), or equivalently we may consider it as a tensor in \( (\mathbb{C}^2)^{\otimes k} \). But the theory can be developed in any commutative ring, in particular in a finite field. We will make occasional use of that.
Matchgate signatures satisfy a remarkable set of identities, called the Matchgate Identities. We will show that these identities are a necessary and sufficient condition for a vector of values to be the signature of a matchgate. We state the Matchgate Identities, or MGI.

**Theorem 4.12.** Let $\Gamma$ be the signature of a matchgate with $k$ external nodes. For any length-$k$ bitstrings $\alpha, \beta \in \{0, 1\}^k$, let $P = \{p_1, \ldots, p_l\}$, where $p_1 < \ldots < p_l$, be the subset of $[k]$ such that $p_i$ is the $i$-th bit where $\alpha$ and $\beta$ differ. Then, the signature $\Gamma$ satisfies:

$$\sum_{i=1}^{l} (-1)^i \Gamma^{\alpha \oplus e_{p_1}} \Gamma^{\beta \oplus e_{p_i}} = 0,$$

where $e_j$ denotes a length-$k$ bitstring with a 1 in the $j$-th index, and 0 elsewhere.

We will identify $P$ with its characteristic sequence, which is $\{0, 1\}^k$, the bitwise XOR of $\alpha$ and $\beta$.

**Exercise:** Show that, for matchgates of arity at most 3, the Parity Condition is sufficient. For arity 4, there is a single MGI that is sufficient together with the Parity Condition.

A perfect matching has an even number of vertices. It follows that $\text{PerfMatch}(G^\alpha) = 0$, whenever $G^\alpha$ has an odd number of vertices. Thus, a matchgate signature satisfies the following:

**Parity Condition:** For any matchgate signature $\Gamma$, either for all $\alpha$ of odd Hamming weight, or for all $\alpha$ of even Hamming weight, $\Gamma^\alpha = 0$.

The Parity Condition is a consequence of MGI.

**Theorem 4.13.** If a vector $\Gamma$ obeys the MGI, then it also obeys the Parity Condition.

**Proof.** For a contradiction assume $\Gamma^\alpha \neq 0$ and $\Gamma^\beta \neq 0$, for some $\alpha$ and $\beta$ of even and odd Hamming weight respectively. We define $\tilde{\Gamma}$ by $\tilde{\Gamma}^\gamma = \Gamma^{\gamma \oplus \alpha}$. Since $\gamma \oplus \gamma' = (\gamma \oplus \alpha) \oplus (\gamma' \oplus \alpha)$, if the vector $\Gamma$ obeys the MGI then the vector $\tilde{\Gamma}$ also obeys the MGI. Also $\tilde{\Gamma}^{00\ldots0} = \Gamma^\alpha \neq 0$ and $\tilde{\Gamma}^{\alpha \oplus \beta} = \Gamma^\beta \neq 0$. Note that $\alpha \oplus \beta$ has an odd Hamming weight.

Let $\beta' = \{p_1, \ldots, p_l\}$ be of minimum odd Hamming weight such that $\tilde{\Gamma}^{\beta'} \neq 0$. Now invoke the MGI on the bitstrings $00\ldots0 \oplus e_{p_1}$ and $\beta' \oplus e_{p_1}$. That gives

$$0 = -\tilde{\Gamma}^{00\ldots0} \tilde{\Gamma}^{\beta'} + \sum_{i=2}^{l} (-1)^i \tilde{\Gamma}^{00\ldots0 \oplus e_{p_1} \oplus e_{p_i}} \tilde{\Gamma}^{\beta' \oplus e_{p_1} \oplus e_{p_i}}.$$

We have $\tilde{\Gamma}^{00\ldots0} \tilde{\Gamma}^{\beta'} \neq 0$. If $l = 1$ then the sum $\sum_{i=2}^{l}$ is vacuous, and we have a contradiction. So $l \geq 2$ and we consider each term in the sum $\sum_{i=2}^{l}$. Observe that for every $2 \leq i \leq l$, $\beta' \oplus e_{p_1} \oplus e_{p_i}$ has an odd Hamming weight less than that of $\beta'$, hence $\tilde{\Gamma}^{\beta' \oplus e_{p_1} \oplus e_{p_i}} = 0$. Thus the sum $\sum_{i=2}^{l}$ is zero but $\tilde{\Gamma}^{00\ldots0} \tilde{\Gamma}^{\beta'} \neq 0$, a contradiction. \qed
Nonetheless, because the Parity Condition is a set of linear equations, and the MGI are quadratic, in practice the Parity Condition is a good criterion to apply first.

**Exercise:** Show that, for matchgates of arity at most 3, the Parity Condition also implies MGI. For signatures of arity 4 satisfying the Parity Condition, there is a single MGI that is sufficient to imply all MGI.

MGI were first introduced by Valiant in [?] in the context of proving certain 2-input 2-output quantum gate cannot be realized by a matchgate. It was shown that 2-input 2-output matchgates must satisfy certain identities which are named the Matchgate Identities. These identities are actually concerned with *characters* of matchgates. These so-called characters are defined directly in terms of Pfaffians, and their underlying matchgates need not be planar by definition. In the case of 2-input 2-output matchgates, these character values constitute a 4 by 4 matrix, called a character matrix. Subsequently [?, ?, ?] this theory is generalized to matchgates of an arbitrary number of external nodes. The ultimate result is that there is an equivalence of matchgate characters (of not necessarily planar matchgates) and matchgate signatures (of planar matchgates). Furthermore Matchgate Identities characterize matchgate signatures. Theorem 4.12 states that MGI is a necessary condition. Theorem 4.18 states that MGI is also sufficient and will be proved in Subsection 4.3.2.

The Matchgate Identities are inspired by a set of identities, called Pfaffian Signature Identities. These identities have the same form as MGI but apply to Pfaffian Signatures. The original discovery and proof of MGI are based on the Pfaffian Signature Identities, but we will present with permission a short and elegant bijective proof of MGI due to Jerrum [?].

If we fix any Pfaffian orientation for $G$, it defines a directed graph $\vec{G}$. Then $\vec{G}^\alpha$, which is obtained from $\vec{G}$ by removing some vertices and their incident edges according to $\alpha$, is also Pfaffian-oriented. This is because we only remove zero or more vertices on the outer face, and the removal of these vertices and their incident edges does not create any non-outer face. Thus a single fixed Pfaffian orientation for $G$ induces a set of Pfaffian orientations, one for each $G^\alpha$. We assume a Pfaffian orientation for $G$ is fixed, and each $G^\alpha$ inherits the induced Pfaffian orientation.

As the orientation in $\vec{G}$ induces a Pfaffian orientation for all $G^\alpha$, we can naturally refer to the oriented graph $\vec{G}^\alpha = \vec{G}^\alpha$. The skew-symmetric matrix $B(\vec{G}^\alpha)$ is obtained from $B(\vec{G})$ (see (4.4)) by removing the appropriate rows and columns indicated by the 1’s in $\alpha$. We abbreviate $\text{Pf}(B(\vec{G}^\alpha))$ as $\text{Pf}^\alpha_{\vec{G}}$. Where $\vec{G}$ is clear, we just write $\text{Pf}^\alpha$. With a given Pfaffian orientation on the plane graph $G$, and a given labeling of its $k$ external nodes in clockwise order, we define the Pfaffian Signature of $\vec{G}$ as follows.

**Definition 4.14.** The Pfaffian Signature of $\vec{G}$ is the vector $(\text{Pf}^\alpha)$ indexed by $\alpha \in \{0, 1\}^k$.

Each Pfaffian Signature entry $\text{Pf}^\alpha$ is a sum of Pfaffian terms (see (4.5) and (4.6)) for $\vec{G}^\alpha$. 

108
Each non-zero Pfaffian term \( \text{Pf}_{\vec{G}}(M) \) corresponds to a perfect matching \( M \in \mathcal{M}(G^\alpha) \).

By the property of a Pfaffian orientation, \( \text{Pf}(B(\vec{G})) = \pm \text{PerfMatch}(G) \) is a term by term equation. Thus, for every \( \alpha \in \{0,1\}^k \), there exists \( \epsilon(\alpha) \in \{-1,1\} \), such that for all \( M \in \mathcal{M}(G^\alpha) \),

\[
\text{Pf}_{\vec{G}}(M) = \epsilon(\alpha) \Gamma_{G^\alpha}(M). \tag{4.12}
\]

The sign \( \epsilon(\alpha) \) is the same for every \( M \in \mathcal{M}(G^\alpha) \).

**Theorem 4.15.** Let \( \vec{G} \) be a plane graph with a Pfaffian orientation and \( k \) external nodes. For any length-\( k \) bitstrings \( \alpha, \beta \in \{0,1\}^k \), let \( P = \{p_1, \ldots, p_l\} \), where \( p_1 < \ldots < p_l \), be the subset of \( [k] \) whose characteristic sequence is \( \alpha \oplus \beta \). Then,

\[
\sum_{i=1}^l (-1)^i \text{Pf}^{\alpha \oplus e_{p_i}} \text{Pf}^{\beta \oplus e_{p_i}} = 0. \tag{4.13}
\]

Theorem 4.15 will follow from the Grassmann-Plücker Identities over Pfaffian minors of a matrix. We state the following definition of the Grassmann-Plücker Identities for a skew-symmetric matrix \( A \). In writing \( \text{Pf}(i_1, i_2, \ldots, i_K) \) we mean the Pfaffian of the \( K \times K \) matrix whose rows and columns are the \( i_1, i_2, \ldots, i_K \)-th rows and columns of \( A \), in that order. The order matters: \( \text{Pf}(i_1, i_2, \ldots) = -\text{Pf}(i_2, i_1, \ldots) \), for instance. In particular, if there are two identical rows and columns, the Pfaffian is 0. When we write \( \text{Pf}(i_1, i_2, \ldots, \hat{i}_k, \ldots, i_K) \), the \( \hat{i}_k \) means that \( i_k \) is explicitly excluded from that list.

**Theorem 4.16 (The Grassmann-Plücker Identities).** Let \( I = \{i_1, i_2, \ldots, i_K\} \), \( J = \{j_1, j_2, \ldots, j_L\} \) be subsets of indices of a skew-symmetric \( A \), where \( i_1 < i_2 < \ldots < i_K \) and \( j_1 < j_2 < \ldots < j_L \). Then

\[
\sum_{\ell=1}^L (-1)^{\ell-1} \text{Pf}(j_\ell, i_1, \ldots, i_K) \text{Pf}(j_1, \ldots, \hat{j}_\ell, \ldots, j_L) + \sum_{k=1}^K (-1)^{k-1} \text{Pf}(i_1, \ldots, \hat{i}_k, \ldots, i_K) \text{Pf}(i_k, j_1, \ldots, j_L) = 0 \tag{4.14}
\]

Theorem 4.16 has the following short proof \([?, ?]\) originally from \([?]\).

**Proof of Theorem 4.16.** From the definition of Pfaffian:

\[
\text{Pf}(j_\ell, i_1, \ldots, i_K) = \sum_{k=1}^K (-1)^{k-1} \text{Pf}(j_\ell, i_k) \text{Pf}(i_1, \ldots, \hat{i}_k, \ldots, i_K)
\]

\[
\text{Pf}(i_k, j_1, \ldots, j_L) = \sum_{\ell=1}^L (-1)^{\ell-1} \text{Pf}(i_k, j_\ell) \text{Pf}(j_1, \ldots, \hat{j}_\ell, \ldots, j_L)
\]

and also \( \text{Pf}(j_\ell, i_k) + \text{Pf}(i_k, j_\ell) = 0 \). The proof is completed by substituting these into (4.14). \( \square \)
There is another form of these identities which is more closely related to the Pfaffian Signature Identities (4.13).

**Theorem 4.17.** Let $A, I, J$ be as in Theorem 4.16. For a subset $S$ of indices of $A$, we write $\text{Pf}(S)$ when $S$ is listed in increasing order. Let $I \Delta J = \{k_1, \ldots, k_m\}$ (listed in increasing order) be the symmetric difference of $I$ and $J$. Then

$$\sum_{s=1}^{m} (-1)^{s-1} \text{Pf}(I \Delta \{k_s\}) \text{Pf}(J \Delta \{k_s\}) = 0 \quad (4.15)$$

**Proof of Theorem 4.17.** We prove Theorem 4.17 by Theorem 4.16.

Consider a term in (4.14), and let $x$ be the element that is moved from the index set of one Pfaffian to another. If $x \in I \cap J$, clearly the term is 0. It follows that there is a one-to-one correspondence between the remaining terms in (4.14) and the terms in (4.15). All we need to show is that each such term in (4.14) has the same sign as its counterpart in (4.15).

Suppose $x \in J - I$. In that case, the term in (4.14) is

$$(-1)^z \text{Pf}(x, i_1, \ldots, i_K) \text{Pf}(j_1, \ldots, \hat{x}, \ldots, j_L)$$

where $z$ is the number of elements in $J$ less than $x$. We write $z = a + b$, where

$$a = |\{y \mid y \in J - I, y < x\}|$$

$$b = |\{y \mid y \in J \cap I, y < x\}|$$

and we also define

$$c = |\{y \mid y \in I - J, y < x\}|.$$

We can put the indices in $\text{Pf}(x, i_1, \ldots, i_K)$ in increasing order by moving $x$ along until it is in its proper place. To do this we move $x$ exactly $b + c$ entries. Thus

$$\text{Pf}(x, i_1, \ldots, i_K) = (-1)^{b+c} \text{Pf}(I \cup \{x\})$$

and so it follows that

$$(-1)^z \text{Pf}(x, i_1, \ldots, i_K) = (-1)^{a+c} \text{Pf}(I \cup \{x\}).$$

It is clear that $a + c$ is precisely the number of those in $I \Delta J$ less than $x$, exactly the sign in front of the corresponding term in (4.15).

The argument for the case $x \in I - J$ is symmetric. \qed

Now we are ready to prove Theorem 4.15.
Proof of Theorem 4.15. We prove Theorem 4.15 by Theorem 4.17. For a matchgate $G$, let $\alpha$ and $\beta$ be two bitstrings of length $k$, where $k$ is the number of external nodes in $G$. The $i$-th bit of $\alpha$, denoted as $\alpha_i$, corresponds to the $i$-th external node in $G$ in clockwise order.

Let $U$ be the set of all internal (that is, not external) nodes in $G$. We define $I = \{v_i \mid \alpha_i = 0\} \cup U$, where $v_i$ is the label of the $i$-th external node in $G$. Similarly let $J = \{v_i \mid \beta_i = 0\} \cup U$. Observe that $I \Delta J = \{v_i \mid \alpha_i \neq \beta_i\}$. It follows that there is a term-for-term correspondence between (4.15) of Theorem 4.17 and (4.13) of Theorem 4.15. \hfill \Box

Now we present the bijective proof of MGI, Theorem 4.12, by Jerrum.

We first note that MGI trivially hold if $l$ is odd. In this case, $\alpha$ and $\beta$ have opposite parity, and so does $\alpha \oplus e_{p_i}$ and $\beta \oplus e_{p_i}$ for every $p_i$, and thus one of $G^{\alpha \oplus e_{p_i}}$ and $G^{\beta \oplus e_{p_i}}$ has an odd number of nodes, and hence MGI trivially holds by the Parity Condition. So we may assume $l$ is even.

As before we denote by $\mathcal{M}(G)$ the set of all perfect matchings of the graph $G$. We will establish a weight preserving bijection

$$\bigcup_{i \text{ odd}} \left[ \mathcal{M}(G^{\alpha \oplus e_{p_i}}) \times \mathcal{M}(G^{\beta \oplus e_{p_i}}) \right] \longleftrightarrow \bigcup_{j \text{ even}} \left[ \mathcal{M}(G^{\alpha \oplus e_{p_j}}) \times \mathcal{M}(G^{\beta \oplus e_{p_j}}) \right]$$

More precisely, note that the MGI can be stated as

$$\sum_{1 \leq i \leq l, \ i \text{ odd}} \Gamma^{\alpha \oplus e_{p_i}} \Gamma^{\beta \oplus e_{p_i}} = \sum_{1 \leq j \leq l, \ j \text{ even}} \Gamma^{\alpha \oplus e_{p_j}} \Gamma^{\beta \oplus e_{p_j}}. \tag{4.16}$$

The product $\Gamma^{\alpha \oplus e_{p_i}} \Gamma^{\beta \oplus e_{p_i}}$ is a sum over all pairs $(M, M') \in \mathcal{M}(G^{\alpha \oplus e_{p_i}}) \times \mathcal{M}(G^{\beta \oplus e_{p_i}})$ of the product of their weight $w(M)w(M')$.

Take any $(M, M') \in \mathcal{M}(G^{\alpha \oplus e_{p_i}}) \times \mathcal{M}(G^{\beta \oplus e_{p_i}})$, consider their XOR $M \oplus M'$. Since $M$ and $M'$ are matchings in $G$, $M \oplus M'$ consists of a set of alternating paths and alternating cycles of even length. Since $\alpha_{p_i} \neq \beta_{p_i}$, there is an alternating path starting at $p_i$ in $M \oplus M'$. This alternating path must end in some external node $p_j$. By planarity of $G$, there must be an even number of external nodes between $i$ and $j$ in cyclic order, among $\{p_1, p_2, \ldots, p_l\} \setminus \{p_i, p_j\}$. This is because each external node in this list is the initial node of an alternating path in $M \oplus M'$, and must end at a node on this list. Alternating paths in $M \oplus M'$ do not cross paths in a plane graph $G$. An alternating path starting at any $p_s$ where $s$ is between $i$ and $j$ in cyclic order, either $i < s < j$ or $j < s < i$, must end in a node $p_t$ with $i < t < j$ or $j < t < i$ respectively.

If we flip the edges of $M$ and $M'$ along the alternating path from $p_i$ to $p_j$ we obtain another pair $(\tilde{M}, \tilde{M}') \in \mathcal{M}(G^{\alpha \oplus e_{p_j}}) \times \mathcal{M}(G^{\beta \oplus e_{p_j}})$. It is clearly an involution, i.e., this mapping is its own inverse. This sets up a bijective map between $\bigcup_{i \text{ odd}} \left[ \mathcal{M}(G^{\alpha \oplus e_{p_i}}) \times \mathcal{M}(G^{\beta \oplus e_{p_i}}) \right]$ and $\bigcup_{j \text{ even}} \left[ \mathcal{M}(G^{\alpha \oplus e_{p_j}}) \times \mathcal{M}(G^{\beta \oplus e_{p_j}}) \right]$.

It is clear that the product of weight $w(M)w(M')$, which is just a product of all edge weights of the alternating paths and cycles of $M \oplus M'$ (together with $w(e)^2$ from length 2
cycles" for every edge $e \in M \cap M'$, is identical to $w(M)w(M')$. The bijective map preserves the weight in the two sides of (4.16). This proves Theorem 4.12.

4.3.2 MGI Imply Matchgate-Realizable

Theorem 4.12 showed that the signature of any matchgate must satisfy the Matchgate Identities. In this subsection, we show that any $2^{(C \otimes k_2)} = C^{2k}$ satisfying the Matchgate Identities can be realized as the signature of a matchgate with $k$ external nodes. Thus MGI are not only necessary but also sufficient conditions for matchgate signatures.

Consider a length $2^k$ vector $\Gamma$ indexed by $\{0, 1\}^k$ satisfying MGI. If it is the all-zeros vector then it is trivially realizable. So assume there is at least one non-zero value.

Preprocessing  Assume $\Gamma_\beta \neq 0$, for some $\beta \in \{0, 1\}^k$. Define $\Gamma_\alpha = \Gamma_\alpha \oplus \overline{\beta} / \Gamma_\beta$, where $\overline{\beta} = \beta \oplus 11 \ldots 1$. Thus, $\Gamma_1^{11\ldots1} = 1$, and $\Gamma'$ also satisfies MGI. In this section we will create a matchgate $G'$ with signature $\Gamma'$. Given such a $G'$, we can create a matchgate $G$ with signature $\Gamma$ as follows: First we add two new non-external nodes $u, v$ to $G'$ and an edge $\{u, v\}$ of weight $\Gamma_\beta$. Those two nodes are not connected to any other nodes—in effect they contribute exactly a factor $\Gamma_\beta$ to each perfect matching term. Then, if the $i$-th bit of $\beta$ is 0, we add a new edge $\{v_i, v'_i\}$ of weight one to the $i$-th external node $v_i$, and making $v'_i$ the new $i$-th external node. It follows that the signature of $G$ is exactly $\Gamma$.

Construction  We now show that we can realize $\Gamma$ satisfying MGI and $\Gamma_1^{11\ldots1} = 1$. Let $K_k$ denote the complete graph on $k$ vertices. The labels of $K_k$ are ordered $1 < 2 < \ldots < k$, and correspond to the bit positions in the index for $\Gamma$. We place the nodes of $K_k$ on a convex curve, as illustrated in Fig. 4.4. The nodes are arranged in clockwise order by their index, and two edges cross each other geometrically in the drawing of the graph if their labels form an overlapping pair as defined before algebraically. (We assume the $k$ nodes are placed in general position, so that any pair of crossing edges intersect at a unique point. There are exactly $\binom{k}{2}$ such intersection points.) For each $\alpha$ of Hamming weight $k - 2$, note that $K_k^{\alpha}$ has exactly one edge left. For each such $\alpha$, set the weight of the unique edge in $K_k^{\alpha}$ to be $\Gamma_\alpha$. This defines a weight for every edge of $K_k$.

Equality with Pfaffian  Clearly our embedding of $K_k$ is not planar for a general $k \geq 4$. We first prove the following equality: Let $Pf(K_k^{\alpha})$ be the Pfaffian value of the skew-symmetric matrix representing $K_k^{\alpha}$ where the nodes of $K_k^{\alpha}$ have the induced order from $1 < 2 < \ldots < k$. Then for all $\alpha \in \{0, 1\}^k$:

$$Pf(K_k^{\alpha}) = \Gamma^{\alpha}. \quad (4.17)$$

It follows that the $\binom{k}{2}$ edge weights of $K_k$ determine the $2^k$ values of any $\Gamma$ satisfying MGI.
Equation (4.17) holds trivially for any $\alpha$ of Hamming weight $k - l$, for any odd $l$. The left-hand side is a Pfaffian of a matrix with an odd number of rows and columns, hence 0, and the right-hand side is 0 by the Parity Condition and Theorem 4.13. Now we consider the case when $l$ is even.

Clearly (4.17) holds for any $\alpha$ of Hamming weight greater than or equal to $k - 2$. By assumption $\Gamma$ satisfies the Matchgate Identities (4.10). Inductively, consider $\alpha$ with Hamming weight $k - l$ for some even $l > 2$. Let $\{p_1, \ldots, p_l\}$ be the set of indices listed in increasing order $p_1 < \ldots < p_l$, where $\alpha$ has the bit 0. These are the bit positions where $\alpha$ differs from $1^k$. Consider the MGI on $\alpha \oplus e_{p_1}$ and $1^k \oplus e_{p_1}$:

$$
\Gamma^{\alpha \Gamma^{11\ldots l}} = \sum_{i=2}^{l} (-1)^i \Gamma^{\alpha \oplus e_{p_1} \oplus e_{p_i}} \Gamma^{1^k \oplus e_{p_1} \oplus e_{p_i}}
$$

As $\Gamma^{11\ldots l} = 1$, we see that $\Gamma^{\alpha}$ is defined by higher Hamming weight terms.

Thus all lower Hamming weight terms of $\Gamma$ are determined by those of weight $k - 2$, or equivalently the \( \binom{k}{2} \) edge weights of $K_k$. However, by Theorem 4.17, the Pfaffian values also satisfy exactly the same identities as MGI. By induction, it follows that $\text{Pf}(K_k^\alpha) = \Gamma^{\alpha}$ for all $\alpha$. We have proved (4.18).

**Planarizing $K_k$** We want to show next that there exists a planar matchgate $G$ with signature $\Gamma_G = \Gamma$. We construct such a $G$ from $K_k$. Consider the convex embedding of $K_k$. For $k \geq 4$ it has some edge crossings, as shown in Fig. 4.4. The planar graph $G$ is created by replacing each crossing of a pair of edges in the embedded $K_k$ by a crossover gadget from Fig. 4.5. The crossover gadget is itself a matchgate $X$ with the following signature:

$$
X^{0000} = 1, \quad X^{0101} = 1, \quad X^{1010} = 1, \quad X^{1111} = -1
$$

and for all other $\beta \in \{0, 1\}^4$, $X^\beta = 0$. We note that even though geometrically this gadget is only symmetric under a rotation of $\pi$ (but not $\pi/2$), its signature is invariant under a cyclic permutation, and thus functionally it is symmetric under a rotation of $\pi/2$. Now we replace every crossing of a pair of edges in the embedded $K_k$ by a copy of $X$. For example,
Figure 4.5: The crossover gadget. The external nodes are those labeled, and all edge weights are 1, except the edge labeled $-1$.

Figure 4.6: The graph from Fig. 4.4 with the crossovers replaced by crossover gadgets from Fig. 4.5.

this replacement by the crossover gadget changes Fig. 4.4 to Fig. 4.6. If an edge $\{i, j\}$ in $K_k$ crosses some other edges (this happens for every non-adjacent $i$ and $j$ in the cyclic sense), then this replacement breaks the edge $\{i, j\}$ into several parts. If $\{i, j\}$ crosses $t \geq 0$ other edges, then it is replaced by $t$ copies of the crossover gadget, connected by $t+1$ edges outside of crossover gadgets. Of course one copy of the crossover gadget is used for both edges of a pair of crossing edges in this replacement (see Fig. 4.7). Define $I$ to be the set of all edges in $G$ that are not part of a crossover gadget. Then each edge $\{i, j\}$ in $K_k$ defines a unique subset of edges in $I$, which we call the $i$-$j$-passage. It is clear that $I$ is a disjoint union of these $i$-$j$-passages, over all $\binom{k}{2}$ pairs $1 \leq i < j \leq k$. Finally we choose one edge in each $i$-$j$-passage to have the weight $\Gamma_{[k]-\{i,j\}}$, namely the edge weight of $\{i, j\}$ in $K_k$. To be specific, we will choose this edge to be the one adjacent to $i$, the lower indexed external node of $\{i, j\}$. All other edges of $I$ are assigned weight one (see Fig. 4.7). This defines our planar matchgate $G$ with external nodes $1 < 2 < \ldots < k$.

We claim that $\Gamma_G = \Gamma$.

114
Figure 4.7: The “planarized” $K_5$ with edge weights. The unlabeled edges have weight 1. For notational simplicity, in the figure we use the notation $w(i, j)$ for $w(\{i, j\})$.

Fix any $\alpha \in \{0, 1\}^k$. For any $S \subseteq I$, define $\mathcal{M}_S(G^\alpha)$ to be the subset of all perfect matchings $M' \in \mathcal{M}(G^\alpha)$ such that $M' \cap I = S$. Every perfect matching $M \in \mathcal{M}(K^\alpha_k)$ defines a collection of $i$-$j$-passages, one for each $\{i, j\} \in M$. Let $S(M)$ be the union of these $i$-$j$-passages. Clearly the perfect matching $M \in \mathcal{M}(K^\alpha_k)$ can be recovered from $S(M)$, and is unique for the given $S(M)$. There is a 1-1 correspondence between $M$ and $S(M)$. As an example, we consider $M = \{(1, 3), (2, 5)\} \in \mathcal{M}(K^\alpha_8^{00010})$. The set $S(M)$ for $G^{00010}$ are the thick edges in Fig. 4.8.

We will show that, for the purpose of computing the signature entry $\Gamma^\alpha_G$, we only need to consider those perfect matchings $M' \in \mathcal{M}(G^\alpha)$ that satisfy the following property:

**Property:** There exists an $M \in \mathcal{M}(K^\alpha_k)$, such that

$$M' \cap I = S(M).$$

(4.19)

This is a consequence of the properties of the crossover gadget. If $i$ is an external node in $G^\alpha$, then any $M' \in \mathcal{M}(G^\alpha)$ must contain a unique edge $e'$ incident to $i$. There is a unique $j$, which is another external node in $G$, such that $e'$ belongs to the $i$-$j$-passage. Then by the properties of the crossover gadgets along this $i$-$j$-passage, we may assume $M'$ contains all edges of this $i$-$j$-passage, saturating $j$. In particular $j$ belongs to $G^\alpha$. All other $M'$ collectively contribute 0, since the evaluation of the crossover gadget $X$ will be 0. More formally, in the computation of $\Gamma^\alpha_G = \sum_{M' \in \mathcal{M}(G^\alpha)} \prod_{e' \in M'} w(e')$, we classify all $M' \in \mathcal{M}(G^\alpha)$ according to $M' \cap I$. If $S \neq S(M)$ for any $M \in \mathcal{M}(K^\alpha_k)$, then

$$\sum_{M' \in \mathcal{M}(G^\alpha): M' \cap I = S} \prod_{e' \in M'} w(e') = 0.$$ 

(4.20)

115
Figure 4.8: The thick edges comprise $S(M)$ for $G^{00010}$, where $M = \{\{1,3\}, \{2,5\}\}$.

In fact, for any $M' \in \mathcal{M}(G^\alpha)$ such that $M' \cap I = S$ which is not $S(M)$ for any $M \in \mathcal{M}(K_k^\alpha)$, it must be the case that at some crossover gadget $X$, $S$ induces an external removal pattern $eta \notin \{0000, 0101, 1010, 1111\}$. Then $X^\beta = 0$, and (4.20) follows.

Thus we restrict to those perfect matchings $M' \in \mathcal{M}(G^\alpha)$ that satisfy the property (4.19). For any $M \in \mathcal{M}(K_k^\alpha)$, it is clear that

$$\sum_{M' \in \mathcal{M}(S(M))} \prod_{e' \in M'} w(e') = (-1)^{c(M)} \prod_{e \in M} w(e),$$

where $c(M)$ counts the number of copies of $X$ where the external removal pattern is $\beta = 1111$. Thus $c(M)$ is exactly the number of overlapping pairs in $M$. It follows that

$$\Gamma_G^\alpha = \sum_{M' \in \mathcal{M}(G^\alpha)} \prod_{e' \in M'} w(e')$$

$$= \sum_{S \subseteq I} \sum_{M' \in \mathcal{M}(G^\alpha)} \prod_{e' \in M'} w(e')$$

$$= \sum_{M \in \mathcal{M}(K_k^\alpha)} \sum_{M' \in \mathcal{M}(S(M))} \prod_{e' \in M'} w(e')$$

$$= \text{Pf}(K_k^\alpha).$$

The last equality is because each Pfaffian term in $\text{Pf}(K_k^\alpha)$ has exactly the same sign as in (4.21). Hence $\Gamma_G = \Gamma$ follows from this and (4.17).

**Theorem 4.18.** Any $\Gamma \in (\mathbb{C}^2)^{\otimes k}$ satisfying the MGI is the signature of a matchgate with $k$ external nodes. The matchgate has $O(k^4)$ nodes. If $\Gamma^{11...1} = 1$, which is achievable by a
normalization for every nonzero $\Gamma$, there exists a skew-symmetric matrix $M \in \mathbb{C}^{k \times k}$ such that $\Gamma^\alpha = \text{Pf}(M^\alpha)$, where $M^\alpha$ is the matrix obtained from $M$ by deleting all rows and columns belonging to the subset denoted by $\alpha$.

Thus, after a normalization of $\Gamma^{11...1} = 1$, a matchgate signature of arity $k$ has a general parametrized form as the sub-Pfaffians of a $k \times k$ skew-symmetric matrix.

For example, consider the matchgate in Figure 4.3 for the problem 2-COLOR-COUNTING in Section 4.2. Its signature is $\Gamma = (1, 0, 0, -1, 0, 1, -1, 0, 0, -1, 1, 0, -1, 0, 0, 1)^T \in \{0, 1\}^4$. The signature is given by the sub-Pfaffians of

$$
\begin{bmatrix}
0 & -1 & +1 & -1 \\
+1 & 0 & -1 & +1 \\
-1 & +1 & 0 & -1 \\
+1 & -1 & +1 & 0
\end{bmatrix}
$$

After assigning appropriate edge weights to $K_4$ according to this matrix, we obtain the matchgate by replacing the single crossover of edges $\{1, 3\}$ and $\{2, 4\}$ by the crossover gadget in Fig. 4.5. This results in the matchgate depicted in Fig. 4.3.

### 4.3.3 Symmetric Signatures by Matchgates

We say a matchgate signature is **even** if it is the signature of an even matchgate, i.e., a matchgate with an even number of nodes. An even signature has nonzero values only for indices of even Hamming weight. We define an **odd** signature similarly. Recall that a symmetric arity-$k$ signature can be denoted as $[z_0, z_1, \ldots, z_k]$, where $z_i$ is the value of the signature for an index of Hamming weight $i$. The symmetric signatures that obey the MGI have a very concise description, which we prove next.

**Theorem 4.19.** If $[z_0, \ldots, z_k]$ is an even symmetric matchgate signature, then $z_i = 0$ for all odd $i$, and there exist $r_1$ and $r_2$ not both zero such that for all even $i \geq 2$:

$$
r_1z_{i-2} = r_2z_i.
$$

Conversely, every sequence of values satisfying these conditions is an even symmetric matchgate signature. The statement for odd symmetric signatures is analogous.

**Proof.** By the Parity Condition, all odd parity entries of the signature of an even matchgate are zero. Consider any even $i$ and $j$, where $0 \leq i < j \leq k$. We invoke the MGI for $\alpha = 1^i10^{k-i-1}, \beta = 1^01^{j-i-1}0^{k-j}$, so that $\alpha \oplus \beta = 0^i1^{j-i}0^{k-j}$. We use the exponentiation notation here to denote repetition. The bitstring $\alpha$ has an odd Hamming weight $i+1$ and $\beta$ has an odd Hamming weight $j-1$. Note that $i$ and $j$ being both even implies that $j-i-1 \geq 1$. Using the fact that $\Gamma$ is symmetric, the MGI under $\alpha, \beta$ can be simplified:

$$
\sum_{s=1}^{j-i} (-1)^s \Gamma^{\alpha\oplus e_{i+s}} \Gamma^{\beta\oplus e_{i+s}} = -z_i z_j + \sum_{s=2}^{j-i} (-1)^s z_{i+2} z_{j-2} = 0.
$$

(4.22)
There are an odd number of terms in the sum \( \sum_{s=2}^{j-i} (-1)^s z_{i+2} z_{j-2} \) and the terms alternate their signs and begin with a +, so we conclude that

\[
z_i z_j = z_{i+2} z_{j-2}.
\]

In particular, if \( i \) is even and \( 0 \leq i \leq k - 4 \), then

\[
z_i z_{i+4} = z_{i+2}^2.
\]

If \( z_{i+2} \neq 0 \), then both \( z_i \neq 0 \) and \( z_{i+4} \neq 0 \). This means that if any even indexed entry that is not the first or the last even indexed entry (call it a non-extremal entry) is nonzero, then all even indexed entries are nonzero. In this case, the geometric progression is established, with common ratio \( z_{i+2}/z_i = z_{i+4}/z_{i+2} \), for even \( 0 \leq i \leq k - 4 \).

Suppose all non-extremal even indexed entries are zero. If \( k \leq 3 \) then the theorem is self-evident. Suppose \( k \geq 4 \). Let \( k^* \leq k \) be the maximum even index. Then \( k^* \geq 4 \) and we have

\[
z_0 z_{k^*} = z_2 z_{k^*-2}.
\]

Thus \( z_2 = 0 \). It follows that \( z_0 z_{k^*} = 0 \) and therefore at most one extremal even indexed entry can be nonzero. It is also easy to verify that a sequence satisfying the conditions of this theorem satisfies MGI, and hence is a matchgate signature. (We will give a direct construction shortly.) This completes the proof for even signatures. The proof for odd signatures is similar. The theorem follows.

The explicit list of symmetric signatures in Theorem 4.11 follows from Theorem 4.19.

We have already demonstrated how to build a planar matchgate realizing any MGI-satisfying signature, through a planarizing procedure. However it is instructive to give a direct construction for the symmetric case. For the symmetric signature \( \text{EXACT-ONE-1} = [0, 1, 0, \ldots, 0] \) of arity \( n \) that represents a perfect matching, and its reversal \( \text{EXACT-ONE-0} = [0, \ldots, 0, 1, 0] \), we can design a matchgate based on a star graph. For \( \text{EXACT-ONE-0} \) we take a star graph with a single internal vertex and \( n \) dangling edges. For \( \text{EXACT-ONE-1} \) we can flip the 0’s and 1’s by extending each dangling edge to be a path of length 2. For weighted versions \( [0, \lambda, 0, \ldots, 0] \) and \( [0, \ldots, 0, \lambda, 0] \) we can assign the weight \( \lambda \) to all the edges incident to the single internal vertex.

These two cases correspond to some special cases in the list in Theorem 4.11 with \( a = 0 \) or \( b = 0 \) (\( a = 0 \) in item 2, \( b = 0 \) in item 3, and \( a \) or \( b = 0 \) in item 4.) Now consider a symmetric signature in Theorem 4.11 other than the two cases above. For such a signature, it is difficult to imagine a construction that is planar and also geometrically symmetric for all pairs of external nodes \( 1 \leq i < j \leq k \), if \( k \geq 4 \). In the following construction, the matchgates are not geometrically symmetric for all pairs of external nodes, but functionally they are, in terms of the signatures.

We present two closely related matchgate constructions, one for even symmetric signatures, and the other for odd. The constructions for both these cases work regardless if the signature has odd or even arity.
To construct a symmetric even matchgate \( G \) of arity \( k \), we first take \( k \) triangles with vertices \( \{a_i, b_i, c_i\} \) (1 \( \leq i \leq k \)). The edges \( \{a_i, b_i\} \) and \( \{a_i, c_i\} \) have weight \( x \), and \( \{b_i, c_i\} \) has weight \( y \). We link them in a cycle, identifying \( c_i \) with \( b_{i+1} \), where the index is counted modulo \( k \). The matchgate \( G \) has \( k \) external nodes \( \{a_1, \ldots, a_k\} \), and a total of \( 2k \) nodes. In Fig. 4.9 we have an example of a planar matchgate for an even, arity-6 signature. For a symmetric odd signature of arity \( k \), we delete one external node in a matchgate for a symmetric even signature of arity \( k + 1 \). This is illustrated in Fig. 4.10 for arity \( k = 5 \).

For an even matchgate of \( 2k \) nodes, we only need to consider its values at entries of even Hamming weight. Consider any \( \alpha \in \{0, 1\}^k \) of even Hamming weight. \( \alpha_i = 0 \) iff \( a_i \) remains in \( G^\alpha \). If \( \alpha = 1^k \), then \( G^\alpha \) is a cycle of length \( k \). Since \( \alpha \) also has even Hamming weight, \( k \) must be even, and there are exactly two perfect matchings of a cycle of even length, each having weight \( y^{k/2} \).

Now assume \( \alpha \neq 1^k \). Then \( \alpha \) cyclically alternates between consecutive 0’s (called a 0-run) and consecutive 1’s (called a 1-run). Each \( a_i \) that remains in \( G^\alpha \) must be matched to either \( b_i \) (we call it left-match) or \( c_i = b_{i+1} \) (we call it right-match), both with weight \( x \). Consider any 0-run. It is clear that either all \( a_i \) within this 0-run left-match or all right-match. Next consider a 1-run of \( m \) 1’s; it is between two (not necessarily distinct) 0-runs, since \( \alpha \neq 1^k \). If \( m \) is even, then the path of \( m \) edges all with weight \( y \) forces the two neighboring 0-runs to take either both left-match or both right-match. Moreover, both possibilities are realizable, and in each case the 1-run contributes a weight \( y^{m/2} \). If \( m \) is odd, then the path of \( m \) edges forces the two neighboring 0-runs to take opposite types of left-match and right-match. Again both possibilities are realizable; in one case the 1-run contributes a weight \( y^{(m-1)/2} \), and in another case it contributes a weight \( y^{(m+1)/2} \). Furthermore, for two 1-runs \( 1^m \) and \( 1^{m'} \) both of odd length and are consecutive in the sense that the only 1-runs in between are of even length, they contribute a combined weight \( y^{(m+m')/2} \). Since \( \alpha \) has an even Hamming weight \( |\alpha| \), there is an even number of 1-runs of odd length. Hence together the 1-runs contribute a weight \( y^{|\alpha|/2} \). There are exactly two perfect matchings in \( G^\alpha \), each uniquely determined by the left-match or right-match choice of any particular \( a_i \) in \( G^\alpha \). It follows that the signature value is \( \Gamma^\alpha = 2x^{k-|\alpha|}y^{|\alpha|/2} \). Clearly by choosing \( x \) and \( y \) suitably, we can realize an arbitrary even symmetric signature other than weighted Exact-One-1 and Exact-One-0.

The construction for odd symmetric signatures is to remove one external node in the matchgate for an even symmetric signature of arity one higher. If the even symmetric signature of arity \( k+1 \) has the form \([z_0, z_1, \ldots, z_{k+1}]\), then the construction gives the signature \([z_1, \ldots, z_{k+1}]\) of arity \( k \). The proof is complete.

### 4.3.4 Symmetric Signatures Transformable to Matchgates

In Section 4.2 we have seen a number of natural problems that can be solved by holographic reductions to matchgates. For an individual problem, specified by suitable local constraint functions expressed as generator and recognizer signatures, the task is to find suitable matchgates and a holographic transformation such that both the generator and
Figure 4.9: A matchgate for an even, symmetric, arity-6 signature.

Figure 4.10: A matchgate for an odd, symmetric, arity-5 signature.
recognizer signatures can be transformed to matchgate signatures. This is called the Simultaneous Realizability Problem. This task for any individual problem is specified by families of algebraic equations. These families of equations are typically exponential in size. Searching for their solutions is what Valiant called "the enumeration" of "freak objects" in his Accidental Algorithm paper [183]. Dealing with such algebraic equations can be difficult.

While finding exotic solutions such as the signature for \( \#_7\text{PL-RTW-MON-3CNF} \) can be artistically satisfying, the situation with ever more complicated algebraic constraints on such signatures for any given problem can quickly overwhelm such an artistic approach as well as a computer search. Furthermore, failure to find such solutions to a particular algebraic system yields no proof that such solutions do not exist, and it generally does not give us any insight as to why. The theory of matchgates provides a more systematic understanding. This theory is quite complete for symmetric signatures. We will show that the Simultaneous Realizability Problem can be solved in polynomial time. Using this theory we can see why the modulus 7 is the modulus that works for counting PL-RTW-MON-3CNF. Underlying this success is the fact that 7 is \( 2^3 - 1 \), and for any odd prime \( p \), the number 2 is a quadratic residue in \( \mathbb{Z}_q \) for any prime factor \( q \) of the Mersenne number \( 2^p - 1 \). Generalizing this, we will show that \( \#_{2^{k-1}}\text{PL-RTW-MON-}k\text{CNF} \) is in P for all \( k \geq 3 \) (the problem is trivial for \( k \leq 2 \)). Furthermore, no suitable signatures exist for any modulus other than factors of \( 2^k - 1 \) for counting PL-RTW-MON-\( k \text{CNF} \).

We start with an equivalent set of normal forms from Theorem 4.11 that is more suitable for holographic transformations. If \( a \) and \( b \) are both non-zero, then all four forms in Theorem 4.11 can be expressed as a sum of two tensor products of the following form: There exist nonzero \( \alpha \) and \( \beta \in \mathbb{C} \) such that

\[
f = [f_0, f_1, \ldots, f_n] = [\alpha, -\beta]^n \pm [\alpha, \beta]^n. \tag{4.23}
\]

Clearly any signature of the form in (4.23) takes one of the four forms in Theorem 4.11. Conversely, for form 1 (arity \( n = 2k \)) and form 2 (arity \( n = 2k + 1 \)) in Theorem 4.11 we can take + for the ± sign in (4.23) and

\[
\alpha = \frac{a^{k/n}}{2^{1/n}} \quad \text{and} \quad \beta = \alpha \sqrt{\frac{b}{a}}.
\]

We can verify that all entries at odd Hamming weight are zero, and the entry at Hamming weight \( 2i \) is \( 2\alpha^{n-2i}\beta^{2i} = a^{k-i}b^i \), for \( 0 \leq i \leq k \). For form 3 (arity \( n = 2k + 1 \)) and form 4 (arity \( n = 2k + 2 \)) in Theorem 4.11 we can take − for the ± sign and

\[
\alpha = \frac{a^{(k+1)/n}}{2^{1/n}b^{1/(2n)}} \quad \text{and} \quad \beta = \alpha \sqrt{\frac{b}{a}}.
\]

We can verify that all entries at even Hamming weight are zero, and the entry at Hamming weight \( 2i + 1 \) is \( 2\alpha^{n-2i-1}\beta^{2i+1} = a^{k-i}b^i \), for \( 0 \leq i \leq k \).

However, if \( a \) or \( b \) is zero, but not both zero, there are some additional signatures in Theorem 4.11 that cannot be expressed as (4.23). These cases are as follows:

\[
f = [0, \lambda, 0, \ldots, 0] \quad \text{or its reversal} \quad f' = [0, \ldots, 0, \lambda, 0] \quad \tag{4.24}
\]
for some $\lambda \neq 0$. These can be expressed as

$$f = \lambda \sum_{k=1}^{n} \left\{ [1, 0] \otimes (k-1) \otimes [0, 1] \otimes [1, 0] \otimes (n-k) \right\}$$

(4.25)

and

$$f' = \lambda \sum_{k=1}^{n} \left\{ [0, 1] \otimes (k-1) \otimes [1, 0] \otimes [0, 1] \otimes (n-k) \right\}.$$  

(4.26)

Of course the case $a = b = 0$ in Theorem 4.11 is the identically 0 signature and corresponds to taking $\alpha = \beta = 0$ in (4.23) and $\lambda = 0$ in (4.25) and (4.26).

Now we can give a closed form expression for symmetric signatures that are transformable to matchgate signatures. We state this for recognizer signatures.

**Theorem 4.20.** A symmetric signature $f = [f_0, f_1, \ldots, f_n]$ is transformable to a matchgate signature as a recognizer, i.e., $f = \Gamma M^{\otimes n}$ for some $M = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in GL_2(\mathbb{C})$ and a matchgate signature $\Gamma$ iff it takes one of the following forms:

1. $\exists \lambda, \alpha, \beta \in \mathbb{C}$ and $\epsilon = \pm 1$, such that
   $$f = \lambda \left( [\alpha, \beta] M \right)^{\otimes n} + \epsilon \left( [\alpha, -\beta] M \right)^{\otimes n}.$$  

   Thus, for $0 \leq k \leq n$, the entry of $f$ at weight $k$ is
   $$f_k = \lambda \left[ (a_\alpha + b_\beta)^{n-k}(c_\alpha + d_\beta)^k + \epsilon (a_\alpha - b_\beta)^{n-k}(c_\alpha - d_\beta)^k \right].$$

2. $\exists \lambda \in \mathbb{C}$, such that
   $$f = \lambda \sum_{k=1}^{n} \left\{ [a, c] \otimes (k-1) \otimes [b, d] \otimes [a, c] \otimes (n-k) \right\}.$$  

   Thus, for $0 \leq k \leq n$, the entry of $f$ at weight $k$ is
   $$f_k = \lambda \left[ (n-k)a^{n-k-1}b^k + \epsilon ka^{n-k}c^{k-1}d \right].$$

3. $\exists \lambda \in \mathbb{C}$, such that
   $$f = \lambda \sum_{k=1}^{n} \left\{ [b, d] \otimes (k-1) \otimes [a, c] \otimes [b, d] \otimes (n-k) \right\}.$$  

   Thus, for $0 \leq k \leq n$, the entry of $f$ at weight $k$ is
   $$f_k = \lambda \left[ (n-k)b^{n-k-1}a^k + \epsilon kb^{n-k}d^{k-1}c \right].$$
(We take the convention that $x^0 = 1$ and $0x^{0-1} = 0$, even if $x = 0$.)

We have a corresponding statement for a symmetric signature transformable to a matchgate signature as a generator: $g = (M^{-1})^\otimes n \Gamma'$ for some $M \in GL_2(\mathbb{C})$ and a matchgate signature $\Gamma'$. These formulae can be obtained from those in Theorem 4.20 by substituting $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ for $\begin{bmatrix} a & c \\ b & d \end{bmatrix}$.

From now on, a signature that is transformable to the signature of a matchgate is also called a matchgate signature (or realizable as a matchgate signature). This conforms to the modern concept of tensors, which are given coordinate-free, and not chained to a particular basis. However, to avoid ambiguity, for a signature that is realizable directly as the signature of a matchgate without a transformation, we will call it a standard (matchgate) signature.

By the relationship between second order linear recurrences and the expressions obtained here, we have

**Theorem 4.21.** A symmetric signature $[x_0, x_1, \ldots, x_n]$ is transformable to a matchgate signature under some holographic transformation $\lambda \neq 0$ such that

$$ax_k + bx_{k+1} + cx_{k+2} = 0$$

for $0 \leq k \leq n - 2$.

The expression of a matchgate realizable signature as the solution to a second order linear recurrence relation has a uniqueness which we state explicitly below.

**Lemma 4.22.** Suppose a sequence $(x_0, x_1, \ldots, x_n)$, where $n \geq 3$, has the following form: $x_i = \lambda \alpha^i + \mu \beta^i$, ($\lambda \mu \neq 0, \alpha \neq \beta$), then the representation is unique. That is, if $x_i = \lambda' \alpha'^i + \mu' \beta'^i$, ($0 \leq i \leq n$, $n \geq 3$), then $\lambda' = \lambda, \mu' = \mu, \alpha' = \alpha, \beta' = \beta$ or $\lambda' = \mu, \mu' = \lambda, \alpha' = \beta, \beta' = \alpha$.

**Lemma 4.23.** Suppose a sequence $(x_0, x_1, \ldots, x_n)$, where $n \geq 3$, has the following form: $x_i = \lambda \alpha^i - \mu \alpha^i$, ($\lambda \neq 0$), then the representation is unique. That is, if $x_i = \lambda' \alpha'^i - \mu' \alpha'^i$, ($0 \leq i \leq n$, $n \geq 3$), then $\lambda' = \lambda, \mu' = \mu, \alpha' = \alpha$.

Also, the two cases in Lemma 4.22 and 4.23 are mutually exclusive, depending on whether the eigenvalues of the linear recurrence are distinct or not.

The following observation is due to Valiant [?].

**Proposition 4.24.** If a generator or recognizer signature is realizable by a matchgate under a holographic transformation $M$, then it is also realizable by a matchgate under $M'$, where $M' = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} M$, for any nonzero $x, y \in \mathbb{C}$, or $M' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} M$.

Let $G$ be a recognizer signature of arity $n$ realizable under a holographic transformation $M$ by a matchgate with the standard signature $\Gamma$, then $G = \Gamma M^{\otimes n}$. If $\Gamma$ is any standard signature, then so are $\Gamma \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$ and $\Gamma \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes n$. For example, a matchgate with signature
is obtained from the given matchgate for \( \Gamma \) by appending an edge of weight one to each external node, which effectively flips the truth values 0 and 1. The Proposition can also be proved easily using MGI. The proof for generators is similar.

The Proposition implies that there is an equivalence relation \( \sim \) among the holographic transformations, where one can multiply each row by a nonzero constant, or exchange the two rows of the transformation matrix \( M \). Under this equivalence relation, \( \text{GL}_2(\mathbb{C})/\sim \) is a two dimensional manifold. We call this the basis manifold \( \mathcal{B} \). From now on we identify a basis transformation with its equivalence class containing it. When it is permissible, we use the dehomogenized coordinates \( \begin{bmatrix} 1 & \xi \\ 1 & \eta \end{bmatrix} \) to represent a point (i.e., a basis class) in \( \mathcal{B} \).

Theorem 4.20 gives a complete characterization of all the symmetric realizable matchgate signatures. These tell us exactly what signatures can be realized over some bases. However, to construct a holographic algorithm, one needs to realize some generators and recognizers simultaneously. In terms of \( \mathcal{B} \), a given generator (recognizer) defines a (possibly empty) subvariety which consists of all the bases over which it is realizable. The simultaneous realizability is equivalent to a non-empty intersection of these subvarieties. Thus we have to go beyond Theorems 4.20. For every signature which is realizable according to Theorem 4.20, we need to determine the subvariety where it is realizable.

**Definition 4.25.** Let \( f \) be a symmetric recognizer or generator signature of arity \( n \). Then \( B_{\text{rec}}(f) \) (respectively \( B_{\text{gen}}(f) \)) is the set of all bases in \( \mathcal{B} \) for which the symmetric signature \( f \) for a recognizer (respectively a generator) is realizable. Formally,

\[
B_{\text{rec}}(f) = \{ M \in \mathcal{B} \mid f = \Gamma M^{\otimes n} \text{ for some matchgate signature } \Gamma \}, \\
B_{\text{gen}}(f) = \{ M \in \mathcal{B} \mid f = (M^{-1})^{\otimes n} \Gamma' \text{ for some matchgate signature } \Gamma' \}.
\]

Since the identically zero signature is realizable in every basis, we will assume the signature is not identically zero in the following discussion.

Lemmas 4.26 to 4.30 give a complete and mutually exclusive list of symmetric matchgate realizable signatures for recognizers that are not identically zero. The corresponding set of results for symmetric matchgate realizable generators can be obtained by substituting \( \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \) for \( \begin{bmatrix} a & c \\ b & d \end{bmatrix} \).

Recall that a symmetric signature \( [x_0, x_1, \ldots, x_n] \) is non-degenerate iff \( \text{rank} \begin{bmatrix} x_0 & \cdots & x_{n-1} \\ x_1 & \cdots & x_n \end{bmatrix} = 2 \). The signature is identically 0 iff \( \text{rank} \begin{bmatrix} x_0 & \cdots & x_{n-1} \\ x_1 & \cdots & x_n \end{bmatrix} = 0 \). It has rank 1 iff it can be expressed as \([a^n, a^{n-1}b, \ldots, b^n] \), for \( a, b \in \mathbb{C} \) and not both 0. Lemma 4.26 deals with degenerate signatures.

**Lemma 4.26.** For \( a, b \in \mathbb{C} \) and not both 0,

\[
B_{\text{rec}}([a^n, a^{n-1}b, \ldots, b^n]) = \left\{ \begin{bmatrix} a & b \\ x & y \end{bmatrix} \in \mathcal{B} \mid x, y \in \mathbb{C} \right\}.
\]
Remark: Every non-zero unary signature is of this form.

In the following we assume the signature is non-degenerate. For the case of arity \( n = 2 \), we have

**Lemma 4.27.**

\[
B_{rec}([x, y, z]) = \left\{ \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in B \mid \begin{array}{l}
xd^2 - 2ybd + zb^2 = 0, \quad xc^2 - 2yac + za^2 = 0 \\
or \quad xcd - y(ad + bc) + zab = 0
\end{array} \right\}.
\]

We can prove this lemma by noting that, for arity 2, the only requirement to be a standard signature of a matchgate is the parity constraint. The MGI is satisfied under the parity condition in this case.

In the following we assume the signature has arity \( n \geq 3 \), and non-degenerate. In this case, we note that the constants \( a, b, c \) in Theorem 4.21 are unique up to a scalar factor. In fact if there are two linearly independent triples \((a, b, c)\), then the following matrix

\[
\begin{bmatrix}
x_0 & x_1 & \ldots & x_{n-2} \\
x_1 & x_2 & \ldots & x_{n-1} \\
x_2 & x_3 & \ldots & x_n
\end{bmatrix}
\]

has rank at most one. The first row and the last row are not both zero, otherwise the signature is identically zero (by \( n \geq 3 \)). It follows that the matrix

\[
\begin{bmatrix}
x_0 & x_1 & \ldots & x_{n-1} \\
x_1 & x_2 & \ldots & x_n
\end{bmatrix}
\]

also has rank 1, hence the signature is degenerate.

The next lemma corresponds to non-degenerate signatures in Theorem 4.21 where \( ac = 0 \) and \( b = 0 \). For \( b = c = 0 \) (and \( a \neq 0 \)) we have \([0, \ldots, 0, \lambda, \mu]\), where \( \lambda \neq 0 \) by non-degeneracy. Its reversal \([\mu, \lambda, 0, \ldots, 0]\) corresponds to \( a = b = 0 \) (and \( c \neq 0 \)).

**Lemma 4.28.** Let \( \lambda \neq 0 \). Then,

\[
B_{rec}([0, \ldots, 0, \lambda, \mu]) = \left\{ \begin{bmatrix} 0 & 1 \\ n\lambda & \mu \end{bmatrix} \right\}.
\]

Similarly for its reversal,

\[
B_{rec}([\mu, \lambda, 0, \ldots, 0]) = \left\{ \begin{bmatrix} 1 & 0 \\ \mu & n\lambda \end{bmatrix} \right\}.
\]

The reversal signature \([\mu, \lambda, 0, \ldots, 0]\) can also be viewed as a special case in Lemma 4.31 if we were to allow \( \alpha = 0 \) there, where for \( \alpha = 0 \) and \( i = 0 \), we take the convention that \( i\alpha^{i-1} = 0 \), and also \( \alpha^i = 1 \). However, for the listing to be mutually exclusive we disallow \( \alpha = 0 \) in Lemma 4.31.

Next we consider the case of non-degenerate signatures in Theorem 4.21 where \( ac = 0 \) and \( b \neq 0 \).
Lemma 4.29. For $\lambda \mu \neq 0$,
\[
B_{rec}([\lambda, \lambda, \ldots, \lambda^{n-1}, \lambda^n + \mu]) = \left\{ \begin{bmatrix} 1 & \alpha + \omega \\ 1 & \alpha - \omega \end{bmatrix} \bigg| \omega^n = \pm \frac{\mu}{\lambda} \right\}.
\]
Similarly for its reversal,
\[
B_{rec}([\lambda^n + \mu, \lambda^{n-1}, \ldots, \lambda, \lambda]) = \left\{ \begin{bmatrix} \alpha + \omega & 1 \\ \alpha - \omega & 1 \end{bmatrix} \bigg| \omega^n = \pm \frac{\mu}{\lambda} \right\}.
\]

Here $\alpha = 0$ is GEN-EQ, and corresponds to the case $a = c = 0$ and $b \neq 0$. For $\alpha \neq 0$, the reversal signature $[\lambda^n + \mu, \lambda^{n-1}, \ldots, \lambda, \lambda]$ can also be written as $\lambda'[1, \alpha', \ldots, \alpha^n] + [\mu, 0, \ldots, 0]$, where $\lambda' = \lambda^n$ and $\alpha' = 1/\alpha$, and then can be viewed as a special case in Lemma 4.30 if we were to allow $\beta = 0$ there; although for mutual exclusion we disallow $\beta = 0$ in Lemma 4.30.

Now we consider $ac \neq 0$ in Theorem 4.21. We use the fact that the triple $(a, b, c)$ in the statement of Theorem 4.21 is unique up to a scalar factor. We have a unique characteristic equation $cX^2 + bX + a = 0$, which has two roots $\alpha$ and $\beta$. In particular the forms 1, 2 and 3 from Theorem 4.20 are mutually exclusive. Suppose $b^2 \neq 4ac$, i.e., $\alpha \neq \beta$, we have the following lemma:

Lemma 4.30. For $\lambda \mu \neq 0$, $\alpha \beta \neq 0$, and $\alpha \neq \beta$,
\[
B_{rec}([\lambda \alpha^i + \mu \beta^i]|i = 0, 1, \ldots, n]) = \left\{ \begin{bmatrix} 1 + \omega & \alpha + \beta \omega \\ 1 - \omega & \alpha - \beta \omega \end{bmatrix} \bigg| \omega^n = \pm \frac{\mu}{\lambda} \right\}.
\]

If $b^2 = 4ac \neq 0$, i.e., the characteristic roots coincide $\alpha = \beta \neq 0$, we have the following lemma:

Lemma 4.31. For $\lambda \neq 0$, and $\alpha \neq 0$,
\[
B_{rec}([\lambda i \alpha^{i-1} + \mu \alpha^i]|i = 0, 1, \ldots, n]) = \left\{ \begin{bmatrix} 1 \\ \mu \end{bmatrix} \bigg| \frac{\alpha}{n\lambda + \mu\alpha} \right\}.
\]

Definition 4.32. Simultaneous Realizability Problem (SRP):

**Input:** A set of symmetric signatures for generators and/or recognizers.

**Output:** A common basis of holographic transformation to matchgate signatures if any exists; “NO” if they are not simultaneously realizable as matchgate signatures by holographic transformations.

**SRP Algorithm:**

For every signature $[x_0, x_1, \ldots, x_n]$, check if it satisfies Theorem 4.20.

If not, output “NO” and halt.
Otherwise find $B_{gen}([x_0, x_1, \ldots, x_n])$ or $B_{rec}([x_0, x_1, \ldots, x_n])$ according to one of the lemmas.

Check if these subvarieties have a non-empty intersection.

**Theorem 4.33.** The SRP Algorithm solves SRP in polynomial time.

*Proof.* Checking whether every input signature satisfies Theorem 4.20 can obviously be done in polynomial time. To find the right form and then the right lemma for a signature which satisfies Theorem 4.21 can also be done in polynomial time as they are mutually exclusive.

Every subvariety of bases from Lemma 4.26 to 4.31 for recognizers, and similarly for generators, is of one of three kinds: a finite set of points (of linear size), a line or a quadratic curve. More precisely, expressing things in terms of the manifold $B$ shows the following. For Lemma 4.26 we get a line with $\xi = \text{constant}$ (in the dehomogenized coordinates notation defining $B$). For Lemma 4.27 we get a union of two sets. The first is finite, where both the dehomogenized coordinates $\xi$ and $\eta$ satisfy a quadratic polynomial (and by projective closure). The second set is defined by an equation of the form $A\xi \eta + B(\xi + \eta) + C = 0$ (and by projective closure), where $A, B, C$ are known constants. The other cases are all similar.

For example, for Lemma 4.29, we get at most $n$ points from the equation $\omega^n = \text{constant}$.

To sum up, the SRP Algorithm is a polynomial time algorithm for SRP. \qed

### 4.3.5 $\#_7$PL-RTW-MON-3CNF Redux

In Section 4.2 we had encountered the problem $\#_7$PL-RTW-MON-3CNF. Valiant [?] also introduced another problem $\#_7$Pl-3/2BIP-VC, which counts the number of vertex covers for a planar $(3,2)$-regular bipartite graph, modulo 7. They were solved in polynomial time by holographic algorithms based on matchgates. He called these accidental algorithms. In this subsection, we apply the theory developed in Section 4.3 to find such algorithms almost mechanically. This theory provides more insight and understanding as to what can or cannot be accomplished. With this machinery we are able to generalize to the problems Pl-RTW-MON-\(k\)CNF (the same problem for \(k\)CNF formulae for a general \(k\)) and Pl-\(k\)/2BIP-VC (the same problem for planar \((k,2)\)-regular bipartite graphs). We show that there is a unique modulus $2^k - 1$ for which we can design such a holographic algorithm based on matchgates, which counts the number of solutions modulo $2^k - 1$. In the case of $k = 3$, this shows why 7 is special.

$\#_{2^k-1}$PL-RTW-MON-kCNF

**Input:** A planar $k$CNF Boolean formula where each variable appears positively and in exactly two clauses (Planar, Read-Twice, Monotone, $k$CNF.)

**Output:** Count the number of satisfying assignments modulo $2^k - 1$.

When $k = 3$ this is $\#_7$PL-RTW-MON-3CNF from Section 4.2. As noted earlier, the counting problem $\#PL$-RTW-MON-3CNF is $\#P$-complete already for $k = 3$. 127
To solve the problem by a holographic algorithm, we wish to replace each variable by a generator with the signature \([1, 0, 1]\), and each clause by a recognizer with the signature \([0, 1, 1, \ldots, 1]\) (with \(k\) 1’s). The symmetric signature \([1, 0, 1]\) is the binary \texttt{EQUALITY} function \(\#_2\) and corresponds to a consistent truth assignment on two edges leading to clauses, and \([0, 1, 1, \ldots, 1]\) (with \(k\) 1’s) is the Boolean \texttt{OR} function of arity \(k\) corresponding to a clause. If we connect the generators and recognizers in a natural way, by Theorem 1.5 and Theorem 4.7, this would solve \#PL-RTW-MON-\(k\)CNF in polynomial time, if the signatures are simultaneously realizable by matchgates under a holographic transformation.

Then the question boils down to whether there is a basis in \(B\) where \([1, 0, 1]\) for a generator and \([0, 1, 1, \ldots, 1]\) (with \(k\) 1’s) for a recognizer can be simultaneously realized. For this, we use our machinery.

From Lemma 4.29, taking \(\lambda = 1\), \(\mu = -1\) and \(\alpha = 1\) for \([\lambda \alpha^k + \mu, \lambda \alpha^{k-1}, \ldots, \lambda \alpha, \lambda]\), we have

\[
B_{\text{rec}}([0, 1, 1, \ldots, 1]) = \left\{ \begin{bmatrix} 1 + \omega & 1 \\ 1 - \omega & 1 \end{bmatrix} : \omega^k = \pm 1 \right\}.
\]

We look for some \(\omega^k = \pm 1\), such that \(\begin{bmatrix} 1 + \omega & 1 \\ 1 - \omega & 1 \end{bmatrix} \in B_{\text{gen}}([1, 0, 1])\), which is obtained from Lemma 4.27 by substituting \(\begin{bmatrix} a & c \\ -c & a \end{bmatrix}\) for \(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\). Explicitly,

\[
B_{\text{gen}}([x, y, z]) = \left\{ \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in B \left| \begin{array}{c}
xa^2 + 2yac + zc^2 = 0, 
xb^2 + 2ybd + zd^2 = 0 \\
\text{or } xab + y(ad + bc) + zcd = 0
\end{array} \right. \right\}. \tag{4.27}
\]

Thus, we want \((1 + \omega)^2 + 1 = (1 - \omega)^2 + 1 = 0\) or \((1 + \omega)(1 - \omega) + 1 = 0\). The first case is impossible, and in the second case we require \(\omega^2 = 2\). Together with the condition \(\omega^k = \pm 1\), we have \(2^k - 1 = 0\). From this we can already see that for every prime \(p | 2^k - 1\), \#\text{PL-RTW-MON-}k\text{CNF} is computable in polynomial time. In particular this is true for every Mersenne prime \(2^q - 1\). (Note that \(\omega^2 = 2\) means that 2 is a quadratic residue.) If we develop the theory over a finite field \(\mathbb{Z}_p\), we can obtain corresponding versions of Lemma 4.27 and Lemma 4.29, and conclude that in order to have these signatures be simultaneously realizable over \(\mathbb{Z}_p\), we must have \(p | 2^k - 1\).

More generally we have:

**Theorem 4.34.** There is a polynomial time algorithm for \#\(2^{k-1}\)PL-RTW-MON-\(k\)CNF. Furthermore, any modulus \(m\) for which the appropriate signatures exist must be a divisor of \(2^k - 1\).

**Proof.** Our discussion above already shows that the modulus \(2^k - 1\) is the best we can do. (Formally speaking we should present a generalization such as Theorem 1.5 over a ring such as \(\mathbb{Z}_{2^k - 1}\), which we will omit here.) We now give the polynomial algorithms in two cases:

**Case 1: \(k\) is even.**

Over the complex numbers \(\mathbb{C}\), from (4.27), the generator form of Lemma 4.27, and Lemma 4.29,
we can see that a generator for \([1, 0, 1]\) and a recognizer for \([1 + \epsilon 2^{k/2}, 1, 1, \ldots, 1]\) (where there are \(k\) 1’s, and \(\epsilon = \pm 1\)) are simultaneously realizable in the basis \([\begin{array}{c} 1 + \sqrt{2} \\ 1 - \sqrt{2} \end{array} \ 1]\).

Setting \(\epsilon = 1\) and replacing each variable by a generator and each clause by a recognizer with the corresponding signatures, we obtain a matchgrid \(\Omega\). If we replace the generators and recognizers by their corresponding matchgates, we get a weighted planar graph \(G\). Then Theorem 1.5 gives

\[
\text{Holant}_\Omega = \text{PerfMatch}(G). \tag{4.28}
\]

We will denote this value by \(X\).

From the left-hand side of (4.28) we know that \(X\) is an integer because every entry in the signatures of generators \([1, 0, 1]\) and recognizers \([1 + 2^{k/2}, 1, 1, \ldots, 1]\) is an integer. Furthermore we have

\[
X \equiv \#\text{PL-RTW-MON-kCNF} \pmod{1 + 2^{k/2}}.
\]

From the right-hand side of (4.28) we know that \(X\) can be computed in polynomial time using the FKT algorithm for perfect matchings of a planar graph. The planar graph has weights from the subfield \(\mathbb{Q}(\sqrt{2}) \subset \mathbb{C}\), which poses no problem to the Pfaffian evaluation of FKT in polynomial time in bit complexity.

Therefore \(\#_{2^{k/2}+1}\text{PL-RTW-MON-kCNF}\) can be computed in polynomial time. Similarly, setting \(\epsilon = -1\), we can compute \(\#_{2^{k/2}-1}\text{PL-RTW-MON-kCNF}\) in polynomial time.

Since \(\gcd(2^{k/2} + 1, 2^{k/2} - 1) = 1\) and \(2^{k} - 1 = (2^{k/2} + 1)(2^{k/2} - 1)\), we can apply Chinese remaindering to get a polynomial time algorithm for \(\#_{2^{k}-1}\text{PL-RTW-MON-kCNF}\).

**Case 2: \(k\) is odd.**

Consider the ring \(\mathbb{Z}_{2^k-1}\). (Formally we can develop the theory over such a ring, and consider invertible elements and matrices for the basis manifold, which we will omit.) Let \(r = 2^{(k+1)/2} \in \mathbb{Z}_{2^k-1}\). Then \(r\) satisfies \(r^2 = 2\) in \(\mathbb{Z}_{2^k-1}\). We observe that \(r^k - 1 = (2^k)^{(k+1)/2} - 1 = 0\) in \(\mathbb{Z}_{2^k-1}\).

Therefore over this ring \(\mathbb{Z}_{2^k-1}\) and with the basis \([\begin{array}{c} 1 + r \\ 1 - r \end{array} \ 1]\), the generator \([1, 0, 1]\) and the recognizer \([0, 1, 1, \ldots, 1]\) (with \(k\) 1’s) are simultaneously realizable by matchgates according to (4.27) and Lemma 4.29. As a result, we have a polynomial time algorithm for \(\#_{2^k-1}\text{PL-RTW-MON-kCNF}\).

Now we consider another problem.

\(\#_{2^k-1}\text{PL-k/2BIP-VC}\)

**Input:** A planar regular bipartite graph \(G = (U, V, E)\), where every \(u \in U\) has degree \(k\) and every \(v \in V\) has degree 2 (Planar \((k, 2)\)-regular bipartite graph).

**Output:** Count the number of vertex covers modulo \(2^k - 1\). 129
Consider an arbitrary subset \( S \subseteq V \) of vertices from the right. Every vertex \( u \in U \) on the left either has all its \( k \) adjacent vertices in \( S \), in which case there are exactly two choices to extend \( S \) at \( u \) to a vertex cover, or has some of its \( k \) adjacent vertices not in \( S \), in which case there is exactly one choice to extend \( S \) at \( u \) to a vertex cover. Thus, following the general recipe for holographic algorithms, we want a generator with signature \([1, 0, 1]\) and a recognizer with signature \([1, 1, \ldots, 1, 2]\) (with \( k \) 1’s), to be simultaneously realized over some basis.

From Lemma 4.29, taking \( \lambda = \mu = \alpha = 1 \) for \([\lambda, \lambda \alpha, \ldots, \lambda \alpha^{n-1}, \lambda \alpha^n + \mu]\), we have

\[
B_{rec}([1, 1, \ldots, 1, 2]) = \left\{ \begin{bmatrix} 1 & 1 + \omega \\ 1 & 1 - \omega \end{bmatrix} \right| \omega^k = \pm 1 \right\}.
\]

Then the same derivation for \( \#_{2^{k-1}} \text{Pl-RTW-Mon-}k\text{CNF} \) gives a polynomial time algorithm for \( \#_{2^{k-1}} \text{PL-}k/2\text{BIP-VC} \).

**Theorem 4.35.** There is a polynomial time algorithm for \( \#_{2^{k-1}} \text{PL-}k/2\text{BIP-VC} \). Furthermore, any modulus \( m \) for which the appropriate signatures exist must be a divisor of \( 2^k - 1 \).

There is a 1-1 correspondence between vertex covers and independent sets, by taking complement. Thus we also have a polynomial time algorithm for \( \#_{2^{k-1}} \text{PL-}k/2\text{BIP-IS} \), which is to count the number of independent sets modulo \( 2^k - 1 \) for planar \((k, 2)\)-regular bipartite graph. More directly, we can proceed as before: For any planar \((k, 2)\)-regular bipartite graph \( G = (U, V, E) \), for any \( S \subseteq V \), to extend it to an independent set, every \( u \in U \) has either two choices if all \( k \) neighbors of \( u \) are not in \( S \), or one choice (not to be included) otherwise. It follows that we should simultaneously realize a generator with signature \([1, 0, 1]\) and a recognizer with signature \([2, 1, \ldots, 1, 1]\) (with \( k \) 1’s), and the rest are the same.

Our general machinery not only can find the required signatures when they exist, but also can prove certain desired signatures cannot be simultaneously realized. As an example, one may wish to extend the previous two problems to allow more than Read-twice as in \#\text{Pl-RTW-Mon-}k\text{CNF} \), where \( \ell > 2 \). This calls for a simultaneous realizability of the \text{EQUALITY} function of arity \( \ell \), \((\ell) = [1, 0, \ldots, 0, 1] \) (where there are \((\ell - 1)\) 0’s), and the Boolean \text{OR} function on \( k \) inputs, \([0, 1, 1, \ldots, 1]\) (where there are \( k \) 1’s). This can be shown to result in an empty intersection on \( \mathcal{B} \).

In this book we focus on the exact complexity of counting problems. There is an active research area to study the approximate complexity of counting problems, and to the extent possible to reach classification theorems \([?, ?, ?, ?, ?, ?]\). It appears that for approximate counting, there are more than just two levels of complexity, those that are solvable by a polynomial time approximation algorithm and those that are NP-hard to approximate. In particular the problem to approximately count the number of independent sets for bipartite graphs \((\#\text{BIS})\) appears to be a representative problem for a distinct class of intermediate complexity. This is in contrast to \#\text{Pl-}k/2\text{BIP-IS} which is solvable in polynomial time exactly.
Chapter 5

2-Spin Systems on Regular Graphs

In this chapter, we consider the following class of counting problems. An input \( k \)-regular graph \( G = (V, E) \) is given, where every \( e \in E \) is labeled with a symmetric edge function \( g : \{0, 1\}^2 \rightarrow \mathbb{C} \). We allow self-loops and parallel edges in \( G \), where each self-loop contributes 2 to the degree of the incident vertex. We present a dichotomy theorem for spin systems on \( k \)-regular graphs. For simplicity and for what turns out to be sufficient in terms of logical dependence from later chapters, we will restrict our proofs to \( k = 3 \) and \( k = 4 \). As a spin system, the vertices of the graph represent Boolean variables. The edge function \( g \) takes 0-1 inputs from its incident vertices and outputs arbitrary (algebraic) values in \( \mathbb{C} \). The problem is to compute the partition function

\[
Z(G) = \sum_{\sigma: V \rightarrow \{0, 1\}} \prod_{(u, v) \in E} g(\sigma(u), \sigma(v)).
\]

For example, if \( g \) is the Boolean \( \text{OR} \) function represented by the matrix

\[
\begin{bmatrix}
0 & 1 \\
1 & 1
\end{bmatrix}
\]

then \( Z(G) \) counts the number of vertex covers on a \( k \)-regular graph \( G \).

A spin system can also be defined for a general domain \([q]\), for \( q > 2 \), where each vertex takes one of \( q \) possible values. But in Part I of this book we focus on the Boolean domain \( q = 2 \). A 2-spin system is a special case of \#CSP where there is a single symmetric binary constraint function. A 2-spin system on a \( k \)-regular graph is the further restriction that every variable appears \( k \) times. Equivalently, we may view this as a bipartite Holant problem \( \text{Holant}(=k | g) \). We treat the binary \( g \) as a generator (contravariant) on the right side, and the \( \text{EQUALITY} (=k) \) of arity \( k \) as a recognizer (covariant) on the left side. The main difference between a general Holant problems and \#CSP is that \( \text{EQUALITY} \) functions of arbitrary arity are presumed to be present in \#CSP. For spin systems on \( k \)-regular graphs, we assume the presence of \( =_k \), and a single symmetric edge function \( g \). When \( \text{EQUALITY} \) gates of arbitrary arity are freely available in possible inputs, it is technically easier to prove \#P-hardness. Two important techniques, called \textit{stretching} and \textit{thickening}, require the availability of \( \text{EQUALITY} \) gates of arbitrary arity (equivalently, vertices of arbitrarily high degrees) to carry out. Stretching is to replace an edge by a path, which requires binary \( \text{EQUALITY} \) gates. Thickening is to replace an edge by multiple parallel edges, which requires
Equality of arbitrary arity. Proving \#P-hardness becomes more challenging in the degree restricted case. Furthermore there are indeed cases within this class of counting problems where the problem is \#P-hard for general graphs, but is solvable in polynomial time when restricted to \( k \)-regular graphs.

We denote the symmetric edge function \( g \) by \([x; y; z]\), where \( x = g(0, 0) \), \( y = g(0, 1) = g(1, 0) \) and \( z = g(1, 1) \). Functions will also be called gates or signatures. (For Vertex Cover, the corresponding function is the OR gate, and is denoted by the signature \([0; 1; 1]\).) The dichotomy theorem presented in this chapter gives a complexity classification of \( Z(G) \) on \( k \)-regular graphs \( G \) in terms of the signature \( g = [x; y; z] \), where \( x, y, z \in \mathbb{C} \). Note that a self-loop with a binary edge function \( g = [x; y; z] \) amounts to a unary vertex weight function \([x; z] \).

First, if \( y = 0 \), it is easy to compute \( Z(G) \) in polynomial time. Assume \( y \neq 0 \). We may normalize \( g \) and set \( y = 1 \). The main theorem of this chapter is the following:

**Theorem 5.1.** Suppose \( a, b \in \mathbb{C} \), \( k \geq 3 \), and let \( X = ab \) and \( Y = a^k + b^k \). Then \( Z(G) \) with \( g = [a, 1, b] \) for input \( k \)-regular graphs \( G \) is \#P-hard, except in the following 4 cases, for which the problem is in \( P \).

1. \( X = 1 \).
2. \( X = Y = 0 \).
3. \( X = -1 \) and \( Y = 0 \).
4. \( X = -1 \) and \( Y^2 = 4(-1)^k \).

If we restrict the input to planar \( k \)-regular graphs, then these four cases are solvable in \( P \), as well as a fifth case \( Y^2 = 4X^k \). The problem remains \#P-hard in all other cases.

Given a binary symmetric function \( g \), if \#CSP\( (g) \) is tractable, then of course the 2-spin system \( Z(G) \) is also tractable with \( g \) on \( k \)-regular graphs. Cases 1 to 4 correspond to tractable cases of \#CSP by the product type \( \mathcal{P} \) (Definition 3.2), and the affine type \( \mathcal{A} \) (Definition 3.6), possibly after a holographic transformation. Furthermore, over planar graphs, there is an additional set of tractable cases which are \#P-hard over general (non-planar) \( k \)-regular graphs, and these consist entirely of suitable holographic transformations from matchgate computable problems. Finally everything else is \#P-hard, even for planar \( k \)-regular graphs. (For \#CSP problems, this planar tractable case will be classified in Chapter ??).

We can parameterize the space by \((a, b)\) as well as equivalently by \((X, Y)\). The region defined by cases 1 to 4 in Theorem 5.1 will be called Region I. In terms of the categories of tractability for \#CSP (cf. Theorem 3.7), case 1 is degenerate, case 2 together with \( y = 0 \) in \([x; y; z]\) before normalization correspond to the product type, and cases 3 and 4 correspond to affine binary signatures, including possibly a suitable holographic transformation. Region II is defined to be \( Y^2 = 4X^k \) but \( X \not\in \{0, \pm 1\} \). It can also be specified by \( Y^2 = 4X^k \), minus
the intersection with Region I. It is worth noting that \( Y^2 = 4X^k \) specifies precisely those signatures that under a suitable holographic transformation both \( = k \) and \( g \) are realizable as standard matchgate signatures. The remaining space is called Region III.

**Definition 5.2.** In terms of \((a, b)\), Region I consists of

1. \( ab = 1 \) (degenerate),
2. \( a = b = 0 \) (disequality),
3. \( ab = -1 \) and \( a^{2k} = b^{2k} \) (combining cases 3 and 4 in Theorem 5.1).

Region II consists of \( a^k = b^k \), minus the intersection with Region I. Region III consists of all remaining \((a, b) \in \mathbb{C}^2\).

The union of Regions I and II is specified by \( X = 1 \) or \( Y^2 = 4X^k \) or \((X, Y) = (-1, 0)\). In terms of \((a, b)\), this union is \( ab = 1 \) or \( a^k = b^k \) or \([ab = -1 \text{ and } a^{2k} = b^{2k}]\).

We can restate this theorem in terms of \(a\) and \(b\). Note that \( Y^2 = 4X^k \) is equivalent to \( a^k = b^k \).

**Theorem 5.3.** Suppose \( a, b \in \mathbb{C} \) and \( k \geq 3 \). Then \( Z(G) \) with \( g = [a, 1, b] \) for input \( k\)-regular graphs \( G \) is \#P-hard except in the following categories, for which the problem is in \( P \).

1. \( ab = 1 \).
2. \( a = b = 0 \).
3. \( ab = -1 \) and \( a^{2k} = b^{2k} \).

If we restrict the input to planar \( k\)-regular graphs, then these three categories are solvable in \( P \), as well as a fourth category \( a^k = b^k \). The problem remains \#P-hard in all other cases.

**Exercise:** Prove that Theorem 5.1 and Theorem 5.3 are equivalent.

Prove that case 3 of Theorem 5.3 is equivalent to \( ab = -1 \) and \( a^{4k} = 1 \), as well as \( ab = -1 \) and \( b^{4k} = 1 \). Prove that if \( k \) and \( k' \) are relatively prime, then the tractability conditions of \#CSP([\(a, 1, b]\)) in Theorem 3.7 and of \( Z(G) \) in Theorem 5.3 for both \( k\)-regular and \( k'\)-regular graphs are equivalent.

We will prove this dichotomy for the cases \( k = 3 \) and \( k = 4 \). For technical reasons, the situations for odd and even \( k \) are slightly different. But we will proceed in a similar fashion in both cases. We will first establish tractability, in Region I for general (not necessarily planar) graphs, and in Region II for planar graphs. The main task is to show everything else is \#P-hard. For 3-regular graphs, we will first prove that these problems are \#P-hard in
Region II for general graphs (where the problem is tractable when restricted to planar graphs as input). Then we move on to prove \#P-hardness within Region III, where the problem is \#P-hard even for planar graphs. For 4-regular graphs the order is reversed. The main theme in the \#P-hardness proof is to show that we can interpolate all unary (or “virtually unary”) signatures. After that, in essence, we appeal to ideas and results from the framework of Holant* problems.

In order to prove this theorem, several new proof techniques are introduced.

1. We present a general approach to interpolate all unary signatures. This involves a recursive iteration gadget, a starter gadget (in some cases), and an analysis of eigenvalues and eigenvectors.

2. We introduce a method to construct gadgets that carry out iterations at a higher dimension, and then collapse it to a lower dimension for the purpose of constructing (virtual) unary signatures. This involves a starter gadget, a recursive iteration gadget, and a finisher gadget. We prove a lemma that guarantees that among polynomially many iterations, some subset of them will be sufficient for interpolation to succeed.

3. Algebraic symmetrization. We derive an expression of \( Z(G) \) over \( k \)-regular graphs as a graph polynomial, whose degree is reduced by taking symmetry into account. This symmetry enables us to draw connections between different problems, and also reduces the size and complexity of various polynomials which will be analyzed.

4. Eigenvalue Shifted Pairs. These are coupled pairs of gadgets whose transition matrices differ by \( \delta I \) where \( \delta \neq 0 \). They have shifted eigenvalues, and by analyzing where they may fail individually, we can show that jointly they succeed.

5. Anti-gadgets. These are also coupled pairs of gadgets, but in which one gadget has the algebraic effect of cancelling out a portion of the other gadget, resulting in a “virtual gadget” with desirable algebraic properties.

6. Gadget syzygy. These are trios of gadgets that in their alignment we can prove that jointly they succeed.

7. Symmetrizer gadgets. Some of our constructions produce signatures with desirable properties, but they are not necessarily symmetric. If symmetric signatures are required, then gadgets can be used to introduce symmetry in the underlying graph, thereby guaranteeing that the signatures produced from the construction are also symmetric.

When the dichotomy was first proved \([?, ?, ?, ?]\), an additional important idea was to use Tarski’s decidability theorem for real closed field \([?, ?, ?]\) and prove the desired result by a substantial amount of symbolic computation. In the present chapter, we have at least formally eliminated symbolic computation that must require machine verification. However
many constructions in this Chapter are found computationally. The fact that they work as stated can be verified by hand with some work, and they have also been verified by computer programs. Symbolic computation remains a valuable tool for exploration and as a proof technique.

Our spin system is the same problem as Holant\(=\_k[a, 1, b]\), which we will abbreviate as Hol\(_k(a, b)\). In this Edge-Vertex incidence graph view, the input instances are \((2, k)\)-regular bipartite graphs. Throughout this chapter, all \((2, k)\)-regular bipartite graphs are arranged with generators (contravariant tensors) on the degree-2 side (right-hand side) and recognizers (covariant tensors) on the degree-\(k\) side (left-hand side).

For convenience, in this chapter we will further designate dangling edges as \textit{leading} edges if they are internally connected to a binary generator \(g = [a, 1, b]\), or \textit{trailing} edges if they are internally connected to a recognizer \(=\_k\). We will draw \(\mathcal{F}\)-gates with leading edges protruding from the left and trailing edges from the right. These \(\mathcal{F}\)-gates can be composed by merging leading edges of one \(\mathcal{F}\)-gate with trailing edges of another, respecting the underlying recognizer and generator bipartite structure. Suppose an \(\mathcal{F}\)-gate has \(m\) leading edges and \(n\) trailing edges. Then the signature of the \(\mathcal{F}\)-gate can be organized as a \(2^m \times 2^n\) matrix, where the rows and columns are indexed by the 0-1 assignments to the leading and trailing edges respectively. If the number of trailing edges in one \(\mathcal{F}\)-gate matches the number of leading edges in another, then a new \(\mathcal{F}\)-gate can be formed by merging these edges, and the associated matrix is obtained by multiplying the two original matrices together. In particular, an \(\mathcal{F}\)-gate with only leading edges would be viewed as a column vector, representing a generator, and then merging with an \(\mathcal{F}\)-gate having a matching number of trailing edges corresponds to pre-multiplication by the associated matrix. In this way we can view an \(\mathcal{F}\)-gate with \(m\) leading edges and \(n\) trailing edges as transforming \(\mathcal{F}\)-gates with arity-\(n\) generator signatures into \(\mathcal{F}\)-gates with arity-\(m\) generator signatures. Furthermore, the \(\mathcal{F}\)-gates in this chapter will transform symmetric signatures to symmetric signatures. This implies that there exists an equivalent \(m + 1 \times n + 1\) matrix which operates directly on vectors written in symmetric signature notation. We will henceforth identify the transition matrix with the \(\mathcal{F}\)-gate itself.

We define some specific types of \(\mathcal{F}\)-gates. An \textit{arity-\(r\) starter gadget} is an \(\mathcal{F}\)-gate with no

(a) A recursive gadget  (b) A starter gadget

Figure 5.1: Examples of binary recursive and starter gadgets
trailing edges and $r$ leading edges, representing an arity-$r$ generator. An *arity-$r$ recursive gadget* is an $F$-gate with $r$ trailing edges and $r$ leading edges.

When we speak of an $F$-gate (where $F = R \cup G$) or gadget $M$ being constructible in the context of Holant$(R \mid G)$, we mean that the underlying graph of $M$ has a bipartite structure, where vertices in one partition are labelled with signatures from $R$ and vertices in the other partition from $G$. In other words, $M$ can appear within an input instance of Holant$(R \mid G)$ without violating any structural bipartite constraints. When we say a generator (or recognizer) is constructible in the context of Holant$(R \cup \{g\} \mid G)$, we mean that an $F$-gate is constructible in that context, and that all dangling edges are internally incident to generators (or recognizers). When we say that a sequence of signatures is efficiently constructible in the context of some Holant problem, we mean that a sequence of $F$-gates having those signatures can be constructed in that context, and that the first $n$ of these can be built in time polynomial in $n$. When we say that a generator $g$ (or recognizer $r$) can be efficiently simulated in the context of Holant$(R \mid G)$, we mean that Holant$(R \mid G \cup \{g\}) \leq^P \text{Holant}(R \mid G)$ or Holant$(R \cup \{r\} \mid G) \leq^P \text{Holant}(R \mid G)$, respectively.

We use Arg to denote the principal value of a complex argument; i.e., $\text{Arg}(z) \in (-\pi, \pi]$ for all nonzero $z \in \mathbb{C}$, such that $z = |z| e^{i \text{Arg}(z)}$. We say that a square nonsingular matrix $M$ has finite projective order if there exists a positive integer $j$, such that $M^j = \lambda I$ for some (nonzero) $\lambda \in \mathbb{C}$. Otherwise, $M$ has infinite projective order. A nonsingular $M$ has finite projective order iff for some positive integers $i \neq j$, $M^i = \lambda M^j$ for some (nonzero) $\lambda \in \mathbb{C}$.

### 5.1 3-Regular graphs

#### 5.1.1 Tractability in Regions I and II

Recall that for 3-regular graphs, Region I is specified by $X = 1$ or $(X, Y) \in \{(0, 0), (-1, 0), (-1, \pm 2i)\}$. Region II is specified by $Y^2 = 4X^3$ but $X \not\in \{0, \pm 1\}$. The union of Regions I and II is specified by $X = 1$ or $Y^2 = 4X^3$ or $(X, Y) = (-1, 0)$.

The next lemma introduces the technique of algebraic symmetrization. We show that over 3-regular graphs, the Holant value is expressible as a polynomial $P(X, Y)$ with integer coefficients, where $X = ab$ and $Y = a^3 + b^3$. This change of variable, from $(a, b)$ to $(X, Y)$, is important in two ways. First, it allows us to derive tractability and hardness results more easily. It draws a connection between problems that may appear unrelated, but the tractability, and respectively hardness, of one problem implies the other. Second, it facilitates the proof of hardness by reducing the degree of the polynomials involved. Such a reduction in degree also makes symbolic computation more feasible, although we do not use this particular approach here.

**Lemma 5.4.** Let $G$ be a 3-regular graph. Then there exists a polynomial $P_G(\cdot, \cdot)$ with two variables and integer coefficients such that for any signature grid $\Omega$ having underlying graph
If the edge function is \([a, 1, b]\), then \(Z(G) = P_G(ab, a^3 + b^3)\).

**Proof.** Consider any 0-1 vertex assignment \(\sigma\) with a nonzero valuation. If \(\sigma'\) is the complement assignment switching every 0 and 1 in \(\sigma\), then for \(\sigma\) and \(\sigma'\), we have the sum of valuations \(a^i b^j + a^j b^i\) for some \(i\) and \(j\). Here \(i\) (resp. \(j\)) is the number of edges connecting two degree 3 vertices both assigned 0 (resp. 1) by \(\sigma\). We note that \(a^i b^j + a^j b^i = (ab)^\min(i,j)(a^{j-i} + b^{i-j})\).

We prove \(i \equiv j \mod{3}\) inductively. For the all-0 assignment, this is clear since in this case \(i\) is the number of edges, \(j = 0\), and the number of edges is divisible by 3 for a 3-regular graph. Now starting from any assignment \(\sigma\), if we switch the assignment on one vertex from 0 to 1, it is easy to verify that it changes the valuation from \(a^i b^j\) to \(a^{i'} b^{j'}\), where \(i - j = i' - j' + 3\). As every 0-1 assignment is obtainable from the all-0 assignment by a sequence of switches, the conclusion \(i \equiv j \mod{3}\) follows.

Now \(a^i b^j + a^j b^i = (ab)^\min(i,j)(a^3 + b^3)\), for some \(\ell \geq 0\) and a simple induction

\[
a^{3(\ell+1)} + b^{3(\ell+1)} = (a^{3\ell} + b^{3\ell})(a^3 + b^3) - (ab)^3(a^{3(\ell-1)} + b^{3(\ell-1)})
\]

shows that the partition function \(Z(G)\) is a polynomial \(P_G(ab, a^3 + b^3)\) with integer coefficients.  

**Corollary 5.5.** If \(X = -1\) and \(Y \in \{0, \pm 2i\}\), then \(\text{Hol}_3(a, b)\) is in P.

**Proof.** The problems \(\text{Hol}_3(1, -1)\), \(\text{Hol}_3(-i, -i)\), and \(\text{Hol}_3(i, i)\) are all solvable in P (these fall within the families \(F_1\), \(F_2\), and \(F_3\) from Chapter 3); \(X = -1\) for each of these three problems, whereas the value of \(Y\) is 0, 2i, and -2i respectively. Since the value \(Z(G)\) of any 3-regular signature grid is completely determined by \((X, Y)\), and the polynomial \(P_G(\cdot, \cdot)\) (which depends only on the underlying graph \(G\)), \(\text{Hol}_3(a, b)\) is computable in polynomial time for any \(a\) and \(b\) such that \(ab = -1\) and \(a^3 + b^3 \in \{0, \pm 2i\}\) (i.e. \(ab = -1\) and \(a^{12} = 1\)).

We now list all the cases where \(\text{Hol}_3(a, b)\) is computable in polynomial time.

**Theorem 5.6.** If any of the following four conditions is true, then \(\text{Hol}_3(a, b)\) is solvable in polynomial time:

1. \(X = 1\),
2. \(X = Y = 0\),
3. \(X = -1\) and \(Y \in \{0, \pm 2i\}\)
4. \(Y^2 = 4X^3\) and the input is restricted to planar graphs.

**Proof.** If \(X = 1\) then the signature \([a, 1, b]\) is degenerate and the Holant can be computed in polynomial time. If \(X = Y = 0\) then \(a = b = 0\), and a 2-coloring algorithm can be employed on the edges. If \(X = -1\) and \(Y \in \{0, \pm 2i\}\) then we are done by Corollary 5.5. If we restrict
the input to planar graphs and $4X^3 = Y^2$ (equivalently, $a^3 = b^3$), holographic algorithms with matchgates from Chapter 4 can be applied. Indeed, if $ab = 0$ then both $a = b = 0$ and this is case 2 which is tractable even for general graphs. Suppose $ab \neq 0$, and let $\omega = a/b$. We apply a holographic transformation defined by $T = \begin{bmatrix} 1 & 0 \\ 0 & \omega \end{bmatrix}$, then $(=3)T^{\otimes3} = (=3)$, and $(T^{-1})^{\otimes2}[1, \frac{1}{\omega}, \frac{b}{a}] = [1, \frac{a}{b}, 1]$, using $a^3 = b^3$. Then under a further holographic transformation by $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, both $(=3)$ and any $[1, c, 1]$ are simultaneously realizable by matchgates. (This can be directly verified, but a more systematic approach is to use the theory developed in Chapter 4, in particular Theorem 4.33.)

The main task in Section 5.1 is to prove that all remaining problems are \#P-hard for 3-regular graphs.

### 5.1.2 Planar Tractable Region II is \#P-hard for General Graphs

In this subsection we get acquainted with our technique of proving hardness for Holant problems over regular graphs. Our specific application at the moment is to prove that $\text{Hol}_3(a, b)$ is \#P-hard assuming $a^3 = b^3$ and $X \notin \{-1, 0, 1\}$. (By Theorem 5.6, $\text{Hol}_3(a, b)$ is tractable under these assumptions when the input is restricted to planar graphs.) We first make the following simplifying observation: instead of working under $a^3 = b^3$, we may as well assume $a = b$.

**Lemma 5.7.** Suppose $Y^2 = 4X^3$ (equivalently, $a^3 = b^3$). Then there exist $a', b' \in \mathbb{C}$ such that $a' = b'$, $a'b' = X$, and $(a')^3 + (b')^3 = Y$.

**Proof.** If $X = 0$ then $Y = 0$ and $a = b = 0$ so we are already done. Otherwise, $X \neq 0$ implies $a \neq 0$ and we may set $a' = b' = b^2a^{-1}$ so that $a'b' = (b^2a^{-1})^2 = b^4a^{-2} = ab = X$ and $(a')^3 + (b')^3 = b^6a^{-3} + b^6a^{-3} = a^3 + b^3 = Y$. 

A recurring theme we will see for proving hardness in this chapter is:

1. obtain an infinite set of pairwise linearly independent signatures,
2. if necessary, project the signatures down to a lower dimension while retaining pairwise linear independence, and
3. use an interpolation argument to simulate any signature in that lower dimensional space.

The ability to simulate any signature in the lower dimensional space will lead to a reduction to prove \#P-hardness. By “lower dimension” we usually mean unary signatures (i.e. a signature sequence of the form $[x_j, y_j]$ for $j \geq 0$). However, in some circumstances (such as for $\text{Hol}_3(a, a)$), it is impossible to do so without sacrificing pairwise linear independence:
Lemma 5.8. Let $F$ be a set of signatures, each having the property that its value remains unchanged under the complement of its inputs. Then any $F$-gate has this same property.

**Proof.** For any $f \in F$, if $\sigma$ and $\sigma'$ are complementary assignments to its inputs, then they take the same value $f|_\sigma = f|_{\sigma'}$. If $\sigma$ and $\sigma'$ are complementary assignments to the dangling edges of an $F$-gate, by complementing the assignments to all internal edges, we have a 1-1 correspondence of all terms defining the signature sum of the $F$-gate under $\sigma$ with those under $\sigma'$, and each pair of corresponding terms have the same value.

Corollary 5.9. If $F = \{[a,1,a],=_k\}$ for some positive integer $k$, then any arity-1 or arity-2 $F$-gate has a signature of the form $[x,x]$ or $[x,y,x]$ respectively.

**Proof.** Immediate from Lemma 5.8.

Given this knowledge, we aim to produce a sequence of “virtual unary” signatures rather than actual unary signatures in such cases. Specifically, we will obtain pairwise linearly independent signatures of the form $[x_j,y_j,x]$ instead of $[x_j,y_j]$. Due to the symmetry exhibited in these binary signatures, we will be able to carry out interpolation without projecting them down to a lower dimension.

**Unary recursive construction**

Given a unary recursive gadget $M$, we can chain $j$ copies of $M$ together (merging the leading edges of one copy with the trailing edges of the next) to arrive at an arity-2 $F$-gate with signature $M^j$. We want such a sequence $\{M^j\}_{j \geq 0}$ of signatures to be pairwise linearly independent.

**Lemma 5.10.** Let $M$ be a nonsingular square matrix. Then $M$ has infinite projective order if and only if $\{M^j\}_{j \geq 0}$ are pairwise linearly independent. In particular, if $M$ is the transition matrix for some arity $r$ recursive gadget, then $M$ has infinite projective order if and only if $\{M^j\}_{j \geq 0}$ is a sequence of pairwise linearly independent signatures.

**Proof.** Immediate from definitions.

The eigenvalues of $M$ give us a convenient way to check that this sequence of $F$-gates really has the property that the signatures $M^j$ are pairwise linearly independent. In other words, we are reducing the problem of generating a sequence of pairwise linearly independent signatures to a test in terms of eigenvalues.

**Lemma 5.11.** Let $M \in \mathbb{C}^{r \times r}$ be a nonsingular matrix, where the ratio of two of its eigenvalues is not a root of unity. Then $\{M^j\}_{j \geq 0}$ is a sequence of pairwise linearly independent matrices.
Proof. We write $M = T J T^{-1}$ in its Jordan Normal Form, where $T$ is invertible, and $J$ is a triangular matrix with all eigenvalues of $M$ appearing on its diagonal. Suppose that $M$ has finite projective order. Then for some $m > 0$ we have $M^m = \lambda I$ for some nonzero $\lambda \in \mathbb{C}$. Then $T^{-1} J^m T = \lambda I$ hence $J^m = \lambda I$. So $\alpha^m = \beta^m$ for all eigenvalues $\alpha$ and $\beta$ of $M$. This is a contradiction. We conclude that $M$ has infinite projective order and by Lemma 5.10, $\{M^j\}_{j \geq 0}$ is a sequence of pairwise linearly independent signatures.

Suppose $a = b$. Depending on the value of $a$ and our gadget $M$, the sequence $\{M^j\}_{j \geq 0}$ may or may not be pairwise linearly independent, but we will show that for every $[a, 1, a]$ that does not correspond to a tractable case, there exists some unary recursive gadget with this property. There is a technical detail in the interpolation step; we will want these $F$-gates to take the place of generator signatures in a signature grid to preserve the bipartite structure of the graph. To address this, we tack on an extra degree-two vertex labeled with the generator $[a, 1, a]$, making the full gadget a binary generator.

Lemma 5.12. Suppose $a = b$ and $a^4 \notin \{0, 1\}$. Then there exists a unary recursive gadget $M$ for which the generator sequence $\{M^j g\}_{j \geq 0}$ is pairwise linearly independent, where $g = \begin{bmatrix} a \\ 1 \\ a \end{bmatrix}$ is a generator $F$-gate having a single vertex.

Proof. Suppose the real part of $a$ is nonzero, and consider gadget $M_1$,

$$M_1 = \begin{bmatrix} a^3 + a^2 & a^2 + a \\ a^2 + a & a + a^2 \end{bmatrix}. $$

The eigenvalues of $M_1$ are $a(a+1)^2$ and $a(a^2-1)$, both of which are nonzero. Their ratio is $\frac{a+1}{a-1}$, which is not a root of unity, as $|a+1| \neq |a-1|$. Note that $|a+1| = |a-1|$ consists of exactly those values $a$ on the imaginary line $i\mathbb{R}$.

Now assume the real part of $a$ is zero, and consider gadget $M_2$,

$$M_2 = \begin{bmatrix} a^3 + 1 & a^2 + a \\ a^2 + a & a^3 + 1 \end{bmatrix}. $$

The eigenvalues are $(a+1)(a-1)^2$ and $(a+1)(a^2+1)$, both nonzero. The ratio of these is $\frac{a^2-2a+1}{a^2+1}$. Since $a$ is purely imaginary, $a^2+1$ is real-valued. Since $a$ is furthermore nonzero and belong to $i\mathbb{R}$, $|a^2+1-2a| > |a^2+1|$ and thus $\frac{a^2-2a+1}{a^2+1}$ is not a root of unity.

140
By Lemma 5.11, we have a unary recursive gadget $M$ for which $\{M^j\}_{j \geq 0}$ is a sequence of pairwise linearly independent signatures. Pairwise linear independence is preserved under invertible linear transformations defined by $g$ on the set $\{M^j\}_{j \geq 0}$. As $M^j \mapsto M^j g$ is an invertible linear transformation, we conclude that $\{M^j g\}_{j \geq 0}$ is also pairwise linearly independent.

Exercise: Compute the signature matrices of Gadgtes $M_1$ and $M_2$ for edge function $[a, 1, b]$. Note that $M_1 - M_2 = (ab - 1)I$. Setting $a = b$ as in Lemma 5.12, verify that the signature matrices are as stated.

Interpolation and reduction

Suppose we have a sequence of pairwise linear independent $\{[x_j, y_j, x_j]\}_{j \geq 0}$ of generators, we show how to simulate any generator of the form $[x, y, x]$.

Lemma 5.13. Let $k \geq 3$ be fixed and suppose that in the context of $\text{Holant}(a, b)$ we can efficiently construct a sequence of pairwise linearly independent generators of the form $\{[x_j, y_j, x_j]\}_{j \geq 0}$. Then for any $x, y \in \mathbb{C}$ we can efficiently simulate generator $[x, y, x]$.

Proof. Let $\Omega$ be a bipartite signature grid for the problem $\text{Holant}(\mathcal{R} \ | \ \mathcal{G} \cup \{[x, y, x]\})$, and suppose the generator $[x, y, x]$ occurs $n$ times in $\Omega$. We stratify the Holant sum according to how many times the generator $[x, y, x]$ evaluates to $x$. Collecting these terms together, we write $\text{Holant}_\Omega = \sum_{j=0}^{n} c_j x^j y^{n-j}$, where $c_j$ is the sum over all edge assignments of the product of signature evaluations other than the $n$ copies of $[x, y, x]$, subject to the restriction that exactly $j$ of the edge-pairs connecting to $[x, y, x]$ are assigned both $0$s or both $1$s.

We construct a sequence of bipartite signature grids $\Omega_i$ for the problem $\text{Holant}(\mathcal{R} \ | \ \mathcal{G})$, by replacing each occurrence of $[x, y, x]$ with a copy of an $\mathcal{F}$-gate with signature $[x_i, y_i, x_i]$, for $i = 0, 1, \ldots, n$. (By omitting at most one element from our sequence and relabeling, we assume that $y_i \neq 0$ for all $i$.) Note that the bipartite structure is maintained, and the Holant of $\Omega_i$ evaluates to $\sum_{j=0}^{n} c_i x_i^j y_i^{n-j}$, where the same $c_j$'s appear in this sum, for all $i$. By oracle queries to $\text{Holant}(\mathcal{R} \ | \ \mathcal{G})$, we obtain a Vandermonde system where $c_j$ are the unknowns.

$$\begin{bmatrix}
    y_0^n \cdot \text{Holant}_{\Omega_0} \\
    y_1^n \cdot \text{Holant}_{\Omega_1} \\
    \vdots \\
    y_n^n \cdot \text{Holant}_{\Omega_n}
\end{bmatrix} =
\begin{bmatrix}
    x_0^0 y_0^0 & x_0^1 y_0^{-1} & \cdots & x_0^n y_0^{-n} \\
    x_1^0 y_1^0 & x_1^1 y_1^{-1} & \cdots & x_1^n y_1^{-n} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_n^0 y_n^0 & x_n^1 y_n^{-1} & \cdots & x_n^n y_n^{-n}
\end{bmatrix}
\begin{bmatrix}
    c_0 \\
    c_1 \\
    \vdots \\
    c_n
\end{bmatrix}.$$ 

Note that this Vandermonde system has full rank iff $[x_j, y_j, x_j]$ are pairwise linearly independent, namely $x_j/y_j$ are pairwise distinct. Solving this linear system for $c_j$ $(0 \leq j \leq n)$ yields the answer to $\text{Holant}_\Omega = \sum_{j=0}^{n} c_j x^j y^{n-j}$. 

141
Now we will use the interpolation result to simulate the signature \([1,0,1]\), in order to get a reduction from \(\text{Holant}(=,=,=,\ldots | [a,1,a])\) to prove \(\#P\)-hardness.

**Lemma 5.14.** Suppose the parameters \((a,b)\) are in Region II, equivalently, suppose \(X,Y \in \mathbb{C}, a^3 = b^3\), and \(X \notin \{0,1,-1\}\). Then \(\text{Hol}_3(a,b)\) is \(\#P\)-hard.

**Proof.** By Lemmas 5.7 and 5.4 we may assume that \(a = b\). Since \(X \notin \{0, \pm 1\}\) and \(a = b\) we know \(a^4 = (ab)^2 \notin \{0,1\}\). We have a sequence of pairwise linearly independent generator signatures from Lemma 5.12, and by Corollary 5.9 they all have the form \([x,y,x]\). Therefore by Lemma 5.13 we can efficiently simulate generator signature \([1,0,1]\). We also know from the dichotomy theorem for Boolean \(#CSP\) problems, Theorem 3.7, that \(#CSP([a,1,a])\) is \(\#P\)-hard. We have a reduction from \(#CSP([a,1,a])\), which is by definition \(\text{Holant}(=,=,=,\ldots | [a,1,a])\). We have a reduction from \(\text{Holant}(=,=,=,\ldots | [a,1,a])\), which is by definition \(\text{Holant}([1,0,1])\).

In the first step, we can use \(=,=,=,\ldots\) on the right-hand-side and \(=,=,\ldots\) on the left-hand-side to simulate \(=,=\), for any \(k\). The second step is by interpolation. \(\square\)

**Exercise:** For any \(k \geq 1\), construct a gadget that has signature \(=,=\) on the LHS, using \(=,=,=,\ldots\) on the RHS and \(=,=,\ldots\) on the LHS. For what values of \(d\) and \(d'\) in place of 2 and 3 can one construct \(=,=\) for all \(k \geq 1\)?

### 5.1.3 Hardness in Region III where \((X,Y) \notin \mathbb{R} \times \mathbb{R}\)

Recall that Region III is the complement of the union of Regions I and II, which is specified by \(X \neq 1\) and \(Y^2 \neq 4X^3\) (equivalently \(a^3 \neq b^3\)) and \((X,Y) \neq (-1,0)\).

So far, we have seen one complete iteration of the general strategy we use in this chapter. We devised a scheme to produce an infinite set of pairwise linearly independent (virtual) unary signatures, found some gadgets to implement that scheme, used interpolation to simulate any such (virtual) unary signature, and applied that result to complete a reduction from a known \(\#P\)-hard problem to prove \(\#P\)-hardness. We will continue to develop the theory to make our tools applicable to wider classes of problems. Specifically, in this subsection we will prove \(\#P\)-hardness for \(\text{Hol}_3(a,b)\), subject to the restrictions that \((X,Y) \notin \mathbb{R} \times \mathbb{R}, a^3 \neq b^3, \) and \(X \neq 1\). To prove \(\#P\)-hardness for Region III, we must use planar gadgets and use an initial problem to reduce from that is \(\#P\)-hard even for planar graphs.

**Unary recursive construction with starter gadgets**

As before, we first seek to produce a sequence of pairwise linearly independent signatures. We expand on the unary recursive construction by introducing a *starter gadget*. The following
Lemma 5.15. Let $M \in \mathbb{C}^{2 \times 2}$ be a nonsingular matrix with infinite projective order, and let $s \in \mathbb{C}^2$ be a vector which is not orthogonal to any row eigenvector of $M$, i.e., the dot product with any row eigenvector of $M$ is nonzero. Then $\{M^j s\}_{j \geq 0}$ is a sequence of pairwise linearly independent vectors in $\mathbb{C}^2$.

Proof. Suppose that $M$ is diagonalizable and we write $M = T^{-1}JT$ where $J = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$ and the rows of $T$ are the row eigenvectors of $M$. As $M$ is nonsingular, we have $\alpha \beta \neq 0$. We know by assumption that $Ts = \begin{bmatrix} x \\ y \end{bmatrix}$ for some $x, y \in \mathbb{C}$ where $xy \neq 0$. Then $M^j s = T^{-1} \begin{bmatrix} \alpha^j & 0 \\ 0 & \beta^j \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = T^{-1} \begin{bmatrix} x y^j \\ 0 \end{bmatrix}$. The vectors $\begin{bmatrix} \alpha^j \\ \beta^j \end{bmatrix}$ are pairwise linearly independent since $\frac{\alpha^j}{\beta^j}$ is not a root of unity, and this property is preserved by the invertible linear transformation $T^{-1} \begin{bmatrix} x \\ 0 \end{bmatrix}$, so $\{M^j s\}_{j \geq 0}$ is a sequence of pairwise linearly independent vectors.

Now suppose that $M$ is not diagonalizable. Then we write the Jordan Normal Form of $M$ as $M = T^{-1}JT$ where $J = \begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix}$ and the first row of $T$ is a row eigenvector corresponding to $\alpha$. We know by assumption that $Ts = \begin{bmatrix} x \\ y \end{bmatrix}$ for some $x, y \in \mathbb{C}$ where $x \neq 0$. Then $M^j s = T^{-1} J^j Ts = \alpha^j T^{-1} \begin{bmatrix} 0 & 1 \\ 1 & \alpha^{-1} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \alpha^j T^{-1} v_j$, where the vectors $v_j = \begin{bmatrix} x \\ \alpha^{-1} x + y \end{bmatrix}$ are pairwise linearly independent. This uses the fact that $x \neq 0$. This pairwise linear independence is preserved by the invertible linear transformation $T^{-1}$, thus after a scalar multiple $\alpha^j$, the vectors $\alpha^j T^{-1} v_j$ are also pairwise linearly independent. $\square$

Corollary 5.16. Let $M$ be a unary recursive gadget with infinite projective order, and let $s$ be a unary starter gadget whose signature is not orthogonal to any row eigenvector of $M$. Then the sequence $\{M^j s\}_{j \geq 0}$ has pairwise linearly independent generator signatures.

Proof. Immediate. $\square$

Surprisingly, a set of general-purpose starter gadgets can be made for this construction as long as $ab \neq 1$ and $a^3 \neq b^3$, so we refine this corollary. It is easy to see that $ab = 1$ is a tractable case, as $[a, 1, b] = [a, 1, a^{-1}]$ is degenerate. Recall that $a^3 = b^3$ essentially corresponds to Region II.

Theorem 5.17. In the context of Hol$_3(a, b)$ let $a, b \in \mathbb{C}$ such that $ab \neq 1$ and $a^3 \neq b^3$, and let $M$ be a unary recursive gadget with infinite projective order. Then there exists a starter gadget $s$ (also in the context of Hol$_3(a, b)$) such that $\{M^j s\}_{j \geq 0}$ is a sequence of pairwise linearly independent unary generator signatures.

Proof. Let $M_3, M_4, M_5$, and $M_6$ be defined as in Figure 5.3. We calculate their signature
Note that these matrices operate directly on symmetric signatures, e.g., the middle column of $M_5$ correspond to the sum of the assignments of 01 and 10 on the trailing edges.

The $2 \times 2$ matrices $[M_4M_5v \ M_4v]$ and $[M_4M_6v \ M_4v]$ both have determinant $(a^3 - b^3)(ab - 1)^2$, whereas $\det([M_4M_5v \ M_4M_6v]) = (a^3 - b^3)(ab - 1)^3$, so the vectors $M_4v$, $M_4M_5v$, and $M_4M_6v$ are pairwise linearly independent provided that $ab \neq 1$ and $a^3 \neq b^3$. The set of vectors that are orthogonal by dot product to at least one row eigenvector of the given matrix $M \in \mathbb{C}^{2 \times 2}$ can be expressed as a union of at most two 1-dimensional linear subspaces. By the Pigeonhole Principle, this union cannot contain 3 pairwise linearly independent vectors. Hence at least one element of $\{M_4M_5v, M_4M_6v, M_4v\}$ is not orthogonal to any row eigenvector of $M$. The corresponding starter gadget can be used with $M$ in a recursive construction and we get a sequence of pairwise linearly independent unary generator signatures by Corollary 5.16.

**Exercise:** Verify the signature matrices for gadgets $M_3$, $M_4$, $M_5$ and $M_6$ are as stated. Notice that these matrices are in a form that can be directly applied to symmetric signatures. Give a simple explanation that the middle columns of $M_4$ and $M_6$ are both 0.

We make two remarks:

1. For a nonzero vector $s \in \mathbb{C}^2$, and a nonsingular matrix $M \in \mathbb{C}^{2 \times 2}$, the condition that $s$ is not a column eigenvector of $M$ is equivalent to $\det[s \ M s] \neq 0$, which is also equivalent to the condition that $s$ is not orthogonal by dot product to any row eigenvector of $M$.

2. A sufficient condition for a nonsingular matrix $M \in \mathbb{C}^{2 \times 2}$ to have infinite projective order is that the ratio of the eigenvalues is not a root of unity, which in turn is implied by the two eigenvalues having different norms.
Interpolation with starter gadgets

Since Theorem 5.17 already produced a sequence of unary generators (provided a suitable unary recursive gadget is given), there is no need to project it down to a lower dimension. Thus, the plan is to show that we can interpolate any unary generator, along similar lines as Lemma 5.13. However, we will prove it for recognizer signatures of the form \([x, 0, y]\) instead, since this allows us to reuse this lemma later on. Note that we can easily convert from generators of the form \([x, y]\) to recognizers of the form \([x, 0, y]\) by attaching a single vertex with recognizer \(=3\).

**Lemma 5.18.** Let \(k \geq 3\) be fixed and suppose that in the context of \(\text{Holant}_k(a, b)\), we can efficiently construct a sequence of pairwise linearly independent recognizers of the form \([x_j, 0, y_j]\) for \(j \geq 0\). Then for any \(x, y \in \mathbb{C}\) we can efficiently simulate recognizer \([x, 0, y]\).

**Proof.** By ignoring at most one signature in the sequence \([x_j, 0, y_j]\) and relabeling the indices, we assume that each \(y_j\) is nonzero. Let \(\Omega\) be a bipartite signature grid for the problem \(\text{Holant}(\mathcal{R} \cup \{[x, 0, y]\} | \mathcal{G})\). Suppose recognizer \([x, 0, y]\) occurs \(n\) times in \(\Omega\). Dropping terms that evaluate to zero, we stratify the Holant sum according to how many times the signature \([x, 0, y]\) evaluates to \(x\). Collecting these terms together, we write \(\text{Holant}_\Omega = \sum_{j=0}^{n} c_j x^j y^{n-j}\), where \(c_j\) is the sum over all edge assignments of the product of signature evaluations other than the \(n\) copies of \([x, 0, y]\), subject to the restriction that exactly \(j\) of the edge-pairs connecting to \([x, 0, y]\) are both assigned 0, and exactly \(n-j\) of the edge-pairs connecting to \([x, 0, y]\) are both assigned 1.

We construct a sequence of bipartite signature grids \(\Omega_i\) for the Holant problem \(\text{Holant}(\mathcal{R} | \mathcal{G})\), by replacing each occurrence of \([x, 0, y]\) with a copy of an \(\mathcal{F}\)-gate with signature \([x_i, 0, y_i]\), for \(i = 0, 1, \ldots, n\). This has the effect of replacing each occurrence of \([x, 0, y]\) by pairwise linearly independent signatures. Note that the bipartite structure is maintained. By oracle queries to \(\text{Holant}(\mathcal{R} | \mathcal{G})\), we obtain a nonsingular Vandermonde system where \(c_j\) are the unknowns. Solving this linear system for \(c_j\) \((0 \leq j \leq n)\) yields the answer to \(\text{Holant}_\Omega = \sum_{j=0}^{n} c_j x^j y^{n-j}\). \(\Box\)

When we apply Theorem 5.17, we will show that the eigenvalues of \(M\) have unequal norm. The following lemma gives a concrete way to test this, by examining the characteristic polynomial of \(M\).

**Lemma 5.19.** If the roots of a polynomial \(x^2 - Bx + C \in \mathbb{C}[x]\) have the same norm, then \(B|C| = \overline{BC}\) and \(B^2\overline{C} = \overline{B^2C}\). If further \(B \neq 0\) and \(C \neq 0\), then \(\text{Arg}(B^2) = \text{Arg}(C)\).

**Proof.** If the roots have equal norm, then for some \(a, b \in \mathbb{C}\) with \(|a| = |b| = 1\), and nonnegative \(r \in \mathbb{R}\), we can write \(x^2 - Bx + C = (x-ra)(x-rb)\), so \(B|C| = r(a+b)r^2 = r(a^{-1} + b^{-1})r^2ab = \overline{BC}\). Squaring both sides and dividing by \(C\), we have \(B^2\overline{C} = \overline{B^2C}\) (note that this equality still holds when \(C = 0\)). Multiplying \(B|C| = \overline{BC}\) by \(B\) we get \(B^2|C| = |B^2C|\), and if \(B\) and \(C\) are both nonzero then \(\frac{B^2}{|B^2|} = \frac{|C|}{|C|}\), that is, \(\text{Arg}(B^2) = \text{Arg}(C)\). \(\Box\)
Eigenvalue Shifted Pairs and Other Gadgetry

Given any fixed $a, b \in \mathbb{C}$ for which $\text{Hol}_3(a, b)$ happens to be $\#P$-hard, it is usually a straightforward process of trial and error of using Lemma 5.19 to find a unary recursive gadget with infinite projective order, so as to satisfy the conditions of Theorem 5.17. This would be fine if we were only aiming to prove hardness of a small finite number of problems. However, given some particular unary recursive gadget, there will inevitably be some settings of $a$ and $b$ which do not work. It is natural to add more gadgets to the mix, but the coefficients of the characteristic polynomial of $M$ are themselves polynomials in $a$ and $b$. Also, the degrees of these polynomials can get rather high and somewhat cumbersome (depending on the size of the gadget $M$). Algebraic symmetrization can help by reducing the complexity of the polynomials involved, but then we are still left with the question: given two polynomials $p(X, Y)$ and $q(X, Y)$, under which settings of $X, Y \in \mathbb{C}$ is it the case that $\text{Arg}(p(X, Y)) = \text{Arg}(q(X, Y))$? Moreover, given many such pairs of polynomials, how can we be assured that all relevant $X$ and $Y$ are “covered” by one or more gadgets?

We now introduce a technique called Eigenvalue Shifted Pairs, or ESP for short, to address this problem. Suppose we have unary recursive gadgets $M$ and $M'$, and suppose the eigenvalues of $M$ are those of $M'$ shifted by some complex number $\delta$. Then unless $\delta$ “lines up just right”, at least one of the two matrices will have eigenvalues with unequal norm, which is just what we need.

**Definition 5.20.** A pair of nonsingular square matrices $M$ and $M'$ is called an Eigenvalue Shifted Pair (ESP) if $M' = M + \delta I$ for some nonzero $\delta \in \mathbb{C}$, and $M$ has distinct eigenvalues.

Clearly for such a pair, $M'$ also has distinct eigenvalues. For example, the transition matrices of gadgets $M_1$ and $M_2$ (Figure 5.4) differ only by $ab - 1$ along the main diagonal, and form an Eigenvalue Shifted Pair for all $a, b \in \mathbb{C}$, except when $ab = 1$ (these are the same gadgets we saw earlier in Figure 5.2). Before we make use of this ESP, we state a technical lemma so we can describe the conditions under which an ESP doesn’t work.

**Lemma 5.21.** Suppose $\alpha, \beta, \delta \in \mathbb{C}$, $|\alpha| = |\beta|$, $\alpha \neq \beta$, $\delta \neq 0$, and $|\alpha + \delta| = |\beta + \delta|$. Then there exist $r, s \in \mathbb{R}$ such that $r \delta = \alpha + \beta$ and $s \delta^2 = \alpha \beta$.

**Proof.** After a rotation in the complex plane, namely multiplying $\alpha, \beta, \delta$ by some $e^{i\theta}$, we can assume $\alpha = \bar{\beta}$, and then since $\alpha + \beta, \alpha \beta \in \mathbb{R}$ we just need to prove $\delta \in \mathbb{R}$. Then $(\alpha + \delta)(\alpha + \bar{\delta}) = |\alpha + \delta|^2 = |\beta + \bar{\delta}|^2 = (\beta + \delta)(\bar{\beta} + \bar{\delta}) = (\bar{\alpha} + \delta)(\bar{\alpha} + \bar{\delta})$ and we distribute to get $\alpha \bar{\alpha} + \delta \bar{\delta} + \alpha \bar{\delta} + \bar{\alpha} \delta = \bar{\alpha} \bar{\delta} + \delta \bar{\delta} + \alpha \delta + \bar{\alpha} \delta$. Canceling repeated terms and factoring, we have $(\bar{\alpha} - \alpha)(\bar{\delta} - \delta) = 0$, and since $\alpha \neq \beta = \bar{\alpha}$ we get $\bar{\delta} = \delta$ therefore $\delta \in \mathbb{R}$.

**Corollary 5.22.** Let $M$ and $M'$ be an Eigenvalue Shifted Pair of 2 by 2 matrices. If $M$ and $M'$ have the property that both eigenvalues have the same norm, then $\text{tr}(M)/\delta \in \mathbb{R}$ and $\text{det}(M)/\delta^2 \in \mathbb{R}$. 
Proof. Let \( \alpha \) and \( \beta \) be the eigenvalues of \( M \), so \( \alpha + \delta \) and \( \beta + \delta \) are the eigenvalues of \( M' \). Suppose that \( |\alpha| = |\beta| \) and \( |\alpha + \delta| = |\beta + \delta| \). By Definition 5.20, \( \alpha \neq \beta \) and \( \delta \neq 0 \). Then by Lemma 5.21, there exist \( r \) and \( s \in \mathbb{R} \) such that \( \text{tr}(M) = \alpha + \beta = r\delta \) and \( \text{det}(M) = \alpha \beta = s\delta^2 \), hence \( \text{tr}(M)/\delta \in \mathbb{R} \) and \( \text{det}(M)/\delta^2 \in \mathbb{R} \).

We now apply an ESP to prove that for most settings \((a, b)\), the problem \( \text{Hol}_3(a, b) \) is \#P-hard.

Lemma 5.23. Suppose \( X \) and \( Y \) are not both real, \( X \neq \pm 1 \), \( X^2 + X + Y \neq 0 \), and \( 4(X - 1)^2(X + 1) \neq (Y + 2)^2 \). Then either gadget \( M_1 \) or gadget \( M_2 \) has nonzero eigenvalues with distinct norm.

Proof. The transition matrices are \( M_1 = \begin{bmatrix} a^2 + ab & a + b^2 \\ a^2 + b & ab + b^2 \end{bmatrix} \) and \( M_2 = \begin{bmatrix} a^2 + 1 & a + b^2 \\ a^2 + b & b^4 + 1 \end{bmatrix} \), so \( M_1 = M_2 + (X - 1)I \), and the eigenvalue shift is nonzero. \( \text{tr}(M_2) = Y + 2 \). Checking the determinants, \( \text{det}(M_2) = (X - 1)^2(X + 1) \neq 0 \) and \( \text{det}(M_1) = (X - 1)(X^2 + X + Y) \neq 0 \). Also, \( \text{tr}(M_2)^2 - 4\text{det}(M_2) = (Y + 2)^2 - 4(X - 1)^2(X + 1) \neq 0 \), so the eigenvalues of \( M_2 \) are distinct. Therefore by Corollary 5.22, either \( M_1 \) or \( M_2 \) has nonzero eigenvalues of distinct norm unless

\[
Y + 2 = \text{tr}(M_2) = r(X - 1) \quad \text{and} \quad (X - 1)^2(X + 1) = \text{det}(M_2) = s(X - 1)^2
\]

for some \( r, s \in \mathbb{R} \). Then we would have \( X = s - 1 \in \mathbb{R} \) and \( Y = r(X - 1) - 2 \in \mathbb{R} \). 

The case where \( X \in \mathbb{R} \) and \( Y \in \mathbb{R} \) will require different techniques, so we put it aside for now and focus on the remaining three cases in this subsection where \( X \) and \( Y \) are not both real, and the following conditions hold. Note that \( X = 1 \) is tractable by Theorem 5.6 and we exclude that.

1. \( X = -1 \) or
2. \( X^2 + X + Y = 0 \) or
3. \( 4(X - 1)^2(X + 1) = (Y + 2)^2 \)

147
Figure 5.5: A unary recursive gadget

Having a specific condition such as \( X = -1 \) or \( X^2 + X + Y = 0 \) introduces fortuitous cancellations, making it possible to use individual gadgets rather than ESPs. The following gadgets were selected by a trial and error process, using a computer to assist with the search.

Gadgets \( M_1 \) and \( M_7 \) can be used to deal with the \( X = -1 \) case. Recall that any setting of \( a \) and \( b \) such that \( X = -1 \) and \( Y = \pm 2i \) is tractable by Theorem 5.6.

**Lemma 5.24.** If \( X = -1, Y \neq \pm 2i, \) and \( Y \notin \mathbb{R}, \) then either gadget \( M_1 \) or gadget \( M_7 \) has a transition matrix with nonzero eigenvalues with distinct norm.

**Proof.** Suppose \( |Y| \neq 2, Y \notin \mathbb{R}. \) Under the assumption \( X = -1, \) \( \det(M_1) = -2Y \neq 0 \) and \( \text{tr}(M_1) = Y - 2, \) so \( \text{tr}(M_1) \cdot |\det(M_1)| - \text{tr}(M_1) \cdot \det(M_1) = (Y - 2) \cdot |2Y| + (\overline{Y} - 2)(2Y) = 2Y \cdot |Y| - 4|Y| + 2|Y|^2 - 4Y = 2(|Y| - 2)(|Y| + Y) \neq 0. \) Thus \( \text{tr}(M_1) \cdot |\det(M_1)| \neq \text{tr}(M_1) \cdot |\det(M_1)| \) and by Lemma 5.19, \( M_1 \) has (nonzero) eigenvalues with distinct norm.

Now suppose \( |Y| = 2, \) but \( Y \neq \pm 2i \) and \( Y \notin \mathbb{R}. \)

\[
M_7 = \begin{bmatrix}
a^6 + 3a^3 + 3ab + b^3 & a^4 + 2a^2b + ab^2 + a + b^5 + 2b^2 \\
a^5 + a^3b + 2a^2 + 2ab^2 + b^4 + b & a^3 + 3ab + b^6 + 3b^3
\end{bmatrix}.
\]

Under the assumption \( X = -1, \) we have \( \det(M_7) = -16Y \neq 0. \) Substituting \( \overline{Y} = 4/Y, \)

\[
\text{tr}(M_7)^2 \det(M_7) - \text{tr}(M_7)^2 \det(M_7) = -16(Y - \overline{Y}) \cdot (-16 + 8Y \overline{Y} + 8Y^2 \overline{Y} + Y^3 \overline{Y} + 8Y \overline{Y}^2 + Y^2 \overline{Y}^2 + Y \overline{Y}^3).
\]

\[
= -64(Y - \overline{Y})(Y^2 + 4)(Y^2 + 8Y + 4)\
\]

\[
\neq 0.
\]

Note that \( |Y^2 + 4| \leq 8 < |8Y| = 16. \) Hence \( \text{tr}(M_7)^2 \det(M_7) \neq \text{tr}(M_7)^2 \det(M_7) \) and the eigenvalues of \( M_7 \) (which are nonzero) have distinct norm by Lemma 5.19.

Note that if \( X^2 + X + Y = 0 \) then \( X \in \mathbb{R} \) implies \( Y \in \mathbb{R}. \) So in the following lemma, the assumption that \( X \) and \( Y \) are not both real numbers amounts to \( X \notin \mathbb{R}. \)

**Lemma 5.25.** If \( X^2 + X + Y = 0 \) and \( X \notin \mathbb{R} \) then the transition matrix of gadget \( M_8 \) has nonzero eigenvalues with distinct norm.
Proof.  
\[
M_8 = \begin{bmatrix}
    a^6 + 2a^4b + a^3 + 3a^2b^2 + ab^4 & a^4 + 3a^2b + 2ab^3 + b^5 + b^2 \\
    a^5 + 2a^3b + a^2 + 3ab^2 + b^4 & a^4b + 3a^2b^2 + 2ab^4 + b^6 + b^3
\end{bmatrix}
\]

Then the determinant is the polynomial
\[
X^6 - 6X^5 - X^4Y + 16X^4 + 11X^3Y - 10X^3 + 5X^2Y^2 - 7X^2Y - X^2 + XY^3 - 4XY^2 - 3XY - Y^3 - Y^2.
\]

Amazingly, with the condition \(X^2 + X + Y = 0\), this polynomial factors into
\[
-X^2(X - 1)^5.
\]

**Exercise:** Verify that the displayed polynomial is congruent to \(-X^2(X-1)^5 \mod X^2+X+Y\).

Similarly, the trace, which is \(-2X^3 + 6X^2 + 3XY + Y^2 + Y\), also factors into \(X(X - 1)^3\). Since \(X \not\in \mathbb{R}\), we have \(\det(M_8) \neq 0\), \(\text{tr}(M_8) \neq 0\), and \((1 - X) \det(M_8) = \text{tr}(M_8)^2\). Therefore \(\text{Arg}(\det(M_8)) \neq \text{Arg}(\text{tr}(M_8)^2)\) for otherwise \(X\) is real, and we conclude by Lemma 5.19 that the eigenvalues of \(M_8\) (which are nonzero) have distinct norm.

The condition \(4(X - 1)^2(X + 1) = (Y + 2)^2\) is somewhat resilient to individual unary recursive gadgets, so we take advantage of another interesting coincidence: two gadgets with transition matrices that have identical trace. In the following lemma, we may assume \(X^2 + X + Y \neq 0\) due to Lemma 5.25. We will also assume \(X \not\in \mathbb{R}\) in this lemma for technical reasons. The case of \(X \in \mathbb{R}\) and \(Y \not\in \mathbb{R}\) will be easy to clean up afterwards.
Lemma 5.26. Suppose $X \not\in \mathbb{R}$, and further assume that $X^2 + X + Y \neq 0$ and $4(X - 1)^2(X + 1) = (Y + 2)^2$. Then the transition matrix of either unary recursive gadget $M_9$ or unary recursive gadget $M_{10}$ has nonzero eigenvalues with distinct norm.

Proof. The transition matrices are

$$M_9 = \begin{bmatrix} a^6 + 4a^4b + 2a^2b^2 + 2ab + b^3 & a^4 + 3a^2b + 2ab^2 + b^3 \\ a^5 + 2a^3b + a^2 + 3ab^2 + b^3 & a^3 + 2b^2 + 2ab + b^3 + b \end{bmatrix},$$

$$M_{10} = \begin{bmatrix} a^6 + 4a^4b + 2a^2b^2 + 2ab + b^3 & a^4 + a^3b^2 + a^2b + 2ab^3 + a + b^5 + b^2 \\ a^5 + 2a^3b + a^2 + 3ab^2 + b^3 & a^3 + a^2b^2 + 2ab + b^3 \end{bmatrix}.$$ 

Let $R$ denote $(Y + 2)^2 - 4(X - 1)^2(X + 1)$. The main diagonals of $M_9$ and $M_{10}$ are identical, so $\text{tr}(M_9) = \text{tr}(M_{10})$. Furthermore, $\text{tr}(M_9) = \text{tr}(M_{10}) - R = (X - 1)(2X^2 + Y)$. Clearly $X - 1 \neq 0$, since $X \not\in \mathbb{R}$. If $2X^2 + Y = 0$, then substituting $Y = -2X^2$ into $4(X - 1)^2(X + 1) = (Y + 2)^2$ would yield $X(X - 1)^2(X + 1) = 0$, again impossible by $X \not\in \mathbb{R}$. So $\text{tr}(M_9) \neq 0$. Next, $\det(M_{10}) = (X - 1)^3(X + 1)(X^2 + X + Y)$ and $\det(M_9) = \det(M_9) - R(X - 1)^2 = (X - 1)^3(X + 4)(X^2 + X + Y)$, so these are both nonzero. If both $M_9$ and $M_{10}$ have eigenvalues with equal norm, then applying Lemma 5.19 twice, $\text{Arg}(|\text{det}(M_9)|) = \text{Arg}(\text{tr}(M_9)^2) = \text{Arg}(\text{tr}(M_{10})^2) = \text{Arg}(\text{det}(M_{10}))$. However, this would imply $\text{Arg}(X + 4) = \text{Arg}(X + 1)$ and $X \in \mathbb{R}$, so we conclude that either $M_9$ or $M_{10}$ has nonzero eigenvalues with distinct norm.

We have already seen a gadget that can handle the remaining case.

Lemma 5.27. Suppose $X \in \mathbb{R}$ and $Y \not\in \mathbb{R}$, and further assume that $4(X - 1)^2(X + 1) = (Y + 2)^2$. Then the transition matrix of unary recursive gadget $M_7$ has nonzero eigenvalues with distinct norm.

Proof. Note that $X < -1$, lest $Y$ be real valued. Let $R$ denote $(Y + 2)^2 - 4(X - 1)^2(X + 1)$ as before. We find that $\text{tr}(M_7) = \text{tr}(M_7) - R = 2X(X - 1)^2 \neq 0$. We calculate that $\det(M_7) = (X - 1)^3(X + 2X^2 + X^3 + 2Y)$, and we claim this is also nonzero. Otherwise, we would have $Y = -(X + 2X^2 + X^3)/2$, and from $X \in \mathbb{R}$ we would have $Y \in \mathbb{R}$, a contradiction. Now $\text{tr}^2(M_7) = 4X^2(X - 1)^4 \in \mathbb{R}$, so $\text{tr}^2(M_7)/\det(M_7) \not\in \mathbb{R}$ follows from $X \in \mathbb{R}$ and $Y \not\in \mathbb{R}$. We conclude by Lemma 5.19 that $M_7$ has nonzero eigenvalues with distinct norm.

Now we sum up the results of the previous lemmas. Recall that if $X = 1$ or $(X, Y) = (-1, \pm 2i)$ then $\text{Hol}_3(a, b)$ is tractable.

Lemma 5.28. Suppose that it is not the case that both $X$ and $Y$ are real valued. Also assume that $X \neq 1$ and $(X, Y) \neq (-1, \pm 2i)$. Then in the context of $\text{Hol}_3(a, b)$ there exists a unary recursive gadget $M$ for which the eigenvalues are both nonzero and have distinct norm.

Proof. We know by Lemma 5.23 that either gadget $M_1$ or $M_2$ has a transition matrix with nonzero eigenvalues of distinct norm, except in the following cases, where we will use other gadgets to fill this requirement.
1. \( X = -1, \)
2. \( X^2 + X + Y = 0, \) or
3. \( 4(X - 1)^2(X + 1) = (Y + 2)^2. \)

First suppose \( X = -1. \) Then \( Y \) is not real valued and \( Y \neq \pm 2i, \) so we get a transition matrix of the required form by Lemma 5.24. If \( X^2 + X + Y = 0 \) we have \( X \notin \mathbb{R} \) (otherwise we would have \( X, Y \in \mathbb{R} \)), and we have the requirement satisfied by Lemma 5.25. Now we may assume \( X^2 + X + Y \neq 0 \) and suppose \( 4(X - 1)^2(X + 1) = (Y + 2)^2. \) If \( X \notin \mathbb{R} \) then we are done by Lemma 5.26, otherwise \( X \in \mathbb{R} \) but \( Y \notin \mathbb{R} \) so we are done by Lemma 5.27.

**Corollary 5.29.** Suppose \( a, b \in \mathbb{C} \) are such that it is not the case that both \( X \) and \( Y \) are real valued. Also assume \( X \neq 1 \) and \( 4X^3 \neq Y^2 \) (equivalently, \( a^3 \neq b^3 \)). Then in the context of \( \text{Hol}_3(a, b) \) we can efficiently simulate any recognizer signature of the form \([x, 0, y]\).

**Proof.** For any such \( a, b \in \mathbb{C} \), we have a unary recursive gadget \( M \) by Lemma 5.28 whose transition matrix has nonzero eigenvalues of distinct norm. Then \( M \) has infinite projective order, and by Theorem 5.17, we can efficiently construct a sequence of pairwise linearly independent unary generators \([x_j, y_j]\). By attaching an additional vertex labeled with recognizer \( =_3 \), these become pairwise linearly independent signatures of the form \([x_j, 0, y_j]\). Therefore by Lemma 5.18 we can efficiently simulate any recognizer of the form \([x, 0, y]\).

**Applying unary signatures to prove \#P-hardness**

The ability to simulate unary signatures (or “virtual unary” signatures) will give us a reduction from \( \text{Vertex Cover} \). Note that counting \( \text{Vertex Cover} \) on \( k \)-regular graphs is just \( \text{Holant}(\mathbb{C} \mathbb{E}_k) [0, 1, 1] \). Counting \( \text{Vertex Cover} \) is \#P-hard even when the input is restricted to \( k \)-regular planar graphs. We remark that for \( k \geq 6 \), any \( k \)-regular planar graph must have self-loops or multiple edges. This follows from Euler’s formula \( V - E + F = 2 \).

**Lemma 5.30.** Counting \( \text{Vertex Covers} \) on \( k \)-regular planar multigraphs (with self-loops and parallel edges) is \#P-hard for \( k \geq 3 \).

**Proof.** We will give a reduction from counting \( \text{Vertex Covers} \) on 3-regular planar graphs, which is known to be \#P-hard \([?]\). If \( k \) is odd, we can easily reduce from the given 3-regular planar graph \( G = (V, E) \) to a \( k \)-regular planar graph by using a gadget consisting of a single vertex incident with \((k - 1)/2 \) self-loops and one dangling edge. Connect \( k - 3 \) copies of this gadget to every vertex \( v \in V \). Due to the self-loops, every copy of this gadget must have its vertex included in the vertex cover, but the graph induced by removing those vertices and their incident edges is identical to \( G \), so the number of vertex covers is the same.

Now we reduce to a \( k \)-regular planar graph where \( k \) is even. The gadget is similar, consisting of a single vertex incident with \((k - 2)/2 \) self-loops and two dangling edges. Being
3-regular, every connected component of G has an even number of vertices, so considering
the vertices in pairs and applying \( k - 3 \) copies of this gadget to each pair of vertices, we
are done by the same reasoning as before. However, the resulting graph is not necessarily
planar, so we will argue that the vertices in \( V \) can be paired in such a way that the gadgets
can be introduced while preserving planarity. We do this in two steps. First, we show that
there exists a pairing of vertices in which the distance between each pair is at most 2. After
that we show that planarity is preserved when we add edges of the gadgets according to such
a pairing.

We may assume that \( G \) is connected, so fix any spanning tree \( T \) of \( G \), and note that \( T \)
has an even number of vertices as \( G \) does and the degree of every vertex in \( T \) is at most 3.
It is then sufficient to argue that such a pairing exists in \( T \), because distance measured in \( G \)
in no more than that measured in \( T \). If there exist adjacent vertices \( u \) and \( v \) in \( T \) such that
\( \deg(u) = 1 \) and \( \deg(v) \leq 2 \), then we pair \( u \) with \( v \). Otherwise, every leaf node is incident
with a vertex of degree 3, implying that there exists a vertex \( t \) which is adjacent to at least
two leaf nodes \( u \) and \( v \), and we pair these two leaf nodes together. (To see that such a vertex
\( t \) exists, fix any node \( r \) and make \( T \) a rooted tree with root \( r \). Let \( u \) be a leaf of maximum
depth. As the number of vertices is even, hence more than one, \( u \neq r \) and \( u \) has a parent \( t \). We have \( \deg(t) = 3 \), since no neighbor of \( u \) has degree \( \leq 2 \). Thus \( t \) has another child \( v \).
This \( v \) must be a leaf by the maximum depth of \( u \).) In any case, after removing \( u \) and \( v \) and
their incident edges, we are left with an induced subtree \( T' \) with an even number of vertices
and for which every vertex has degree at most 3. This implies a simple recursive algorithm
that finds a pairing where paired vertices have distance (on tree \( T \)) at most 2.

Given such a pairing, we may further assume that there are no pairs \((u, v)\) and \((s, t)\)
such that the distance between both pairs is 2 and \((u, s, v, t)\) is a path; otherwise we may
instead pair \( u \) with \( s \) and \( v \) with \( t \), resulting in two pairs of adjacent vertices. Now for every
pair \((u, v)\) of paired vertices, fix a minimal length path between them. Draw all gadgets in a
neighborhood of the path connecting each pair. Clearly this can be done preserving planarity
if \((u, v)\) has distance 1. For \((u, v)\) having distance 2, let the chosen path be \((u, s, v)\), then
\( s \) must have degree 3, otherwise \( s \) must be paired with some \( t \) and \((u, s, v, t)\) or \((t, u, s, v)\)
is a path. Let the third edge incident to \( s \) be \((s, w)\). All gadgets for \((u, v)\) are drawn on
the opposite side of \((s, w)\). Note that \( s \) is paired with some vertex \( t \) through a path \((s, w, t)\)
not involving \( u \) or \( v \), and causes no edge crossings with the gadgets drawn between \( u \) and \( v \).
Moreover, gadgets that are drawn along a path involving \( u \) or \( v \) also introduce no crossings
with the gadgets between \( u \) and \( v \) (they do not connect to \( s \), and the drawing was defined
in such a way that they do not cross the path \((u, s, v))\). Hence planarity is preserved.

The ability to simulate all unary signatures is what will allow us to prove \#P-hardness,
although technically we will use “virtual unary” recognizers of the form \([x, 0, y]\). The next
lemma says that if we are considering \( \text{Hol}_k(a, b) \), then other than on a 1-dimensional curve
\( ab = 1 \) and an isolated point \((a, b) = (0, 0)\) we get a reduction from \textsc{Vertex Cover}
by using virtual unary recognizers. As we have seen, the problem \( \text{Hol}_k(a, b) \) is tractable on the
curve \( ab = 1 \) and at \((a, b) = (0, 0)\).
Lemma 5.31. Suppose $k \geq 3$ and $(a, b) \in \mathbb{C}^2$ satisfies $ab \neq 1$ and $(a, b) \neq (0, 0)$, and assume that for any fixed $m \geq 1$, and any $x_j, y_j \in \mathbb{C}$ $(1 \leq j \leq m)$,

$$\text{Holant}(\{=, [x_1, 0, y_1], \ldots, [x_m, 0, y_m] \mid [a, 1, b]) \leq_T \text{Holant}(=, [a, 1, b]).$$

Then

$$\text{Holant}(=, [0, 1, 1]) \leq_T \text{Holant}(=, [a, 1, b])$$

and $\text{Holant}(=, [a, 1, b])$ is $\#P$-hard.

Proof. Assume $ab \neq 1$ and $(a, b) \neq (0, 0)$. Since $\text{Hol}_k(0, 1)$ is the same problem as $\text{Holant}(=, [0, 1, 1])$, i.e., counting VERTEX COVER on $k$-regular graphs, which is $\#P$-hard, we only need to show how to simulate the generator signature $[0, 1, 1]$. We split this into three cases, and use a chain of three reductions, each involving a gadget in Figure 5.7. Each gadget will have recognizer signatures of the type $[x; 0; y]$ assigned to the labeled vertices and the generator $[a; 1; b]$ assigned to the unlabeled vertices.

1. $ab \neq 0$ and $ab \neq -1$
2. $ab = 0$
3. $ab = -1$

If $ab \neq 0$ and $ab \neq -1$, then we use gadget 3 in Figure 5.7, and we set its signatures to be

$$\alpha = [-b/(ab - 1), 0, a/(ab - 1)] \sim [-b, 0, a],$$
$$\beta = [-a^{-2}, 0, b^{-1}(1 + ab)^{-1}], \quad \text{and}$$
$$\gamma = [(ab + 1)/(1 - ab), 0, -a^2(ab + 1)/(1 - ab)] \sim [1, 0, -a^2].$$

Here we use the notation $\sim$ to mean up to a nonzero multiplier. Calculating the resulting signature of Gadget 3 using matrix product, we find that it is $[0, 1, 1]$ as desired.

If $ab = 0$ then assume without loss of generality that $a = 0$ and $b \neq 0$ (we can do this because $\text{Hol}_k(0, 1) = \text{Hol}_k(1, 0)$ — also note that $\text{Hol}_k(1, 0)$ is the problem of counting INDEPENDENT SET on $k$-regular graphs). This time we use Gadget 1, setting $\gamma = [b, 0, b^{-1}]$. Then Gadget 1 simulates a generator signature $[b^{-1}, 1, 2b]$. But since this signature fits the criteria of case 1 above, we are done by a reduction from case 1.
Similarly, if \( ab = -1 \), then Gadget 2 exhibits a generator signature of the form \([0, 1, 5/(2a)]\) by choosing the signatures \( \gamma = [1/(6a), 0, -a/24] \) and \( \beta = [-3/a, 0, a] \). Since \( 5/(2a) \) is nonzero, we are done by a reduction from case 2.

**Corollary 5.32.** Suppose \( a, b \in \mathbb{C} \) are such that \( X \) and \( Y \) are not both real valued. Also assume \( X \neq 1 \) and \( 4X^3 \neq Y^2 \) (equivalently, \( a^3 \neq b^3 \)). Then \( \text{Hol}_3(a, b) \) is \#P-hard, even when restricted to planar graphs as input.

**Proof.** Immediate from Corollary 5.29 and Lemma 5.31. Note that all constructions preserve planarity.

The proof of Lemma 5.31 is a good demonstration of the power of having all unary signatures. However, the particular choices of the (virtual) unary signatures \( \alpha, \beta, \gamma \) may appear mysterious. Some explanations will demystify this. As indicated earlier, each signature of the form \([x; 0; y]\) as a recognizer is obtained by connecting a copy of \( \equiv_3 \) to a unary generator \([x; y]\). The more detailed and direct depiction of Gadgets 1 to 3 are in Figure 5.8 with unary signatures. We have seen a similar proof in the dichotomy for Holant* problems, where there were gadgets that resemble “growing weeds”. The matrix form of a signature \( u = [x, 0, y] \) is diagonal \( D_u = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \). Gadget 3 (in either Fig 5.7 (c) or Fig 5.8 (c)) has the signature \((gD_\alpha g)D_\beta (gD_\gamma g)D_\beta (gD_\alpha g)\), expressed as a matrix product, where \( g = \begin{bmatrix} a & 1 \\ 1 & b \end{bmatrix} \). If \( D = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \) then \( gDg = \begin{bmatrix} a^2x + y & ax + by \\ ax + by & x + b^2y \end{bmatrix} \), and it can be made a diagonal matrix if we choose \((x, y) \sim (b, a)\), which is what we did for \( D_\alpha \). On the other hand, if we take \( y = -a^2x \) then \( gDg \) has the form \( g_1 = x(1 - ab)\begin{bmatrix} 0 & a \\ 1 & 1 + ab \end{bmatrix} \), which is what we did for \( D_\gamma \).

Our goal is to realize \( \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \). With a diagonal \( D' \sim \begin{bmatrix} 1/a & 0 \\ 0 & 1/(1 + ab) \end{bmatrix} \), \( D'g_1D' = D'(gD_\gamma g)D' \) takes the form \( g_2 = \lambda \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \). We need the outside factors \( gD_\alpha g \) to provide a proper bipartite form, so that Gadget 3 constructed in the end is a generator. Thus the strategy is to choose \( \gamma \) such that \( gD_\gamma g \) has the form \( g_1 \), choose \( \alpha \) such that \( gD_\alpha g \) takes a diagonal form \( D'' \). Assume \( D'' \) is invertible, we choose \( D_\beta \sim (D'')^{-1}D' \). Note that we can choose \( D_\beta \) to be an arbitrary diagonal matrix of the form \( \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \), in particular any given \((D'')^{-1}D'\), and because these matrices are diagonal they commute \((D'')^{-1}D' = D'(D'')^{-1}\). Thus \((gD_\alpha g)D_\beta (gD_\gamma g)D_\beta (gD_\alpha g) = D'(gD_\gamma g)D' \) takes the from \( g_2 \).
5.1.4 Hardness in Region III where \((X, Y) \in \mathbb{R} \times \mathbb{R}\)

Recall that Region III is specified by \(X \neq 1, Y^2 \neq 4X^3\) (equivalently \(a^3 \neq b^3\)) and \((X, Y) \neq (-1, 0)\). In this subsection we further assume \((X, Y) \in \mathbb{R} \times \mathbb{R}\).

Interpolation

We now introduce a new type of gadgets called finisher gadgets, and adapt our interpolation techniques to the scenario with binary recursive gadgets and finisher gadgets. A finisher gadget has two sets of 2 dangling edges all connected internally to recognizers. It can transform a binary symmetric generator to a binary symmetric recognizer. Symmetry is guaranteed by the symmetry of the underlying graph of the gadget. They have the additional property that one set of 2 dangling edges (depicted as protruding from the left) are connected to a single \(=k\). These finisher gadgets will be represented by \(3 \times 3\) matrices with middle row identically 0, and thus rank at most 2. This extension is necessary to treat the case in Region III where \((X, Y) \in \mathbb{R} \times \mathbb{R}\), and it is also a natural place to introduce the idea of iterating at a higher dimensional space, which affords more flexibility, and then apply a projection down to a lower dimension. The most crucial lemma is Lemma 5.35, which amounts to an exchange of quantifiers. But first we establish tools for binary recursive gadgets as we did for unary recursive gadgets.

**Lemma 5.33.** Let \(M \in \mathbb{C}^{3 \times 3}\) be a nonsingular binary recursive gadget with infinite projective order. Let \(s \in \mathbb{C}^3\) be a binary starter gadget such that \(s\) is not orthogonal to any row eigenvector of \(M\). Then \(\{M^js\}_{j \geq 0}\) is a sequence of pairwise linearly independent generator signatures.

Proof. The form of \(M^js\) defines generator signatures, where both \(s\) and \(M\) are expressed for symmetric signatures.

Let \(M = T^{-1}JT\), where \(J\) is the Jordan Normal Form for \(M\), and let \(u, v, w \in \mathbb{C}\) be

![Diagram](image)

(a) A finisher gadget (b) Another finisher gadget

Figure 5.9: Examples of finisher gadgets
determined by $[u \ v \ w]^T = Ts$. Suppose $J$ is diagonal, let $B = T^{-1} \text{diag}(u, v, w)$, then

$$M^j_s = T^{-1} J^j T s = T^{-1} \begin{bmatrix}
\alpha^j & 0 & 0 \\
0 & \beta^j & 0 \\
0 & 0 & \gamma^j
\end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = T^{-1} \begin{bmatrix} u & 0 & 0 \\
0 & v & 0 \\
0 & 0 & w \end{bmatrix} \begin{bmatrix} \alpha^j \\ \beta^j \\ \gamma^j \end{bmatrix} = B \begin{bmatrix} \alpha^j \\ \beta^j \\ \gamma^j \end{bmatrix}.$$ 

There is a pair of eigenvalues of $M$ for which the ratio is not a root of unity (otherwise, we would have some $j$ for which $M^j = \lambda I$ for some nonzero $\lambda$). Then the sequence of vectors $[\alpha^j \ \beta^j \ \gamma^j]^T$ are pairwise linearly independent, and since pairwise linear independence is preserved under multiplication by the nonsingular matrix $B$ (note $uvw \neq 0$), we conclude that $\{M^j_s\}_{j \geq 0}$ is pairwise linearly independent.

If $M$ has two Jordan blocks, then we have

$$M^j_s = T^{-1} J^j T s = T^{-1} \begin{bmatrix}
\alpha^j & 0 & 0 \\
\alpha^j & \beta^j & 0 \\
0 & 0 & \gamma^j
\end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

where $uw \neq 0$. If $M$ has one Jordan block, then

$$M^j_s = T^{-1} J^j T s = T^{-1} \begin{bmatrix}
\alpha^j & 0 & 0 \\
\alpha^j & \beta^j & 0 \\
\alpha^j & \beta^j & \gamma^j
\end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

where $u \neq 0$. Either way, we have

$$M^j_s = T^{-1} J^j T s = T^{-1} \begin{bmatrix}
\alpha^j & 0 & 0 \\
\alpha^j & \beta^j & 0 \\
* & * & *
\end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} u \alpha^j \\ u \beta^j + v \alpha^j \\ * \\ * \\ * \\ * \\ * \end{bmatrix} = u \alpha^j \begin{bmatrix} 1 \\ j/\alpha + v/u \\ 1 \\ * \\ * \end{bmatrix}.$$ 

The determinant of the 2 by 2 matrix $\det \begin{bmatrix} j/\alpha + v/u & i/\alpha^i + v/u \end{bmatrix} = \frac{i-j}{\alpha} \neq 0$. Hence $\{M^j_s\}_{j \geq 0}$ is a sequence of pairwise linearly independent signatures.

We will not construct general-purpose binary starter gadgets (as we did for unary starter gadgets). Instead, we will test starter gadgets individually, using the following Lemma.

**Lemma 5.34.** Suppose $M \in \mathbb{C}^{m \times m}$ and $s \in \mathbb{C}^m$. If $\det([s \ M \ M^2 \ M^3 \ldots \ M^{m-1} \ s]) \neq 0$ then $s$ is not orthogonal to any row eigenvector of $M$.

**Proof.** Suppose $s$ is orthogonal to a row eigenvector $v$ of $M$ with eigenvalue $\lambda$. Then since $v M^j s = \lambda^j vs = 0$ we have $v[s \ M \ M^2 \ M^3 \ldots \ M^{m-1} \ s] = 0$, but $v \neq 0$ so this is a contradiction.

The next lemma is crucial.
Lemma 5.35. Suppose {mi }i≥0 is a series of pairwise linearly independent column vectors
in C3 . Let F ′ , F ′′ , and F ′′′ ∈ C3×3 be matrices of rank 2 where ker(F ′ ), ker(F ′′ ), and ker(F ′′′ )
are linearly independent. Then for every n, there exists some S ⊆ {i | 0 ≤ i ≤ n3 }, and
some F ∈ {F ′ , F ′′ , F ′′′ }, such that |S| ≥ n and vectors in {F mi | i ∈ S} are pairwise linearly
independent.
Proof. Let j > i ≥ 0 be integers, let N = [mi mj ] ∈ C3×2 , then rank(N ) = 2 and
dim(Im(N )) = 2. Since ker(F ′ ), ker(F ′′ ), and ker(F ′′′ ) are linearly independent of C3 ,
we can let F ∈ {F ′ , F ′′ , F ′′′ } such that ker(F ) ∩ Im(N ) = {0}. Then F N ∈ C3×2 has rank
2, hence F mi and F mj are linearly independent.
Each F ∈ {F ′ , F ′′ , F ′′′ } defines a coloring of the set K = {0, 1, . . . , n3 } as follows: color
i ∈ K with the linear subspace spanned by F mi . Assume for a contradiction that for each
F ∈ {F ′ , F ′′ , F ′′′ } there are not n pairwise linearly independent vectors among {F mi : i ∈
K}. Then, including possibly the 0-dimensional space {0}, there can be at most n distinct
colors assigned by each F ∈ {F ′ , F ′′ , F ′′′ }. By the pigeonhole principle, some i and j with
0 ≤ i < j ≤ n3 must receive the same color for all F ∈ {F ′ , F ′′ , F ′′′ }. This is a contradiction
and we are done.
We remark that Lemma 5.35 amounts to an exchange of quantifiers. As shown in the
first paragraph of the proof, the antecedent of the lemma says that for every pair of distinct
mi and mj from a sequence of pairwise linearly independent {mi }i≥0 , there exists a finisher
F ∈ {F ′ , F ′′ , F ′′′ } such that F mi and F mj are linearly independent. This is a ∀∃ type
condition. Then the consequent of the lemma says that there is a single F ∈ {F ′ , F ′′ , F ′′′ }
that produces an arbitrarily large set of pairwise linearly independent {F mi | i ∈ S}. This
is an ∃∀ type statement (there exist a large S among the first nO(1) many i’s and a single
F , such that F works for all mi from S.) The cost of this exchange of quantifiers is that we
may have to look at polynomially more vectors in the original list {mi }i≥0 , and we may not
know in advance which subset S would work. As we will see both caveats are harmless in
our intended application of this lemma.
In the following lemma we use binary finisher gadgets. Recall that such a gadget has two
sets of 2 dangling edges each, both sets are connected internally to recognizer signatures,
and both dangling edges of one set (depicted as protruding from the left) are connected
internally to a single Equality signature. The 2 dangling edges of the other set (depicted
as protruding from the right) are to be merged with the 2 leading edges of a binary recursive
gadget. Thus these finisher gadgets have 3 by 3 transition matrcies that have an identically
0 middle row. They transform a generator signature to a recognizer signature.
Lemma 5.36. Suppose that the following gadgets can be built using complex-valued signatures
from a finite generator set G and a finite recognizer set R.
1. A nonsingular binary recursive gadget M with infinite projective order.
2. A binary starter gadget s where det([s M s M 2 s]) ̸= 0.
157


Figure 5.10: Gadgets for constructing finisher gadgets

3. Three binary finisher gadgets with rank 2 transition matrices $F_1, F_2, F_3 \in \mathbb{C}^{3 \times 3}$, for which $\ker(F_1)$, $\ker(F_2)$, and $\ker(F_3)$ are linearly independent.

Then for any $x, y \in \mathbb{C}$, $\text{Holant} (\mathcal{R} \cup \{[x, 0, y]\} | \mathcal{G}) \leq_{\text{P}}^{\text{NP}} \text{Holant} (\mathcal{R} | \mathcal{G})$.

**Proof.** By Lemma 5.34 we know that $s$ is not orthogonal to any row eigenvector of $M$, so \(\{M^j s\}_{j \geq 0}\) is a sequence of pairwise linearly independent vectors by Lemma 5.33. We will efficiently find a subsequence of pairwise linearly independent recognizers from the sequence $\{F_1 M^j s, F_2 M^j s, F_3 M^j s \mid j \geq 0\}$.

By design both sets of dangling edges of each finisher gadget are internally incident to recognizers, making this a list of recognizer signatures of the form $[*, 0, *]$. We can greedily pick a large subset of pairwise linearly independent recognizers from the above list in polynomial time. Start with the empty subset, we test each successive signature for pairwise linear independence with all previously chosen signatures, and append it to our subsequence if it passes the test. Note that a signature is eliminated in this process iff it is a multiple of a previously chosen signature. By Lemma 5.35, we will have collected the first $n$ signatures in this process by considering only $0 \leq j \leq n^3$ among the list. Indeed if at the end of this process we end up having collected less than $n$ pairwise linearly independent recognizers, then for all $0 \leq j \leq n^3$, every $F_1 M^j s, F_2 M^j s$ and $F_3 M^j s$ is a multiple of some chosen signature. But then some two distinct members of the set $S$ of size $|S| \geq n$ guaranteed by Lemma 5.35 must be independently dependent, a contradiction. Then by Lemma 5.18, for any $x, y \in \mathbb{C}$ we can efficiently simulate recognizer $[x, 0, y]$. $\square$

Now we explicitly construct some finisher gadgets.

**Lemma 5.37.** Suppose $X \neq 1$ and $a^3 \neq b^3$. Then in the context of $\text{Hol}_3(a, b)$ there exist three binary finisher gadgets with rank 2 transition matrices $F_1, F_2, F_3 \in \mathbb{C}^{3 \times 3}$, for which $\ker(F_1)$, $\ker(F_2)$, and $\ker(F_3)$ are linearly independent.

**Proof.** We start first with the case where $ab \neq 0$. Using the simplest possible choice for a
finisher gadget $M_{11}$ (Figure 5.10), we get $M_{11} = \begin{bmatrix} a & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & b \end{bmatrix}$. We calculate that

$$M_5 = \begin{bmatrix} a^3 & 2a & b \\ a^2 & ab + 1 & b^2 \\ a & 2b & b^3 \end{bmatrix}.$$  

We build two more finisher gadgets $F'$ and $F''$ using gadgets $M_{11}$ and $M_5$ so that $F' = M_{11}M_5$ and $F'' = M_{11}M_5^2$. Since $M_{11}$ has rank 2 and $M_5$ is nonsingular (note $\det(M_5) = ab(ab-1)^3$), it follows that $F'$ and $F''$ also have rank 2. Now to show that the kernels of $M_{11}$, $F'$ and $F''$ are linearly independent, we check that the cross products of the first and last row vectors of $M_{11}$, $F'$, and $F''$ (denoted respectively by $v$, $v'$, and $v''$) are linearly independent. Note that $M_{11}$ has an identically 0 middle row, so the cross product of the 1st row and the 3rd row spans the kernel $\{x \in \mathbb{C}^3 \mid M_{11}x = 0\}$. The same is true for $F'$ and $F''$. The cross products of the first and last row vectors of $M_{11}$, $F'$, and $F''$ are respectively, $[0, 1-ab, 0], (ab-1)^2[2b^2, -ab(ab+1), 2a^2]$, and

$$(ab-1)^3[2b(a^2b^3 + a^2 + ab^2 + b^4), -ab(a^3b^3 + 2a^3 + 2a^2b^2 + ab + 2b^3), 2a(a^4 + a^3b^2 + a^2b + b^2)].$$

As $ab \neq 1$, to see that these 3 vectors are linearly independent, it suffices to verify that the 2 by 2 matrix

$$\begin{bmatrix} 2b^2 & 2a^2 \\ 2(a^2b^3 + a^2 + ab^2 + b^4) & 2a(a^4 + a^3b^2 + a^2b + b^2) \end{bmatrix}$$

is nonsingular. Since $a \neq 0, b \neq 0$, we just need to check

$$\det \begin{bmatrix} b & a \\ a^2b^3 + a^2 + ab^2 + b^4 & a^4 + a^3b^2 + a^2b + b^2 \end{bmatrix} = (ab-1)(a^3 - b^3) \neq 0,$$

so the kernels of $M_{11}$, $F'$, and $F''$ are linearly independent when $ab \neq 0$.

Now suppose $ab = 0$. Since $a^3 \neq b^3$, by symmetry, if $ab = 0$ we may assume without loss of generality that $a \neq 0$ and $b = 0$.

$$M_{12} = \begin{bmatrix} a^6 + 2a^3 + 1 & 2a^4 + 2a^2 & a^2 \\ a^5 + a^2 & 2a^3 + 1 & a \\ a^4 & 2a^2 & 1 \end{bmatrix}.$$  

Composing $M_{11}$ with $M_{12}$, we get a finisher gadget with matrix $M_{11}M_{12}$, which has rank 2 since $M_{11}$ has rank 2 and $\det(M_{12}) = 1$. It is also straightforward to see that $F' = M_{11}M_5$ has rank 2, as $F' = \begin{bmatrix} a^4 + a & 2a^2 & 0 \\ 0 & 0 & 0 \\ a^3 & 2a & 0 \end{bmatrix}$ using $b = 0$. The cross products of the first and last rows of $M_{11}$, $F'$, and $M_{11}M_{12}$ are $[0, 1, 0]$, $[0, 0, 2a^2]$, and $[-2a, 2a^3 + 1, -2a^2(1+a)(a^2-a+1)]$ respectively. Then the matrix of cross products is clearly nonsingular, and we conclude that for any $a, b \in \mathbb{C}$, we have 3 finisher gadgets satisfying our requirement.  

\[\square\]  

159
Corollary 5.38. Let $a, b \in \mathbb{C}$ such that $X \neq 1$ and $a^3 \neq b^3$. Suppose that in the context of $\text{Hol}_3(a, b)$, there exists a nonsingular binary recursive gadget $M$ with infinite projective order and a binary starter gadget $s$ where \( \det([sMs M^2s]) \neq 0 \). Then $\text{Hol}_3(a, b)$ is $\#P$-hard.

\[\text{Proof.} \ \text{Immediate from Lemmas 5.37, 5.36, and 5.31.} \]

Applying Interpolation with anti-gadgets

Now we will construct specific recursive gadgets to handle the special case of Region III when $X$ and $Y$ are real valued: $(X, Y) \in \mathbb{R} \times \mathbb{R}$. We have $X \neq 1$, $Y^2 \neq 4X^3$ (equivalently $a^3 \neq b^3$) and $(X, Y) \neq (-1, 0)$. We will first set $X \neq 0, \pm 1$. There are two cases depending on whether $X + X^2 + Y = 0$. After that we consider $X = -1$ (but $Y \neq 0$) or $X = 0$ (but $Y \neq 0$).

We will introduce the idea of an anti-gadget in the next lemma. In a typical reduction from one problem to another in complexity theory, one often designs various gadgets. A reduction starts from an instance of the first problem and constructs an instance of the second problem which may contain a polynomial number of copies of the gadget. But the notion of containing a negative number of copies of a gadget seems meaningless. However, the effect of including an anti-gadget will indeed be the same as introducing a negative copy of a gadget in a reduction. More precisely, the effect of an anti-gadget, when expressed in algebraic terms, is the same as erasing the presence of some other gadget fragment. It is as if we managed to include a negative copy of a certain gadget.

We will see that anti-gadget greatly simplifies our work in finding a recursive gadget with infinite projective order. Effectively, we can use one gadget to “cancel out” a portion of another gadget, leaving a “virtual gadget” with a diagonal signature matrix — ideal for our purposes, as it is straightforward to analyze.

Lemma 5.39. Let $a, b \in \mathbb{C}$ such that $X, Y \in \mathbb{R}$, and assume that $X \notin \{0, 1, -1\}$, $a^3 \neq b^3$, and $X^2 + X + Y \neq 0$. Then $\text{Hol}_3(a, b)$ is $\#P$-hard.

\[\text{Proof.} \ \text{We begin this proof with binary recursive gadget } M_5 = \begin{bmatrix} a^3 & 2a & b \\ a^2 & ab + 1 & b^2 \\ a & 2b & b^3 \end{bmatrix}. \ \text{Taking}\]

\[\text{Figure 5.11: An anti-gadget pair}\]
determinant, we find $\det(M_5) = X(X - 1)^3 \neq 0$. If we let $s$ be the simplest possible binary starter gadget consisting of a single vertex (Figure 5.3(a), Gadget $M_5$), then $\det([s M_5 s M_5^2 s]) = -(a^3 - b^3)(ab - 1)^4 \neq 0$. If $M_5$ has infinite projective order, then all criteria for Corollary 5.38 are met, and Hol$_3(a, b)$ is #P-hard.

Now suppose that $M_5$ has finite projective order, and we let $i \in \mathbb{Z}^+$ and nonzero $\lambda \in \mathbb{C}$ such that $M_5^i = \lambda I$. Then chaining $i - 1$ copies of gadget $M_5$ together, we have an explicit construction for a gadget with signature $M_5^{i-1} = \lambda M_5^{-1}$.

The gadget $M_{13}$ has a $4 \times 4$ signature matrix obtained as a tensor product of the $2 \times 2$ matrix of $M_1 = \begin{bmatrix} a(a^2 + b) & a + b^2 \\ a^2 + b & (a + b^2)b \end{bmatrix}$. Expressed for symmetric signatures, we have

$$M_{13} = \begin{bmatrix} a^2A^2 & 2aAB & b^2 \\ aA^2 & (ab + 1)AB & bB^2 \\ A^2 & 2bAB & b^2B^2 \end{bmatrix},$$

where $A = a^2 + b$ and $B = a + b^2$.

Then concatenating the gadget with signature $M_5^{i-1}$ and a copy of $M_{13}$, we have

$$M = M_5^{i-1}M_{13} = \lambda \begin{bmatrix} \alpha & 0 & 0 \\ 0 & AB & 0 \\ 0 & 0 & \beta \end{bmatrix},$$

where $\alpha = \frac{A^2}{a} = \frac{(a^2 + b)^2}{a}$ and $\beta = \frac{B^2}{b} = \frac{(a + b^2)^2}{b}$. Note that $AB = (a^2 + b)(a + b^2) = X^2 + X + Y \neq 0$ and $X \neq 0$ by assumption. It follows that $\det(M) = \lambda^3\alpha \beta AB = \lambda^3(AB)^3/X \neq 0$, so $M$ is nonsingular. We prove that $M$ has infinite projective order, by showing that some two eigenvalues among $\{\alpha, \beta, AB\}$ have distinct norm. Suppose $|\alpha| = |\beta|$, then $|\alpha \beta X| = |\alpha|^2|X| \neq |\alpha|^2$ since $X \in \mathbb{R}$ and $X \neq \pm 1$. But $|\alpha \beta X| = |AB|^2$ and so $|\alpha| \neq |AB|$. Finally, we check that $\det([s M_5 s M_5^2 s]) = \lambda^3(a^3 - b^3)(X - 1)^3(X^2 + X + Y) \neq 0$, where $s = [a \ 1 \ b]^T$. Hence we apply Corollary 5.38 again and conclude that Hol$_3(a, b)$ is #P-hard. \hfill \square

The following test is a straightforward way to verify that a matrix has infinite projective order, by looking at the characteristic polynomial.

**Lemma 5.40.** If all roots of the complex polynomial $x^3 - Bx^2 + Cx - D$ have the same norm, then $\|C\|^2 = |B|^2D$.

**Proof.** If the roots have equal norm, then for some $a, b, c \in \mathbb{C}$ with $|a| = |b| = |c| = 1$, and nonnegative $r \in \mathbb{R}$ we can write $x^3 - Bx^2 + Cx - D = (x - ra)(x - rb)(x - rc)$, so $B = r(a + b + c)$, $C = r^2(ab + bc + ca)$, and $D = r^3abc$. Then

$$\|C\|^2 = r^2(ab + bc + ca)r^4|ab + bc + ca|^2 = r(a + b + c)r^2|a + b + c|^2r^3abc = |B|^2D,$$

where we used the fact that $|ab + bc + ca| = |ab + bc + ca| \cdot |a^{-1}b^{-1}c^{-1}| = |a^{-1} + b^{-1} + c^{-1}| = |a + b + c| = |a + b + c|$. \hfill \square
Lemma 5.41. Suppose \( X, Y \in \mathbb{R}, X^2 + X + Y = 0, a^3 \neq b^3, \) and \( X \notin \{0, 1, -1\}. \) Then \( \text{Hol}_3(a, b) \) is \#P-hard.

Proof. We first verify that \( \det(M_5) = X(X - 1)^3 \neq 0. \) Next we show that \( M_5 \) has infinite projective order. The characteristic polynomial of \( M_5 \) is \( x^3 - Bx^2 + Cx - D, \) where \( B = X + Y + 1 = 1 - X^2, C = (X - 1)(X^2 + X + Y) = 0, \) and \( D = X(X - 1)^3. \) Thus \( B, C, D \in \mathbb{R} \) and by Lemma 5.40 it is sufficient to observe \( C^3 - B^3D = X(X - 1)^6(X + 1)^3 \neq 0. \) We also have \( \det([s M_5 s M_5^2 s]) = -(a^3 - b^3)(ab - 1)^4 \neq 0 \) as required by Corollary 5.38, where \( s \) is the single-vertex starter gadget.

We are left with two cases: \( X = -1 \) (but \( Y \neq 0 \)) or \( X = 0 \) (but \( Y \neq 0 \)).

Lemma 5.42. Suppose \( X = -1, a^3 \neq b^3, \) and \( Y \in \mathbb{R} \) but \( Y \neq 0. \) Then \( \text{Hol}_3(a, b) \) is \#P-hard.

Proof. As before, \( \det(M_5) = X(X - 1)^3 \neq 0. \) Looking at the characteristic polynomial \( x^3 - Bx^2 + Cx - D \) of \( M_5, \) we have \( C^3 - B^3D = -16Y^3 \neq 0 \) so by Lemma 5.40 there are eigenvalues for which the ratio is not a root of unity and the matrix has infinite projective order. Once again, \( \det([s M_5 s M_5^2 s]) = -(a^3 - b^3)(ab - 1)^4 \neq 0, \) where \( s \) is the single-vertex starter gadget, and we are done by Corollary 5.38.

Lemma 5.43. Suppose \( X = 0 \) and \( Y \in \mathbb{R} \) but \( Y \neq 0. \) Then \( \text{Hol}_3(a, b) \) is \#P-hard.

Proof. Let \( M = M_{14}. \) Under the condition \( X = 0, \) we have \( \det(M) = Y^3 \neq 0, \) and also

\[
C^3 - B^3D = 2Y^6(3L^2 + 1) \neq 0,
\]

where we let \( L = Y^2 + 6Y + 8 \) and the characteristic polynomial of \( M \) is \( x^3 - Bx^2 + Cx - D. \) Also, \( \det([s M s M^2 s]) = 2Y^2(a^3 - b^3)(Y + 3), \) where \( s \) is the single-vertex starter gadget, so we are done by Corollary 5.38 unless \( Y = -3 \) (note \( a^3 \neq b^3 \) because either \( a \) or \( b \) is zero, but not both).
For $X = 0$ and $Y = -3$ we let $M = M_{15}$ and then we have $\det(M) = 6 \neq 0$, $C^3 - B^3D = 21 \neq 0$, and
\[
\det([s Ms M^2s]) = -(X - 1)^6(X + 1)(a^3 - b^3)(Y + 2) \neq 0.
\]

Finally we are ready to prove the main theorem of Section 5.1: Theorem 5.1 for the case $k = 3$.

**Proof of Theorem 5.1 restricted to $k = 3$**

*Proof.* Tractability (for Region I and for Region II on planar inputs) is given by Theorem 5.6. Intractability is given by Lemma 5.14 for Region II. For Region III where $X$ and $Y$ are not both real valued, we have Corollary 5.32. For Region III where $X$ and $Y$ are both real, we appeal to Lemmas 5.43, 5.42, 5.41, and 5.39 as follows: For $X = 0$ we have Lemma 5.43, for $X = -1$ we have Lemma 5.42. For the remaining cases we have Lemma 5.39 if $X^2 + X + Y \neq 0$, and Lemma 5.41 if $X^2 + X + Y = 0$. \qed

## 5.2 4-Regular Graphs

### 5.2.1 Tractability in Regions I and II

Recall that for 4-regular graphs, Region I is specified by $X = 1$ or $(X, Y) \in \{(0, 0), (-1, 0), (-1, \pm 2)\}$. Region II is specified by $Y^2 = 4X^4$ but $X \not\in \{0, \pm 1\}$. The union of Regions I and II is specified by $X = 1$ or $Y^2 = 4X^4$ or $(X, Y) = (-1, 0)$.

We will show that $X$ and $Y$ capture the complexity of $\text{Hol}_4(a, b)$, just as we did for $\text{Hol}_3(a, b)$. This will be slightly more involved, since in the case of $\text{Hol}_4(a, b)$ it is not true that the Holant is an integer polynomial in $X$ and $Y$; there can be an extra factor $a^2 + b^2$. This factor appears precisely when the number of vertices in the graph is odd (which can never happen in a 3-regular graph). We start with a version of Lemma 5.4 adapted to any $k$.

**Lemma 5.44.** Let $G$ be a $k$-regular graph with $n$ vertices. Then there exists a polynomial $P_G(\cdot, \cdot)$ with two variables and integer coefficients such that, for any signature grid $\Omega$ having underlying graph $G$, $Z(G) = P_G(ab, a^3 + b^3)$ (if $n$ is even) and $Z(G) = (a^{k/2} + b^{k/2})P(ab, a^k + b^k)$ (if $n$ is odd), for any edge function $[a, 1, b]$.

*Proof.* In the Edge-Vertex bipartite view, the partition function $Z(G)$ is the value $\text{Holant}_\Omega$ in the Holant problem $\text{Holant}(=a_k| [a, 1, b])$. First note that if $k$ and $n$ are both odd then no $k$-regular graph exists on $n$ vertices, as $kn$ is twice the number of edges, so $k$ must be even whenever $n$ is odd.

163
Consider any \(\{0, 1\}\)-vertex assignment \(\sigma\). If \(\sigma'\) is the complement assignment switching all 0’s and 1’s in \(\sigma\), then the sum of valuations for \(\sigma\) and \(\sigma'\) in Holant\(\Omega\) is \(a^i b^j + a^j b^i\) where \(i\) (resp. \(j\)) is the number of edges connecting two degree \(k\) vertices both assigned 0 (resp. 1) by \(\sigma\). We note that \(a^i b^j + a^j b^i = (ab)^{\min(i,j)} (a^{i-j} + b^{i-j})\).

For the all-0 assignment, \(i - j = kn/2\). Now starting from any assignment \(\sigma\), if we switch the assignment on one vertex \(v\) from 0 to 1, it is easy to verify that it changes the valuation from \(a^i b^j\) to \(a^j b^i\), where \(i - j = i' - j' + k\). This takes into account the changes among all edges incident to \(v\), including self loops at \(v\), if any. Every \(\{0, 1\}\) assignment \(\sigma\) is obtainable from the all-0 assignment by a sequence of switches, hence \(i - j \equiv kn/2 \pmod{k}\).

If \(n\) is even, then \(kn/2 \equiv 0 \pmod{k}\). Thus, for every assignment \(\sigma\), we have \(i - j \equiv 0 \pmod{k}\). Now \(a^i b^j + a^j b^i = (ab)^{\min(i,j)} (a^{k+} + b^{k-})\), for some integer \(\ell \geq 0\), and a simple induction

\[
\sum_{i=0}^{\ell} a^{k(i+1)} b^{k(i+1)} = (a^k + b^k) - (ab)^{k(k-1)} + b^{k(k-1)}
\]

shows that Holant\(\Omega\) is a polynomial \(P_G(ab, a^k + b^k)\) with integer coefficients.

If \(n\) is odd, then in particular \(k\) is even. For every assignment \(\sigma\), we have \(i - j \equiv k/2 \pmod{k}\). Now \(a^i b^j + a^j b^i = (ab)^{\min(i,j)} (a^{k/2+k\ell} + b^{k/2+k\ell})\), for some integer \(\ell \geq 0\). We verify that at \(\ell = 0\) and \(\ell = 1\), \(a^{k/2+k\ell} + b^{k/2+k\ell}\) becomes \(a^{k/2} + b^{k/2}\) and \(a^{3k/2} + b^{3k/2} = (a^{k/2} + b^{k/2}) (a^k + b^k - (ab)^{k/2})\) respectively, both of which are of the form: a product of \(a^{k/2} + b^{k/2}\) with an integer polynomial in \((ab, a^k + b^k)\). Then an easy induction

\[
a^{k/2+k(\ell+1)} + b^{k/2+k(\ell+1)} = (a^{k/2+k\ell} + b^{k/2+k\ell}) (a^k + b^k) - (ab)^{k(k-1)} + b^{k(k-1)}
\]

shows that \(a^{k/2+k\ell} + b^{k/2+k\ell}\) is of this form for all \(\ell \geq 0\).

\[\square\]

**Corollary 5.45.** Let \(G\) be any \(k\)-regular graph with \(n\) vertices, where \(k\) is even and \(n\) is odd, and let \(\Omega\) be any signature grid having underlying graph \(G\) and every edge labeled \([a, 1, b]\). If \(a^{k/2} + b^{k/2} = 0\), then Holant\(\Omega\) = 0.

**Corollary 5.46.** Let \(a', b' \in \mathbb{C}\) such that \(a' b' = X\) and \((a')^k + (b')^k = Y\). Then Holant\(_k\)(a, b) is polynomial time computable (respectively, is \#P-hard) if and only if Holant\(_k\)(a', b') is polynomial time computable (is \#P-hard).

**Proof.** Consider an input instance \(\Omega\) for Holant\(_k\)(a, b). If \(n\) is even then the claim is obvious. Suppose there are an odd number of vertices, and let \(\Omega'\) be a copy of \(\Omega\) with an additional vertex having \(k/2\) self-loops. Then

\[
\text{Holant}_{\Omega'} = (a^{k/2} + b^{k/2}) \text{Holant}_{\Omega}. \tag{5.1}
\]

If \(a^{k/2} + b^{k/2} \neq 0\) we can easily obtain Holant\(\Omega\) from Holant\(_{\Omega'}\) and vice versa. If \(a^{k/2} + b^{k/2} = 0\) then both Holant\(\Omega\) = 0 (by Corollary 5.45) and Holant\(_{\Omega'}\) = 0 (by (5.1)). Thus the complexity of Holant\(_k\)(a, b) is captured by input instances with an even number of vertices, for which the Holant can be expressed as a function of \(X\) and \(Y\).

\[\square\]
Now we specialize to $k = 4$. We first list the tractable cases of $\text{Hol}_4(a, b)$.

**Theorem 5.47.** For any $a, b \in \mathbb{C}$, $\text{Hol}_4(a, b)$ is solvable in polynomial time if any of the following four conditions are true.

1. $X = 1$.
2. $X = Y = 0$.
3. $X = -1$ and $Y \in \{0, \pm 2\}$.
4. $Y^2 = 4X^4$ and the input is restricted to planar graphs.

**Proof.** If $X = 1$ then the signature $[a, 1, b]$ is degenerate and the Holant can be computed in polynomial time. If $X = Y = 0$, then $a = b = 0$ and a 2-coloring argument can be applied to calculate the Holant. If $X = -1$, then applying a holographic transformation under basis $T = \begin{bmatrix} 1 & 0 \\ 0 & a^{-1} \end{bmatrix}$, we get $(T^{-1})^\otimes 2 g = [a, a, -a]^{\top}$ and $(=4) T^\otimes 4 = [1, 0, 0, 0, a^{-4}]$, where $g = [a, 1, -a^{-1}]^{\top}$ (note that the form of this $g$ corresponds to the assumption $X = -1$). Multiplying the signature $[a, a, -a]$ by $a^{-1}$ does not change the complexity of the problem, so Holant($=4$) $g$ is equivalent in complexity to Holant($[[1, 0, 0, 0, a^{-4} ] | [1, 1, -1]$), and the problem Holant($\{[1, 0, 0, 0, a^{-4}], [1, 1, -1]\}$) is tractable if $a^4 \in \{1, -1, i, -i\}$, by families $\mathcal{F}_1$ and $\mathcal{F}_3$ in Chapter 3. It is easy to see that $a^4 = \pm 1$ or $\pm i$ is equivalent to $Y = a^4 + (-a^{-1})^4 = a^4 + a^{-4}$ evaluates to $\pm 2$ or $0$. Finally, if $Y^2 = 4X^4$, then $a^4 = b^4$ and holographic algorithms using matchgates can be applied when the input graph is planar (see Chapter 4 and the proof of Theorem 5.6).

**5.2.2 Hardness in Region III**

Recall that for 4-regular graphs, Region III is the complement of the union of Regions I and II, which is specified by $X \neq 1$ and $Y^2 \neq 4X^4$ (equivalently $a^4 \neq b^4$) and $(X, Y) \neq (-1, 0)$. To prove $\#P$-hardness in Region III, we must use planar gadgets, and reduce from a problem that is $\#P$-hard for planar graphs.

For $\text{Hol}_3(a, b)$, we were able to make use of constructions consisting of a unary starter gadget together with iterations of a unary recursive gadget. However in the context of $\text{Hol}_4(a, b)$ the signature of any $\mathcal{F}$-gate has an even arity. This can be seen by adding up modulo 2 all arities of signatures in the $\mathcal{F}$-gate, and noting that every internal edge is counted twice in this sum. So we cannot construct unary starter gadgets. However, we would still like to make use of some nice unary recursive gadgets, such as those we will see in Lemma 5.48. One alternative is to leave off the unary starter gadget, and consider the iterated unary recursive gadgets, which have two dangling edges, as constructions of binary signatures. However, these signatures, like unary recursive gadgets, have one dangling edge $e$ internally incident to a generator and one dangling edge $e'$ internally incident to a recognizer, and are
not symmetric. To get around this difficulty, we introduce some gadgets for the sole purpose of transforming the construction into a sequence of symmetric binary signatures. Symmetry of the signatures constructed will be guaranteed by the symmetry in the underlying graph fragment ($F$-gate).

In the next lemma we introduce a new idea of gadget construction—syzygy—a trio of gadgets that work in concert, and their alignment produces their success.

**Lemma 5.48.** Suppose either $X \not\in \mathbb{R}$, or both $Y \not\in \mathbb{R}$ and $X \neq 0,1$. Then in the context of Hol$_4(a,b)$, there exists a unary recursive gadget $M$ and a gadget $A$ such that $\{M^jA\}_{j \geq 0}$ is a sequence of pairwise linearly independent symmetric binary generator signatures.

**Proof.** The matrices for these gadgets are as follows.

- $M_{16} = \begin{bmatrix} a^4 + a^2b^2 & a^2b + b^3 \\ a^3 + ab^2 & a^2b^2 + b^4 \end{bmatrix}$
- $M_{17} = \begin{bmatrix} a^4 + ab & a + b^3 \\ a^3 + b & ab + b^4 \end{bmatrix}$
- $M_{18} = \begin{bmatrix} a^8 + a^6b^2 + a^5b + a^4 + 2a^3b^3 + a^2b^2 + ab^5 & a^6b + a^4b^3 + 2a^3b^2 + a^2b^5 + 2ab^4 + b^7 \\ a^7 + 2a^4b + a^5b^2 + 2a^3b^3 + a^2b^4 + ab^6 & a^5b + a^2b^2 + 2a^3b^3 + b^4 + ab^5 + a^2b^6 + b^8 \end{bmatrix}$

We begin by calculating the trace and determinant of each gadget in the trio of gadgets $M_{16}, M_{17}$ and $M_{18}$.

- $\text{tr}(M_{16}) = 2X^2 + Y$
- $\text{tr}(M_{17}) = 2X + Y$
- $\text{tr}(M_{18}) = (2X^2 + Y)(-X^2 + 2X + Y + 1)$
- $\text{det}(M_{16}) = X(X - 1)(2X^2 + Y)$
- $\text{det}(M_{17}) = (X - 1)(X^3 + X + Y)$
- $\text{det}(M_{18}) = X(X - 1)^2(X^3 + X + Y)(2X^2 + Y)$

We note that $M_{18} = M_{17}M_{16}$. 

166
We will show that if any one of these traces or determinants is zero, then another one of the three gadgets has nonzero eigenvalues of distinct norms. Since \( X \neq 0,1 \) in any case, the only possibilities for any of these traces or determinants to be zero are:

\[
2X^2 + Y = 0, \quad \text{or} \quad 2X + Y = 0, \quad \text{or} \quad X^3 + X + Y = 0, \quad \text{or} \quad -X^2 + 2X + Y + 1 = 0.
\]

First we note that if any of these four equations is satisfied, then \( X \not\in \mathbb{R} \), for either it is by direct hypothesis of the lemma, or it is by \( Y \not\in \mathbb{R} \) and then by the equation.

Now we consider the four equations. If \( 2X^2 + Y = 0 \) then \( \text{tr}(M_{17}) = -2X(X - 1) \neq 0 \), \( \det(M_{17}) = X(X - 1)^3 \neq 0 \), and the ratio \( \frac{\text{tr}^2(M_{17})}{\det(M_{17})} = \frac{4X}{X - 1} \not\in \mathbb{R}^+ \). If \( 2X + Y = 0 \) then \( \text{tr}(M_{18}) = -2X(X - 1)^2(X + 1) \neq 0 \), \( \det(M_{18}) = 2X^3(X - 1)^4(X + 1) \neq 0 \), and the ratio \( \frac{\text{tr}^2(M_{18})}{\det(M_{18})} = \frac{2(X^2 + 1)}{X} \not\in \mathbb{R}^+ \). If \( X^3 + X + Y = 0 \) then \( \text{tr}(M_{16}) = -X(X - 1)^2 \neq 0 \), \( \det(M_{16}) = -X^2(X - 1)^3 \neq 0 \), and the ratio \( \frac{\text{tr}^2(M_{16})}{\det(M_{16})} = -X + 1 \not\in \mathbb{R}^+ \). Finally, suppose \( -X^2 + 2X + Y + 1 = 0 \). Then \( \text{tr}(M_{16}) = (X - 1)(3X^2 + 1) \neq 0 \), \( \det(M_{16}) = X(X - 1)^2(3X^2 + 1) \neq 0 \), and the ratio \( \frac{\text{tr}^2(M_{16})}{\det(M_{16})} = \frac{3X^2 + 1}{X} \not\in \mathbb{R}^+ \). In each case, by Lemma 5.19 one of the gadgets \( M_{16}, M_{17} \) or \( M_{18} \) has nonzero eigenvalues of distinct norm.

Now suppose each trace and determinant is nonzero, and suppose all three gadgets fail to have eigenvalues with distinct norm. Then Lemma 5.19 implies that \( \frac{\text{tr}^2(M_i)}{\det(M_i)} \in \mathbb{R}^+ \) for each \( M_i \). In particular, we claim that in this case \( X \neq -1 \). Otherwise, by hypothesis \( Y \not\in \mathbb{R} \) since \( X = -1 \not\in \mathbb{R} \), and yet \( \text{tr}(M_{16}) = Y + 2 \), \( \det(M_{16}) = 2(Y + 2) \), and \( \frac{\text{tr}^2(M_{16})}{\det(M_{16})} = (Y + 2)/2 \in \mathbb{R}^+ \); this is a contradiction.

We write \( \frac{\text{tr}^2(M_{18})}{\det(M_{18})} \frac{\det(M_{17})}{\text{tr}^2(M_{17})} = \frac{(-X^2 + 2X + Y + 1)^2}{(2X + Y)^2} = \left(1 - \frac{(X - 1)(X + 1)}{2X + Y}\right)^2 \). The cancelation and its factorization as a square all follow from the fact that \( \det(M_{18}) = \det(M_{16}) \det(M_{17}) \). Hence \( \frac{(X - 1)(X + 1)}{2X + Y} \) is real-valued. Furthermore \( \frac{(X - 1)(X + 1)}{2X + Y} \) is nonzero since \( X \neq 1 \) in either case of the hypothesis and \( X \neq -1 \) by the proof above. Thus \( \frac{2X + Y}{(X - 1)(X + 1)} \) is well defined and real-valued. We also know that \( \frac{2X^2 + Y}{X(X - 1)} = \frac{\text{tr}^2(M_{16})}{\det(M_{16})} \in \mathbb{R}^+ \), so there exist \( r \) and \( s \in \mathbb{R} \) (both nonzero) such that

\[
Y = r(X - 1)(X + 1) - 2X, \quad \text{(5.2)}
\]

Setting these two expressions of \( Y \) equal to each other and refactoring, we get

\[
(X - 1)(r - (s - 2 - r)X) = 0,
\]

therefore \( r = (s - 2 - r)X \), but since \( r \neq 0 \), we get \( X \in \mathbb{R} \), and by (5.2), we also get \( Y \in \mathbb{R} \). This contradiction establishes that one of the gadgets \( M_{16}, M_{17}, \) or \( M_{18} \) has nonzero eigenvalues of distinct norms. By Lemma 5.11, the powers of one of \( M_{16}, M_{17}, \) or \( M_{18} \) are pairwise linearly independent. Equivalently it has infinite projective order by Lemma 5.10.

Now we show that in each case, for the chosen unary recursive gadget \( M \), there exists a gadget \( A \) such that \( M^jA \) are signature matrices of symmetric binary generator signatures,
for all nonnegative integer \( j \). We take \( g \) to be a gadget consisting of a single generator vertex with matrix \( g = \begin{bmatrix} a \\ 1 \\ b \end{bmatrix} \). If \( M = M_{16} \) or \( M = M_{17} \), then by the symmetry in the gadget design each function \( M^j g \) is a symmetric binary generator signature. Hence we take \( A = g \) in these cases. For gadget \( M_{18} \), we set \( A = M_{17} g \), and again we get the desired graph symmetry. We know \( g \) is nonsingular as \( X \neq 1 \). For any case where we need to apply \( M_{18} \), we also have \( M_{18} \) is nonsingular. (This includes the case when \( 2X + Y = 0 \), where we explicitly verified that \( \det(M_{18}) \neq 0 \), and when we supposed all three traces and determinants are nonzero.) However \( M_{18} = M_{17} M_{16} \), so it follows that \( M_{17} g \) is also nonsingular in such a case. In any case, applying the nonsingular linear transformation defined by \( A : M^j \mapsto M^j A \), does not change the fact that we have a sequence of pairwise linearly independent signatures, and these are symmetric binary generators by design.

In the next three lemmas we will do the same thing as Lemma 5.48 to produce a sequence of pairwise linearly independent symmetric binary generator signatures, in the case \( X \in \mathbb{R} \). Note that we will avoid tractable cases; thus we will exclude the case \( X = 1 \) or \( (X, Y) = (0, 0), (1, 0) \) or \( (-1, \pm 2) \). In Lemmas 5.49 and 5.50, the generators will be constructed using the technique of a binary recursive gadget and a starter gadget, rather than a unary recursive gadget which is symmetrized as in Lemma 5.48. Thus these generator signatures will be expressed as column vectors in \( \mathbb{C}^3 \) rather than the matrix form in \( \mathbb{C}^{2\times 2} \) as in Lemma 5.48. However in terms of their functionality as generator signatures there is no difference.

**Lemma 5.49.** Suppose \( X \in \mathbb{R} - \{-1, 0, 1\} \) and \( a^4 \neq b^4 \). Then there is a nonsingular binary recursive gadget \( M \in \mathbb{C}^{3\times 3} \) and a starter gadget \( s \in \mathbb{C}^3 \) for which \( \{M^j s\}_{j\geq 0} \) is a sequence of pairwise linearly independent symmetric binary generator signatures.

**Proof.**

\[
M_{19} = \begin{bmatrix} a^4 & 2a & b^2 \\ a^3 & 1 + ab & b^3 \\ a^2 & 2b & b^4 \end{bmatrix}
\]

\[
M_{20} = \begin{bmatrix} a^4 & 2a^2b & b^2 \\ a^3 & ab + a^2b^2 & b^3 \\ a^2 & 2ab^2 & b^4 \end{bmatrix}
\]
We calculate \( \det(M_{19}) = X^2(X - 1)^3 \neq 0 \) and
\[
\det([s M_{19}s M_{19}^2s]) = -(a^4 - b^4)(X + 1)(X - 1)^4 \neq 0,
\]
where \( s \) is the single vertex starter gadget. If gadget \( M_{19} \) also has infinite projective order then we are done by Lemma 5.34 and Lemma 5.33, so assume otherwise. Then for some integer \( i \geq 0 \) and nonzero \( \lambda \in \mathbb{C} \) we have \( M_{19}^i = \lambda M_{19}^{-1} \), and we construct the gadget \( M = M_{19}^i M_{20} = \lambda M_{19}^{-1} M_{20} = \lambda \left[ \begin{array}{ccc} 1 & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & 1 \end{array} \right] \). This is really an application of the idea of an anti-gadget. Then \( M^j s = \lambda^j \text{diag}(1, X^j, 1)[a \ b]^T = \lambda^j[a \ X^j \ b]^T \). Since \( X \in \mathbb{R} - \{-1, 0, 1\} \), we have that \( \{M^j s\}_{j \geq 0} \) is a sequence of pairwise linearly independent symmetric binary generator signatures. \( \square \)

**Lemma 5.50.** Suppose \( X = -1 \) and \( Y \in \mathbb{R} - \{-2, 0, 2\} \). Then there is a nonsingular binary recursive gadget \( M \in \mathbb{C}^{3 \times 3} \) and starter gadget \( s \in \mathbb{C}^3 \) for which \( \{M^j s\}_{j \geq 0} \) is a sequence of pairwise linearly independent symmetric binary generator signatures.

**Proof.**
\[
M_{21} = \begin{bmatrix}
  a^8 + a^5b + a^6b^2 + a^3b^3 & 2a^5 + 4a^3b^2 + 2ab^4 & a^3b + ab^3 + a^2b^4 + b^6 \\
  a^7 + a^4b + a^5b^2 + a^2b^3 & a^4 + a^5b + 2a^2b^2 + 2a^3b^3 + b^4 + ab^5 & a^3b^2 + ab^4 + a^2b^5 + b^7 \\
  a^6 + a^3b + a^4b^2 + ab^3 & 2a^4b + 4a^2b^3 + 2b^5 & a^3b^3 + ab^5 + a^2b^6 + b^8
\end{bmatrix}
\]

If \( X = -1 \), then the characteristic polynomial of gadget \( M_{21} \) is
\[
x^3 - (Y + 2)(Y - 2)x^2 - 2(Y + 2)^2(Y - 2)x - 8(Y + 2)^2(Y - 2),
\]
so the determinant is nonzero. By Lemma 5.40, if all roots of the characteristic polynomial have the same norm, then \( C^3 = B^3 D \) and this amounts to \( (Y + 2)^2 = -(Y + 2)(Y - 2) \), but then \( Y(Y + 2) = 0 \), which is not true. Finally, applying Lemma 5.34 to \( M_{21} \) and the single-vertex starter gadget \( s \), we get \( \det([s M_{21}s M_{21}^2s]) = 16Y^2(a^4 - b^4)(Y + 2) \neq 0 \). \( \square \)

**Lemma 5.51.** Suppose \( X = 0 \) and \( Y \in \mathbb{C} - \{0\} \). Then there exists a unary recursive gadget \( M \) and gadgets \( A \) and \( B \) such that \( \{AM^jB\}_{j \geq 0} \) is a sequence of pairwise linearly independent symmetric binary generator signatures.
Figure 5.16: Unary recursive gadgets

Proof.

\[
M_{22} = \begin{bmatrix}
1 + a^4 & a + b^3 \\
 a^3 + b & 1 + b^4
\end{bmatrix}
\]

\[
M_{23} = \begin{bmatrix}
2a^4 + a^8 + 2ab + a^5b + a^3b^3 + b^4 & a^5 + 2a^2b + a^3b^2 + b^3 + 2ab^4 + b^7 \\
 a^3 + a^7 + 2a^4b + 2ab^2 + a^2b^3 + b^5 & a^4 + 2ab + a^3b^3 + 2b^4 + ab^5 + b^8
\end{bmatrix}
\]

By $X = 0$ and $Y \neq 0$, without loss of generality we assume $a = 0$, and $b \neq 0$.

First we show the existence of a unary recursive gadget $M$ which has infinite projective order. Under the condition $X = 0$, det$(M_{22}) = 1$. If gadget $M_{22}$ has infinite projective order then we define $M = M_{22}$ and we are already done. So suppose otherwise. Then for some integer $i \geq 0$ and nonzero $\lambda \in \mathbb{C}$ we have $M_{22} = \lambda M_{22}^{-1}$, and we consider the gadget $M = M_{22}^{-1} M_{23} = \lambda b^4 \begin{bmatrix} 1 & b^{-1} \\ 0 & 1 \end{bmatrix}$, which is nonsingular. Then $M^j = \lambda^j b^{4j} \begin{bmatrix} 1 & jb^{-1} \\ 0 & 1 \end{bmatrix}$ for any positive integer $j$, so $M$ has infinite projective order.

Now we show the existence of appropriate gadgets $A$ and $B$ to symmetrize the construction. If $M = M_{22}$, then we set $A$ to be the identity matrix $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (representing an empty gadget with no vertex and a single straight through edge) and $B$ to be a single generator vertex with one leading edge to be merged with the trailing edge of the recursive gadget, and the other edge becoming a dangling edge. The induced graph symmetry implies that $AM^j B$ is a symmetric binary generator. Furthermore, $A$ and $B$ are nonsingular. If $M = M_{22} M_{23}$, then we set $A$ to $M_{23}$ and $B$ to a single generator vertex with one leading edge to be merged with the trailing edge of the recursive gadget, and the other edge becoming a dangling edge. Again, the induced graph symmetry implies that $AM^j B$ is a symmetric binary generator. Furthermore, $B$ is nonsingular, as $X \neq 1$, and $A$ is nonsingular, since if we are using this $M$ as our unary recursive gadget, then we already know that $M_{22} M_{23}$ is nonsingular, therefore $M_{23}$ is nonsingular. In any case, the linear transformation $M^j \mapsto AM^j B$ preserves pairwise linear independence, and we have constructed a sequence of pairwise linearly independent symmetric binary generator signatures.

Now we show the existence of a set of general-purpose finisher gadgets. In these finisher gadgets, all dangling edges are internally incident to vertices which are assigned a recognizer.
signature (=4). Furthermore, one set of 2 dangling edges are incident to the same (=4) vertex. Hence, we are aiming to interpolate recognizer signatures of the form [*,0,*]. These finisher gadgets operate under the assumption that the “incoming” gadget has a binary symmetric generator signature. Our finisher gadgets will be expressed as 3 by 3 matrices (of rank 2), since the incoming signatures are symmetric. The essential property is that for a trio of finisher gadgets, their 1-dimensional kernels are linearly independent.

Lemma 5.52. Let \(a, b \in \mathbb{C}\) such that \(ab \neq 1\), and \(a^4 \neq b^4\). Then in the context of \(\text{Hol}_4(a,b)\), there exist three binary finisher gadgets with rank 2 matrices \(F, F', F'' \in \mathbb{C}^{3 \times 3}\), such that their middle rows are all zero, and the kernels of \(F, F',\) and \(F''\) are linearly independent.

Proof.

\[
M_{24} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix},
\quad
M_{25} = \begin{bmatrix}
a^8 + 2a^5b + a^2b^2 & a^7 + 2a^6b + ab^2 & a^6 + 2a^4b + b^2 \\
2a^5 + 4a^3b^2 + 2ab^4 & a^4 + a^5b + 2a^2b^2 + 2a^3b^3 + b^4 + ab^5 & 2a^4b + 4a^2b^2 + 2b^5 \\
a^2 + 2ab^3 + b^6 & a^2b + 2ab^4 + b^7 & a^2b^2 + 2ab^5 + b^8
\end{bmatrix}
\]

We will break this proof down into two cases: \(X \neq 0\) and \(X = 0\). For both cases, we will show the existence of three finisher gadgets for which the (1-dimensional) kernels are linearly independent. Every finisher gadget we utilize in this proof has the property that the middle row of the corresponding transition matrix contains only zeros, hence we can get a nonzero vector \(v\) that spans its kernel by calculating the cross product of the first and last rows (provided that the matrix has rank 2). Suppose \(F, F',\) and \(F''\) are finisher gadgets and let \(v_1, v_2,\) and \(v_3\) denote the respective cross products of the first and last rows of each transition matrix. Then we use \(\text{cross}(F, F', F'')\) to denote the matrix whose rows are \(v_1, v_2,\) and \(v_3\) (in that order). We can test the linear independence of \(\{v_1, v_2, v_3\}\) by verifying the determinant \(\det(\text{cross}(F, F', F'')) \neq 0\).

Suppose \(X \neq 0\). Let \(F = M_{24}, F' = M_{24}M_{19}, F'' = M_{24}M_{19}^2\). Since \(M_{24}\) has rank 2 and \(\det(M_{19}) = X^2(X - 1)^3 \neq 0\), it follows that \(F, F',\) and \(F''\) all have rank 2. We calculate that \(\det(\text{cross}(F, F', F'')) = 4X^2(a^4 - b^4)(X - 1)^4 \neq 0\), so the kernels are linearly independent.
Suppose $X = 0$. Without loss of generality we have $a = 0$ and $b \neq 0$ since $a^4 \neq b^4$. Let $F = M_{24}$, $F' = M_{24}M_{25}$, $F'' = M_{24}M_{25}^2$. Then $M_{24}$ has rank 2,

$$\det(M_{25}) = (ab - 1)^3(a^3 + b)(a^2 + b^2)^2(a + b)^2, \text{ and}$$
$$\det(\text{cross}(F, F', F'')) = 4(ab - 1)^4(a^2 - b^2)(a^3 + b)(a^2 + b^7)(a + b)^2,$$

Under the condition $a = 0$ these polynomials simplify to $-b^{12} \neq 0$ and $-4b^{28} \neq 0$. Therefore $F$, $F'$, and $F''$ all have rank 2 and the kernels of $F$, $F'$, and $F''$ are linearly independent. □

**Corollary 5.53.** Let $a, b \in \mathbb{C}$ be parameters in Region III, namely, $(a, b)$ are such that $X \neq 1$, $Y^2 \neq 4X^4$ (equivalently $a^4 \neq b^4$), and $(X, Y) \neq (-1, 0)$. Then $\text{Hol}_4(a, b)$ is $\#P$-hard even when restricted to planar inputs.

**Proof.** We have produced a sequence of pairwise linearly independent symmetric binary generator signatures by the following lemmas: If $X \not\in \mathbb{R}$, then we have Lemma 5.48 that uses the syzygy construction. Next we assume $X \in \mathbb{R}$ but $X \neq 1$. If further $X \notin \{0, -1\}$, then with $a^4 \neq b^4$, we can apply Lemma 5.49. Suppose $X = 0$, then for all $Y \in \mathbb{C} - \{0\}$, we have Lemma 5.51. Finally suppose $X = -1$. In this case if $Y \not\in \mathbb{R}$, then we appeal to Lemma 5.48 again. If $X = -1$ and $Y \in \mathbb{R}$, then by the hypothesis of this Corollary, $Y \in \mathbb{R} - \{-2, 0, 2\}$. Then we apply Lemma 5.50.

Now we apply Lemma 5.52 and Lemma 5.35 to the sequence of pairwise linearly independent symmetric binary generator signatures just produced, and obtain a sequence of pairwise linearly independent symmetric binary recognizer signatures of the form $\{[x_1, 0, y_1], \ldots, [x_m, 0, y_m]\}$.

Finally we apply Lemma 5.31. The conclusion of Corollary follows. We observe that all reductions use only planar gadgets, and Lemma 5.30 is valid for planar inputs, thus the $\#P$-hardness proved holds over planar instances of $\text{Hol}_4(a, b)$. □

### 5.2.3 Planar Tractable Region II is $\#P$-hard for General Graphs

Recall that Region II is specified by $Y^2 = 4X^4$ (equivalently $a^4 = b^4$) but $X \not\in \{0, \pm 1\}$.

It will be convenient to transform the problem to a setting where generator signatures are of the form $[a, 1, a]$. This can be done with a holographic transformation, and we can also reuse some of our work, using the observation that a sequence of pairwise linearly independent signatures remains so after a holographic transformation.

**Lemma 5.54.** Let $f$ and $f'$ be $\mathcal{F}$-gates of arity $r$, where each dangling edge is internally incident to a vertex labeled with a generator signature. Then the property that $f$ and $f'$ are linearly independent is invariant under holographic reductions.

**Proof.** Let $T \in \mathbb{C}^{2 \times 2}$ be the basis for a holographic reduction. View the signatures $f$ and $f'$ as column vectors in $\mathbb{C}^2$ in unsymmetric signature notation. After a holographic reduction
under basis $T$, we get the signatures $\hat{f}$ and $\hat{f}'$, where $[\hat{f} \hat{f}'] = (T^{-1})^{\otimes r}[f f']$. Since $T$ is invertible, the $2^r \times 2^r$ block matrix $[f f']$ has rank 2 if and only if the $2^r \times 2^r$ block matrix $[\hat{f} \hat{f}']$ has rank 2, i.e. $f$ and $f'$ are linearly independent if and only if $\hat{f}$ and $\hat{f}'$ are linearly independent.

**Lemma 5.55.** Suppose $a^4 = b^4$. Then there exists $c \in \mathbb{C}$ such that the problem Holant($=4$ $[a, 1, b]$) is equivalent to the problem Holant($[1, 0, 0, 0, e] \mid [c, 1, c]$) under a holographic reduction, where $ab = c^2$ and $e = \pm 1$. Conversely, given a problem of the form Holant($[1, 0, 0, 0, e] \mid [c, 1, c]$) where $c \in \mathbb{C}$ and $e = \pm 1$, there is a holographic reduction to an equivalent problem Holant($=4$ $[a, 1, b]$) for which $ab = c^2$.

**Proof.** If $ab = 0$, then $a = b = 0$ by $a^4 = b^4$. In this case, the lemma is trivial. Suppose otherwise, then both $a \neq 0$ and $b \neq 0$.

Let $\omega$ be a 16th root of unity such that $a \omega^4 = b$. By the holographic transformation by $T = \begin{bmatrix} \omega^{-1} & 0 \\ 0 & \omega \end{bmatrix}$, we have $(T^{-1})^{\otimes 2}[a, 1, b]^T = [a \omega^2, 1, b \omega^{-2}]^T = [a \omega^2, 1, a \omega^2]^T$, and $[1, 0, 0, 0, 1]^T[T^{\otimes 2}] = [\omega^{-4}, 0, 0, 0, \omega^4]$, so Hol$_4(a, b)$ is equivalent to Holant($[\omega^{-4}, 0, 0, 0, \omega^4] \mid [a \omega^2, 1, a \omega^2]$). Multiplying each entry of a signature by a nonzero value does not change the complexity of the problem, and $\omega^8 \in \{\pm 1\}$, so the problem is equivalent to Holant($[1, 0, 0, 0, \pm 1] \mid [a \omega^2, 1, a \omega^2]$).

Conversely, consider a problem of the form Holant($[1, 0, 0, 0, e] \mid [c, 1, c]$) where $c \in \mathbb{C}$ and $e = \pm 1$. If $e = 1$, then we are already done, so assume $e = -1$. Let $\omega = e^{\pi i/8}$, and performing a holographic reduction by $T = \begin{bmatrix} \omega^{-1} & 0 \\ 0 & \omega \end{bmatrix}$, we arrive at Holant($[-i, 0, 0, 0, -i] \mid [\omega^2c, 1, \omega^{-2}c]$), which is equivalent to Holant($=4$ $[a, 1, b]$), where $a = \omega^2c$, $b = \omega^{-2}c$, $ab = c^2$, and $a^4 = b^4$.

We observed in Lemma 5.8 that if all signatures in $\mathcal{F}$ are complement-invariant, then any $\mathcal{F}$-gate that we build with them is also complement-invariant. Although we can’t get the same conclusion in the setting of Holant($[1, 0, 0, 0, -1] \mid [a, 1, a]$), there is a similar structural property to observe.

**Lemma 5.56.** Let $\mathcal{F} = \{[1, 0, 0, 0, -1], [a, 1, a]\}$. Then any $\mathcal{F}$-gate with an even number of degree-4 vertices has the property that its value remains unchanged under the complement of its inputs.

**Proof.** Let $f \in \mathcal{F}$. If $\sigma$ and $\sigma'$ are complementary assignments to the edges of $f$, then their values satisfy $f \mid_\sigma = \pm f \mid_{\sigma'}$, where the sign is $-1$ if $f$ is the function $[1, 0, 0, 0, -1]$, and the sign is $+1$ if $f = [a, 1, a]$. If $\sigma$ and $\sigma'$ are complementary assignments to the dangling edges of an $\mathcal{F}$-gate, by complementing the assignments to all internal edges, we have a 1-1 correspondence of all terms defining the signature sum of the $\mathcal{F}$-gate under $\sigma$ with those under $\sigma'$, and since the number of occurrences of $[1, 0, 0, 0, -1]$ is even, each pair of complementary assignments has the same value.
The following lemma is practically identical to Lemma 5.49, except we carry out the proof in Region II. Note we do the analysis with a 2 by 2 iteration matrix due to the extra degree of symmetry we have in this setting.

**Lemma 5.57.** Consider the problem Holant([1, 0, 0, 0, e] | [a, 1, a]), where $a^2 \in \mathbb{R} - \{0, \pm 1\}$ and $e = \pm 1$. Then there exists a binary starter gadget $s$ and a binary recursive gadget $M$ such that $\{M^js\}_{j \geq 0}$ is a sequence of pairwise linearly independent generator signatures of the form $\{[x_j, y_j, x_j]\}_{j \geq 0}$.

**Proof.** In this proof, we will use $M_{19}$, $M_{20}$, and the single-vertex starter gadget $s$.

$$M_{19} = \begin{bmatrix} a^4 & 2ae & a^2 \\ a^3 & e + a^2e & a^3 \\ 1 & 2ae & a^4 \end{bmatrix}, \quad M_{20} = \begin{bmatrix} a^4 & 2a^3e & a^2 \\ a^3 & a^2e + a^4 & a^3 \\ a^2 & 2a^3e & a^4 \end{bmatrix}, \quad s = \begin{bmatrix} a \\ 1 \\ a \end{bmatrix}.$$  

Let $C = \{[x y z]^T : x, y \in \mathbb{C}\}$. Then $s \in C$ and $C$ is closed under $M_{19}$ and $M_{20}$. For this reason, all $F$-gate signatures in this proof will be in $C$ and we can restrict our attention to the first two entries of each signature. As such, we form $s'$, $M'$, and $M''$ as stand-ins for $s$, $M_{19}$, and $M_{20}$ (respectively) as follows,

$$s' = \begin{bmatrix} a \\ 1 \end{bmatrix}, \quad M' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a^4 & 2ae & a^2 \\ a^3 & e + a^2e & a^3 \\ a^2 & 2ae & a^4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a^4 + a^2 & 2ae \\ 2a^3e & e + a^2e \end{bmatrix},$$

$$M'' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a^4 & 2a^3e & a^2 \\ a^3 & a^2e + a^4 & a^3 \\ a^2 & 2a^3e & a^4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a^4 + a^2 & 2a^3e \\ 2a^3e & a^2e + a^4 \end{bmatrix}.$$  

Note that $M'' = M'[\begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}]$. First, $\det([sM's]) = a(a^2 - 1)(a^2 + e) \neq 0$, so if $M'$ has infinite projective order then we are done by Lemma 5.34 and Lemma 5.15. Now assume otherwise, and since $\det(M') = a^2e(a^2 - 1)^2 \neq 0$ and $M'$ has finite projective order, there is some integer $i \geq 0$ and nonzero $\lambda \in \mathbb{C}$ for which $M'^i = \lambda M'^{i-1}$. Then we construct the gadget $M = M'^i M'' = \lambda M'^{i-1} M'' = \lambda [\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ a \end{bmatrix}]$. Then $M' s = \lambda^i \text{diag}(1, a^{2j})[a \cdot 1]^T = \lambda^j [a a^{2j}]^T$, hence $\{M^j s\}_{j \geq 0}$ is a sequence of pairwise linearly independent vectors.  

**Lemma 5.58.** Consider the problem Holant([1, 0, 0, 0, e] | [c, 1, c]), where $c^2 \notin \mathbb{R}$ and $e = \pm 1$. Then there exists a unary recursive gadget $M$ and gadget $A$ such that $\{M^j A\}_{j \geq 0}$ is a sequence of pairwise linearly independent generator signatures of the form $\{[x_j, y_j, x_j]\}_{j \geq 0}$.

**Proof.** By Lemma 5.55 let $a, b \in \mathbb{C}$ such that Holant([1, 0, 0, 0, e] | [c, 1, c]) is equivalent to Holant([c, 1, b]) and $ab = c^2$. Then by Lemma 5.48, since $X = ab = c^2 \notin \mathbb{R}$, we have a recursive unary gadget $M$ and gadget $A$ such that $\{M^j A\}_{j \geq 0}$ is a sequence of pairwise linearly independent symmetric binary generator signatures (in the context of Holant(=4 [a, 1, b])).

174
By Lemmas 5.54 and 5.55, this gadget construction also produces a sequence of pairwise linearly independent binary generator signatures in the context of Holant([1, 0, 0, 0, e] | [c, 1, c]). All of the gadgets from Lemma 5.48 have an even number of degree-4 vertices, so by Lemma 5.56 (and Lemma 5.8) the signatures of the F-gates we constructed are invariant under the complement of their inputs.

The following lemma proves the \#P-hardness for Hol_4(a, b) in Region II for general graphs.

**Lemma 5.59.** Let \( a, b \in \mathbb{C} \) such that \( a^4 = b^4 \) and \( ab \notin \{-1, 0, 1\} \). Then Hol_4(a, b) is \#P-hard.

**Proof.** By Lemma 5.55 there exists \( c \in \mathbb{C} \) and \( e = \pm 1 \) such that \( ab = c^2 \) and Hol_4(a, b) is equivalent to Holant([1, 0, 0, 0, e] | [c, 1, c]). Then \( c^2 = ab \notin \{0, \pm 1\} \) and by Lemmas 5.58 and 5.57 we have a sequence of pairwise linearly independent generator signatures of the form \( \{[x_j, y_j, x_j]\}_{j \geq 0} \). By Lemma 5.13, we can interpolate generator signature [1, 1, 1], which we use in the following chain of reductions.

\[
\text{Hol}_3(\text{ec}, \text{ec}) \leq_{PT} \text{Holant}([1, 0, 0, e] | [c, 1, c]) \\
\leq_{PT} \text{Holant}([1, 0, 0, 0, e] | \{[c, 1, c], [1, 1, 1]\}) \\
\leq_{PT} \text{Holant}([1, 0, 0, 0, e] | [c, 1, c]) \\
\leq_{PT} \text{Holant}(=_4 | [a, 1, b])
\]

To justify this chain of reductions, observe that any instance of \( \text{Hol}_3(\text{ec}, \text{ec}) = \text{Holant}(=_3 | \text{ec}, 1, \text{ec}) \) must have an even number of recognizer vertices \( =_3 \). The first step follows from a holographic reduction under basis \( \begin{bmatrix} 1 & 0 \\ 0 & e \end{bmatrix} \), so that Hol_3(ec, ec) is equivalent to Holant([1, 0, 0, e] | [ec, e, ec]), as \( e^2 = 1 \), which has the same complexity as Holant([1, 0, 0, e] | [c, 1, c]). Note the number of recognizer vertices is unchanged, thus still even. For the next step of the reduction, any pair of vertices with signature [1, 0, 0, e] can be simulated by a pair of vertices with signature [1, 0, 0, 0, e]. Simply introduce a vertex with generator signature \( [1, 1, 1] = [1, 1] \otimes 2 \) and make it adjacent to both recognizer vertices. (This step of the reduction is in general not planar.) The last two steps are carried out by interpolation and by a holographic reduction. Since \( ec \notin \{0, \pm 1, \pm i\} \), by Lemma 5.14 (also by Theorem 5.1 for the case \( k = 3 \) already proved at the end of Section 5.1) we know that Hol_3(ec, ec) is \#P-hard and we are done.

**Proof of Theorem 5.1 restricted to \( k = 4 \)**

**Proof.** Tractability (for Regions I and II) is given by Theorem 5.47. Intractability over Region II for non-planar graphs is given by Lemma 5.59. Intractability over Region III even for planar graphs is proved in Corollary 5.53.

175
We conclude this Chapter by giving an equivalent statement of Theorem 5.1 (hence also equivalent to Theorem 5.3), the dichotomy theorem for spin systems on regular-$k$ graphs. Note that in this book we have only proved the cases when $k = 3$ and $k = 4$, and that will be the only cases we will use in what follows. A full proof of Theorem 5.3 can be found in [?]. The following equivalent statements are in a transformational form, which are more conceptual than Theorem 5.3, and will be consistent with the general dichotomy expression in Theorem 7.19 in Chapter 7.

Recall the definition of functions of product type $\mathcal{P}$ (Definition 3.2), and functions of affine type $\mathcal{A}$ (Definition 3.6).

**Theorem 5.60.** Suppose $x, y, z \in \mathbb{C}$ and $k \geq 3$. Then $Z(G)$ for a spin system on $k$-regular graphs $G$ with edge function $g = [x, y, z]$ is $\#P$-hard except in the following cases, when the problem is in $P$: There exists a holographic transformation $T$ such that

1. $g \in T\mathcal{P}$ and $(=k)T^{\otimes k} \in \mathcal{P}$, or
2. $g \in T\mathcal{A}$ and $(=k)T^{\otimes k} \in \mathcal{A}$.

If we restrict the input to planar $k$-regular graphs, then these are solvable in $P$, as well as when there exists a holographic transformation $T$ such that $g \in T\mathcal{M}$ and $(=k)T^{\otimes k} \in \mathcal{M}$, where $\mathcal{M}$ denotes all matchgate signatures. The problem remains $\#P$-hard in all other cases.

**Proof.** We first show that condition 1 is equivalent to $g \in \mathcal{P}$. It is obvious that any $g \in \mathcal{P}$ satisfies condition 1 with $T = I$. Suppose $g$ satisfies condition 1 with $T$. Since $(=k)T^{\otimes k} \in \mathcal{P}$ is symmetric and $k \geq 3$, it follows that $(=k)T^{\otimes k}$ is a GEN-EQ, expressible as $[a, 0]^{\otimes k} + [0, b]^{\otimes k}$, where $ab \neq 0$ because $(=k)T^{\otimes k}$ is non-degenerate. So $(=k)(TD)^{\otimes k} = (=k)$ where $D = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$. Let $T_1 = TD = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$. Then

$$[\alpha, \beta]^{\otimes k} + [\gamma, \delta]^{\otimes k} = [1, 0]^{\otimes k} + [0, 1]^{\otimes k}.$$ 

Dot product with $[\beta, \alpha]$ and then with $[\delta, \gamma]$ imply that $\beta \delta [1, 0]^{\otimes (k-2)} + \alpha \gamma [0, 1]^{\otimes (k-2)} = 0$. Since $k \geq 3$, we get $\alpha \gamma = \beta \delta = 0$. So $T_1$ is either diagonal or anti-diagonal since it is non-singular, thus $T_1\mathcal{P} = \mathcal{P}$. We are also given $g \in T\mathcal{P}$. Since $D$ is diagonal, $D\mathcal{P} = \mathcal{P}$. Thus $g \in T\mathcal{P} = T D \mathcal{P} = T_1\mathcal{P} = \mathcal{P}$.

Clearly if $y = 0$ in $g = [x, y, z]$ then $g$ is a GEN-EQ, and $g \in \mathcal{P}$. Suppose $y \neq 0$ and we normalize it and write $g = [a, 1, b]$. Then in Theorem 5.3 conditions 1 and 2 correspond to $g$ being degenerate and $g$ being a binary DISEQUALITY. Together these cases are exactly $g \in \mathcal{P}$.

In the following we assume $g \notin \mathcal{P}$. It is easy to verify that condition 3 in Theorem 5.3 is equivalent to $a^{4k} = 1$ and $b = -1/a$. If so, we take $T = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}$. Then $(=k)T^{\otimes k} \in \mathcal{A}$, being in the family $\mathcal{F}_1$ of Section 3.2. Furthermore $(T^{-1})^{\otimes 2}(g) = [a, a, ba^2] = a[1, 1, -1] \in \mathcal{A}$. 

176
Now suppose there exists some $T$ such that $(=k)T^\otimes k \in \mathcal{A}$ and $g \in T\mathcal{A}$. Denote $\mathcal{F}_{123} = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ (defined in Section 3.2), then since the signature $(=k)T^\otimes k$ is non-degenerate and symmetric, $(=k)T^\otimes k \in \mathcal{F}_{123}$. By its form, there exists a matrix $M \in \left\{ I, \begin{bmatrix} 1 & 1 \\
-1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & i \\
i & -i \end{bmatrix} \right\}$ such that $(=k)(TM^{-1})^\otimes k$ has the form $\lambda \left\{ [1,0]^\otimes k + iv[0,1]^\otimes k \right\}$ for some $\lambda \neq 0$ and $r \in \{0, 1, 2, 3\}$. Let $\omega = e^{i \frac{\pi}{4k}}$, then $\omega^k = i^r$. Let $T_1 = TM^{-1}\begin{bmatrix} 1 & 0 \\
o & \omega \end{bmatrix}^{-1}$, then $(=k)T_1^\otimes k = \lambda \left\{ [1,0]^\otimes k + [0,1]^\otimes k \right\} = \lambda(=k)$. We write $T_1 = \begin{bmatrix} \alpha \\
\gamma \\
\delta \end{bmatrix}$, then
\[
[\alpha, \beta]^\otimes k + [\gamma, \delta]^\otimes k = \lambda \left\{ [1,0]^\otimes k + [0,1]^\otimes k \right\}.
\]
Dot product with $[-\beta, \alpha]$ and then with $[-\delta, \gamma]$ imply that $\alpha \gamma = \beta \delta = 0$ as before. So $T_1$ has the form $\alpha \begin{bmatrix} 1 & 0 \\
0 & \rho \end{bmatrix}$ or $\alpha X \begin{bmatrix} 1 & 0 \\
0 & \rho \end{bmatrix}$, where $\alpha \neq 0$, $X = \begin{bmatrix} 0 & 1 \\
1 & 0 \end{bmatrix}$, and $\rho^k = 1$. It follows that $T = \alpha X' \begin{bmatrix} 1 & 0 \\
0 & \omega \end{bmatrix} M$ where $X' = I$ or $X$, $\omega' = \omega \rho$ and $\omega'^{4k} = 1$.

It is also given that $g \in T\mathcal{A}$. Since $(T^{-1})^\otimes 2 g$ is a symmetric signature, $(T^{-1})^\otimes 2 g \in \mathcal{F}_{123}$. By the form of $M$, $M\mathcal{F}_{123} = \mathcal{F}_{123}$. Hence $(\begin{bmatrix} 1 & 0 \\
0 & \omega \end{bmatrix}^{-1}X')^\otimes 2 [a, 1, b] \in \mathcal{F}_{123}$. Then we have $[a\omega', 1, b/\omega'] \in \mathcal{F}_{123}$, or $[b\omega', a/\omega'] \in \mathcal{F}_{123}$. By their forms in Section 3.2, and having assumed $g \notin \mathcal{P}$, we have $ab = -1$ and $a^k = 1$.

Concerning $k$-regular planar graphs, we show that there exists a holographic transformation $T$ such that $[a, 1, b] \in T\mathcal{M}$ and $(=k)T^\otimes k \in \mathcal{M}$ iff $a^k = b^k$. Suppose $a^k = b^k$. Let $T = DH_2$, where $D = \begin{bmatrix} 1 & 0 \\
0 & \omega \end{bmatrix}$, $\omega = \sqrt{\frac{\pi}{2}}$ with $\omega^2 = 1$, and $H_2 = \begin{bmatrix} 1 & 1 \\
1 & -1 \end{bmatrix}$. Then $(=k)D^\otimes k = [1,0]^\otimes k + \omega^{-k}[0,1]^\otimes k$, and $(=k)T^\otimes k$ is a nonzero multiple of $[1,1]^\otimes k \pm [1,-1]^\otimes k \in \mathcal{M}$. Also $(D^{-1})^\otimes 2 [a, 1, b] = [a, \omega, b\omega^2] = [a, \omega, a]$, and $(T^{-1})^\otimes 2 [a, 1, b]$ has the matrix form
\[
\begin{bmatrix}
\frac{1}{2} & 1 & 1 \\
1 & -1 & 1 \\
\omega & a & \omega \end{bmatrix} = \begin{bmatrix}
\frac{1}{2} & 1 & 1 \\
1 & -1 & 1 \\
0 & a & \omega \end{bmatrix},
\end{bmatrix}
\]
which is in $\mathcal{M}$.

Now suppose $T = \begin{bmatrix} \alpha \\
\gamma \\
\delta \end{bmatrix}$, and both $(=k)T^\otimes k \in \mathcal{M}$ and $[a, 1, b] \in T\mathcal{M}$. By the parity requirements for a matchgate signature $(=k)T^\otimes k = [\alpha, \beta]^\otimes k + [\gamma, \delta]^\otimes k \in \mathcal{M}$ we have either
\[
\alpha^k + \gamma^k = \alpha^{k-2}\beta^2 + \gamma^{k-2}\delta^2 = 0,
\]
(5.3)
or
\[
\alpha^{k-1}\beta + \gamma^{k-1}\delta = \alpha^{k-3}\beta^3 + \gamma^{k-3}\delta^3 = 0.
\]
(5.4)
First we claim that none of $\alpha, \beta, \gamma, \delta$ can be 0. Suppose $\alpha = 0$, then $\gamma = 0$ or $\gamma^{-1}\delta = 0$. $T$ is non-singular, it implies that $\delta = 0$, and $\beta \gamma \neq 0$. But clearly for such a $T$, $(=k)T^\otimes k \notin \mathcal{M}$ by Theorem 4.11, since $k \geq 3$. Hence $\alpha \neq 0$. By the invariance of $\mathcal{M}$ under $X = \begin{bmatrix} 0 & 1 \\
1 & 0 \end{bmatrix}$, the proof that $\alpha \neq 0$ also gives $\beta \neq 0$. By the symmetry between $[\alpha, \beta]$ and $[\gamma, \delta]$ we also have $\gamma \delta \neq 0$.

Then dividing in either set of the equations (5.3) and (5.4), we get $\frac{\alpha^2}{\beta^2} = \frac{\gamma^2}{\delta^2}$. By $\det(T) = \alpha \delta - \beta \gamma \neq 0$, we get
\[
\frac{\alpha}{\beta} = \frac{-\gamma}{\delta}.
\]
(5.5)
From (5.3) we get \((\alpha/\gamma)^k = -1\). From (5.4), we get \((\alpha/\gamma)^{k-1} = -\delta/\beta\), and together with (5.5) we get \((\alpha/\gamma)^k = 1\). Thus \((\alpha/\gamma)^{2k} = 1\) from either (5.3) or (5.4).

It follows that \(T = \begin{bmatrix} \alpha & \beta \\ \alpha \omega & -\beta \omega \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \omega \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}\) for some \(\omega^{2k} = 1\).

Now we consider \((T^{-1})^\otimes^2[a, 1, b] \in \mathcal{M}\), where \(T^{-1}\) takes the form \(\begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}\), where \(xy \neq 0\) and \(\rho^{2k} = 1\). Then we can calculate the binary signature \((T^{-1})^\otimes^2[a, 1, b]\) by its matrix form to get
\[
(T^{-1})^\otimes^2[a, 1, b] = [x^2(a + 2\rho + b\rho^2), xy(a - b\rho^2), y^2(a - 2\rho + b\rho^2)].
\]

By the parity requirements for a matchgate signature, and \(xy \neq 0\), we get either \(a - b\rho^2 = 0\) or \(a + 2\rho + b\rho^2 = a - 2\rho + b\rho^2 = 0\). The latter leads to \(\rho = 0\), a contradiction, thus \(a = b\rho^2\). Taking \(k\)-th power, we get \(a^k = b^k\). \qed
Chapter 6

Holant Problems and \#CSP

We have by now encountered three representative frameworks for counting problems: Holant Problems, \#CSP and Spin Systems. \#CSP can be viewed as the special case of Holant Problems where the local constraint function set is assumed to contain \textsc{Equality} of all arities. If we allow other auxiliary functions to be free, such as all unary functions, we get other restricted Holant Problems, in this case Holant* Problems. Spin Systems can be seen as a further specialization of \#CSP. At the same time Holant Problems can be viewed as \#CSP where each variable appears twice, syntactically.

Their relationship goes deeper than this. We have already seen in Chapter 3 and 5 how results from Holant* Problems can be used to derive dichotomy theorems for Boolean \#CSP and 2-Spin Systems. In this Chapter we will go the other way. We use results from Boolean \#CSP and 2-Spin Systems to derive several results on Holant problems. We introduce, and prove a dichotomy for, a class of Holant problems called Holant\textsubscript{c} Problems, an intermediate class between Holant and Holant* Problems. We also prove a dichotomy for a version of Boolean \#CSP, called \#CSP\textsubscript{d} where each variable appears a multiple of \(d\) times, for some integer \(d \geq 1\). Finally we discuss a concrete problem called the \textsc{Eulerian Orientation} problem, which has a very simple expression as a Holant Problem. While interesting in their own right, the primary purpose of these results is to pave the way for the dichotomy theorem for Holant problems without auxiliary functions in Chapter 7.

In addition to the new theorems, this chapter also introduces a new transformational perspective not only as a proof technique but also how a dichotomy is conceptually understood and even stated. Some group theoretic ideas are naturally introduced in this setting.

6.1 A Dichotomy for a Single Ternary Signature

In this section, we consider the complexity of Holant\((f)\), where \(f = [f_0, f_1, f_2, f_3]\) is a symmetric signature of arity 3. We also denote by Pl-Holant\((f)\) the restriction of Holant\((f)\) to planar graphs. If \(f\) is degenerate, then Holant\((f)\) is trivially tractable, since a degenerate signature
factors as a tensor product of unary signatures, and the signature grid simply decomposes into isolated edges. In the following we assume that $f$ is non-degenerate. In Chapter 2 (Subsection 2.3.1), we have shown that any non-degenerate signature $f = [f_0, f_1, f_2, f_3]$ can be expressed in one of the following three forms of parameterization (with the convention that $\alpha^0 = 1$, and $k\alpha^{k-1} = 0$ if $k = 0$, even when $\alpha = 0$):

1. $f_k = \alpha^{3-k}\beta^k + \gamma^{3-k}\delta^k$, where $\det \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \neq 0$;

2. $f_k = c k \lambda^{k-1} + d\lambda^k$, where $c \neq 0$;

3. $f_k = c(3-k)\lambda^{2-k} + d\lambda^{3-k}$, where $c \neq 0$.

The first form corresponds to the case when the characteristic equation of a second order linear recurrence has two distinct roots and we call it the generic case. The second form, and its reversal in the third form, correspond to the double-root case. We can omit the formal proof for the third form since it is symmetric to the second form.

For the generic case, we can apply a holographic transformation using $T = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$. We have the following reduction:

$$\text{Holant}(f) \equiv_T \text{Holant}(=2|f)$$

$$\equiv_T \text{Holant}([1, 0, 1]^T \otimes^2 | (T^{-1}) \otimes^3 [f_0, f_1, f_2, f_3])$$

$$\equiv_T \text{Holant}(g[1, 0, 0, 1]),$$

where $g = [1, 0, 1]^T \otimes^2$. The crucial step is of course $[f_0, f_1, f_2, f_3] = T \otimes^3 [1, 0, 0, 1]$.

Therefore for the generic case, the complexity of Holant($f$) is equivalent to the spin system on 3-regular graphs Holant($g \#_3$), where $g$ is a symmetric binary signature, for which Theorem 5.1 is a dichotomy, and the case $k = 3$ has been proved in Chapter 5. Theorem 5.3 is an equivalent statement of Theorem 5.1. Theorem 5.60 is a more conceptual form of Theorem 5.1 and 5.3. We state these theorems together for $k = 3$, for easy reference.

**Theorem 6.1.** The problem Holant([x, y, z]|1, 0, 0, 1]) is $\#P$-hard for all $x, y, z \in \mathbb{C}$ except in the following cases, where the problem is solvable in polynomial time: (1) $xz = y^2$; (2) $y = 0$; (3) $x = z = 0$; or (4) $x^{12} = y^{12}$ and $xz = -y^2$. Furthermore, in addition to these cases Pl-Holant($f$) is solvable in polynomial time if $x^3 = z^3$, and remains $\#P$-hard in all other cases.

Equivalently, Holant([x, y, z]|1, 0, 0, 1]) is in $P$ if there exists $T \in \text{GL}_2(\mathbb{C})$ such that

1. $[x, y, z]T^{\otimes^2} \in \mathcal{P}$ and $(=3) \in T\mathcal{P}$, or
2. $[x, y, z]T^{\otimes^2} \in \mathcal{A}$ and $(=3) \in T\mathcal{A}$.

Otherwise it is $\#P$-hard. In addition, Pl-Holant($f$) is in $P$ if there exists $T \in \text{GL}_2(\mathbb{C})$ such that $[x, y, z]T^{\otimes^2} \in \mathcal{M}$ and $(=3) \in T\mathcal{M}$. It remains $\#P$-hard in all other cases.
Remark: The signature \([x, y, z]\) in case (1) is degenerate, which is a special case of product type. Case (2) and (3) are of product type, being binary generalized EQUALITY and DISEQUALITY. These are the only cases for a binary symmetric signature to be of product type. It is easy to show that assuming \(y \neq 0\), condition (4) can be equivalently stated as \((4') x^6 = z^6\) and \(xz = -y^2\). Case (4) includes what is transformable to an affine signature by \(T = \begin{bmatrix} 1 & 0 \\ 0 & \omega \end{bmatrix}\) where \(\omega^3 = 1\) and \(T\) keeps \((=3)\) invariant. One can easily verify that the only matrices \(T\) that keep \((=3)\) invariant are of the form \(\begin{bmatrix} \omega & 0 \\ 0 & \omega' \end{bmatrix}\) or \(\begin{bmatrix} 0 & \omega \\ \omega' & 0 \end{bmatrix}\), where \(\omega^3 = \omega'^3 = 1\), and such matrices \(T\) preserve \(\mathcal{P}\). The tractable cases of Theorem 6.1 can also be equivalently stated as follows: There exists some \(T\) such that \([x, y, z]T^{\otimes 2} \in \mathcal{M} \cup \mathcal{P}\) (additionally in \(\mathcal{M}\) for Pl-Holant\((f)\)) and \((T^{-1})^{\otimes 3}(=3) = (=3)\).

To get a dichotomy for Holant\(([f_0, f_1, f_2, f_3])\), we next deal with the double-root case. We have the following lemma:

**Lemma 6.2.** Let \(f_k = ck\lambda^{k-1} + dk^k\), where \(c \neq 0\) and \(0 \leq k \leq 3\). If \(\lambda = \pm i\), then Holant\(([f_0, f_1, f_2, f_3])\) is in \(P\); in fact the Holant value is always 0 in this case. If \(\lambda \neq \pm i\), then Holant\(([f_0, f_1, f_2, f_3])\) is \#P-hard. Additionally, Pl-Holant\((f)\) is in \(P\) if there exists orthogonal \(T \in O_2(\mathbb{C})\) such that \((=2)T^{\otimes 2} = (=2) \in \mathcal{M}\) and \(T^{\otimes 3}[f_0, f_1, f_2, f_3] = [0, 1, 0, 0] \in \mathcal{M}\), and it remains \#P-hard otherwise.

**Proof.** If \(\lambda = \pm i\), the signature \(f = [f_0, f_1, f_2, f_3]\) satisfies the recurrence relation \(f_{k+2} = 2\lambda f_{k+1} + f_k\), where \(k = 0, 1\). This is a generalized Fibonacci signature. Thus Holant\((f)\) is solvable in \(P\). Note that for \(\lambda = \pm i\), \(f\) can also be written as \(Z[u, v, 0, 0]\) or \(Z[0, 0, v, u]\) where \(Z = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\) is defined in (2.11). In fact for \(\lambda = i\) we may take \(u = -(f_0 + 3if_1)/2\) and \(v = (f_0 + f_1)/2\). Similarly for \(\lambda = -i\). Under a holographic transformation by \(Z\), Holant\((f) \equiv_T \text{Holant}(\neq 2) (Z^{-1})^{\otimes 3} f\). As \((Z^{-1})^{\otimes 3} f = [u, v, 0, 0]\) or \([0, 0, v, u]\), in any bipartite signature grid for Holant\((\neq 2) (Z^{-1})^{\otimes 3} f\), the LHS \((\neq 2)\) requires exactly the same number of 0’s and 1’s, while the RHS requires strictly more, or strictly less, 0’s than 1’s. Thus the Holant value is always 0.

In the following we assume \(\lambda \neq \pm i\). We can view Holant\((f)\) as the bipartite Holant\((=2 | f)\). Recall from (2.8) that an orthogonal transformation \(T \in O_2(\mathbb{C})\) keeps the binary EQUALITY (=2) invariant: \((=2)T^{\otimes 2} = (=2)\).

Let \(M = \begin{bmatrix} 1 & \frac{d-1}{c+1}\lambda \\ \lambda & \frac{d+1}{c+1}\lambda \end{bmatrix}\), then the signature \([f_0, f_1, f_2, f_3]\) can be expressed as

\[(f_0, f_1, f_1, f_2, f_1, f_2, f_2, f_3)^T = M^{\otimes 3}(1, 1, 1, 0, 1, 0, 0, 0)^T.\]

Let \(Q = \frac{1}{\sqrt{1 + \lambda^2}}\begin{bmatrix} 1 & \lambda \\ \lambda & -1 \end{bmatrix}\), then \(Q\) is orthogonal. In exactly the same way as in (2.17) from Chapter 2 (replacing \(\lambda\) for \(\alpha\)), the orthogonal transformation \(Q\) transforms \(f\) to \([z, 1, 0, 0]\) for some \(z \in \mathbb{C}\), up to a scalar. Then the complexity of Holant\((f)\) is the same as Holant\(([z, 1, 0, 0])\).
For \( z = 0 \) in the signature \([z, 1, 0, 0]\), it is the problem of counting perfect matchings on 3-regular graphs. Because \([0, 1, 0, 0] \in \mathcal{M}\), Pl-Holant\((f)\) is in P in this case. Over general 3-regular graphs \( \text{Holant}([0, 1, 0, 0]) \) is \#P-hard. 

In the following we assume \( z \neq 0 \). We show that Pl-Holant\(([z, 1, 0, 0])\) is \#P-hard. We can realize a new ternary signature \( f' = [z^3 + 3z, z^2 + 1, z, 1] \) by connecting \([z, 1, 0, 0]'s\) in a triangle gadget as illustrated in Figure 6.1. So if Pl-Holant\((f')\) is \#P-hard then Pl-Holant\( ([z, 1, 0, 0]) \) is also \#P-hard. To see that the gadget in Figure 6.1 has the signature \([z^3 + 3z, z^2 + 1, z, 1]\) note that all three vertices are assigned the weighted matching signature \([z, 1, 0, 0]\) and can have at most one incident edge assigned 1. Thus, for example, on external input 000, there are three ways to assign exactly one internal edge the value 1 giving the value \( 3z \), and one way to assign all internal edges the value 0 giving the value \( z^3 \). The other entries of \( f' \) can be computed similarly.

In tensor product notation this signature is

\[
f' = [z^3 + 3z, z^2 + 1, z, 1]^T = \frac{1}{2} \left( \begin{bmatrix} z + 1 \\ z^2 + 1 \end{bmatrix} \otimes^3 + \begin{bmatrix} z - 1 \\ z^2 - 1 \end{bmatrix} \otimes^3 \right).
\]

Then the following reduction chain holds:

\[
\text{Pl-Holant}(f') \equiv_T \text{Pl-Holant}(\frac{z^3 + 3z}{z^2 + 1}, z, 1) \equiv_T \text{Pl-Holant}([z^2 + 2z + 2, z^2 - 2z + 2, z^2 - 2z + 2] =_3) \tag{6.1}
\]

where the second step is a holographic reduction using \( T = \begin{bmatrix} z^2 + 1 \\ z^2 - 1 \end{bmatrix} \in \text{GL}_2(\mathbb{C}) \), and \( T \) transforms \( (=2) \) to \( (=2)T_{\otimes^2} \), namely \( TIT = \begin{bmatrix} z^2 + 2z + 2 \\ z^2 - 2z + 2 \end{bmatrix} \).

Now we can apply Theorem 6.1 to the Holant problem in (6.1), and check against the tractable cases.

1. \([z^2 + 2z + 2, z^2, z^2 - 2z + 2] \) is non-degenerate. This is because it is the transformed binary signature from the non-degenerate \( (=2) \) under the non-singular matrix \( T \in \text{GL}_2(\mathbb{C}) \).
2. \( z^2 \neq 0 \), since by assumption \( z \neq 0 \).
3. It cannot be the case that \( z^2 + 2z + 2 = z^2 - 2z + 2 = 0 \), again by \( z \neq 0 \).
4. We verify the equivalent condition (4') stated in the Remark after Theorem 6.1, that there is no solution to \( (z^2 + 2z + 2)^6 = (z^2 - 2z + 2)^6 \) and \( (z^2 + 2z + 2)(z^2 - 2z + 2) + z^4 = 0 \). This is by a direct computation.
Therefore Holant\((f')\) is \#P-hard, and so is Holant\(([z, 1, 0, 0])\) for all \(z \in \mathbb{C}\).

Next we consider the planar case, with \(z \neq 0\). We still have the equivalence \(\text{Pl-Holant}(f) \equiv_T \text{Pl-Holant}([z, 1, 0, 0])\) by the orthogonal transformation \(Q\).

To show that \(\text{Pl-Holant}([z, 1, 0, 0])\) is also \#P-hard for \(z \neq 0\), we need to rule out the additional planar tractable case \((z^2 + 2z + 2)^3 = (z^2 - 2z + 2)^3\), if we still use the gadget in Figure 6.1. This condition is
\[
3z^4 + 16z^2 + 12 = 0. \tag{6.2}
\]
If \(3z^4 + 16z^2 + 12 \neq 0\) then \(\text{Pl-Holant}([z, 1, 0, 0])\) is \#P-hard by the gadget in Figure 6.1.

There are 4 exceptional values of \(z\), for which we prove it separately as follows. Note that for these values, \(z^2\) is real and \(z^2 < 0\).

Figure 6.2: Both vertices are assigned the signature \([z, 1, 0, 0]\).

In addition to the gadget in Figure 6.1, we can construct a gadget in \(\text{Pl-Holant}([z, 1, 0, 0])\) with a binary signature \([z^2 + 2, z, 1]\) by Figure 6.2. Thus the bipartite planar holant problem \(\text{Pl-Holant}([z^2 + 2, z, 1]|f')\) is reducible to \(\text{Pl-Holant}([z, 1, 0, 0])\). Under the same transformation \(T = [z^3 - z, 1] \in \text{GL}_2(\mathbb{C})\), we get an equivalent problem \(\text{Pl-Holant}([X, Y, Z] | [1, 0, 0, 1])\), where \([X Y Z] = T^T \begin{bmatrix} z^2 + 2 \\ z \end{bmatrix}
\)
thus
\[
X = z^4 + 2z^3 + 5z^2 + 6z + 3, \\
Y = z^4 + 3z^2 - 1, \\
Z = z^4 - 2z^3 + 5z^2 - 6z + 3.
\]

Now again we can apply Theorem 6.1 to this problem, and verify that no tractable condition is satisfied for any root of (6.2). For reader’s convenience we include this verification as follows.

(1) Non-degeneracy of \([X, Y, Z]\) follows from the fact that it is transformed from a non-degenerate binary signature by \(T\).

(2) \(Y = 0\) is incompatible with (6.2).

(3) It cannot be the case that \(X = Z = 0\) assuming (6.2). In fact \(X = Z\), which is equivalent to \(z^3 + 3z = 0\), is already incompatible with (6.2).

(4’) There is no solution to \(XZ = -Y^2\) and \(X^6 = Z^6\). Luckily we don’t have to check \(X^6 = Z^6\); \(XZ = -Y^2\) is already incompatible with (6.2). We can write \(X = A + zB\) and \(Z = A - zB\) where \(A = z^4 + 5z^2 + 3 \in \mathbb{R}\) and \(B = 2z^2 + 6 \in \mathbb{R}\) by (6.2). Then \(XZ + Y^2 = 0\) is equivalent to \(A^2 + Y^2 = z^2B^2\), yet \(A^2 + Y^2 > 0\) and \(z^2B^2 < 0\) by \(z^2 < 0\) that follows from (6.2).

183
(5) Finally for the planar condition $X^3 = Z^3$, we already know that $X = Z$ is incompatible with (6.2). We can also verify that $X^2 + XZ + Z^2 = 0$ is also incompatible. In fact $X^2 + XZ + Z^2 = 3A^2 + z^2B^2$. By (6.2), $3A = -(z^2 + 3)$ and $B = 2(z^2 + 3)$, and therefore $(3A)^2 + 3z^2B^2 = (z^2 + 3)(12z^2 + 1)$, which is not zero, again by (6.2).

This completes the proof of the theorem. □

By Theorem 6.1 and Lemma 6.2, we have a dichotomy theorem for Holant([f₀, f₁, f₂, f₃]). From this, we can get a further dichotomy for all bipartite Holant problems on 2-3 regular graphs Holant([g₀, g₁, g₂][f₀, f₁, f₂, f₃]). The reduction is standard. For any non-degenerate [g₀, g₁, g₂], we can find a transformation $M$, such that $[g₀, g₁, g₂] = [1, 0, 1]M^{⊗2}$, as it was shown in (2.18) in Chapter 2. Then the bipartite problem Holant([g₀, g₁, g₂][f₀, f₁, f₂, f₃]) is transformed to the equivalent problem Holant($M^{⊗3}$[f₀, f₁, f₂, f₃]), for which we can apply our dichotomy.

We state the following Theorem 6.3 and 6.4 in a more conceptual form, in terms of what is transformable to various tractable classes $\mathcal{A}$, $\mathcal{P}$ and $\mathcal{M}$.

**Theorem 6.3.** Let $f = [f₀, f₁, f₂, f₃]$ be a ternary symmetric signature. Holant($f$) is $\#P$-hard unless $f$ satisfies one of the following conditions, in which case the problem is in $P$:

1. $f$ is degenerate;
2. There exists $T ∈ \text{GL}_2(\mathbb{C})$ such that $f = T^{⊗3} (=₃)$, and the signature $(=₂)T^{⊗2} ∈ \mathcal{A} \cup \mathcal{P}$;
3. For $λ ∈ \{2i, -2i\}$, $f_2 − λf_1 − f_0 = 0$ and $f_3 − λf_2 − f_1 = 0$.

Additionally Pl-Holant($f$) is in $P$ if there exists $T ∈ \text{GL}_2(\mathbb{C})$ such that $(=₂)T^{⊗2} ∈ \mathcal{M}$ and $f ∈ T\mathcal{M}$. Pl-Holant($f$) is $\#P$-hard otherwise.

As shown in Section 2.2, $f = [f₀, f₁, f₂, f₃]$ satisfies the condition in item (3) in Theorem 6.3 iff $f$ is a generalized Fibonacci gate with parameter $λ = ±2i$. This happens iff $f$ has the form $f_k = ckμ^{k-1} + du^k$ (0 ≤ $k$ ≤ 3 and $μ = ±i$). Furthermore this happens iff $(Z^{-1})^{⊗3}f$ has the form $[x, y, 0, 0]$ or its reversal $[0, 0, y, x]$, where $Z = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$. (See also the remark after Theorem 3.5.) Thus, item (3) in Theorem 6.3 can also be stated as $(Z^{-1})^{⊗3}f$ has the form $[x, y, 0, 0]$ or its reversal $[0, 0, y, x]$, while $(=₂)Z^{⊗2} = 2(=₂)$, a constant multiple of the binary DISEQUALITY function. Signatures of the form $Z[x, y, 0, 0]$ or $Z[0, 0, y, x]$ are special cases of the so-called vanishing signatures; they will be discussed in Chapter 7.

Thus we can state condition (3) of Theorem 6.3 equivalently as $(3')$: $(Z^{-1})^{⊗3}f$ has the form $[*,* ,0, 0]$ or its reversal $[0, 0,*,$].

The next theorem is concerned with 2-3 regular graphs.

**Theorem 6.4.** Let $f = [f₀, f₁, f₂, f₃]$ and $g = [g₀, g₁]$ be symmetric signatures of arity 3 and 2 respectively. Holant($g|f$) is $\#P$-hard unless $f$ and $g$ satisfy one of the following conditions, in which case the problem is in $P$:
1. \( f \) or \( g \) is degenerate;

2. There exists \( T \in \text{GL}_2(\mathbb{C}) \) such that \( f = T^{\otimes 3}(=3) \), and the signature \( gT^{\otimes 2} \in \mathcal{A} \cup \mathcal{P} \);

3. There exists \( T \in \text{GL}_2(\mathbb{C}) \) such that \( f = T^{\otimes 3}[1,1,0,0] \), and \( gT^{\otimes 2} \) is of form \([0,*,*] \).

Additionally \( \text{Pl-Holant}(g \mid f) \) is in \( \mathcal{P} \) if there exists \( T \in \text{GL}_2(\mathbb{C}) \) such that \( gT^{\otimes 2} \in \mathcal{M} \) and \( f \in T.\mathcal{M} \). \( \text{Pl-Holant}(g \mid f) \) is \( \#\mathcal{P} \)-hard otherwise.

Theorem 6.3 is a special case of Theorem 6.4 with \( g = (=2) \). We show that the respective conditions in item (3) are equivalent when \( g = (=2) \). Suppose \( f \) is non-degenerate and satisfies the condition (3) of Theorem 6.3. We will write it in the form (3'), namely \((Z^{-1})^{\otimes 3}f = [x,y,0,0]\) or its reversal \([0,0,y,x]\), for some \( x,y \in \mathbb{C} \). Suppose \( f = Z^{\otimes 3}[x,y,0,0] \); the other case is similar. By being non-degenerate, \( y \neq 0 \). Using this we can find an upper triangular matrix \( R = \begin{bmatrix} \frac{y}{x} & 1 \\ 0 & \frac{y}{x} \end{bmatrix} \in \text{GL}_2(\mathbb{C}) \) such that \( R^{\otimes 3}[1,1,0,0] = [x,y,0,0] \). Then we may take \( T = ZR \) for the required matrix in item (3) of Theorem 6.4. Note that for an upper triangular matrix \( R \), we have \([0,1,0]R^{\otimes 2} = [0,*,*] \). In the other direction, suppose \( f \) is non-degenerate and satisfies the condition in item (3) of Theorem 6.4 with \( T = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \). The fact that \((=2)T^{\otimes 2}\) has the form \([0,*,*] \) implies that \( a^2 + b^2 = 0 \), and so \( b = \epsilon a \) where \( \epsilon = \pm 1 \). Since \( f \) is non-degenerate, we have \( \det T \neq 0 \), thus \( a \neq 0 \) and \( d \neq \epsilon c \). Using this we can find an upper triangular matrix \( R = \begin{bmatrix} \frac{a}{c} & -\frac{b}{c} \\ 0 & \frac{a}{c} \end{bmatrix} \) such that \( TR^{-1} = Z = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \) (for \( \epsilon = +1 \)), or \( TR^{-1} = Z' = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \) (for \( \epsilon = -1 \)). Then \((Z^{-1})^{\otimes 3}f \) has the form \([*,*,0,0] \) or \([0,0,*,*]\).

We can also derive Theorem 6.4 from Theorem 6.3. Suppose \( f \) and \( g \) are non-degenerate. Write \( g = (=2)M \) as in (2.18) from Chapter 2. Then \( \text{Holant}(g \mid f) \equiv_\mathcal{T} \text{Holant}(=2 \mid Mf) \). The condition in item (2) of Theorem 6.3 is the existence of some \( T \) such that \( Mf = T^{\otimes 3}(=3) \), and \((=2)T^{\otimes 2}\) is in \( \mathcal{A} \cup \mathcal{P} \). It translates to the condition in item (2) of Theorem 6.4 by taking \( M^{-1}T \) as the transformation. For item (3), the condition that \( Z^{-1}Mf \) has the form \([c,d,0,0]\) (where \( d \neq 0 \) by non-degeneracy) is equivalent to the condition that for some upper triangular \( R \), we have \( R^{-1}Z^{-1}Mf = [1,1,0,0] \). If we can set \( T = M^{-1}ZR \), then \( T^{-1}f = [1,1,0,0] \) and \( gT = (=2)MT = (=2)ZR = 2(=2)R \) has the form \([0,*,*]\). In the other direction, suppose for some \( T \), we have \( f = T[1,1,0,0] \) and \( gT = [0,a,b] \). By non-degeneracy, \( \det T \neq 0 \) and \( a \neq 0 \). Then we may take an upper triangular \( R = \begin{bmatrix} 1 & -\frac{b}{2a} \\ 0 & 1 \end{bmatrix} \). Then \([0,a,b]R = a(=2) \) which we can verify by its matrix product form \( R\begin{bmatrix} 0 \\ a \\ b \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ a \\ b \\ 0 \end{bmatrix} \). Then \( g = (=2)M \) with \( M \) taken to be a scalar multiple of \( ZR^{-1}T^{-1} \), and \( \text{Holant}(g \mid f) \equiv_\mathcal{T} \text{Holant}(=2 \mid Mf) \). With \( T^{-1}f = [1,1,0,0] \) and \( R^{-1}[1,1,0,0] = [*,*,0,0] \), it follows that \( Mf \) has the form \([*,*,0,0]\). We have proved that the condition in item (3) of Theorem 6.4 corresponds precisely to the condition in item (3) of Theorem 6.3 after transforming \( g \) to \( gM^{-1} = (=2) \). The same equivalence of Theorem 6.3 and Theorem 6.4 for the planar case can be shown.
The third cases in both Theorem 6.3 and Theorem 6.4 are a manifestation of a phenomenon called *vanishing*, which we will discuss in detail in Chapter 7. There is also a symmetric statement to case (3) in Theorem 6.4: Holant\((g|f)\) is in P, if there is a \(2 \times 2\) matrix \(T\) such that \(f = T^{{\otimes}3}[0, 0, 1, 1]\) and \(g{T^{{\otimes}2}}\) is of form \([*, *, 0]\). This is equivalent to case (3) in Theorem 6.4 if we use \(T\left[\begin{array}{cc}0 & 1 \\ 1 & 0\end{array}\right]\) as \(T\).

### 6.2 Reductions Between Holant and \#CSP

In this section, we consider Holant\((f)\), where \(f = [f_0, f_1, f_2, f_3]\) and \(g = [g_0, g_1, g_2]\) are non-degenerate, and \(f\) is in the generic case of having distinct eigenvalues. We establish an equivalence between this problem and a \#CSP problem, under a mild condition, and thus giving a dichotomy via the \#CSP dichotomy of Chapter 3. This can be viewed as extending Theorems 6.3 and 6.4 in Section 6.1 for a single ternary signature to a set of signatures containing a non-degenerate ternary signature of the generic case. We will consider the double root case in Section 6.3.

For \(f\) in the generic case, we can apply a holographic reduction to transform \(f\) to \((=3)\). Therefore we only need to consider Holant problems of the form Holant\((g)\), where \(g\) is a non-degenerate symmetric binary signature. Furthermore, we observed after Theorem 6.1 that the ternary \(\text{EQUALITY}\) signature \((=3)\) is invariant under any transformation from the following set,

\[
\mathcal{T}_3 = \left\{ \left[ \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right], \left[ \begin{array}{cc} 1 & 0 \\ 0 & \omega \end{array} \right], \left[ \begin{array}{cc} 1 & 0 \\ 0 & \omega^2 \end{array} \right] \right\}, \tag{6.3}
\]

where \(\omega = \omega_3 = e^{2\pi i/3}\). Thus, for any \(T \in \mathcal{T}_3\),

\[
\text{Holant}(\{g\} \cup \mathcal{G}|\{=3\} \cup \mathcal{F}) \equiv_T \text{Holant}(\{g{T^{{\otimes}2}}\} \cup \mathcal{G}T|\{=3\} \cup T^{-1}\mathcal{F}).
\]

As a result, we can normalize \(g\) by a holographic reduction \(T \in \mathcal{T}_3\). We define a symmetric binary signature \(g = [g_0, g_1, g_2]\) to be normalized if either \(g_0 = 0\) or it is not the case that \(g_2\) is \(g_0\) times a \(t\)-th primitive root of unity, for some integer \(t = 3t'\) where gcd\((t', 3) = 1\). If \(g\) is not normalized, then \(g_0 \neq 0\) and \(g_2 = g_0\omega^t\), where \(\omega_t = e^{2\pi i/t}\), \(t = 3t'\), gcd\((t', 3) = 1\) and gcd\((s, t) = 1\). Write \(1 = 3u + t'v\) for some integers \(u\) and \(v\), then \(\omega_t = \omega^u\omega_0^v\) and \(\omega_t^2 = \omega^k\omega_0^v\), for some integers \(k\) and \(l\), where \(k \equiv sv \mod 3\), and gcd\((k, 3) = 1\). Hence we may take \(k = 1\) or \(2\). After applying the transformation \(\left[ \begin{array}{cc} 1 & k \\ 0 & \omega \end{array} \right] \in \mathcal{T}_3\), the signature \([g_0, g_1, g_2]\) becomes \([g_0, g_1\omega^k, g_0\omega^v]\), which is normalized. So in the following, we only deal with normalized \(g\). In one case, we also need to normalize a unary signature \([u_0, u_1]\). We say \([u_0, u_1]\) is normalized if either \(u_0 = 0\) or \(u_1\) is not a multiple of \(u_0\) by a \(t\)-th primitive root of unity, for some integer \(t = 3t'\) where gcd\((t', 3) = 1\). Again we can normalize the unary signature by a suitable \(T \in \mathcal{T}_3\). We note that a normalized signature is still normalized after a nonzero scalar multiple.
Theorem 6.5. Let $\mathcal{F}$ and $\mathcal{G}$ be arbitrary sets of complex-valued signatures. Let $g = [g_0, g_1, g_2]$ be a normalized and non-degenerate symmetric signature. And in the case of $g_0 = g_2 = 0$, we further assume that $\mathcal{G}$ contains a unary signature $[u_0, u_1]$, which is normalized and $u_0u_1 \neq 0$. Then

$$\text{Holant}([g] \cup \mathcal{G}|{-3} \cup \mathcal{F}) \equiv_f \text{CSP}([g] \cup \mathcal{F} \cup \mathcal{G}).$$

Consequently, by Theorem 3.7, $\text{Holant}([g] \cup \mathcal{G}|{-3} \cup \mathcal{F})$ is $\#P$-hard unless $\{g\} \cup \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P}$ or $\{g\} \cup \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{A}$, in which cases the problem is in $\mathcal{P}$.

We prove Theorem 6.5 in the rest of this section. We note that we do not require the signatures in $\mathcal{F}$ and $\mathcal{G}$ to be symmetric. This dichotomy establishes an interesting link between Holant and $\#\text{CSP}$. The dichotomy on Holant problems established in Theorem 6.5 uses results from $\#\text{CSP}$. The applicability of Theorem 6.5 is not limited by the assumption on signature normalization of $[g_0, g_1, g_2]$; for a non-normalized binary signature, we can first normalize it and then apply the theorem. The additional assumption of the existence of a non-zero unary signature circumvents a technical difficulty.

One direction of the reduction in Theorem 6.5, from Holant to $\#\text{CSP}$, is straightforward by definition. Thus the main claim in Theorem 6.5 is a reduction from $\#\text{CSP}$ to these bipartite Holant problems. The approach is to construct the binary Equality function $(=_2)$ on the left-hand side (LHS) in $\text{Holant}([g] \cup \mathcal{G}|{-3} \cup \mathcal{F})$. As soon as we have $(=_2)$ on LHS, together with $(=_3)$ on the right-hand side (RHS), we get Equality gates of all arities $(=_k)$ on RHS. Also with the help of $(=_2)$ on LHS we can transfer $\mathcal{F}$ to LHS. Then we can simulate all of $\#\text{CSP}([g] \cup \mathcal{F} \cup \mathcal{G})$.

If the problem $\text{Holant}([g]|{-3})$ is already $\#\text{P}$-hard, then for any $\mathcal{F}$ and $\mathcal{G}$, the more general problem $\text{Holant}([g] \cup \mathcal{G}|{-3} \cup \mathcal{F})$ is also $\#\text{P}$-hard, and $\#\text{CSP}([g] \cup \mathcal{F} \cup \mathcal{G})$ is reducible to any $\#\text{P}$-hard problem. So we only need to consider the case $\text{Holant}([g]|{-3})$ is not $\#\text{P}$-hard. For this, we again apply Theorem 6.1 to $g = [g_0, g_1, g_2]$. The first tractable case $g_1^2 = g_0 g_2$ is degenerate, which does not apply here. The following three lemmas deal with the remaining three tractable cases respectively.

Let us consider the case $g_0^2 = g_1^2$ and $g_0 g_2 = -g_1^2$. Assume $g_1 \neq 0$, we can scale it to $[a, 1, b]$, where $a^2 = 1$ and $ab = -1$. Since $[a, 1, b]$ is assumed to be normalized, it follows that $a^4 = 1$.

**Lemma 6.6.** Let $\mathcal{F}$ and $\mathcal{G}$ be two sets of signatures. For all pairs of $a$ and $b$ satisfying $a^4 = 1$ and $ab = -1$, $\text{Holant}([a, 1, b] \cup \mathcal{G}|{-3} \cup \mathcal{F})$ is $\#\text{P}$-hard unless $\mathcal{F} \cup \mathcal{G} \subseteq \mathcal{A}$, in which case it is in $\mathcal{P}$.

**Proof.** We first prove that when $a^4 = 1$ and $ab = -1$,

$$\text{Holant}([a, 1, b] \cup \mathcal{G}|{-3} \cup \mathcal{F}) \equiv_f \text{CSP}([a, 1, b] \cup \mathcal{F} \cup \mathcal{G}).$$

To get this, it is sufficient to construct $(=_2)$ on LHS.

**Case 1:** $a = \pm i$. 

187
Figure 6.3: A gadget construction for the binary Disequality function ($\neq_2$) on LHS. Degree 3 vertices have signature ($=_3$), degree 2 vertices have signature $[1, \pm i, 1]$.

Figure 6.4: Another gadget construction for the binary Disequality function ($\neq_2$) on LHS. Degree 3 vertices have signature ($=_3$), degree 2 vertices have signature $[1, 1, 1]$.

It is equivalent to consider $\text{Holant}(\{1, \pm i, 1\} \cup G \{=_3\} \cup F)$ because they only differ by a constant factor. We can construct $[1, 1]$ on RHS by connecting the two edges of $[1, \pm i, 1]$ on LHS with two edges of a $=_3$ on RHS. With the gadget in Figure 6.3, we can construct the binary Disequality function ($\neq_2$) on LHS up to a nonzero scalar. Together with the $=_3$ on RHS, we can have $=_3$ on LHS. Connecting this LHS $=_3$ with $[1, 1]$ on RHS, we obtain the binary Equality function ($=_2$) on LHS.

Case 2: $a = \pm 1$.

It is equivalent to consider $\text{Holant}(\{1, \pm 1, -1\} \cup G \{=_3\} \cup F)$. With the gadget in Figure 6.4, we can construct $=_2$ on LHS, and thus $=_3$ on LHS, using the given $=_3$ on RHS. Furthermore, we can construct $[1, -1]$ and $[1, 0, -1]$ on both sides. E.g., $[1, -1]$ is constructed as a fragment of the gadget in Figure 6.4; we first obtain $[1, -1]$ on RHS by connecting two edges of $=_3$ with a copy of $[1, \pm 1, -1]$, and later obtain $[1, -1]$ on LHS as well. Using $[1, -1]$ and $=_3$ on both sides we get $[1, 0, -1]$ on both sides. By connecting $[1, -1]$ with $[1, 0, -1]$, we can realize $[1, 1]$ on both sides, and consequently $=_2$ on both sides.

By Theorem 3.7, $\#\text{CSP}([a, 1, b] \cup F \cup G)$ is $\#P$-hard unless $[a, 1, b] \cup F \cup G \subseteq \mathcal{P}$ or $[a, 1, b] \cup F \cup G \subseteq \mathcal{A}$. Since $[a, 1, b] \in \mathcal{A} - \mathcal{P}$, we conclude that the only possible case which is not $\#P$-hard is $F \cup G \subseteq \mathcal{A}$. This is also sufficient for tractability. The proof is complete.

Back to $\text{Holant}(\{g\} \{=_3\})$ with $g = [g_0, g_1, g_2]$, for the tractable case $g_1 = 0$ in Theorem 6.1, by non-degeneracy, we can scale it to be $[1, 0, a]$, where $a \neq 0$. Then we have the following lemma:

**Lemma 6.7.** Let $F$ and $G$ be two sets of signatures, and let $a \neq 0$ be a complex number. We assume $[1, 0, a]$ is normalized. Then we have the following dichotomy:
If \( a^4 = 1 \), then \( \text{Holant}([1, 0, a] \cup \mathcal{G} \{=3\} \cup \mathcal{F}) \) is \#P-hard unless \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P} \) or \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{A} \), in which cases it is in \( \mathcal{P} \).

If \( a^4 \neq 1 \), then \( \text{Holant}([1, 0, a] \cup \mathcal{G} \{=3\} \cup \mathcal{F}) \) is \#P-hard unless \( \mathcal{F} \subseteq \mathcal{P} \), in which case it is in \( \mathcal{P} \).

*Proof.* As above, it is sufficient to construct \((=2)\) on LHS, and establish the equivalence
\[
\text{Holant}([1, 0, a] \cup \mathcal{G} \{=3\} \cup \mathcal{F}) \equiv_T \text{CSP}([1, 0, a] \cup \mathcal{F} \cup \mathcal{G}).
\]
Note that for \( a^4 \neq 1 \), the signature \([1, 0, a] \notin \mathcal{A} \) and thus \( \text{CSP}([1, 0, a] \cup \mathcal{F} \cup \mathcal{G}) \) being not \#P-hard implies that \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P} \).

We will use the gadget in Figure 6.5 in our proof. We can use it to realize \([1, 0, a^{3k+1}]\) for any \( k \in \mathbb{N} \) on LHS.

If \( a \) is not a root of unity, then we will be able to interpolate all signatures of the form \([1, 0, x]\) where \( x \in \mathbb{C} \) on LHS. This uses a Vandermonde system as we did in Chapter 3. In particular, we will be able to interpolate \((=2) = [1, 0, 1]\) on LHS.

Next we can assume that \( a \) is a \( t \)-th primitive root of unity, that is \( a = \omega_t^b \) for some \( b \) relatively prime to \( t \), where \( \omega_t = e^{2\pi i/t} \). If \( t \) is not a multiple of 3, then we can find an integer \( k \), such that \( 3k + 1 \equiv 0 \pmod{t} \). Therefore, we can realize \((=2) = [1, 0, 1]\) on LHS.

Now we consider the case of \( t = 3l' \), where \( l' \geq 1 \) and \( \gcd(t',3) = 1 \). Since \([1, 0, a]\) is normalized, we have a further condition that \( l > 1 \). For this case, we cannot construct \((=2)\) on LHS directly. Instead we will further apply a holographic reduction. Also in this case, we have \( a^4 \neq 1 \), so we want to prove that \( \text{Holant}([1, 0, a] \cup \mathcal{F} \{=3\} \cup \mathcal{G}) \) is \#P-hard unless \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P} \). The fact that the problem is in \( \mathcal{P} \) when \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P} \) is obvious by Theorem 3.7, since \([1, 0, a] \in \mathcal{P} \).

Since \( 2t' \) is not a multiple of 3, there exist some integers \( k \) and \( s \), such that \( 3k + 1 = 2st' \). Since
\[ a^{3k+1} = a^{2st'} = \omega_3^{2bst'} = \omega_3^{2bs} \]
we can realize \([1, 0, \omega_3^{2bs}]\) on LHS. So
\[
\text{Holant}([1, 0, \omega_3^{2bs}] \cup \mathcal{G} \{=3\} \cup \mathcal{F}) \leq_T \text{Holant}([1, 0, a] \cup \mathcal{G} \{=3\} \cup \mathcal{F}).
\]
Therefore it suffices to prove that Holant([1, 0, \omega_{3l}^{2bs}] \cup \mathcal{G}|\{=3\} \cup \mathcal{F}) is \#P-hard if \mathcal{F} \cup \mathcal{G} \not\subseteq \mathcal{P}.

We apply a holographic reduction \( T = \begin{bmatrix} 1 & 0 \\ 0 & \omega_{3l}^{bs} \end{bmatrix} \), and get

\[
\text{Holant}([1, 0, \omega_{3l}^{2bs}] \cup \mathcal{G}|\{=3\} \cup \mathcal{F}) \equiv_{T} \text{Holant}([\{=2\} \cup \mathcal{GT}|[1, 0, 0, \omega_{3l}^{bs(l-1)}] \cup T^{-1}\mathcal{F}).
\]

We then use the gadget in Figure 6.6 to realize \([1, 0, 0, \omega_{3l}^{bs(l-1)}] = [1, 0, 0, 1] = (=3)\) on RHS. Together with (=2) on LHS this gives all EQUALITIES on both sides. As a result

\[
\#\text{CSP}([1, 0, 0, \omega_{3l}^{bs(l-1)}] \cup \mathcal{GT} \cup T^{-1}\mathcal{F}) \leq_T \text{Holant}([\{=2\} \cup \mathcal{GT}|[1, 0, 0, \omega_{3l}^{bs(l-1)}] \cup T^{-1}\mathcal{F}).
\]

Since \( l > 1, [1, 0, 0, \omega_{3l}^{bs(l-1)}] \not\in \mathcal{A} \), as both \( b \) and \( s \) are relatively prime to 3. So the problem is \#P-hard unless \( \mathcal{GT} \cup T^{-1}\mathcal{F} \subseteq \mathcal{P} \). Since \( T \) and \( T^{-1} \) are diagonal matrices, it is equivalent to the statement \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P} \). This completes the proof. □

For the last tractable case \( g_0 = g_2 = 0 \) in Theorem 6.1, we can scale it to \((\neq_2) = [0, 1, 0]\).

**Lemma 6.8.** Let \( \mathcal{F} \) and \( \mathcal{G} \) be two sets of signatures, and \( a \neq 0 \) be a complex number. We assume \([1, a] \) is normalized. Then we have the following dichotomy:

- If \( a^4 = 1 \), then \( \text{Holant}([\{0, 1, 0\}, [1, a] \cup \mathcal{G}|\{=3\} \cup \mathcal{F}) \) is \#P-hard unless \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P} \) or \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{A} \), in which cases it is in \( P \).

- If \( a^4 \neq 1 \), then \( \text{Holant}([\{0, 1, 0\}, [1, a] \cup \mathcal{G}|\{=3\} \cup \mathcal{F}) \) is \#P-hard unless \( \mathcal{F} \cup \mathcal{G} \subseteq \mathcal{P} \), in which case it is in \( P \).

**Proof.** By connecting one copy of \([1, a] \) and two copies of \([0, 1, 0]\) to a (=3) on RHS, we can realize \([a, 0, 1]\), or equivalently \([1, 0, 1/a]\), on LHS. Now Lemma 6.8 follows from Lemma 6.7. □

This completes the proof of Theorem 6.5.

### 6.3 Holant\(^c\) Problems

In Holant\(^*\) Problems, all unary functions are assumed to be freely available. Two special unary functions, Is-ZERO \( \Delta_0 = [1, 0] \) and Is-ONE \( \Delta_1 = [0, 1] \), are particularly meaningful. They correspond to setting a variable to the constant zero or one. This is called pinning, and by Lemma 3.13 these are available for all \#CSP problems, namely \( \Delta_0 \) and \( \Delta_1 \) are implied by the presence of EQUALITIES of all arities. Note that if we have \( \Delta_0 \) and \( \Delta_1 \), then we can construct all sub-signatures of any given signature.

We define a class of restricted Holant problems where only the two special unary functions \( \Delta_0 \) and \( \Delta_1 \) are assumed to be freely available. This is called the Holant\(^c\) Problems.
Definition 6.9. Given a set of signatures $\mathcal{F}$, we use $\text{Holant}^c(\mathcal{F})$ to denote $\text{Holant}(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$.

In this section, we prove a dichotomy theorem for Holant$^c$ problems with complex valued symmetric signatures, Theorem 6.12. The proof uses the dichotomies proved in the previous two sections. In order to use them, we first prove in Lemma 6.10 that we can realize a non-degenerate ternary symmetric signature except in some trivial cases. From this ternary signature we immediately have #P-hardness if it is not in one of the tractable cases in Theorem 6.3. For a tractable ternary signature, we use Theorem 6.5 to extend the dichotomy theorem to a set of signatures. In Theorem 6.5, we only considered the generic case of the ternary function. The double-root case is handled here in Lemma 6.11. After that we will be ready to prove Theorem 6.12.

Lemma 6.10. Given any set of symmetric signatures $\mathcal{F}$ which contains $\Delta_0$ and $\Delta_1$, we can construct a non-degenerate ternary symmetric signature $f$, except in the following two cases:

1. All non-degenerate signatures in $\mathcal{F}$ have arity at most 2;
2. In $\mathcal{F}$, all unary signatures are of the form $[x, 0]$ or $[0, x]$; all binary signatures are of the form $[x, 0, y]$ or $[0, x, 0]$; and all signatures of arity greater than 2 are of the form $[x, 0, \ldots, 0, y]$.

Proof. Suppose case 1 does not hold, and let $f = [f_0, f_1, \ldots, f_n] \in \mathcal{F}$ be a non-degenerate signature of arity at least 3. Since we have $\Delta_0$ and $\Delta_1$, we can construct all sub-signatures of any signature in $\mathcal{F}$. (A signature is also considered a sub-signature of itself.) If there exists a ternary non-degenerate sub-signature, we are done. Now suppose all ternary sub-signatures of $f$ are degenerate, in particular $n > 3$. By Proposition 2.8, $f$ must be of the form $[f_0, 0, \ldots, 0, f_n]$, where $f_0 f_n \neq 0$. Then, if we have a unary signature or a unary sub-signature of the form $[a, b]$ from $\mathcal{F}$ with $ab \neq 0$, we can connect $n - 3$ copies of $[a, b]$ to $n - 3$ dangling edges of $f$ to get a non-degenerate ternary signature of the form $[x, 0, 0, y]$, and we are done. Otherwise, all unary signatures are of the form $[x, 0]$ or $[0, x]$ and all binary signatures are of the form $[x, 0, y]$ or $[0, x, 0]$. Any non-degenerate signature of arity greater than 2 must be of the form $[x, 0, \ldots, 0, y]$. If a degenerate signature $[a, b]^{\otimes n}$ has $ab \neq 0$, then we can get a unary sub-signature of the form $[a', b']$ with $a'b' \neq 0$. Thus the degenerate signature is also a GEN-EQ of the form $[x, 0, \ldots, 0, y]$ where $x = 0$ or $y = 0$.

Remark: All the exceptional cases listed in case 2 are special cases of functions of product type.

We next consider $\text{Holant}([\{f, g\}])$, where $f$ and $g$ are non-degenerate symmetric signatures of arity 3 and 2 respectively, and the characteristic equation for the second order linear recurrence of $f = [f_0, f_1, f_2, f_3]$ has a double root. By Lemma 6.2, Holant$(f)$ is already #P-hard unless the double eigenvalue is $i$ or $-i$. Therefore we may assume the double root is $\lambda = \pm i$, and $f_{k+2} - \lambda f_{k+1} - f_k = 0$ for $k = 0, 1$. 191
Lemma 6.11. Let \( f = [f_0, f_1, f_2, f_3] \) be a non-degenerate complex signature satisfying \( f_{k+1} - \lambda f_k = 0 \) for \( k = 0, 1 \), where \( \lambda = \pm 2i \). Let \( g = [g_0, g_1, g_2] \) be a non-degenerate binary signature. Then Holant\((g \mid f)\) is \#P-hard, a fortiori Holant\((f, g)\) is also \#P-hard, unless \( g_2 - \lambda g_1 - g_0 = 0 \); if so, the problem Holant\((\{f, g\})\) is in P by generalized Fibonacci gates.

Proof. We prove this result for \( \lambda = 2i \). The other case is similar.

The sequence \( \{f_k\} \) can be written as follows: \( f_k = c k^{n-1} + d k^n \), where \( c \neq 0 \). Thus, we have
\[
f = T^{\otimes 3}[1, 0, 0]^T,
\]
where
\[
T = \begin{bmatrix}
1 & \frac{d-1}{i} \\
0 & c + \frac{d-1}{3}i
\end{bmatrix}.
\]
We can express \( \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix} = M^T g \), as shown in (2.18) of Section 2.3.2, for some \( M = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \). \( M \) is non-singular since \( g \) is non-degenerate. Then we have
\[
g_0 = \alpha^2 + \beta^2, \quad g_1 = \alpha \gamma + \beta \delta, \quad \text{and} \quad g_2 = \gamma^2 + \delta^2.
\]

Thus we have the following reduction
\[
\text{Holant}(g \mid f) = \text{Holant}(\{z\} M^{\otimes 2} \mid T^{\otimes 3}[1, 0, 0]) \equiv_{\tau} \text{Holant}(\{z\} (MT)^{\otimes 3}[1, 0, 0]),
\]
where
\[
MT = \begin{bmatrix}
\alpha + \gamma i & * \\
\beta + \delta i & * 
\end{bmatrix}, \quad \text{and we will call it} \quad \begin{bmatrix} p \\ q \end{bmatrix}.
\]

Our next step is to use an orthogonal matrix to transform \( MT \) to be upper-triangular, if we can. If \( p^2 + q^2 \neq 0 \), then we can find a (complex) orthogonal matrix \( Q = \frac{1}{\sqrt{p^2 + q^2}} \begin{bmatrix} p & q \\ q & -p \end{bmatrix} \) such that \( QMT \) is upper-triangular. As shown in Section 6.1, for a non-singular upper-triangular \( QMT \), \( (QMT)^{\otimes 3}[1, 0, 0] \) has the form \([x, y, 0, 0]\) for some \( y \neq 0 \). We can normalize it to \([z, 1, 0, 0]\).

For any orthogonal matrix \( Q \), the LHS \( \equiv_{\tau} \) is unchanged under the holographic transformation: \( \equiv_{\tau} (Q^{-1})^{\otimes 2} \equiv_{\tau} \). This gives us
\[
\text{Holant}(\{z\} (MT)^{\otimes 3}[1, 0, 0]) \equiv_{\tau} \text{Holant}(\{z\} [z, 1, 0, 0]) \equiv_{\tau} \text{Holant}([z, 1, 0, 0]),
\]
for some \( z \). This shows the equivalence of Holant\((g \mid f)\) with Holant\(([z, 1, 0, 0])\). By Lemma 6.2, Holant\(([z, 1, 0, 0])\) is \#P-hard. Hence Holant\((g \mid f)\) is also \#P-hard.

Finally, \( p^2 + q^2 = 0 \) is \((\alpha + \gamma i)^2 + (\beta + \delta i)^2 = (\alpha^2 + \beta^2) + 2i(\alpha \gamma + \beta \delta) = (\gamma^2 + \delta^2) = 0 \). This is exactly \( g_2 - \lambda g_1 - g_0 = 0 \).

A perceptive reader will have noticed that the very explicit dichotomy statement in Lemma 6.11 should be consistent with, and derivable from, the more conceptual dichotomy statement of Theorem 6.4. We now give such a derivation.
As the discussion after Theorems 6.3 and 6.4 shows, the assumption on \( f \) here is equivalent to \( f \) having the form \( Z[c, d, 0, 0] \), or \( Z[0, 0, d, c] \), with \( d \neq 0 \) by non-degeneracy. We will assume the first case. The second case is similar. We have

\[
\text{Holant}(g \mid f) = \text{Holant}( (=2) \cdot M \mid Z[c, d, 0, 0]) \equiv \text{Holant}(=2 \mid MZ[c, d, 0, 0]),
\]

where \( g = (=2) \cdot M \) as shown in (2.18) for some \( M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \). The property that a ternary symmetric signature is non-degenerate, or is in the generic case (having distinct eigenvalues) is invariant under a non-singular holographic transformation. This is easily seen in terms of the tensor product expression. Hence the property that it has a double root is also invariant. It follows that, by Theorem 6.4 the only possibility for \( \text{Holant}(=2 \mid MZ[c, d, 0, 0]) \) to be tractable is when \( MZ[c, d, 0, 0] \) has the form \( Z[*, *, 0, 0] \) or \( Z[0, 0, *, *] \).

The following fact can be directly verified: For any non-singular matrix \( T \) and \( d \neq 0 \), \( T[c, d, 0, 0] = [\ast, \ast, 0, 0] \) iff \( T \) is an upper triangular matrix. Hence the tractability criterion from Theorem 6.4 is \( Z^{-1}MZ \) or \( XZ^{-1}MZ \) is upper triangular, where \( X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \). Since

\[
\begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} \begin{bmatrix} \alpha + i\gamma & \alpha - i\gamma \\ \beta + i\delta & \beta - i\delta \end{bmatrix} = \begin{bmatrix} (\alpha + i\gamma) - i(\beta + i\delta) & * \\ (\alpha + i\gamma) + i(\beta + i\delta) & * \end{bmatrix},
\]

the two tractable cases can be expressed as

\[
[(\alpha + i\gamma) + i(\beta + i\delta)][(\alpha + i\gamma) - i(\beta + i\delta)] = 0.
\]

This is just \((\alpha + i\gamma)^2 + (\beta + i\delta)^2 = 0\), and it matches exactly with \( g_2 - \lambda g_1 - g_0 = 0 \).

The next theorem is the main result of this section. It gives a dichotomy for \( \text{Holant}^c \) problems. Define

\[
\mathcal{F} \triangleq \{ T \in \text{GL}_2(\mathbb{C}) \mid (=2)T^{\otimes 2}, \Delta_0 T, \Delta_1 T \in \mathcal{A} \}. \tag{6.4}
\]

**Theorem 6.12.** Let \( \mathcal{F} \) be a set of complex symmetric signatures. \( \text{Holant}^c(\mathcal{F}) \) is \#P-hard unless \( \mathcal{F} \) satisfies one of the following conditions, in which case it is tractable:

1. \( \text{Holant}^*(\mathcal{F}) \) is tractable; or
2. There exists a \( T \in \mathcal{F} \) such that \( T^{-1}\mathcal{F} \subseteq \mathcal{A} \).

**Remark:** As we have \( \Delta_0 \) and \( \Delta_1 \), any degenerate \( f = [a, b]^{\otimes n} \in \mathcal{F} \) can be replaced by \([a, b] \), as follows: If \( a = b = 0 \) then \( f \) is identically 0 and applying \( \Delta_0 \) to the signature \( n - 1 \) times gives the unary \([0, 0] \). If \( a, b \) are not both 0, then use \( \Delta_0 \) or \( \Delta_1 \) we can obtain a nonzero multiple of \([a, b] \). Thus we may assume all signatures of arity at least two in \( \mathcal{F} \) are non-degenerate. Also, since we deal with symmetric signatures, by Theorem 3.8, the affine class \( \mathcal{A} \) can be replaced by \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \) in the statement of Theorem 6.12.

The dichotomy criterion in Theorem 6.12 can be made explicit. For case 1 we have the very explicit dichotomy for \( \text{Holant}^* \) problems—Theorem 2.12. This also has a more conceptual but less explicit form in Theorem 3.5. For case 2 we will give a finite list of \( T \in \text{GL}_2(\mathbb{C}) \) in place of \( \mathcal{F} \). We will discuss this in more detail after the proof.
Proof. First of all, if $\mathcal{F}$ is in one of the two listed cases of Lemma 6.10, then Holant$^*$($\mathcal{F}$) is tractable and we are done. Now we can assume that we have a non-degenerate symmetric ternary signature $f = [f_0, f_1, f_2, f_3]$ and the problem is Holant$^c$($\mathcal{F} \cup \{ f \}$).

As discussed in Section 6.1, there are three categories for $f$ and we only need to consider the first two:

1. $f_k = \alpha^{3-k} \beta^k + \gamma^{3-k} \delta^k$, where $\det \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \neq 0$;

2. $f_k = ck\lambda^{k-1} + d\lambda^k$, where $c \neq 0$.

Case 1: $f_k = \alpha^{3-k} \beta^k + \gamma^{3-k} \delta^k$.

In this case, $f = T^{\otimes 3}(=3)$, where $T = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$. Furthermore, we will choose such a $T$ so that $(=2)T^{\otimes 2}$ is normalized. We may accomplish this by replacing $T$ with $TT'$ for some $T' \in \mathcal{T}_3$ defined in (6.3), as explained in Section 6.2. Then we have the following reduction chain,

$$
\text{Holant}^c(\mathcal{F}) \equiv_T \text{Holant}^c(\mathcal{F} \cup \{ f \})
\equiv_T \text{Holant}(\mathcal{F} \cup \{ f, \Delta_0, \Delta_1 \})
\equiv_T \text{Holant}((=2) \mid \mathcal{F} \cup \{ f, \Delta_0, \Delta_1 \})
\equiv_T \text{Holant}((\{=2\}, \Delta_0, \Delta_1) \mid \mathcal{F} \cup \{ f \})
\equiv_T \text{Holant}((\{=2\}T^{\otimes 2}, \Delta_0 T, \Delta_1 T) \mid \{=3\} \cup T^{-1} \mathcal{F}).
$$

Note that $(=2)$ on LHS connected with $\Delta_c$ on RHS gives $\Delta_c$ on LHS, for $c = 0$ and 1. Since $(=2)T^{\otimes 2}$ is a normalized non-degenerate binary signature, we can apply Theorem 6.5.

The only thing we need to verify is that in case $(=2)T^{\otimes 2} = [\alpha^2 + \beta^2, \alpha \gamma + \beta \delta, \gamma^2 + \delta^2] = [0, \alpha \gamma + \beta \delta, 0]$, at least one of $\Delta_0 T = [1, 0]T = [\alpha, \gamma]$ or $\Delta_1 T = [0, 1]T = [\beta, \delta]$ has both entries non-zero. If not, we would have $\alpha \gamma = 0$ and $\beta \delta = 0$, which implies that $(=2)T^{\otimes 2} = [0, \alpha \gamma + \beta \delta, 0] = [0, 0, 0]$, a contradiction. In the case of $(=2)T^{\otimes 2} = [0, *, 0]$, we may further normalize the unary signature $\Delta_0 T$ or $\Delta_1 T$ without affecting $(=2)T^{\otimes 2}$ being normalized. Therefore, by Theorem 6.5, the problem Holant$^c(\mathcal{F})$ is #P-hard unless

$$(=2)T^{\otimes 2} \cup T^{-1} \mathcal{F} \subseteq \mathcal{P}$$

(note that unary $\Delta_0 T, \Delta_1 T$ are automatically in $\mathcal{P}$), or

$$(=2)T^{\otimes 2}, \Delta_0 T, \Delta_1 T \cup T^{-1} \mathcal{F} \subseteq \mathcal{A}.$$

In the first case, Holant$^*$($\mathcal{F}$) is tractable, by the dichotomy theorem for Holant$^*$ problems, Theorem 2.12 (see also Theorem 3.5). In the second case, this is equivalent to having $T \in \mathcal{T}$ satisfying $T^{-1} \mathcal{F} \subseteq \mathcal{A}$.

Case 2: $f_k = ck\lambda^{k-1} + d\lambda^k$, where $c \neq 0$. 

194
In this case, if \( \lambda \neq \pm i \), the problem is \#P-hard by Lemma 6.2 and we are done. Now we consider the case \( \lambda = i \) (the case \( \lambda = -i \) is similar). Consider the following equation

\[
z_{k+2} - 2iz_{k+1} - z_k = 0. \tag{6.5}
\]

We note that \( f = [f_0, f_1, f_2, f_3] \) satisfies this equation for \( k = 0, 1 \). If for all non-degenerate signatures \( \Xi = [z_0, z_1, \ldots, z_n] \) in \( \mathcal{F} \) with arity \( n \geq 2 \) we have the following

**Condition:** \( \Xi \) satisfies Equation (6.5) for \( 0 \leq k \leq n - 2 \),

then, by Theorem 2.12 tractable Class \( \mathcal{B} \) type I(1, 2i) (equivalently case 3 of Theorem 3.5), Holant*\((\mathcal{F})\) is tractable, and we are done. So suppose this is not the case, and there is a non-degenerate signature \( \Xi = [z_0, z_1, \ldots, z_n] \in \mathcal{F} \), for some \( n \geq 2 \), that does not satisfy this Condition. By Lemma 6.11, if any non-degenerate sub-signature \( [z_k, z_{k+1}, z_{k+2}] \) does not satisfy Equation (6.5), then, together with \( f \) which does satisfy (6.5), we know that the problem is \#P-hard and we are done. So we assume every non-degenerate sub-signature \( [z_k, z_{k+1}, z_{k+2}] \) of \( \Xi \) satisfies (6.5). As \( \Xi \) is non-degenerate and does not satisfy the Condition, we have \( n \geq 3 \). Observe that \( \Xi \) does not satisfy the Condition is equivalent to some binary sub-signature of \( \Xi \) does not satisfy (6.5). Hence there exists some binary sub-signature of \( \Xi \) that is degenerate and does not satisfy (6.5).

**Subcase 1:** All binary sub-signatures of \( \Xi \) are degenerate (but \( \Xi \) itself is non-degenerate).

By Proposition 2.8, \( \Xi \) must be of the form \([z_0, 0, \ldots, 0, z_n]\), where \( z_0z_n \neq 0 \).

We claim that there exists a unary sub-signature \([f_k, f_{k+1}]\) of \( f \) with both entries non-zero. If \( f_0f_1 \neq 0 \), then we set \( k = 0 \) and the claim is proved; if \( f_0 = 0, f_1 \neq 0 \), then we have \( f_2 = 2if_1 + f_0 \neq 0 \); if \( f_0 \neq 0, f_1 = 0 \), then we have \( f_2 = 2if_1 + f_0 \neq 0 \) and \( f_3 = 2if_2 + f_1 \neq 0 \). Note that \( f_0 = f_1 = 0 \) is impossible because \( c \neq 0 \).

Connect this unary signature to \( n - 3 \) dangling edges of \([z_0, 0, \ldots, 0, z_n]\), we have a ternary signature \([a, 0, 0, b]\) where \( ab \neq 0 \). We can use this as the non-degenerate ternary signature \( f \) and we have reduced this case to Case 1.

**Subcase 2:** Some binary sub-signatures of \( \Xi \) are non-degenerate (some others are degenerate).

Then we can find a ternary sub-signature \([z_k, z_{k+1}, z_{k+2}, z_{k+3}]\) (or its reversal) where \([z_k, z_{k+1}, z_{k+2}]\) is degenerate and \([z_{k+1}, z_{k+2}, z_{k+3}]\) is non-degenerate and thus satisfies \( z_{k+3} - 2iz_{k+2} - z_{k+1} = 0 \). We have

\[
z_{k+1}^2 = z_kz_{k+2} \quad \text{and} \quad z_{k+2}^2 \neq z_{k+1}z_{k+3}. \tag{6.6}
\]

Thus \( z_{k+1} \) and \( z_{k+2} \) cannot both be zero. First suppose \( z_k \neq 0 \). Then \( z_{k+1} \neq 0 \) for otherwise both \( z_{k+1} = z_{k+2} = 0 \) by \( z_{k+1}^2 = z_kz_{k+2} \). Then \( (z_k, z_{k+1}, z_{k+2}) \) is a geometric progression with a non-zero ratio \( \rho = z_{k+1}/z_k \), and we can assume that \( z_k = 1 \) after a scaling. Then \([z_k, z_{k+1}, z_{k+2}, z_{k+3}] = [1, \rho, \rho^2, 2i\rho^2 + \rho] \), and we have

\[
\begin{bmatrix}
1 \\
\rho \\
0 \\
\rho(\rho + 2i\rho^2 - \rho^3)^{1/3}
\end{bmatrix} \otimes^3 [1, 0, 0, 1] = \begin{bmatrix}
1 \\
\rho \\
0 \\
1
\end{bmatrix} \otimes^3 [1, 0, 0, 1] + (\rho + 2i\rho^2 - \rho^3) \begin{bmatrix}
0 \\
0 \\
0 \\
1
\end{bmatrix} \otimes^3 [1, \rho, \rho^2, 2i\rho^2 + \rho]
\]


Note that $\rho + 2i\rho^2 - \rho^2 \neq 0$ so that the 2-by-2 matrix is non-singular. We already know that $\rho \neq 0$. If $1+2i\rho - \rho^2 = 0$ then $\rho = i$ which would imply that $[z_k, z_{k+1}, z_{k+2}, z_{k+3}] = [1, i, -1, -i]$ is degenerate, a contradiction. Therefore, the ternary sub-signature $[z_k, z_{k+1}, z_{k+2}, z_{k+3}]$ is in the first category and we have reduced the problem to Case 1.

Now suppose $z_k = 0$. Then it must be that $z_{k+1} = 0$ and $z_{k+2} \neq 0$ by (6.6). The signature $[z_k, z_{k+1}, z_{k+2}, z_{k+3}]$ becomes $[0, 0, z_{k+2}, z_{k+3}]$, which is equivalent to $[0, 0, 1, z]$ for some $z \in \mathbb{C}$, and as we proved in Lemma 6.2, it is $\#$P-hard. 

There is a different character to this Holant$^c$ dichotomy, as stated, from previous dichotomies we have proved so far. For example, Theorem 2.12 for Holant$^*$ problems, or Theorem 3.7 for Boolean $\#$CSP are very explicit, so that for any given finite signature set $\mathcal{F}$, it is clear one can decide whether $\mathcal{F}$ satisfies the tractability criteria, in time polynomial in the size of $\mathcal{F}$. In fact, to say that the tractability condition is decidable in polynomial time does not do full justice to the explicitness of these dichotomy theorems; their tractability conditions are expressed essentially in closed form. This level of explicitness is useful, beyond the important issue of decidability in formal logic, or even polynomial time decidability, of the dichotomy criteria. It gives us a most concrete understanding of the dichotomy, as well as it can be used effectively to prove other dichotomy theorems, as shown for example when we applied Theorem 2.12 and Theorem 3.7.

Theorem 6.12 for Holant$^c$ problems has a more conceptual form. It has an existential statement referring to the set $\mathcal{F}$. The conceptual form provides a better understanding of the reason why the dichotomy works. But it brings up the question of how one may decide whether a given $\mathcal{F}$ satisfies the tractability criteria of Theorem 6.12. One should note that for Holant$^*$ problems, Theorem 2.12 has an equivalent form Theorem 3.5, which also has such an existential statement. Our next goal is to show that the Holant$^c$ dichotomy Theorem 6.12 in fact also has a very explicit form. Denote by $\mathcal{F}_{123} = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. We will show that the following set

$$\left\{ I, \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{i} \end{bmatrix}, \begin{bmatrix} \sqrt{2} & 0 \\ i & 1 \end{bmatrix}, \begin{bmatrix} \sqrt{2} & 0 \\ -i & 1 \end{bmatrix} \right\}$$

(6.7)

can be used in place of the infinite set $\mathcal{F}$ in criterion 2 of Theorem 6.12, namely whether for one of these four matrices $T$, we have $T^{-1} \mathcal{F} \subseteq \mathcal{F}_{123}$. For convenience, we denote these four matrices by $I, A, B, C$ respectively. We verify that they all belong to $\mathcal{F}$. Trivially $I \in \mathcal{F}$. For $A$, we have $(\equiv_2) A^{\otimes 2} = [1, 0, i] \in \mathcal{F}_1$. Also we have $\Delta_0 A = [1, 0]$ and $\Delta_1 A = \sqrt{i}[0, 1]$, thus $A \in \mathcal{F}$. For $B$, we have $(\equiv_2) B^{\otimes 2} = [1, i, 1] \in \mathcal{F}_2$, which has matrix form $[\sqrt{2} 0 \] [0 \sqrt{1}] = [1 1]$. Also we have $\Delta_0 B = \sqrt{2}[1, 0]$ and $\Delta_1 B = i[1, -i] \in \mathcal{F}_2$, thus $B \in \mathcal{F}$. For $C$, we can take conjugation in the above calculations for $B$.

The key to this simplification of replacing (6.7) for the finite set $\mathcal{F}$ is a normalizing process afforded by the following group action by $G = GL_2(\mathbb{C})$ and a certain stabilizer subgroup. For signatures $f$ written as covariant tensors, we consider the group action $(f, T) \mapsto fT$ by $G$, where $T \in G$. There is a corresponding group action $(T, f) \mapsto T^{-1} f$ by $G$ on signatures.
written as contravariant tensors. We wish to investigate those $T \in G$ such that

$$(=_2)T, \Delta_0 T, \Delta_1 T \in F_{123}, \quad \text{and} \quad T^{-1}F \subseteq F_{123}. \quad (6.8)$$

Let

$$S = \{ S \in G \mid F_{123}S \subseteq F_{123} \} \quad (6.9)$$

be the stabilizer of $F_{123}$. Clearly $S$ is a subgroup of $G$. In fact, by definitions (6.9) and (6.4), it is clear that $S \subseteq F$, since $(=), \Delta_0, \Delta_1 \in F_{123}$. The purpose of defining this subgroup $S$ is that we will use multiplication on the right from $S$ to normalize the set of transformations.

Let $D = [1 \ 0]$ and $H = \frac{1}{\sqrt{2}} [1 \ 1]$. Also let $X = [0 \ 1]$ and $Z = \frac{1}{\sqrt{2}} [1 \ 1]$. Note that $Z = DH$. Also $D^2 Z = \frac{1}{\sqrt{2}} [1 \ 1] = ZX$, hence $X = Z^{-1} D^2 Z$. It is easy to verify that $D, H \in S$, and hence also $X, Z \in S$, as well as all non-zero scalar multiples of these. We will show that $S$ consists of exactly the non-zero scalar multiples of members of the group $\langle D, H \rangle$ generated by $D$ and $H$,

$$S = \mathbb{C}^* \cdot \langle D, H \rangle. \quad (6.10)$$

Note that $X, Z$ are also in the group $\langle D, H \rangle$ generated by $D$ and $H$.

To show that $S$ is contained in the right-hand side of (6.10), let $T \in S$. Take $(=) \in F_{123}$, then $(=) T^{\otimes 3} \in F_{123}$. Then by the form of $F_{123}$, for some $M \in \langle D, H \rangle$, chosen to be either $I$, or $H^T = H$, or $Z^T = HD$, we have $(=) (TM^{-1})^{\otimes 3} \in F_1$, which is a GEN-EQ. Then either $T^T = TM^{-1}$ or $TM^{-1} X$ is a diagonal matrix $\lambda [1 \ 0]$. Furthermore, by applying it to $(=) \in F_{123}$ we conclude that $(=) T^{\otimes 4} \in F_1$, since it is in $F_{123}$ but not in $F_2 \cup F_3$. It follows that $d$ is a power of $i$, and hence $[1 \ 0]$ is a power of $D$. Thus $T \in \mathbb{C}^* \cdot \langle D, H \rangle$.

Since both generators $D$ and $H$ are symmetric, $\langle D, H \rangle$ is closed under transpose. Hence we also have $S$ is the stabilizer of $F_{123}$ by left action.

$$S = \{ S \in G \mid S F_{123} \subseteq F_{123} \} = \{ S \in G \mid S^{-1} F_{123} \subseteq F_{123} \},$$

where equivalently we can replace $\subseteq$ by $=$.

Define the left and right stabilizer groups of $\mathcal{A}$:

$$\text{LStab}(\mathcal{A}) = \{ T \in \text{GL}_2(\mathbb{C}) \mid T \mathcal{A} \subseteq \mathcal{A} \};$$

$$\text{RStab}(\mathcal{A}) = \{ T \in \text{GL}_2(\mathbb{C}) \mid \mathcal{A} T \subseteq \mathcal{A} \}.$$ 

Then it is easy to verify that both $D, H \in \text{LStab}(\mathcal{A}) \cap \text{RStab}(\mathcal{A})$. It is also clear that both $\text{LStab}(\mathcal{A}) \subseteq S$ and $\text{RStab}(\mathcal{A}) \subseteq S$. Hence $S = \mathbb{C}^* \cdot \langle D, H \rangle = \text{LStab}(\mathcal{A}) = \text{RStab}(\mathcal{A})$.

From now on we will write $\text{Stab}(\mathcal{A})$, or $S$, for this group.

For future reference, we record this as a lemma.

**Lemma 6.13.** The left and right stabilizer groups of $\mathcal{A}$ and the left and right stabilizer groups of $F_{123}$ all coincide, and it is the group $\mathbb{C}^* \cdot \langle D, H \rangle$.  

197
We remark that $S$ is not normalized by $T$. We take $H \in S$, and calculate that $AHA^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1/\sqrt{1} \\ \sqrt{1} & -1 \end{bmatrix}$, which maps $[1, 1] \in A$ to $[1, 1] \begin{bmatrix} 1 & 1/\sqrt{1} \\ \sqrt{1} & -1 \end{bmatrix} = [1 + \sqrt{i}, 1/\sqrt{1} - 1]$ (ignoring the scalar $\frac{1}{\sqrt{2}}$), having two nonzero entries of unequal norm, and hence not in $A$. Therefore $AHA^{-1} \not\in S$.

The following observation is a key point: If $T$ and $T'$ are two elements of $T$ from the same right coset of $S$, i.e., $TS = T'S$, then $T$ satisfies (6.8) iff $T'$ satisfies (6.8). Thus we may use right multiplication from $S$ to normalize $T$ in Theorem 6.12. Formally, we only need to check for $T \in T$ in the second condition in Theorem 6.12 for a complete set of right coset representatives of $S$ in $T$. We claim that the four matrices in (6.7) form such a list.

We show that for every $T \in T$, $T$ belongs to one of the cosets $S$, $AS$, $BS$ and $CS$. To do that we will first narrow down the list of possible signatures for $[1, 0, 1]T^{\otimes 2}$ and $[1, 0]T$, by right multiplication from $S$. By assumption that $T \in T$, both $[1, 0, 1]T^{\otimes 2}$ and $[1, 0]T$ belong to $T_{123}$. Binary signatures of $T_{123}$ are nonzero scalar multiples of $[1, 0, \pm 1], [1, 0, \pm i], [1, \pm 1, -1], [1, \pm i, 1], [0, 1, 0]$.

Unary signatures of $T_{123}$ are nonzero scalar multiples of $[1, \pm 1], [1, \pm i], [1, 0], [0, 1]$.

Using powers of $D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, we may already reduce $[1, 0, -1]$ to $[1, 0, 1]$, reduce $[1, 0, -i]$ to $[1, 0, i]$, and reduce $[1, 1, -1]$ to $[1, -1, -1]$, then to $[1, -i, 1]$, and finally to $[1, i, 1]$. We may use $Z^{-1}$ to reduce $[0, 1, 0]$ to $[1, 0, 1]$, as $[1, 0, 1]Z^{\otimes 2} = [0, 1, 0]$. We may further reduce $[1, 0, i]$ to $[1, i, 1]$ as follows. $[1, 0, i]H^{\otimes 2} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} = \frac{1+i}{2}[1, -i, 1]$, and $[1, -i, 1](D^2)^{\otimes 2} = [1, i, 1]$.

So we may assume $[1, 0, 1]T^{\otimes 2}$ is a nonzero scalar multiple of $[1, 0, 1]$ or $[1, i, 1]$. Note that these two signatures are invariant under reversal. Now we further normalize $[1, 0]T$ by the reversal operation $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which does not affect the normalization already done for $[1, 0, 1]T^{\otimes 2}$. Then it is clear that we may assume $[1, 0]T$ is a nonzero scalar multiple of $\{(1, 0), [1, 1], [1, -1], [1, i]\}$.

In the following, we denote by $\alpha = \frac{1+i}{\sqrt{2}} = \sqrt{i}$. We also omit a nonzero scalar factor if it does not cause any confusion.

Suppose $[1, 0, 1]T^{\otimes 2} = \lambda[1, 0, 1]$. If $[1, 0]T = \mu[1, 0]$, then $T = [1, 0, \pm 1] \in S$, with coset representative $I$. If $[1, 0]T = \mu[1, 1]$, then $T = \begin{bmatrix} 1 & 0 \\ \pm 1 & \pm 1 \end{bmatrix} \in S$, with coset representative $I$. Note that $[1, 1, -1] = \sqrt{2}H$ and $[1, -1, 1] = \sqrt{2}H$. If $[1, 0]T = \mu[1, 1]$, then $T = \begin{bmatrix} 1 & 0 \\ \pm 1 & \pm 1 \end{bmatrix} \in S$, with coset representative $I$. Note that $[1, -1, -1] D^2 = \begin{bmatrix} 1 & 1 \\ 1, 1 \end{bmatrix} \in S$ and $[1, 1, -1] D^2 X \in S$. Finally $[1, 0]T = \mu[i, 1]$ with $[1, 0, 1]T^{\otimes 2} = \lambda[1, 0, 1]$ is impossible, since the latter implies that $T$ is orthogonal up to a nonzero factor, but the dot product $[1, i] : [1, i] = 0$.

Next suppose $[1, 0, 1]T^{\otimes 2} = \lambda[1, i, 1]$. If $[1, 0]T = \mu[1, 0]$, then $T = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \pm 1 \end{bmatrix}$. Clearly with the + sign, $T = B \in BS$, and with the − sign, $T = CD^2 \in CS$. If $[1, 0]T = \mu[1, 1]$,
then \(T = \begin{bmatrix} 1 & -1 \\ \pm i & \mp i \end{bmatrix}\). Note that \(A^{-1} \begin{bmatrix} 1 & -1 \\ \pm i & \mp i \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \in S\), being a scalar multiple of \(Z\), and also \(A^{-1} \begin{bmatrix} 1 & -1 \\ \pm i & \mp i \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \in S\), being a scalar multiple of \(ZX\). If \([1, 0]T = \mu[1, -1]\), then \(T = \begin{bmatrix} 1 & -1 \\ \pm i & \mp i \end{bmatrix}\). Note that \(A^{-1} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \in S\), being a multiple of \(HD^2\), and also \(A^{-1} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \in S\), being a multiple of \(HXD^2\). Finally, if \([1, 0]T = \mu[1, i]\), then \(T = \begin{bmatrix} 1 & i \\ \pm i & \mp i \end{bmatrix}\). Note that \(C^{-1} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \in S\), and \(B^{-1} \begin{bmatrix} 1 & i \\ 0 & -i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ i & -i \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \in S\).

We have proved that for every \(T \in \mathcal{F}\), there exists some \(M \in \{I, A, B, C\}\), such that \(T \in MS\). We remark that the four coset representatives \(\{I, A, B, C\}\) represent distinct right cosets. This amounts to verifying that \(A, B, C, B^{-1}A, C^{-1}A, C^{-1}B\) all do not belong to \(S\). We have \([1, 1]A = [1, \sqrt{i}], [1, 1]B = [\sqrt{2} + i, 1]\) and \([1, 1]C = [\sqrt{2} - i, 1]\), having two nonzero entries with unequal norm or a ratio not a power of \(i\), hence not in \(\mathcal{A}\). Similarly \([0, 1]B^{-1}A = \frac{1}{\sqrt{2}}[0, 1] \begin{bmatrix} 1 & 0 \\ -i & 0 \end{bmatrix} = \frac{1}{\sqrt{2}}[0, 1, \sqrt{2}], [0, 1]C^{-1}A = \frac{1}{\sqrt{2}}[0, 1] \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{\sqrt{2}}[0, 1, \sqrt{2}], [0, 1]C^{-1}B = \frac{1}{\sqrt{2}}[0, 1] \begin{bmatrix} 1 & 0 \\ \sqrt{2}i & 0 \end{bmatrix} = [0, 1, \sqrt{2}], [0, 1]A = \frac{1}{\sqrt{2}}[0, 1] \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = 0\).

We conclude this discussion by restating Theorem 6.12 in the following effective form:

**Theorem 6.14.** Let \(\mathcal{F}\) be a set of complex symmetric signatures. Holant\(^c(\mathcal{F})\) is \#P-hard unless \(\mathcal{F}\) satisfies one of the following conditions, in which case it is tractable:

1. Holant\(^c(\mathcal{F})\) is tractable, for which we have an effective dichotomy—Theorem 2.12; or
2. There exists a matrix \(T\) in the set defined in (6.7), such that \(T^{-1}\mathcal{F} \subseteq \mathcal{A}\).

### 6.4 A Dichotomy for \#CSP\(^d\)

In this section we consider a variant of the standard Boolean \#CSP problem, where every variable occurs a multiple of \(d\) times, for some integer \(d \geq 1\). We will prove a complexity dichotomy, due to Huang and Lu [?], for this class of problems for any set of symmetric signatures on Boolean variables. This variant is interesting in its own right, but more importantly it is a key step in the proof of the general dichotomy in Chapter 7.

**Definition 6.15.** Let \(d \geq 1\) be an integer. Given a set of signatures \(\mathcal{F}\), we use \#CSP\(^d(\mathcal{F})\) to denote the restriction of Boolean \#CSP(\(\mathcal{F}\)) where every variable occurs a multiple of \(d\) times.

For example, \#CSP\(^2\) refers to \#CSP problems where every variable occurs an even number of times. Define

\[\mathcal{T}_d = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & \omega_d^k \end{bmatrix} \mid 0 \leq k < d \right\},\]

where \(\omega_d = e^{2\pi i/d}\). We may apply a holographic transformation by \(T \in \mathcal{T}_d\). Since each variable occurs a multiple of \(d\) times, this transforms \#CSP\(^d(\mathcal{F})\) to the equivalent \#CSP\(^d(T^{-1}\mathcal{F})\).
Consider the problem \( \#\text{CSP}^3([1, \omega_3, \omega_6]) \), where \( \omega_3 = e^{2\pi i/3} = -1 + \sqrt{3}i \) and \( \omega_6 = e^{2\pi i/6} = 1 + \sqrt{3}i \). Taking \( T = \begin{bmatrix} 1 & 0 \\ 0 & \omega_3 \end{bmatrix} \), the problem \( \#\text{CSP}^3([1, \omega_3, \omega_6]) \) is transformed to the equivalent problem \( \#\text{CSP}^3([1, 1, -1]) \) (where \( -1 = \omega_6/\omega_3^2 \)), which is computable in polynomial time, by Theorem 3.7. Note that the standard \( \#\text{CSP}([1, \omega_3, \omega_6]) \) is \#P-hard, by the same theorem.

As another example, \( \#\text{CSP}^2([1, i^{3/2}, i]) \) is computable in polynomial time. We can first write \( \#\text{CSP}^2([1, i^{3/2}, i]) \) as an equivalent bipartite Holant problem \( \text{Holant}([1, 2, \ldots] | [1, i^{3/2}, i]) \). Then we can perform a holographic transformation by \( T = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt[i]{1} \end{bmatrix} \in T_8 \). The LHS \( (=2k)T^{\otimes 2k} = [1, 0]^{\otimes 2k} + i^k [0, 1]^{\otimes 2k} \in \mathcal{A} \). On the RHS, the transformed signature is \( T^{-1}[1, i^{3/2}, i] = [1, i, 1] \in \mathcal{A} \) as well. It follows that the problem is tractable in \( P \). Note that the standard \( \#\text{CSP}([1, i^{3/2}, i]) \) is \#P-hard by Theorem 3.7.

The dichotomy theorem for \( \#\text{CSP}^d \) problems says that these are essentially the only additional tractable cases; all others are \#P-hard.

**Theorem 6.16 (Huang-Lu).** Let \( d \geq 1 \) be an integer, and \( \mathcal{F} \) be a set of symmetric signatures from Boolean variables to \( \mathbb{C} \). Then \( \#\text{CSP}^d(\mathcal{F}) \) is \#P-hard, unless there exists some \( T \in T_{4d} \) such that \( \mathcal{F} \subseteq T\mathcal{P} \) or \( \mathcal{F} \subseteq T\mathcal{A} \), in which case the problem \( \#\text{CSP}^d(\mathcal{F}) \) is in \( P \).

We remark that since \( T \) in the theorem is diagonal, the statement \( \mathcal{F} \subseteq T\mathcal{P} \) can be replaced by simply \( \mathcal{F} \subseteq \mathcal{P} \).

Before we prove this theorem, we recall Theorem 6.5. The condition that the needed binary or unary signatures being normalized can always be achieved by a holographic transformation \( T \in T_3 \) defined in (6.3). Thus we can restate Theorem 6.5 as follows:

**Theorem 6.17.** Suppose \( \mathcal{F} \) and \( \mathcal{G} \) are two sets of signatures, where \( \mathcal{G} \) contains a non-degenerate binary signature \([g_0, g_1, g_2]\). Furthermore, if \( g_0 = g_2 = 0 \) then we assume \( \mathcal{G} \) also contains a unary signature \([a, b]\) with \( ab \neq 0 \). Then

\[
\text{Holant}(\mathcal{G} | [1, 0, 0, 1] \cup \mathcal{F})
\]

is \#P-hard unless there exists some \( T \in T_3 \) such that \( \mathcal{G}T \cup T^{-1}\mathcal{F} \subseteq \mathcal{P} \) or \( \mathcal{G}T \cup T^{-1}\mathcal{F} \subseteq \mathcal{A} \), in which case the problem is in \( P \).

The tractability proof of Theorem 6.16 is straightforward. The hardness proof of Theorem 6.16 will use Theorem 6.17. For general \( \#\text{CSP} \) problems, there is a pinning lemma (Lemma 3.13) which allows the use of the unary \( \Delta_0 \) and \( \Delta_1 \) functions. We will show that in the \( \#\text{CSP}^d \) setting, one can obtain a weaker version of pinning, namely one can pin \( d \) occurrences of variables at a time to the same value, and then we will show that this is sufficient here. The main part of the proof of Theorem 6.16 is to show that the theorem holds when \( \Delta_0 \) and \( \Delta_1 \) are present, which becomes \( \text{Holant}^c \) in essence. The proof uses Theorem 6.17 to repeatedly regularize a given signature set, lest the problem be \#P-hard. Finally the signature set is sufficiently regularized that it must satisfy the tractability criteria if it is not \#P-hard.
We will postpone the proof of the weaker pinning lemma for now, and focus first on the following Theorem 6.18. Note that $T^{-1} \Delta_c$ is a scalar multiple of $\Delta_c$ for an invertible diagonal $T$, thus $T^{-1} \Delta_c \in \mathcal{P} \cap \mathcal{A}$. Hence Theorem 6.18 is also a special case of Theorem 6.16.

**Theorem 6.18.** Let $d \geq 1$ be an integer, and $\mathcal{F}$ be a set of symmetric signatures from Boolean variables to $\mathbb{C}$. Then $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is $\#P$-hard, unless there exists some $T \in \mathcal{T}_{4d}$ such that $\mathcal{F} \subseteq T \mathcal{P}$ or $\mathcal{F} \subseteq T \mathcal{A}$, in which case the problem $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is in $P$.

Again we remark that since $T$ in the theorem is diagonal, $\mathcal{F} \subseteq T \mathcal{P}$ is equivalent to $\mathcal{F} \subseteq \mathcal{P}$. We may also remove identically 0 signatures from $\mathcal{F}$ since it does not affect both the complexity of the problem and the tractability criteria.

Before we prove Theorem 6.18 we make some simple observations: Suppose $M = \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix} \in \text{GL}_2(\mathbb{C})$.

1. Signatures of the form $[f_0, 0, f_2, \ldots]$ or $[0, f_1, 0, \ldots]$ maintain their respective form under a holographic transformation by $M$. In particular this is true for $\Delta_0$ and $\Delta_1$.
2. $f \in \mathcal{P}$ iff $fM \in \mathcal{P}$. This can be seen by the effect of the transformation $M$ on $(\neq_2)$, $(\neq_2)$ and the unary functions.
3. For all $d \geq 1$, $\mathcal{T}_d$ is a cyclic group. $M \in \mathcal{T}_d$ iff $M^{-1} \in \mathcal{T}_d$.
4. For all $d \geq 1$, $(\neq_d)M^{\otimes d}$ is a scalar multiple of $(\neq_d)$ iff $M \in \mathcal{T}_d$.
5. For all $d \geq 1$, $(\neq_d)M^{\otimes d} \in \mathcal{A}$ iff $M \in \mathcal{T}_{4d}$.

**Proof of Theorem 6.18.** We start with the following equivalence

$$\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \equiv_T \text{Holant}(\mathcal{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1\})$$

where we denote $\mathcal{E}_d = \{\neq_d, \neq_{2d}, \ldots\}$. We first prove tractability. Suppose there exists some $T \in \mathcal{T}_{4d}$ such that $T^{-1} \mathcal{F} \subseteq \mathcal{P}$, or $T^{-1} \mathcal{F} \subseteq \mathcal{A}$, then we transform the bipartite Holant problem by $T$.

$$\text{Holant}(\mathcal{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1\}) \equiv_T \text{Holant}(\mathcal{E}_d T \mid T^{-1} \mathcal{F} \cup \{\Delta_0, \Delta_1\}).$$

We have $\mathcal{E}_d T \subset \mathcal{P} \cap \mathcal{A}$, and either all $T^{-1} \mathcal{F} \cup \{\Delta_0, \Delta_1\} \subseteq \mathcal{P}$ or all $T^{-1} \mathcal{F} \cup \{\Delta_0, \Delta_1\} \subseteq \mathcal{A}$. Thus the problem is computable in polynomial time.

We turn to the proof of hardness next. We first prove a general reduction about $\#\text{CSP}^d$ problems. For a function $f$, let $f^d$ be the function such that $f^d(x) = (f(x))^d$.

**Lemma 6.19.** For all $d \geq 1$, $\#\text{CSP}(f^d) \leq_T \#\text{CSP}^d(f)$.

**Proof.** Given an instance of $\#\text{CSP}(f^d)$, we may replace each constraint of $f^d$ by $d$ many copies of $f$, each applied to the same sequence of occurrences of variables. This gives the same answer to the given instance, and the new instance is a valid instance of $\#\text{CSP}^d(f)$ since the number of occurrences of every variable is multiplied by $d$. The lemma is proved. \qed
By Theorem 3.7, both \( \#\text{CSP}(\{0,1,0,0\}) \) and \( \#\text{CSP}(\{0,0,1,0\}) \) are \#P-hard. Then by Lemma 6.19, we obtain the next lemma.

**Lemma 6.20.** For all \( d \geq 1 \), \( \#\text{CSP}^d(\{0,1,0,0\}) \) and \( \#\text{CSP}^d(\{0,0,1,0\}) \) are \#P-hard.

We continue with the proof of Theorem 6.18. Suppose \( \text{Holant}(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \) is not \#P-hard. Connect \( (=d) \) on LHS and \( d - 1 \) copies of \( \Delta_c \) on RHS we get \( \Delta_c \) on LHS, for \( c = 0, 1 \), with which we can obtain all sub-signatures of \( \mathcal{F} \) on RHS. Since we can obtain all sub-signatures of \( \mathcal{F} \), we can and will replace every degenerate (but not identically zero) \( f = [a, b]^{\otimes n} \in \mathcal{F} \) by (a nonzero multiple of) the unary signature \( [a, b] \), without changing the complexity of the problem. If \( \mathcal{F} \) consists of only unary functions then \( \mathcal{F} \subseteq \mathcal{U} \subseteq \mathcal{P} \). If all binary sub-signatures of \( \mathcal{F} \) are degenerate, then by Proposition 2.8, we also have \( \mathcal{F} \subseteq \mathcal{P} \). Therefore we may assume \( \mathcal{F} \) contains a non-degenerate binary sub-signature.

The next claim is crucial: we may assume for any \( f = [f_0, f_1, \ldots, f_n] \in \mathcal{F} \) \((n \geq 1)\) for any two consecutive terms at least one term must be zero,

\[
f_if_{i+1} = 0 \quad (0 \leq i < n). \tag{6.11}
\]

Suppose for some \( f \) and some \( 0 \leq i < n \), \( f_if_{i+1} \neq 0 \). Then by pinning we can realize \([f_i, f_{i+1}]\), which is \([1, a]\) up to a scale, for some \( a \neq 0 \). By connecting \( 3d - 3 \) copies of \([1, a]\) to \( =_d \) on LHS we get \([1, 0, 0, a^{3d-3}] \) on LHS. Now we apply the holographic transformation by \( M = \begin{bmatrix} 1 & 0 \\ 0 & a^{-(d-1)} \end{bmatrix} \), the LHS signature \([1, 0, 0, a^{3d-3}] \) is transformed to \([1, 0, 0, a^{3d-3}]M^{\otimes 3} = [1, 0, 0, 1] \) and the RHS has a non-degenerate binary signature, as well as a unary signature \( M^{-1}\otimes [1, a] = [1, a^d] \) with both entries nonzero. By Theorem 6.17 \( \text{Holant}(\mathcal{E}_d \cup \{\Delta_0, \Delta_1\}) \) is \#P-hard unless there exists some \( M' = MT \) where \( T \in \mathcal{T}_d \), such that \( \mathcal{E}_dM' \cup M'^{-1}\mathcal{F} \) is a subset of \( \mathcal{P} \) or a subset of \( \mathcal{A} \). Since \( T \) is diagonal, if \( M'^{-1}\mathcal{F} \subseteq \mathcal{P} \) then \( \mathcal{F} \subseteq \mathcal{P} \). For the second alternative, \( (=d)M' \in \mathcal{A} \) implies that \( M' \in \mathcal{T}_d \). Thus we have established the claim in (6.11), or else we have finished the proof of Theorem 6.18.

Thus, in particular, the only unary signatures in \( \mathcal{F} \) are (nonzero multiples of) \( \Delta_0 \) or \( \Delta_1 \), the only binary signatures in \( \mathcal{F} \) have the form \([0, *, 0]\) \((\text{DISEQUALITY})\) or \([*, 0, *]\) \((\text{GEN-EQ})\). If all signatures in \( \mathcal{F} \) have arity at most two, then \( \mathcal{F} \subseteq \mathcal{P} \), and Theorem 6.18 is proved.

Therefore we may assume there are signatures in \( \mathcal{F} \) of arity at least three. If they are all \text{GEN-EQ}, then again \( \mathcal{F} \subseteq \mathcal{P} \). So we may assume there is \( f = [f_0, f_1, \ldots, f_n] \in \mathcal{F} \) of arity \( n \geq 3 \) and is not a \text{GEN-EQ}. We claim that for any such \( f \), there are no two consecutive zero entries: \( f_i = f_{i+1} = 0 \). Suppose otherwise. Since it is not a \text{GEN-EQ}, there exists some \( 0 < j < n \) such that \( f_j \neq 0 \). Either \( j < i \), in which case we pick the maximum such \( j < i \); or \( j > i + 1 \), in which case we pick the minimum such \( j > i + 1 \). By (6.11), we get a sub-signature \([0, f_j, 0, 0]\) or \([0, 0, f_j, 0]\). By Lemma 6.20, we conclude that \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \) is \#P-hard.

We can therefore strengthen the claim in (6.11) to say that every signature in \( \mathcal{F} \) other than \text{GEN-EQ} must have the following strictly alternating form

\[
\forall i, (0 \leq i < n), \quad \text{exactly one of } f_i \text{ or } f_{i+1} = 0. \tag{6.12}
\]
Thus other than GEN-Eq, every \( f \in \mathcal{F} \) has the form \([f_0, 0, f_2, \ldots] \) or \([0, f_1, 0, \ldots] \), where all symbols \( f_i \neq 0 \), and \( \mathcal{F} \) contains such a signature of arity at least three. The strengthened claim (6.12) applies to all signatures in \( \mathcal{F} \) of any arity \( n \geq 1 \), other than GEN-Eq.

In particular there exists some sub-signature \([1, 0, a] \) from \( \mathcal{F} \) with \( a \neq 0 \). Suppose there is a non-degenerate GEN-Eq in \( \mathcal{F} \) with an odd arity \( r \geq 3 \). We connect this GEN-Eq on RHS with \( =_{2rd} \) on LHS having an even arity, to create a non-degenerate GEN-Eq on LHS of an odd arity \( 2rd - r \geq 3 \). Then we connect zero or more copies of \([1, 0, a] \) on RHS to it to realize a non-degenerate GEN-Eq on LHS of arity three. Now we apply a diagonal holographic transformation \( M \) to map it to \([1, 0, 0, 1] \) on LHS, while \( M^{-1}[1, 0, a] \) on RHS is non-degenerate. It does not have the form \([0, *, 0] \) since \( M \) is diagonal. This allows us to apply Theorem 6.17. This completes the proof of Theorem 6.18 in this case.

We summarize what we may assume for \( \mathcal{F} \):

Every unary signature in \( \mathcal{F} \) is \( \Delta_0 \) or \( \Delta_1 \). Every signature in \( \mathcal{F} \) of arity greater than one is non-degenerate, and is either (1) a GEN-Eq of even arity, or (2) has a strictly alternating form. And there is at least one signature in \( \mathcal{F} \) of type (2), with arity at least three.

Without loss of generality, we assume \( \mathcal{F} \) contains a sub-signature of the form \([0, 1, 0, a] \), for \( a \neq 0 \). The other form \([a, 0, 1, 0] \) is symmetric.

By Lemma 6.19, we see that

\[
\#\text{CSP}([0, 1, 0, a^d]) \leq \#\text{CSP}^d([0, 1, 0, a]).
\]

If \( \#\text{CSP}^d([0, 1, 0, a]) \) is \#P-hard, then Theorem 6.18 is proved. Suppose \( \#\text{CSP}^d([0, 1, 0, a]) \) is not \#P-hard, then \( \#\text{CSP}([0, 1, 0, a^d]) \) is also not \#P-hard, and by Theorem 3.7, \([0, 1, 0, a^d] \in \mathcal{P} \cup \mathcal{A} \). However \([0, 1, 0, a^d] \notin \mathcal{P} \), so we have \([0, 1, 0, a^d] \in \mathcal{A} \). This forces \( a^d = \pm 1 \) (see the table for \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \) in Section 3.2.) Now we can apply a holographic transformation with \( T = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix} \) to \( \text{Holant}(\mathcal{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1\} \cup \{[0, 1, 0, a]\}) \). Note that \( T \in \mathcal{T}_{ad} \), as \( a^{2d} = 1 \). \((T^{-1})^\otimes 3 \) maps \([0, 1, 0, a] \) to \([0, 1, 0, 1] \) in RHS up to a scalar \( a^{-1/2} \). Being diagonal \( T^{-1} \) maintains \( \Delta_0 \) and \( \Delta_1 \). All \textsc{Equality} functions \((=_{4ad}) \) in \( \mathcal{E}_{4d} \subset \mathcal{E}_d \) on LHS remain unchanged. Thus we still can obtain \( \Delta_0 \) and \( \Delta_1 \) on LHS and thus all sub-signatures of RHS of \( \text{Holant}(\mathcal{E}_d T \mid T^{-1} \mathcal{F} \cup \{\Delta_0, \Delta_1\} \cup \{[0, 1, 0, 1]\}) \). In particular we can realize \((=)_2 = [1, 0, 1] \) on RHS. Then we can realize all \( \mathcal{E}_d = \{=, =_2, \ldots\} \) on LHS from \( \mathcal{E}_{4d} \) on LHS and \((=)_2 \) on RHS. Since we have \((=)_2 \) on both sides now, we can reduce the following non-bipartite Holant problem to the original problem \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \):

\[
\begin{align*}
&\text{Holant}(\mathcal{E}_d \cup T^{-1} \mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, 1]\}) \\
\leq_T \text{Holant}(\mathcal{E}_d T \mid T^{-1} \mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, 1]\}) \\
\equiv_T \text{Holant}(\mathcal{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, a]\}) \\
\equiv_T \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, a]\}) \\
\equiv_T \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})
\end{align*}
\]
Our final step is to prove that the problem in (6.13) is \#P-hard, unless \( T^{-1} \mathcal{F} \subseteq \mathcal{A} \), which will complete the proof of Theorem 6.18.

Since \( T \) is diagonal, the summary on what we may assume for \( \mathcal{F} \) is also valid for \( T^{-1} \mathcal{F} \). Next we show that if there is a non-degenerate GEN-Eq in \( T^{-1} \mathcal{F} \) not in \( \mathcal{A} \), then the problem in (6.13) is \#P-hard. After a scalar multiplication we may assume it is \([1, 0, \ldots, 0, b] \in T^{-1} \mathcal{F}\) of even arity \( 2k \geq 2 \), with \( b \neq 0 \). If \( 2k = 2 \), then it is \([1, 0, b]\). If \( 2k \geq 4 \), then by connecting it to \((=2k-2) \in \mathcal{E}^2\) we can realize \([1, 0, b]\). Now we reduce \#CSP([1, 0, b], [0, 1, 0, 1]) to \#CSP^2([1, 0, b], [1, 0, 1], [0, 1, 0, 1]), which in turn is reducible to the problem in (6.13): Each variable \( x \) which occurs in \( \ell \) constraints in an instance of \#CSP([1, 0, b], [0, 1, 0, 1]) will be made to occur in \( 2\ell \) constraints. Each constraint \([1, 0, b]\) is replaced by a pair of constraints \([1, 0, b] \) and \((=2) = [1, 0, 1]\) on the same two occurrences of variables. Each constraint \([0, 1, 0, 1]\) is replaced by two copies of the same constraint \([0, 1, 0, 1]\) on the same three occurrences of variables. This reduction shows that the problem in (6.13) is \#P-hard unless \{[1, 0, b], [0, 1, 0, 1]\} \( \subset \mathcal{P} \) or \{[1, 0, b], [0, 1, 0, 1]\} \( \subset \mathcal{A} \). As \([0, 1, 0, 1]\) \( \not\in \mathcal{P} \), we get \([1, 0, b] \in \mathcal{A} \), and thus \( b^4 = 1 \). This shows that all non-degenerate GEN-Eq in \( T^{-1} \mathcal{F} \) are in \( \mathcal{A} \). In particular we have shown that all signatures in \( T^{-1} \mathcal{F} \) of arity at most two are in \( \mathcal{A} \). Note that binary signatures in \( T^{-1} \mathcal{F} \) are either GEN-Eq or of the form \([0, *, 0] \in \mathcal{A} \).

Now we prove that all signatures in \( T^{-1} \mathcal{F} \) of the strictly alternating form are also in \( \mathcal{A} \). We may assume it has arity at least three. Suppose \( T^{-1} \mathcal{F} \) has a ternary sub-signature of the form \([0, 1, 0, b] \) \( (b \neq 0) \), after a scalar multiple. We can reduce \#CSP([0, 1, 0, b]) to the problem \#CSP^2([0, 1, 0, b], [0, 1, 0, 1]), which is in turn reducible to the problem in (6.13). The reduction from \#CSP([0, 1, 0, b]) to \#CSP^2([0, 1, 0, b], [0, 1, 0, 1]) is as follows: We replace each constraint \([0, 1, 0, b]\) with a copy of \([0, 1, 0, b]\) and \([0, 1, 0, 1]\). Each variable now occurs an even number of times. Thus the problem \#CSP([0, 1, 0, b]) is \#P-hard unless \([0, 1, 0, b] \in \mathcal{A} \), which forces \( b = \pm 1 \). This argument can be extended to ternary sub-signatures of the form \([b, 0, 1, 0] \) \( (b \neq 0) \). Indeed, within the problem in (6.13) we can construct \((=2) = [0, 1, 0]\) as a sub-signature of \([0, 1, 0, 1]\), and then connecting \((=2)\) to all three inputs of \([b, 0, 1, 0]\) converts it to \([0, 1, 0, b]\), and we conclude that \( b = \pm 1 \). Now extending this pattern to any signature in \( T^{-1} \mathcal{F} \) of the strictly alternating form, we see that either it is in \( \mathcal{A} \) or there is a sub-signature of arity 4 of the form \([1, 0, 1, 0, -1]\) or \([1, 0, -1, 0, -1]\).

By Theorem 6.14, both Holant\(^c([1, 0, 1, 0, -1])\) and Holant\(^c([1, 0, -1, 0, -1])\) are \#P-hard. This involves a check with Theorem 2.12 (or more directly Lemma 2.16), and a check with the four matrices in (6.7) with reference to the table for \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \) in Section 3.2. Hence the problem in (6.13) is also \#P-hard. This completes our proof of Theorem 6.18.

Now we give a suitable pinning lemma for \#CSP\(^d\), to remove the extra assumption that the signature set contains \( \Delta_0 \) and \( \Delta_1 \). This will improve Theorem 6.18 to Theorem 6.16.

**Lemma 6.21.** For any integer \( d \geq 1 \) and any \( \mathcal{F} \), \( \#CSP^d(\mathcal{F}) \equiv_T \#CSP^d(\mathcal{F} \cup \{\Delta_0^{\otimes d}, \Delta_1^{\otimes d}\}) \).

The proof of this lemma is essentially the same as that of Lemma 3.13. In that proof, we
defined \( V_c \) to be the set of variables to which \( \Delta_\ast \) is applied \((c = 0, 1)\). Here we also define \( V_c \) to be the set of variables \( x \) such that \( x \) is among the inputs to at least one occurrence of \( \Delta_c^\otimes \) \((c = 0, 1)\). Again we may assume \( V_0 \cap V_1 = \emptyset \), otherwise there is no satisfying assignment. The total number of occurrences of variables applied to by \( \Delta_c^\otimes \) in each \( V_c \) \((c = 0, 1)\) is a multiple of \( d \). Then we replace all variables in \( V_0 \) and \( V_1 \) by two new variables \( y_0 \) and \( y_1 \) respectively, each occurring a multiple of \( d \) times, which is the total number of occurrences of variables in \( V_0 \) and \( V_1 \) respectively. We now observe that when we remove all occurrences of the constraints \( \Delta_0^\otimes \) and \( \Delta_1^\otimes \), both variables \( y_0 \) and \( y_1 \) occur a multiple of \( d \) times, making it an instance of \( \#\text{CSP}^d(\mathcal{F}) \). This is also true when we replace both \( V_0 \) and \( V_1 \) by a single variable \( y \) in the second step of the proof in Lemma 3.13. The rest of the proof is exactly the same as that of Lemma 3.13.

Using Lemma 6.21, we can show the equivalence of \( \#\text{CSP}^d(\mathcal{F}) \) and \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \), given the fact that we have already proved a dichotomy theorem for the latter, thus deriving a dichotomy theorem for the former.

**Lemma 6.22.** \( \#\text{CSP}^d(\mathcal{F}) \) is in \( P \) (or is \( \#P \)-hard, respectively) iff \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \) is in \( P \) (or is \( \#P \)-hard, respectively).

**Proof.** Clearly \( \#\text{CSP}^d(\mathcal{F}) \leq_T \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \). We have already proved a dichotomy theorem for the latter problem \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \). If it is computable in polynomial time, then so is \( \#\text{CSP}^d(\mathcal{F}) \). If \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \) is \( \#P \)-hard, we show \( \#\text{CSP}^d(\mathcal{F}) \) is also \( \#P \)-hard.

We observe that in all the proofs for \( \#P \)-hardness so far, which ultimately lead to the \( \#P \)-hardness of \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \), we reduce from one of the following three problems to it by a chain of reductions: (a) Holant\(_{\ast} \) \([0, 1, 1, 1]\) (VERTEX COVER for 3-regular graphs), (b) Holant\(_{\ast} \) \([1, 1, 0, 0]\) (MATCHING), or (c) Holant\(_{\ast} \) \([0, 1, 0, 0]\) (PERFECT MATCHING). There are only three reduction methods in this reduction chain, namely, direct gadget construction, polynomial interpolation, and holographic reduction.

For any signature grid \( \Omega \) as an input to one of the three Holant problems listed above, the answer Holant\(_{\ast} \) to the Holant problem is a non-negative integer. Let \( d\Omega \) denote the signature grid which is a disjoint union of \( d \) copies of \( \Omega \). Then the Holant value Holant\(_{\ast} \) \( d\Omega \) = \((\text{Holant}\(_{\ast} \) \Omega \))^\(d \), the \( d \)-th power of the Holant value on \( \Omega \). Thus Holant\(_{\ast} \) \( \Omega \) can be recovered uniquely from Holant\(_{\ast} \) \( d\Omega \), and therefore we may start our reduction from \( d\Omega \) instead of \( \Omega \). In each reduction step, the reduction operation can be carried out in each disjoint copy of \( \Omega \) in parallel. Start from one problem, suppose the reduction chain on the instance \( \Omega \) produced instances \( \Omega_1, \Omega_2, \ldots, \Omega_m \) of the second problem. The same reduction applied to \( d\Omega \) produces instances of the form \( d\Omega_1, d\Omega_2, \ldots, d\Omega_m \). Here \( m \) and \( m' \) are polynomially bounded in the size of the problem instance. Note that the number of steps in the reduction chain is a constant depending only on the problem, but not the size of the problem instance.

Finally we reach a sequence of instances of the form \( d\Omega_i \) for the problem \( \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \), following the \( \#P \)-hardness reduction of this problem. By the syntactic form of \( d\Omega_i \), each \( \Delta_0 \) and \( \Delta_1 \) is applied a multiple of \( d \) times. Thus we can rewrite this instance as
an instance in \( \#\text{CSP}^d(F \cup \{\Delta_0^{\otimes d}, \Delta_1^{\otimes d}\}) \). This shows that \( \#\text{CSP}^d(F \cup \{\Delta_0^{\otimes d}, \Delta_1^{\otimes d}\}) \) is also \#P-hard. By Lemma 6.21, \( \#\text{CSP}^d(F) \) is \#P-hard.

This completes the proof of Theorem 6.16. We conclude this section with two remarks. The first is concerned with a subtlety in the proof of Lemma 6.22. The statement that the two problems \( \#\text{CSP}^d(F) \) and \( \#\text{CSP}^d(F \cup \{\Delta_0, \Delta_1\}) \) are either simultaneously in P or simultaneously \#P-hard technically does not depend on the assumption that P is different from \#P, although the presumed truth of this hypothesis is what makes the theory of this book interesting. The trivial reduction \( \#\text{CSP}^d(F) \leq_{T} \#\text{CSP}^d(F \cup \{\Delta_0, \Delta_1\}) \), shows that if the former is \#P-hard, then so is the latter; and if the latter is in P, then so is the former. The proof of the lemma shows that if the latter is \#P-hard, then so is the former. Now suppose the former is in P. We want to claim that the latter is also in P. Suppose not. Then by the dichotomy theorem already proved, Theorem 6.18, the latter problem \( \#\text{CSP}^d(F \cup \{\Delta_0, \Delta_1\}) \) is \#P-hard. The proof of Lemma 6.22 shows that the former problem \( \#\text{CSP}^d(F) \) is also \#P-hard. Thus \#P collapses to P, and therefore \( \#\text{CSP}^d(F \cup \{\Delta_0, \Delta_1\}) \) is in P anyhow.

The second remark is that the proof of Lemma 6.22 does not maintain planarity in general. This is because when we substitute \( \Delta_0^{\otimes d} \) and \( \Delta_1^{\otimes d} \) for \( d \) occurrences of \( \Delta_0 \) and \( \Delta_1 \) in a signature grid of the form \( d\Omega \), there is no guarantee we can do so while maintaining planarity. In the next section we will see that in some circumstances this is still possible.

## 6.5 Eulerian Orientation

In this section, we prove a specific problem to be \#P-hard. This problem will serve as a new basic problem from which we prove \#P-hardness in the dichotomy theorems that follow. We first define the problem. Recall that an orientation of an undirected graph is to assign a unique direction, out of two possible directions, to each edge.

**Definition 6.23.** Given a graph \( G \), an orientation is an Eulerian orientation if for each vertex \( v \) of \( G \), the number of incoming edges of \( v \) equals the number of outgoing edges of \( v \).

We will prove that the problem of counting the number of Eulerian orientations, denoted by \( \#\text{EULERIAN-ORIENTATIONS} \), or \#EO for short, is \#P-hard over planar 4-regular graphs. There is a natural expression of this problem as a bipartite Holant problem. Given a 4-regular undirected graph \( G \), we assign the binary Disequality signature on each edge, which corresponds to an orientation. Then we assign the Exact-Two signature \([0,0,1,0,0]\) of arity 4 for each vertex of \( G \). This creates a bipartite signature grid out of the Edge-Vertex incidence graph of \( G \). Then \#EULERIAN-ORIENTATIONS can be expressed as Holant(\( \neq_2 \mid [0,0,1,0,0] \)). Over planar graphs, we use the notation Pl-Holant(\( \neq_2 \mid [0,0,1,0,0] \)) to denote this problem.

The Tutte polynomial of an undirected graph is a polynomial in two variables with integer coefficients. It contains a wealth of information about the graph.
Definition 6.24 (cf. [?]). Let \( G = (V, E) \) be an undirected graph, the Tutte polynomial of \( G \) is defined as
\[
T(G; x, y) = \sum_{A \subseteq E} (x - 1)^{k(A)}(y - 1)^{k(A) + |A| - |V|},
\]
where \( k(A) \) denotes the number of connected components of the graph \((V, A)\).

To prove the problem Pl-Holant(\( \neq 2 \mid [0, 0, 1, 0, 0] \)) \#P-hard, we reduce from the problem of evaluating the Tutte polynomial at the point \((3, 3)\), which is \#P-hard for planar graphs.

The following theorems give complexity classifications of the Tutte polynomial over general and planar graphs.

Theorem 6.25 (Jaeger, Vertigan and Welsh). For \( x, y \in \mathbb{C} \), evaluating the Tutte polynomial at \((x, y)\) is \#P-hard over graphs unless
\[
(x - 1)(y - 1) = 1
\]
or
\[
(x, y) \in \{(1, 1), (-1, -1), (0, -1), (i, -i), (-i, i), (\omega, \omega^2), (\omega^2, \omega)\}, \text{ where } \omega = e^{2\pi i/3}.
\]
In each exceptional case, the problem is in polynomial time.

Theorem 6.26 (Theorem 5.1 in [?]). For \( x, y \in \mathbb{C} \), evaluating the Tutte polynomial at \((x, y)\) is \#P-hard over planar graphs unless
\[
(x - 1)(y - 1) \in \{1, 2\} \text{ or } (x, y) \in \{(1, 1), (-1, -1), (\omega, \omega^2), (\omega^2, \omega)\},
\]
where \( \omega = e^{2\pi i/3} \). In each exceptional case, the problem is in polynomial time.

The first step in the reduction from the Tutte polynomial is concerned with a sum of weighted Eulerian orientations on the medial graph of a plane graph. Recall that a plane graph is a planar embedding of a planar graph. We define a medial graph next.

Definition 6.27 (cf. [?]). Given a connected plane graph \( G \), its medial graph \( G_m \) has a vertex \( e' \) for each edge \( e \) of \( G \), and vertices \( e'_1 \) and \( e'_2 \) in \( G_m \) are joined by an edge for each face of \( G \) in which their corresponding edges \( e_1 \) and \( e_2 \) in \( G \) occur consecutively.

Note that \( G \) may have loops and parallel edges, and the vertices \( e'_1 \) and \( e'_2 \) of \( G_m \) in Definition 6.27 may not be distinct. An example of a plane graph and its medial graph are given in Figure 6.7. Note that the medial graph of a plane graph is always a planar 4-regular graph. Las Vergnas [?] connected the evaluation of the Tutte polynomial of a planar graph \( G \) at the point \((3, 3)\) with a sum of weighted Eulerian orientations on a medial graph of \( G \).

Theorem 6.28 (Theorem 2.1 in [?]). Let \( G \) be a connected plane graph and let \( \mathcal{O}(G_m) \) be the set of all Eulerian orientations in the medial graph \( G_m \) of \( G \). Then
\[
2 \cdot T(G; 3, 3) = \sum_{O \in \mathcal{O}(G_m)} 2^{\beta(O)},
\]
where \( \beta(O) \) is the number of saddle vertices in the orientation \( O \), i.e. the number of vertices in which the edges are oriented “in, out, in, out” in cyclic order.

207
Figure 6.7: A plane graph (a), its medial graph (c), and the two graphs superimposed (b).

Although the medial graph depends on a particular embedding of the planar graph $G$, the right-hand side of (6.15) is invariant under different embeddings of $G$. This follows from (6.15) and the fact that the Tutte polynomial does not depend on the embedding of $G$.

Given a signature $g$ of arity 4, we abbreviate $g^{abcd}$ for the value $g(a, b, c, d)$, for $(a, b, c, d) \in \{0, 1\}^4$. Recall the definition of the signature matrix $M_g$ of an arity 4 signature $g$ (Definition 1.4).

$$
M_g = \begin{bmatrix}
g^{0000} & g^{0010} & g^{0001} & g^{0011} \\
g^{0100} & g^{0110} & g^{0101} & g^{0111} \\
g^{1000} & g^{1010} & g^{1001} & g^{1011} \\
g^{1100} & g^{1110} & g^{1101} & g^{1111}
\end{bmatrix}. 
$$

(6.16)

In general when we present $g$ pictorially with four external edges $ABCD$, we may order them counterclockwise. In $M_g$, the rows are indexed by the bits $AB$, and the columns are indexed by the bits $DC$ in the reverse order, both lexicographically. This reversal is for the convenience that the signature matrix of a construction that links two arity 4 signatures is the matrix product of the signature matrices of the two arity 4 signatures.

In particular, the signature matrix of a symmetric arity 4 signature has identical two middle rows and two middle rows. If $f = [f_0, f_1, f_2, f_3, f_4]$ is a symmetric signature of arity 4, then its signature matrix is

$$
M_f = \begin{bmatrix}
f_0 & f_1 & f_1 & f_2 \\
f_1 & f_2 & f_2 & f_3 \\
f_1 & f_2 & f_2 & f_3 \\
f_2 & f_3 & f_3 & f_4
\end{bmatrix}.
$$

(6.17)

Now we can prove

**Theorem 6.29.** \#Eulerian-Orientations is \#P-hard for planar 4-regular graphs.

**Proof.** We reduce the problem of calculating the right-hand side of (6.15) to Pl-Holant($\neq_2 |$...
Then by Theorem 6.26 and Theorem 6.28, we conclude that \#EO is \#P-hard for planar 4-regular graphs.

The right-hand side of (6.15) is the bipartite planar Holant problem \( \text{Pl-Holant}(\not= 2 | f) \), where the signature matrix of \( f \) is

\[
M_f = \begin{bmatrix}
0 & 0 & 0 & 1 \\
0 & 1 & 2 & 0 \\
0 & 2 & 1 & 0 \\
1 & 0 & 0 & 0
\end{bmatrix}.
\]

Note that \( f^{0101} = f^{1010} = 2 \) are the two entries respectively listed in the second row indexed by 01 and the third column indexed by 01, and in the third row indexed by 10 and the second column indexed by 10. We perform a holographic transformation by \( Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \) to get

\[
\text{Pl-Holant}(\not= 2 | f) \equiv_T \text{Pl-Holant}([0, 1, 0](Z^{-1})^\otimes 2 | Z^\otimes 4 f)
\equiv_T \text{Pl-Holant}([1, 0, 1] | \hat{f})
\equiv_T \text{Pl-Holant}(\hat{f}),
\]

where the signature matrix of \( \hat{f} \) is

\[
M_{\hat{f}} = \begin{bmatrix}
2 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 2
\end{bmatrix},
\]

and can be computed as \( Z^\otimes 2 M_f (Z^T)^\otimes 2 \). An alternative way to compute this transformed signature is to write \( Z^\otimes 4 f \) by definition

\[
Z^\otimes 4 \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
+ 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
\right\} = \frac{1}{4} \left\{ \begin{bmatrix} i \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} -1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} + \begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} -1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} + 2 \begin{bmatrix} 1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} \\
+ 2 \begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} + \begin{bmatrix} 1 \\ i \end{bmatrix} \otimes \begin{bmatrix} -1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} \right\}
\]

Then it can be easily calculated that, \( \hat{f}^{0000} = \hat{f}^{1111} = 2, \hat{f}^{0011} = \hat{f}^{0101} = \hat{f}^{1010} = \hat{f}^{1100} = 1, \) and all other entries are 0.

We can perform the same holographic transformation by \( Z \) on our target counting prob-
lem \( \text{Pl-Holant}(\neq_2 \mid [0, 0, 1, 0, 0]) \) and get

\[
\text{Pl-Holant}(\neq_2 \mid [0, 0, 1, 0, 0]) \equiv_T \text{Pl-Holant}([0, 1, 0] (Z^{-1}) \otimes^2 \mid Z \otimes^4 [0, 0, 1, 0, 0]) \\
\equiv_T \text{Pl-Holant}([1, 0, 1] \mid \frac{1}{2} [3, 0, 1, 0, 3]) \\
\equiv_T \text{Pl-Holant}([3, 0, 1, 0, 3]).
\]

We can verify \( Z \otimes^4 [0, 0, 1, 0, 0] = \frac{1}{2} [3, 0, 1, 0, 3] \) as follows: First note that \( Z \otimes^4 [0, 0, 1, 0, 0] \) is symmetric, and we may denote it as \([h_0, h_1, h_2, h_3, h_4]\). Writing \( Z \otimes^4 [0, 0, 1, 0, 0] \) by definition, it is \( \frac{1}{4} \) of the sum over all six permutations of the tensor product \([\mathbf{1}^i] \otimes [\mathbf{1}^i] \otimes [-\mathbf{i}] \otimes [-\mathbf{i}]\).

Hence it is clear that \( h_0 = h_4 = \frac{6}{4} = \frac{3}{2} \) and \( h_2 = 1 \). The entries of odd weight \( h_1 = h_3 = 0 \) because there is an involution \( \pm \mathbf{i} \mapsto \mp \mathbf{i} \).

Our goal is to reduce \( \text{Pl-Holant}(\hat{f}) \) to \( \text{Pl-Holant}([3, 0, 1, 0, 3]) \).

Consider the planar tetrahedron gadget in Figure 6.8. We assign \([3, 0, 1, 0, 3]\) to every vertex and obtain a signature \( 32 \hat{g} \), where the signature matrix of \( \hat{g} \) is

\[
M_{\hat{g}} = \frac{1}{2} \begin{bmatrix}
19 & 0 & 0 & 7 \\
0 & 7 & 5 & 0 \\
0 & 5 & 7 & 0 \\
7 & 0 & 0 & 19 \\
\end{bmatrix}.
\]

This signature matrix requires some computation to obtain. One can observe that all entries with odd Hamming weight from \([0, 1]\)^4 are zero, because \([3, 0, 1, 0, 3]\) has this property. Then, for example, the entry \( \hat{g}^{0000} \) can be computed as follows. Set the external pattern 0000. Then the four outside vertices virtually have the signature \([3, 0, 1, 0]\). The central vertex \( v_0 \) takes nonzero values at eight bit patterns: 0000, 1111, and the six patterns of Hamming weight two. For 0000 at \( v_0 \), the contribution to \( \hat{g}^{0000} \) is 3 times the trace of \([\mathbf{3}^0 \mathbf{0}^1]\), which is 246. Similarly for 1111 at \( v_0 \), the contribution to \( \hat{g}^{0000} \) is 3 times the trace of \([\mathbf{0}^1 \mathbf{1}^0]\), which is 6. For the bit patterns 0101 and 1010 of Hamming weight two at \( v_0 \), the contributions to \( \hat{g}^{0000} \) is twice the trace of \([\mathbf{0}^1 \mathbf{1}^0]\), which is 12. Finally for 0011 at \( v_0 \), and its four cyclic permutations, the contributions to \( \hat{g}^{0000} \) is 4 times the trace of \([\mathbf{0}^1 \mathbf{1}^0]\), which is 40. This gives the total value \( \hat{g}^{0000} = 304 \). Moreover, the signature \( \hat{g} \) is invariant when we flip 0’s with 1’s, because \([3, 0, 1, 0, 3]\) has this property. Thus \( \hat{g}^{1111} = 304 \) as well. Also \( \hat{g} \) is invariant under cyclic permutations due to the symmetry of the tetrahedron gadget. All nonzero values of \( \hat{g} \) can be computed similarly.

Now we show how to reduce \( \text{Pl-Holant}(\hat{f}) \) to \( \text{Pl-Holant}(\hat{g}) \) by interpolation. Consider an instance \( \Omega \) of \( \text{Pl-Holant}(\hat{f}) \). Suppose that \( \hat{f} \) appears \( n \) times in \( \Omega \). We construct from \( \Omega \) a sequence of instances \( \Omega_s \) of \( \text{Holant}(\hat{g}) \) indexed by \( s \geq 1 \). We obtain \( \Omega_s \) from \( \Omega \) by replacing each occurrence of \( \hat{f} \) with the gadget \( N_s \) in Figure 6.9 with \( \hat{g} \) assigned to all vertices. Although \( \hat{f} \) and \( \hat{g} \) are asymmetric signatures, they are invariant under a cyclic permutation of their inputs. Thus, it is unnecessary to specify which edge corresponds to the first input of \( \hat{f} \) and \( \hat{g} \). We call such signatures rotationally symmetric.
To obtain $\Omega_s$ from $\Omega$, we effectively replace $M_f$ with $M_{N_s} = (M_{\hat{g}})^s$, the $s$th power of the signature matrix $M_{\hat{g}}$. Let

$$T = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \Lambda_f = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \Lambda_{\hat{g}} = \begin{bmatrix} 13 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}$$

Then

$$M_f = T\Lambda_f T^{-1} \quad \text{and} \quad M_{\hat{g}} = T\Lambda_{\hat{g}} T^{-1}$$

We can view our construction of $\Omega_s$ as first replacing each $M_f$ by $T\Lambda_f T^{-1}$ to obtain a signature grid $\Omega'$, which does not change the Holant value, and then replacing each $\Lambda_f$ with $\Lambda_{\hat{g}}^s$. We stratify the assignments in $\Omega'$ based on the assignment to $\Lambda_f$. Recall that the rows of $\Lambda_f$ and $\Lambda_{\hat{g}}$ are indexed by 00, 01, 10, 11 and the columns are indexed by 00, 10, 01, 11, in their respective orders. We only need to consider the assignments to $\Lambda_f$ that assign

- $(00, 00)$ $j$ many times,
- $(01, 10)$ or $(11, 11)$ $k$ many times, and
- $(10, 01)$ $\ell$ many times,

where $j + k + \ell = n$, the total number of occurrences of $\Lambda_f$ in $\Omega'$. Let $c_{j\ell}$ be the sum over all such assignments of the products of evaluations from $T$ and $T^{-1}$ but excluding $\Lambda_f$ on $\Omega'$. Then

$$\text{Pl-Holant}_\Omega = \sum_{j+k+\ell=n} 3^j c_{j\ell}$$

and the value of the Holant on $\Omega_s$, for $s \geq 1$, is

$$\text{Pl-Holant}_{\Omega_s} = \sum_{j+k+\ell=n} (13^j 6^k)^s c_{j\ell}. \quad (6.19)$$
This is a linear equation system with unknowns $c_{jk\ell}$, and a coefficient matrix whose rows are indexed by $s$ and columns are indexed by $(j, k)$, where $0 \leq j, k$ and $j + k \leq n$. We take $1 \leq s \leq \binom{n+2}{2}$. This coefficient matrix in the linear system of (6.19) is Vandermonde and has full rank since for any $j, k, j', k' \geq 0$, if $(j, k) \neq (j', k')$ then $13^j 6^k \neq 13^{j'} 6^{k'}$. Therefore, after obtaining the values of Pl-Holant$_\Omega$ by oracle calls to $\#EO$, for $1 \leq s \leq \binom{n+2}{2}$, we can solve the linear system for the unknown $c_{jk\ell}$'s and obtain the value of Pl-Holant$_\Omega$. \qed
Figure 6.9: Recursive construction to interpolate $\hat{f}$. The vertices are assigned $\hat{g}$. 
Chapter 7

Holant Dichotomy for Symmetric Constraints

Most results in previous chapters can be viewed as preparatory for the dichotomy theorems to be achieved in this and the next two chapters. In this chapter we prove a dichotomy theorem valid for all Holant Problems on an arbitrary set of complex-valued symmetric constraint functions on Boolean variables, Theorem 7.19. This includes as special cases the dichotomy theorems for Holant* (when all unary functions are free), Holantc (when the two unary functions Is-Zero [1, 0] and Is-One [0, 1] are free), and #CSP (when all EQUALITIES are free) restricted to symmetric constraints.

For Holant problems without freely available auxiliary functions such as unary signatures, the construction of gadgets becomes more constricted. This difficulty has shown itself in Chapter 5. In particular, the only natural local operations available are to connect two edges, or to make a loop. If we want to prove hardness with some inductive argument based on the arity of constraint functions, there will be two base cases of arity 3 and 4. Note that signatures of arity at most 2 define tractable Holant problems. The arity 3 case is understood from Theorem 6.3 (and also Theorem 6.4), with the main work done in Chapter 5. Therefore it will be natural to study carefully arity 4 signatures.

However before we can develop the theory satisfactorily for arity 4 signatures we must first deal with a phenomenon called vanishing.

7.1 Vanishing Signatures

Definition 7.1. A set of signatures $\mathcal{F}$ is called vanishing if the value $\text{Holant}_\Omega(\mathcal{F})$ is zero for every signature grid $\Omega$. A signature $f$ is called vanishing if the singleton set $\{f\}$ is vanishing.

In this section, we characterize all sets of symmetric vanishing signatures. The following lemma makes the simple observation that vanishing signatures are closed under gadget constructions, and under linear combinations. Let $af + bg$ denote the entry-wise linear combination of two signatures $f$ and $g$ with the same arity $n$ ($a, b \in \mathbb{C}$), i.e. $(af + bg)(x) =$
Lemma 7.2. Let $\mathcal{F}$ be a vanishing signature set. If a signature $f$ can be realized by a gadget using signatures in $\mathcal{F}$, then $\mathcal{F} \cup \{f\}$ is also vanishing. If $f$ and $g$ are two signatures in $\mathcal{F}$ of the same arity, and $a, b \in \mathbb{C}$, then $\mathcal{F} \cup \{af + bg\}$ is vanishing as well.

Proof. That $\mathcal{F} \cup \{f\}$ is vanishing is obvious, since any vertex in a signature grid assigned the function $f$ can be replaced by the $\mathcal{F}$-gate defining it. If $f, g \in \mathcal{F}$, then the Holant value on any signature grid using $\mathcal{F} \cup \{af + bg\}$ can be expressed as a linear combination of Holant values using $\mathcal{F}$, one term for a choice of $f$ or $g$ at every vertex labeled $af + bg$. \qed

Obviously, the identically zero signature, in which all entries are 0, is vanishing. This is trivial. However, there are less trivial examples of vanishing signatures. Notice that the unary signature $[1, i]$ when connected to another $[1, i]$ has a Holant value 0. Consider a signature set $\mathcal{F}$ where every signature is degenerate, and moreover, as a tensor product of unary signatures, more than half of them are $[1, i]$. For any signature grid $\Omega$ with signatures from $\mathcal{F}$, each signature is decomposed into unary signatures, and $\Omega$ is decomposed into pairs of vertices, one for each edge. The total Holant value is the product of the Holant on each pair. Since more than half of the unaries in each signature are $[1, i]$, more than half of the unaries in $\Omega$ are $[1, i]$. Then two $[1, i]$'s must be paired up and hence $\text{Holant}_\Omega = 0$. Thus, all such signatures form a vanishing set. We also observe that this argument holds when $[1, i]$ is replaced by $[1, -i]$.

These signatures described above are generally not symmetric and our present aim is to characterize symmetric vanishing signatures. To this end, we define the following symmetrization operation.

Definition 7.3. Let $S_n$ be the symmetric group of degree $n$. Then for positive integers $t$ and $n$ with $t \leq n$ and unary signatures $v, v_1, \ldots, v_{n-t}$, we define

$$
\text{Sym}_n^t(v; v_1, \ldots, v_{n-t}) = \sum_{\pi \in S_n} \prod_{k=1}^{n} u_{\pi(k)},
$$

where the ordered sequence $(u_1, u_2, \ldots, u_n) = (v, \underbrace{v_1, \ldots, v_1}_t, \ldots, v_{n-t})$.

Note that we include redundant permutations of $v$ in the definition. Identical $v_i$’s also introduce redundant permutations. These redundant permutations simply introduce a nonzero constant factor. However, the allowance of redundant permutations may simplify calculations. An illustrative example of Definition 7.3 is

$$
\text{Sym}_3^2([1, i]; [a, b]) = 2[a, b] \otimes [1, i] \otimes [1, i] + 2[1, i] \otimes [a, b] \otimes [1, i] + 2[1, i] \otimes [1, i] \otimes [a, b] \\
= 2[3a, 2ia + b, -a + 2ib, -3b].
$$

216
Definition 7.4. A nonzero symmetric signature $f$ of arity $n$ has positive vanishing degree $k \geq 1$, denoted by $\text{vd}^+(f) = k$, if $k \leq n$ is the largest positive integer such that there exists $n-k$ unary signatures $v_1, \ldots, v_{n-k}$ satisfying

$$f = \text{Sym}_n^k([1, i]; v_1, \ldots, v_{n-k}).$$

If $f$ cannot be expressed as such a symmetrization form, we define $\text{vd}^+(f) = 0$. If $f$ is the all zero signature, define $\text{vd}^+(f) = n+1$.

We define negative vanishing degree $\text{vd}^-$ similarly, using $-i$ instead of $i$.

In Chapter 2, we showed in (2.10) to (2.12) that any generalized Fibonacci gate $f$ with parameter $\lambda = 2i$, i.e., $f_k = cki^{k-1} + di^k$ (for some $c, d \in \mathbb{C}$, and for all $0 \leq k \leq n$), has the form $f = Z^\otimes n[a, b, 0, \ldots, 0]$, for some $a, b \in \mathbb{C}$, where $Z = [\frac{1}{i} \frac{1}{-i}]$. A similar expression is true if the parameter $\lambda = 2i$; in that case $f = Z^\otimes n[0, 1, 0, \ldots, 0]$. The expression in (2.12) shows that $\text{vd}^+(Z^\otimes n[1, 0, 0, \ldots, 0]) = n$ and $\text{vd}^+(Z^\otimes n[0, 1, 0, \ldots, 0]) = n-1$. For the linear combination $f = Z^\otimes n[a, b, 0, \ldots, 0] = aZ^\otimes n[1, 0, 0, \ldots, 0] + bZ^\otimes n[0, 1, 0, \ldots, 0], \text{Lemma 7.17 in Subsection 7.1.3 will show that } \text{vd}^+(f) = n-1 \text{ if } b \neq 0. \text{ Similarly for } f = Z^\otimes n[0, \ldots, 0, b, a], \text{vd}^-(f) = n-1 \text{ if } b \neq 0.$

Notice that it is possible for a signature $f$ to have both $\text{vd}^+(f)$ and $\text{vd}^-(f)$ nonzero. For example, $f = [1, 0, 1]$ has $\text{vd}^+(f) = \text{vd}^-(f) = 1$.

By the discussion above and Lemma 7.2, we know that for a signature $f$ of arity $n$, if $\text{vd}^\sigma(f) > \frac{n}{2}$ for some $\sigma \in \{+, -\}$, then $f$ is a vanishing signature. Thus all generalized Fibonacci gates of arity $n \geq 3$ with parameter $\lambda = \pm 2i$ are vanishing signatures.

This argument is easily generalized to a set of signatures.

Definition 7.5. For $\sigma \in \{+, -\}$, we define $\mathcal{V}^\sigma = \{f \mid 2 \text{vd}^\sigma(f) > \text{arity}(f)\}$.

Lemma 7.6. For a set of symmetric signatures $\mathcal{F}$, if $\mathcal{F} \subseteq \mathcal{V}^+$ or $\mathcal{F} \subseteq \mathcal{V}^-$, then $\mathcal{F}$ is vanishing.

In Theorem 7.13, we show that these two sets capture all symmetric vanishing signature sets.

7.1.1 Characterization by Recurrence Relations

Now we give another characterization of vanishing signatures.

Definition 7.7. An arity $n$ symmetric signature of the form $f = [f_0, f_1, \ldots, f_n]$ is in $\mathcal{R}_t^+$ for a nonnegative integer $t \geq 0$ if

1. $t > n$; or
2. for any $0 \leq k \leq n-t$, $f_k, \ldots, f_{k+t}$ satisfy the recurrence relation of order $t$

\[
\left(\frac{t}{i}\right) f_{k+t} + \left(\frac{t}{t-1}\right) f_{k+t-1} + \cdots + \left(\frac{t}{0}\right) f_0 = 0. \tag{7.2}
\]

We define $\mathcal{R}_t^-$ similarly but with $-i$ in place of $i$ in (7.2).

It is easy to see that $\mathcal{R}_0^+ = \mathcal{R}_0^-$ is the set of all zero signatures. Also, for $\sigma \in \{+, -\}$, we have $\mathcal{R}_t^\sigma \subseteq \mathcal{R}_t^\sigma$ when $t \leq t'$. By definition, if arity$(f) = n$ then $f \in \mathcal{R}_{n+1}^\sigma$.

Let $f = [f_0, f_1, \ldots, f_n] \in \mathcal{R}_t^+$ with $0 < t \leq n$. Then the characteristic polynomial of its recurrence relation is $(1 + xi)^t$. Thus there exists a polynomial $p(x)$ of degree at most $t - 1$ such that $f_k = i^k p(k)$, for $0 \leq k \leq n$. This statement extends to $\mathcal{R}_{n+1}^+$ since a polynomial of degree $n$ can interpolate any set of $n + 1$ values. Furthermore, such an expression is unique. If there are two polynomials $p(x)$ and $q(x)$, both of degree at most $n$, such that $f_k = i^k p(k) = i^k q(k)$ for $0 \leq k \leq n$, then $p(x)$ and $q(x)$ must be the same polynomial. Now suppose $f_k = i^k p(k)$ ($0 \leq k \leq n$) for some polynomial $p$ of degree at most $t - 1$, where $0 < t \leq n$. Then $f$ satisfies the recurrence (7.2) of order $t$. Hence $f \in \mathcal{R}_t^+$. Thus $f \in \mathcal{R}_t^+$ iff there exists a polynomial $p(x)$ of degree at most $t-1$ such that $f_k = i^k p(k)$ ($0 \leq k \leq n$), for all $1 \leq t \leq n + 1$. For $\mathcal{R}_t^-$, just replace $i$ by $-i$.

**Definition 7.8.** For a nonzero symmetric signature $f$ of arity $n$, it is of positive recurrence degree $0 \leq t \leq n$, denoted by $\text{rd}^+(f) = t$, if and only if $f \in \mathcal{R}_{t+1}^+ - \mathcal{R}_t^+$. We define negative recurrence degree similarly, namely $\text{rd}^-(f) = t$ if and only if $f \in \mathcal{R}_{t+1}^- - \mathcal{R}_t^-$. If $f$ is the all zero signature, we define $\text{rd}^+(f) = \text{rd}^-(f) = -1$.

Note that although we call it the recurrence degree, it refers to a particular recurrence relation. For any nonzero symmetric signature $f$, by the uniqueness of the representing polynomial $p(x)$, it follows that $\text{rd}^\sigma(f) = t$ iff $\deg(p) = t$, where $0 \leq t \leq n$. We remark that $\text{rd}^\sigma(f)$ is the maximum integer $t$ such that $f$ does not belong to $\mathcal{R}_t^\sigma$. Also, for an arity $n$ signature $f$, $\text{rd}^\sigma(f) = n$ if and only if $f$ does not satisfy any recurrence relation (7.2) of order $t \leq n$ for $\sigma \in \{+, -\}$.

**Remark:** Since membership in $\mathcal{R}_t^\pm$ is easily decidable by (7.2), for any symmetric signature $f$, its recurrence degree $\text{rd}^\pm(f)$ can be computed in polynomial time in the size of its symmetric signature. This is one advantage of the concept of recurrence degree $\text{rd}^\pm$ compared to vanishing degree $\text{vd}^\pm$, while conceptually the latter relates more directly to the notion of vanishing. Lemma 7.9 shows that the two concepts are equivalent.

**Lemma 7.9.** Let $f = [f_0, \ldots, f_n]$ be a symmetric signature of arity $n$, not identically 0. Then for any nonnegative integer $0 \leq t < n$ and $\sigma \in \{+, -\}$, the following are equivalent:

1. There exist $t$ unary signatures $v_1, \ldots, v_t$, such that

\[
f = \text{Sym}^{n-t}_n([1, \sigma i]; v_1, \ldots, v_t). \tag{7.3}
\]
(2) \( f \in \mathcal{R}_{t+1}^\sigma \).

Proof. We consider the case \( \sigma = + \) since the case \( \sigma = - \) is similar, and so let \( v = [1, i] \).

We start with (1) \( \implies \) (2) and proceed via induction on both \( t \) and \( n \). For the first base case of \( t = 0 \), \( f = \text{Sym}_n^1(v) = [1, i]^{\otimes n} = [1, i, -1, -i, \ldots, i^n] \), so \( f_{k+1} = if_k \) for all \( 0 \leq k \leq n-1 \) and \( f \in \mathcal{R}_{1}^+ \).

The other base case is that \( t = n - 1 \). Let \( f = \text{Sym}_n^1(v; v_1, \ldots, v_t) = [f_0, \ldots, f_n] \) where \( v_i = [a_i, b_i] \) for \( 1 \leq i \leq t \), and we need to show that \( f \in \mathcal{R}_{n}^+ \). Let

\[
S = \binom{n}{n} i^nf_n + \cdots + \binom{n}{1} i^kf_1 + \binom{n}{0} i^0f_0,
\]

then we need to show that \( S = 0 \). First notice that any entry in \( f \) is a linear combination of terms of the form \( a_{i_1} \cdots a_{i_{n-1-k}} b_{j_1} \cdots b_{j_k} \), where \( 0 \leq k \leq n-1 \), and \( \{i_1, \ldots, i_{n-1-k}, j_1, \ldots, j_k\} = \{1, 2, \ldots, n-1\} \). Thus \( S \) is a linear combination of such terms as well. Now we compute the coefficient of each of these terms in \( S \).

The term \( b_1 \cdots b_k a_{k+1} \cdots a_{n-1} \) appears twice in \( S \), once in \( f_k \) and the other time in \( f_{k+1} \). Consider \( f_k \), the signature value on Hamming weight \( k \). As the signature is symmetric, we may consider the input \( (1, \ldots, 1, 0, \ldots, 0) \). To obtain the product \( b_1 \cdots b_k a_{k+1} \cdots a_{n-1} \) it must be from the terms in the sum (7.1) of Definition 7.3 defining \( \text{Sym}_{n-t}^1([1, \sigma i]; v_1, \ldots, v_t) \) in (7.3) where \( v_1 = [a_1, b_1]^T, \ldots, v_k = [a_k, b_k]^T \) are in the first \( k \) positions, and \( v = [1, i]^T, v_{k+1} = [a_{k+1}, b_{k+1}]^T, \ldots, v_{n-1} = [a_{n-1}, b_{n-1}]^T \) are in the last \( n-k \) positions, both in any permuted way. Thus in \( f_k \), the coefficient is \( k!(n-k)! \). By a similar reason, in \( f_{k+1} \), the coefficient of \( b_1 \cdots b_k a_{k+1} \cdots a_{n-1} \) is \( i(k+1)!(n-k-1)! \). Thus, its coefficient in \( S \) is

\[
\binom{n}{n} i^{k+1}i(k+1)!(n-k-1)! + \binom{n}{k} i^k k!(n-k)! = 0.
\]

The above computation works for any term \( a_{i_1} \cdots a_{i_{n-1-k}} b_{j_1} \cdots b_{j_k} \) due to the symmetry of \( f \), so all coefficients in \( S \) are 0, which means that \( S = 0 \).

Now we carry out a nested induction, first on \( t \geq 0 \) and then on \( n - 1 - t \geq 0 \). Having proved the two base cases, we assume \( t > 0 \), \( n - 1 - t > 0 \), and we inductively assume for any \( 0 \leq t' < t \), or for the same \( t \) and any \( n' \) with \( n' \geq t+1 \) and \( n' < n \), the statement holds. Let

\[
f = [f_0, \ldots, f_n] = \text{Sym}_n^{n-t}(v; v_1, \ldots, v_t),
g = [g_0, \ldots, g_{n-1}] = \text{Sym}_{n-1}^{n-1-t}(v; v_1, \ldots, v_t),
\]

and for any \( j \) (\( 1 \leq j \leq t \)),

\[
h^{(j)} = [h_0^{(j)}, \ldots, h_{n-1}^{(j)}] = \text{Sym}_{n-1}^{n-t}(v; v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_t).
\]
By the induction hypothesis, \( g \) satisfies the recurrence relation of order \( t+1 \), namely \( g \in R_{t+1}^+ \). Also for any \( j \), \( h^{(j)} \) satisfies the recurrence relation of order \( t \), namely \( h^{(j)} \in R_t^+ \subseteq R_{t+1}^+ \).

We have the recurrence relation

\[
\text{Sym}_n^{n-t}(v; v_1, \ldots, v_t) = (n-t)v \otimes \text{Sym}_{n-1}^{n-1-t}(v; v_1, \ldots, v_t) + \sum_{j=1}^{t} v_j \otimes \text{Sym}_{n-1}^{n-t}(v; v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_t)
\]

\[= (n-t)v \otimes g + \sum_{j=1}^{t} v_j \otimes h^{(j)}. \tag{7.4}\]

By equation (7.4), the entry of weight \( k \) in \( f \) for any \( k > 0 \) is

\[f_k = (n-t)g_{k-1} + \sum_{j=1}^{t} b_j h^{(j)}_{k-1}.\]

We can obtain this expression by considering \( f \) on input \((1, \ldots, 1, 0, \ldots, 0)\), and this is sufficient since \( f \) is symmetric. We know that \( \{g_i\} \) and \( \{h^{(j)}_i\} \) satisfy the recurrence relation (7.2) of order \( t+1 \). Thus, their linear combination \( \{f_i\} \) also satisfies the same recurrence relation (7.2) of order \( t+1 \) starting from any positive index \( k > 0 \).

Similarly, from equation (7.4), the entry of weight \( k \) in \( f \) for any \( k < n \) is

\[f_k = (n-t)g_k + \sum_{j=1}^{t} a_j h^{(j)}_k.\]

We can obtain this expression by considering \( f \) on input \((0, \ldots, 0, 1, \ldots, 1)\). Note that \( g \) and all \( h^{(j)} \) have arity \( n-1 \), since \( t < n-1 \), by the same argument again, the recurrence relation (7.2) of order \( t+1 \leq n-1 \) holds for \( f \) starting from index \( k = 0 \) as well.

Now we prove \( (2) \implies (1) \). Notice that we only need to find unary signatures \( \{v_i\} \) for \( 1 \leq i \leq t \) such that \( \text{Sym}_n^{n-t}(v; v_1, \ldots, v_t) \) matches the first \( t+1 \) entries of \( f \). The theorem follows from this since we have shown that \( \text{Sym}_n^{n-t}(v; v_1, \ldots, v_t) \) satisfies the recurrence relation of order \( t+1 \) and any such signature is determined by the first \( t+1 \) entries.

We show that there exist \( v_i = [a_i, b_i] \) (\( 1 \leq i \leq t \)) satisfying the above requirement. Since \( f \) is not identically 0, by equation (7.2), some nonzero term occurs among \( \{f_0, \ldots, f_t\} \). Let \( f_s \neq 0 \), for \( 0 \leq s \leq t \), be the first nonzero term. By a nonzero constant multiplier, we may normalize \( f_s = s!(n-s)! \), and set \( v_j = [0, 1] \), for \( 1 \leq j \leq s \) (which is vacuous if \( s = 0 \)), and set \( v_{s+j} = [1, b_{s+j}] \), with \( b_{s+j} \) to be determined later, for \( 1 \leq j \leq t-s \) (which is vacuous if \( s = t \)). Let \( F \) be the function defined in equation (7.3). Then \( F_k = f_k = 0 \) for \( 0 \leq k < s \)
(which is vacuous if \( s = 0 \)). By expanding the symmetrization function, for \( s \leq k \leq t \), we get

\[
F_k = k! (n-k)! \sum_{j=0}^{k-s} \binom{n-t}{k-s-j} \sigma_j(b_{s+1}, \ldots, b_t) t^k s^{-j},
\]

where \( \sigma_j \) is the elementary symmetric polynomial of degree \( j \) for \( 0 \leq j \leq t-s \). This can be seen as follows: Consider the input \((1, \ldots, 1, 0, \ldots, 0)\). The \( s \) copies of \([0, 1]^\top, v_1, \ldots, v_s\), clearly must be among the first \( k \) tensor factors of a nonzero term in the sum defining (7.3). This contributes a factor \( \binom{k}{s} s! = k!/(k-s)! \). Then we effectively consider the input pattern \((1, \ldots, 1, 0, \ldots, 0)\). By symmetry \( F_k \) is clearly a linear combination of \( \sigma_j \), for \( 0 \leq j \leq k-s \).

We may consider the coefficient of \( b_{s+1} \cdots b_{t+j} \). The \( t-s \) tensor factors \( v_{s+i} \) \((1 \leq i \leq t-s)\) must be distributed in such a way that the first \( j \) are among the first \( k-s \) bit positions corresponding to the 1’s and the other \( t-s-j \) are among the last \( n-k \) bit positions corresponding to the 0’s. Finally \( n-t \) copies of \( v = [1, \ldots, 1] \) are distributed among the remaining positions, which has exactly \( k-s \) many bit positions corresponding to the 1’s. Thus the coefficient is

\[
\binom{k}{s} s! \binom{k-s}{j} j! \binom{n-k}{t-s-j} (t-s-j)! (n-t)! t^{k-s-j} = k! (n-k)! \binom{n-t}{k-s-j} t^{k-s-j}.
\]

By definition, \( \sigma_0 = 1 \) and \( F_s = f_s \). If \( t = s \) we are done. Suppose \( t > s \). Setting \( F_k = f_k \) for \( s+1 \leq k \leq t \), this is a linear equation system on \( \sigma_j \) \((1 \leq j \leq t-s)\), with a triangular matrix and nonzero diagonals. From this, we know that all \( \sigma_j \)'s are uniquely determined by \( \{f_{s+1}, \ldots, f_t\} \). Moreover, \( \{b_{s+1}, \ldots, b_t\} \) are the roots of the equation \( \sum_{j=0}^{t-s} (-1)^j \sigma_j x^{t-j} = 0 \). Thus \( \{b_{s+1}, \ldots, b_t\} \) are also uniquely determined by \( \{f_{s+1}, \ldots, f_t\} \) up to a permutation. \( \square \)

**Corollary 7.10.** If \( f \) is a symmetric signature and \( \sigma \in \{+, -\} \), then

\[
vd^\sigma(f) + rd^\sigma(f) = \text{arity}(f).
\]

Thus we have an equivalent form of \( \mathcal{V}^\sigma \) for \( \sigma \in \{+, -\} \). Namely,

\[
\mathcal{V}^\sigma = \{f \mid 2 \text{rd}^\sigma(f) < \text{arity}(f)\}.
\]

**Remark:** By Corollary 7.10 membership in \( \mathcal{V}^\pm \) is decidable in polynomial time in the size of a symmetric signature.

### 7.1.2 Characterization of Vanishing Signature Sets

Now we show that \( \mathcal{V}^+ \) and \( \mathcal{V}^- \) capture all symmetric vanishing signature sets. To begin, we show that a vanishing signature set cannot contain both types of nonzero vanishing signatures.

221
Lemma 7.11. Let $f_+ \in \mathcal{V}^+$ and $f_- \in \mathcal{V}^-$. If neither $f_+$ nor $f_-$ is the zero signature, then the signature set $\{f_+, f_-\}$ is not vanishing.

Proof. Let $\text{arity}(f_+) = n$ and $\text{rd}^+(f_+) = t$, so $2t < n$. Consider the gadget with two vertices and $2t$ edges between two copies of $f_+$. (See Figure 7.1 for an example of this gadget.) Consider $f_+$ in the symmetrized form as in (7.3). Since $\text{vd}^+(f_+) = n - t$, in each term, there are $n - t$ many $[1, i]$’s and $t$ many unary signatures not equal to (a multiple of) $[1, i]$. The signature of the gadget is a superposition of degenerate signatures, but the only nonzero contributions come from the cases where the $n - 2t$ dangling edges on both sides are all assigned $[1, i]$, while inside, the $t$ copies of $[1, i]$ are paired up perfectly with all $t$ many unary signatures not equal to a multiple of $[1, i]$ from the other side. Notice that for any such contribution, the Holant value of the inside part is always the same constant and this constant is not zero because $[1, i]$ paired up with any unary signature other than (a multiple of) $[1, i]$ is not zero. Then the superposition of all of the permutations is a degenerate signature $[1, i]^{\otimes 2(n-2t)}$ up to a nonzero constant factor.

Similarly, we can do this for $f_-$ of arity $n'$ and $\text{rd}^-(f_-) = t'$, where $2t' < n'$, and get a degenerate signature $[1, -i]^{\otimes 2(n' - 2t')}$, up to a nonzero constant factor. Then form a bipartite signature grid with $n' - 2t'$ vertices on one side, each assigned $[1, i]^{\otimes 2(n-2t)}$, and $n - 2t$ vertices on the other side, each assigned $[1, -i]^{\otimes 2(n' - 2t')}$.

Connect edges between the two sides arbitrarily as long as it is a 1-1 correspondence. The resulting Holant is a power of 2, which is not vanishing.

Lemma 7.12. Every symmetric vanishing signature is in $\mathcal{V}^+ \cup \mathcal{V}^-$.

Proof. Let $f$ be a symmetric vanishing signature. We prove this by induction on $n$, the arity of $f$. For $n = 1$, by connecting $f = [f_0, f_1]$ to itself, we have $f_0^2 + f_1^2 = 0$. Then up to a constant factor, we have either $f = [1, i]$ or $f = [1, -i]$. The lemma holds.

For $n = 2$, first we do a self loop. The Holant value is $f_0 + f_2$. Also, we can connect two copies of $f$, in which case the Holant value is $f_0^2 + 2f_1^2 + f_2^2$. Since $f$ is vanishing, both $f_0 + f_2 = 0$ and $f_0^2 + 2f_1^2 + f_2^2 = 0$. Solving them, we get $f = [1, i, -1] = [1, i]^{\otimes 2}$ or $[1, -i, -1] = [1, -i]^{\otimes 2}$ up to a constant factor.
Now assume \( n > 2 \) and the lemma holds for any signature of arity \( k < n \). Let \( f = [f_0, f_1, \ldots, f_n] \) be a vanishing signature. A self loop on \( f \) gives \( f' = [f'_0, f'_1, \ldots, f'_{n-2}] \), where \( f'_j = f_j + f_{j+2} \) for \( 0 \leq j \leq n - 2 \). Since \( f \) is vanishing, \( f' \) is vanishing as well. By the induction hypothesis, \( f' \in \mathcal{V}^+ \cup \mathcal{V}^- \).

If \( f' \) is a zero signature, then we have \( f_j + f_{j+2} = 0 \) for \( 0 \leq j \leq n - 2 \). This means that the \( f_j \)'s satisfy a recurrence relation with characteristic polynomial \( x^2 + 1 \), so we have \( f_j = at^j + b(-1)^j \) for some \( a \) and \( b \). Then we perform a holographic transformation with \( Z = [1 \ 1] \).

\[
\text{Holant}(-2 \mid f) \equiv T \text{Holant}([1, 0, 1]Z^2 \mid (Z^{-1})^n f) \equiv T \text{Holant}(2[0, 1, 0] \mid \hat{f}),
\]

with the same Holant value, where \( \hat{f} = (Z^{-1})^n f = [a, 0, \ldots, 0, b] \). The Holant problem Holant([0, 1, 0] \mid \hat{f}) is a weighted version of testing if a graph is bipartite. Now consider a graph with only two vertices, both assigned \( f \), and \( n \) edges between them. The value of Holant([0, 1, 0] \mid \hat{f}) of this graph is \( 2ab \). However, we know that it must be vanishing, so \( ab = 0 \). If \( a = 0 \), then \( f \in \mathcal{V}^- \). Otherwise, \( b = 0 \) and \( f \in \mathcal{V}^+ \).

Now suppose that \( f' \) is in \( \mathcal{V}^+ \cup \mathcal{V}^- \) but is not identically zero. We consider \( f' \in \mathcal{V}^+ \) since the other case is similar. Then \( \text{rd}^+(f') = t \), for some \( t \geq 0 \) and \( 2t < n - 2 \). Consider the gadget which has only two vertices, both assigned \( f' \), and has \( 2t \) edges between them. (See Figure 7.1 for an example of this gadget.) It forms a signature of arity \( d = 2(n - 2 - 2t) \). This gadget construction is valid because \( n - 2 > 2t \). By viewing it as a superposition of degenerate signatures as in the proof of Lemma 7.11, this signature is a nonzero multiple of \( [1, i]^{\otimes d} \).

Moreover, \( \text{rd}^+(f') = t \) implies that the entries of \( f' \) satisfy a recurrence of order \( t + 1 \). Replacing \( f'_j \) by \( f_j + f_{j+2} \), we get a recurrence relation for the entries of \( f \) with characteristic polynomial \( (x^2 + 1)(x - i)^{t+1} = (x + i)(x - i)^{t+2} \). Thus, \( f_j = ivp(j) + c(-i)^j \) for some polynomial \( p(x) \) of degree at most \( t + 1 \) and some constant \( c \). It suffices to show that \( c = 0 \). Note that \( t + 1 < 2(t + 1) < n \) by \( t \geq 0 \) and \( 2t < n - 2 \), then the expression \( f_j = ivp(j) \) shows that \( f \in \mathcal{R}_d \) and thus \( \text{rd}^+(f) \leq t + 1 \).

Consider the signature \( h = [h_0, \ldots, h_{n-1}] \) created by connecting \( f \) with a single unary signature \([1, i]\). For any \((n-1)\)-regular graph \( G = (V, E) \) with \( h \) assigned to every vertex, we can define a duplicate graph of \((d+1)|V|\) vertices as follows. First for each \( v \in V \), define vertices \( v', v_1, \ldots, v_d \). For each \( i, 1 \leq i \leq d \), we make a copy of \( G \) on \( \{v_i \mid v \in V \} \), i.e., for each edge \((u, v) \in E \), include the edge \((u_i, v_i) \) in the new graph. Next for each \( v \in V \), we introduce edges between \( v' \) and \( v_i \) for all \( 1 \leq i \leq d \). For each \( v \in V \), assign the degenerate signature \([1, i]^{\otimes d} \) that we just constructed to the vertices \( v' \); assign \( f \) to all the vertices \( v_1, \ldots, v_d \). Assume the Holant of the original graph \( G \) with \( h \) assigned to every vertex is \( H \). Then for the new graph with the given signature assignments, the Holant value is the \( d \)-th power \( H^d \). By our assumption, \( f \) is vanishing, so \( H^d = 0 \). Thus, \( H = 0 \). This holds for any graph \( G \), so \( h \) is vanishing.

Notice that \( h_k = f_k + if_{k+1} \) for any \( 0 \leq k \leq n - 1 \). If \( h \) is identically zero, then \( f_k + if_{k+1} = 0 \) for any \( 0 \leq k \leq n - 1 \), which means \( f = [1, i]^{\otimes n} \) up to a constant factor and we
are done. Suppose \( h \) is not identically zero. By the inductive hypothesis, \( h \in \mathcal{V}^+ \cup \mathcal{V}^- \). We claim \( h \) cannot be from \( \mathcal{V}^- \). This is because, although we do not directly construct \( h \) from \( f \), we can always realize it by the method depicted in the previous paragraph. Therefore the set \{ \( f' \), \( h \) \} is vanishing. As both \( f' \) and \( h \) are nonzero, and \( f' \in \mathcal{V}^+ \), we have \( h \not\in \mathcal{V}^- \), by Lemma 7.11.

Hence \( h \) is in \( \mathcal{V}^+ \). Then there exists a polynomial \( q(x) \) of degree at most \( t' = \left\lceil \frac{n-1}{2} \right\rceil \) such that \( h_k = i^k q(k) \), for any \( 0 \leq k \leq n-1 \). Since \( 2t < n-2 \), we have \( t \leq t' \). On the other hand, \( h_k = f_k + i f_{k+1} \) for any \( 0 \leq k \leq n-1 \), so we have

\[
\begin{align*}
h_k &= f_k + i f_{k+1} \\
&= i^k p(k) + c(-i)^k + i \left( i^{k+1} p(k + 1) + c(-i)^{k+1} \right) \\
&= i^k (p(k) - p(k + 1)) + 2c(-i)^k \\
&= i^k r(k) + 2c(-i)^k \\
&= i^k q(k),
\end{align*}
\]

where \( r(x) = p(x) - p(x + 1) \) is another polynomial of degree at most \( t \). Then we have

\[
q(k) - r(k) = 2c(-1)^k,
\]

which holds for all \( 0 \leq k \leq n-1 \). Notice that the left-hand side is a polynomial of degree at most \( t' \), call it \( s(x) \). However, for all even \( k \in \{0, \ldots, n-1\} \), \( s(k) = 2c \). There are exactly \( \left\lceil \frac{n}{2} \right\rceil > \left\lceil \frac{n-1}{2} \right\rceil = t' \) many even \( k \) within the range \( \{0, \ldots, n-1\} \). Thus \( s(x) = 2c \) for any \( x \). Now we pick \( k = 1 \), so \( 2c = s(1) = -2c \), which implies \( c = 0 \). This completes the proof.

Combining Lemma 7.6, Lemma 7.11, and Lemma 7.12, we obtain the following theorem that characterizes all symmetric vanishing signature sets.

**Theorem 7.13.** Let \( \mathcal{F} \) be a set of symmetric signatures. Then \( \mathcal{F} \) is vanishing if and only if \( \mathcal{F} \subseteq \mathcal{V}^+ \) or \( \mathcal{F} \subseteq \mathcal{V}^- \).

We note that some particular categories of tractable cases in previous dichotomies (case 3 of Theorem 6.3, the case of \( I(1, \pm 2i) \) in Class \( \mathcal{B} \) of Theorem 2.12, and the subcase of type \( I(1, \pm 2i) \) in case 1 of Theorem 6.12) are in \( \mathcal{R}^+ \).

To finish this section, we prove some useful properties regarding vanishing and recurrence degrees. For two symmetric signatures \( f \) and \( g \) such that \( \text{arity}(f) \geq \text{arity}(g) \), let \( \langle f, g \rangle = \langle g, f \rangle \) denote the signature that results after connecting all edges of \( g \) to \( f \). (If \( \text{arity}(f) = \text{arity}(g) \), then \( \langle f, g \rangle \) is a constant, which is considered as a signature of arity 0.)

**Lemma 7.14.** For \( \sigma \in \{+, -\} \), suppose symmetric signatures \( f \) and \( g \) satisfy \( \text{vd}^\sigma (g) = 0 \) and \( \text{arity}(f) - \text{arity}(g) \geq \text{rd}^\sigma (f) \). Then \( \text{rd}^\sigma (\langle f, g \rangle) = \text{rd}^\sigma (f) \).

**Proof.** We consider \( \sigma = + \) since the case \( \sigma = - \) is similar. Let \( \text{arity}(f) = n \), \( \text{arity}(g) = m \), and \( \text{rd}^\sigma (f) = t \). Denote the signature \( \langle f, g \rangle \) by \( f' \).
If $t = -1$, then $f$ is identically 0 and so is $f'$. Hence $\text{rd}^+(f') = -1$.

Suppose $t \geq 0$. Then we have $f_k = i^{k}p(k)$ where $p(x)$ is a polynomial of degree exactly $t$. Also $\text{arity}(f') = n - m \geq t$. We have

$$f'_k = \sum_{j=0}^{m} \binom{m}{j} f_{k+j}g_j$$

$$= i^k \sum_{j=0}^{m} \binom{m}{j} p(k+j)i^j g_j$$

$$= i^k q(k),$$

where $q(k) = \sum_{j=0}^{m} \binom{m}{j} p(k+j)i^j g_j$ is a polynomial in $k$. Notice that $\text{vd}^+(g) = 0$. Then $\text{rd}^+(g) = m$ and $g \notin \mathcal{R}_n$. Thus $\sum_{j=0}^{m} \binom{m}{j} i^j g_j \neq 0$. Then the leading coefficient of degree $t$ in the polynomial $q(k)$ is not zero. However, $\text{arity}(f') \geq t$. Thus $\text{rd}^+(f') = t$ as well. \qed

**Lemma 7.15.** For $\sigma \in \{+, -\}$, let $f$ be a nonzero symmetric signature and suppose that $f'$ is obtained from $f$ by a self loop. If $\text{vd}^\sigma(f) > 0$, then $\text{vd}^\sigma(f) - \text{vd}^\sigma(f') = \text{rd}^\sigma(f) - \text{rd}^\sigma(f') = 1$.

**Proof.** We may assume $\sigma = +$, $\text{arity}(f) = n$, and $\text{rd}^+(f) = t$. Since $f$ is not the zero signature, $t \geq 0$. Also since $\text{vd}^+(f) > 0$, $t = n - \text{vd}^+(f) < n$. By assumption, we have $f_k = i^k p(k)$, where $p(x)$ is a polynomial of degree exactly $t$. Then we have

$$f'_k = f_k + f_{k+2}$$

$$= i^k (p(k) - p(k+2))$$

$$= i^k q(k),$$

where $q(k) = p(k) - p(k+2)$ is a polynomial in $k$. If $t = 0$, then $p(x)$ is a constant polynomial and $q(x)$ is identically zero. Then $\text{rd}^+(f') = -1$ by definition and $\text{rd}^+(f) - \text{rd}^+(f') = 1$ holds. Suppose $t > 0$, then in $q(k)$, the term of degree $t$ has a zero coefficient, but the term of degree $t - 1$ is nonzero. So $q(x)$ has degree exactly $t - 1 \leq n - 2 = \text{arity}(f')$. Thus $\text{rd}^+(f') = t - 1$. Notice that $\text{arity}(f) - \text{arity}(f') = 2$, then $\text{vd}^+(f) - \text{vd}^+(f') = 1$ as well. \qed

Moreover, the set of vanishing signatures is closed under orthogonal transformations. This is because under any orthogonal transformation, the unary signatures $[1, i]$ and $[1, -i]$ are either invariant or transformed into each other, up to a nonzero multiple. Then considering the symmetrized form of any signature, we have the following lemma.

**Lemma 7.16.** For a symmetric signature $f$ of arity $n$, $\sigma \in \{+, -\}$, and an orthogonal matrix $T \in \mathbb{C}^{2 \times 2}$, either $\text{vd}^\sigma(f) = \text{vd}^\sigma(T^{\otimes n} f)$ or $\text{vd}^\sigma(f) = \text{vd}^{-\sigma}(T^{\otimes n} f)$. 

225
7.1.3 Characterization via Holographic Transformation

There is another explanation for the vanishing signatures. Given an \( f \in \mathcal{V}^+ \) with \( \text{arity}(f) = n \) and \( \text{rd}^+(f) = d \), we perform a holographic transformation with \( Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \).

\[
\text{Holant}(=_{2} \mid f) \equiv_{T} \text{Holant}([1,0,1]Z^{n^2} \mid (Z^{-1})^{\otimes n}f) \\
\equiv_{T} \text{Holant}([0,1,0] \mid \hat{f}),
\]

where \( \hat{f} \) is of the form \([\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0]\), and \( \hat{f}_d \neq 0 \). To see this, note that \( Z^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \) and \( Z^{-1}[1,i] = \sqrt{2}[1,0] \). By hypothesis \( f \) has the form \( \text{Sym}_{n-d}([1,i];v_1,\ldots,v_d) \).

Then \( \hat{f} = (Z^{-1})^{\otimes n}f = \text{Sym}_{n-d}([1,0];u_1,\ldots,u_d) \), up to a nonzero scalar factor, where \( u_i = Z^{-1}v_i \) for \( 1 \leq i \leq d \) and \( u_i \) and \( v_i \) are column vectors in \( \mathbb{C}^2 \). From this expression for \( \hat{f} \), it is clear that all entries of Hamming weight greater than \( d \) in \( \hat{f} \) are 0. Moreover, if \( \hat{f}_d = 0 \), then one of the \( u_i \) has to be a multiple of \([1,0]\). This contradicts the degree assumption of \( f \), namely \( \text{vd}^+(f) = n - \text{rd}^+(f) = n - d \) and no higher. We record this as a lemma.

**Lemma 7.17.** Suppose \( f \) is a symmetric signature of arity \( n \). Let \( \hat{f} = (Z^{-1})^{\otimes n}f \). If \( \text{rd}^+(f) = d \), then \( \hat{f} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0] \) and \( \hat{f}_d \neq 0 \). Also \( f \in \mathcal{R}^+_d \) iff all nonzero entries of \( \hat{f} \) are among the first \( d \) entries in its symmetric signature notation.

Similarly, if \( \text{rd}^-(f) = d \), then \( \hat{f} = [0, \ldots, 0, \hat{f}_{n-d}, \ldots, \hat{f}_n] \) and \( \hat{f}_{n-d} \neq 0 \). Also \( f \in \mathcal{R}^-_d \) iff all nonzero entries of \( \hat{f} \) are among the last \( d \) entries in its symmetric signature notation.

For future reference, we also note the following. If \( f = g + h \) is of arity \( n \), where \( \text{rd}^+(g) = d \), \( \text{rd}^-(h) = d' \), and \( d + d' < n \), then after a holographic transformation by \( Z \), \( \hat{f} = (Z^{-1})^{\otimes n}f \) takes the form \([\hat{g}_0, \ldots, \hat{g}_d, 0, \ldots, 0, \hat{h}_{n-d'}, \ldots, \hat{h}_n]\), where \( n - d - d' - 1 \geq 0 \) zeros are present.

In any bipartite graph for \( \text{Holant}([0,1,0] \mid \hat{f}) \), the binary **DISEQUALITY** \( (\neq_2) = [0,1,0] \) on the left imposes the condition that half of the edges must take the value 0 and the other half must take the value 1. On the right side, by \( f \in \mathcal{V}^+ \), we have \( d < n/2 \), thus \( \hat{f} \) requires that less than half of the edges are assigned the value 1. Therefore the Holant is always 0.

Under this transformation, one can observe another interesting phenomenon. For any \( a, b \in \mathbb{C} \),

\[
\text{Holant}([0,1,0] \mid [a, b, 1, 0, 0]) \quad \text{and} \quad \text{Holant}([0,1,0] \mid [0,0,1,0,0])
\]

take exactly the same value on every signature grid. This is because, to contribute a nonzero term in the Holant, exactly half of the edges must be assigned 1. Then for the first problem, the signature on the right can never contribute a nonzero value involving \( a \) or \( b \). Thus the Holant values of these two problems on any signature grid are always the same. Nevertheless, for every \( a, b \in \mathbb{C} \) not both 0, there is no holographic transformation between these two problems. This shows that the converse of the Holant Theorem (Theorem ??) does not hold.
We know that Holant([0, 1, 0] | [0, 0, 1, 0, 0]) counts the number of Eulerian orientations in a 4-regular graph. This problem is \#P-hard (Theorem 6.29) and in fact its planar version Pl-Holant(\# \neq 2 | [0, 0, 1, 0, 0]) is also \#P-hard for 4-regular graph. This will play an important role in proofs later. Translating back to the standard setting, the problem of counting Eulerian orientations in a 4-regular graph is Holant([3, 0, 1, 0, 3]). The problem Holant([0, 1, 0] | [a, b, 1, 0, 0]) corresponds to a certain signature \( f = Z^{\otimes 4}[a, b, 1, 0, 0] \) of arity 4 with recurrence degree 2. It has a different appearance but induces exactly the same Holant value as the signature for counting Eulerian orientations. Therefore, all such signatures are \#P-hard as well. We use this fact later.

### 7.2 Theorem Statement and Proof of Tractability

For a class \( \mathcal{C} \) of signatures we define the notion of \( \mathcal{C} \)-transformable. We will particularly use this notion for \( \mathcal{C} = \mathcal{A} \) or \( \mathcal{C} = \mathcal{P} \).

**Definition 7.18.** A signature \( f \) (respectively a signature set \( \mathcal{F} \)) is \( \mathcal{C} \)-transformable if there exists a holographic transformation \( T \) such that \( f \in T\mathcal{C} \) (respectively the set \( \mathcal{F} \subseteq T\mathcal{C} \)) and \([1, 0, 1]T^{\otimes 2} \in \mathcal{C} \).

Now we can formally state the dichotomy theorem for Holant problems on an arbitrary set of symmetric signatures.

**Theorem 7.19.** Let \( \mathcal{F} \) be any set of symmetric, complex-valued signatures in Boolean variables. Then Holant(\( \mathcal{F} \)) is \#P-hard unless \( \mathcal{F} \) satisfies one of the following conditions, in which case the problem is in P:

1. All non-degenerate signatures in \( \mathcal{F} \) are of arity at most 2;
2. \( \mathcal{F} \) is \( \mathcal{A} \)-transformable;
3. \( \mathcal{F} \) is \( \mathcal{P} \)-transformable;
4. \( \mathcal{F} \subseteq \mathcal{V}^{\sigma} \cup \{ f \in \mathcal{R}_2^{\sigma} | \text{arity}(f) = 2 \} \) for some \( \sigma \in \{+, -\} \);
5. All non-degenerate signatures in \( \mathcal{F} \) are in \( \mathcal{R}_2^{\sigma} \) for some \( \sigma \in \{+, -\} \).

Note that any signature in \( \mathcal{R}_2^{\sigma} \) having arity at least 3 is a vanishing signature. Thus all non-degenerate signatures of arity at least 3 in case 5 are vanishing. While both cases 4 and 5 are largely concerned with vanishing signatures, these two cases differ. In case 4, all signatures in \( \mathcal{F} \), including unary signatures but excluding binary signatures, must be vanishing of a single type \( \sigma \); the binary signatures are only required to be in \( \mathcal{R}_2^{\sigma} \). In contrast, case 5 has no requirement placed on degenerate signatures which include all unary signatures. Then all non-degenerate binary signatures are required to be in \( \mathcal{R}_2^{\sigma} \). Finally all non-degenerate signatures of arity at least 3 are also required to be in \( \mathcal{R}_2^{\sigma} \), which is a strong form of vanishing; they must have a large vanishing degree of type \( \sigma \).
Case 5 is tractable due to Fibonacci gates. In fact every non-degenerate signature in $R_2^\sigma$ is a generalized Fibonacci signature with the parameter $\lambda = \sigma 2i$ in equation (2.5), and after replacing all degenerate signatures with the corresponding unary signatures we can invoke Holant$^*$ tractability. However, in the following we present a unified proof of tractability based on vanishing signatures, which leads to an alternative algorithm for this case.

Proof of Tractability of Theorem 7.19. For any signature grid $\Omega$, Holant$^\Omega$ is the product of the Holant on each connected component, so we only need to compute over connected components.

For case 1, after decomposing all degenerate signatures into unary ones, a connected component of the graph is either a path or a cycle and the Holant can be computed using matrix product and trace. Cases 2 and 3 are tractable because, after a particular holographic transformation, their instances are tractable instances of $\#\text{CSP}(\mathcal{F})$, by the dichotomy theorem for $\#\text{CSP}$, Theorem 3.7. For case 4, any binary signature $g \in R_2^\sigma$ has $\text{rd}^\sigma(g) \leq 1$, and thus $\text{vd}^\sigma(g) \geq 1 - \text{arity}(g)/2$. Any signature $f \in \mathcal{V}^\sigma$ has $\text{vd}^\sigma(f) > \text{arity}(f)/2$. If $\mathcal{F}$ contains a signature $f$ of arity at least 3, then it must belong to $\mathcal{V}^\sigma$. Then by the combinatorial view, more than half of the unary signatures are $[1, \sigma i]$, so Holant$^\Omega$ vanishes. On the other hand, if the arity of every signature in $\mathcal{F}$ is at most 2, then we have reduced to tractable case 1.

Now consider case 5. First, replace each degenerate signature by its unary signature tensor factors. Any remaining signature of arity at least two is non-degenerate and thus belongs to $R_2^\sigma$. Then recursively absorb any unary signature into its neighboring signature. If it is connected to another unary signature, then this produces a global constant factor. If it is connected to a binary signature, then this creates another unary signature. We observe that if $f \in R_2^\sigma$ has $\text{arity}(f) \geq 2$, then for any unary signature $u$, after connecting $f$ to $u$, the signature $\langle f, u \rangle$ still belongs to $R_2^\sigma$. Hence after recursively absorbing all unary signatures in the above process, we have a signature grid where all signatures belong to $R_2^\sigma$. Any signature $f$ that remains and has arity at least 3 belongs to $\mathcal{V}^\sigma$ since $\text{rd}^\sigma(f) \leq 1$ and thus $\text{vd}^\sigma(f) \geq \text{arity}(f) - 1 > \text{arity}(f)/2$. Thus we have reduced to tractable case 4.

7.3 A Sample of Problems

We illustrate the scope of Theorem 7.19 by several concrete problems. Note that some problems are naturally expressed with real weights, but they are linked to other problems that use complex weights. Sometimes the inherent link between two real-weighted problems is provided by a transformation through $\mathbb{C}$.

Problem: $\#\text{VertexCover}$

Input: An undirected graph $G$.  

228
**Output:** The number of vertex covers in $G$.

This classic problem is most naturally expressed as a spin system with edge function being the Boolean OR function, or as a counting CSP problem $\#\text{CSP}([0,1,1])$. As a real-weighted bipartite Holant problem it is $\text{Holant}([0,1,1] \mid \mathcal{E} \mathcal{Q})$, where $\mathcal{E} \mathcal{Q} = \{ (=_k) \mid k \geq 1 \}$ is the set of all EQUALITIES. A vertex assigned an equality signature forces all its incident edges to be assigned the same value; this is equivalent to assigning a value to the vertex. The degree two vertices assigned the binary OR $[0;1;1]$ should be thought of as an edge between its neighboring vertices. These edge-like vertices force at least one of its neighbors to be selected. The number of assignments satisfying these requirements is exactly the number of vertex covers.

To apply Theorem 7.19, we perform a holographic transformation by $T = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$. Note that $T^{-1} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, and

$$
[0, 1, 1] = (0 1 1 1) = [1, 1]^{\otimes 2} + [i, 0]^{\otimes 2} = \{ [1, 0]^{\otimes 2} + [0, 1]^{\otimes 2} \} (T^{-1})^{\otimes 2} \\
= (1 0 0 1)(T^{-1})^{\otimes 2} = (=_2)(T^{-1})^{\otimes 2}.
$$

Thus, a holographic transformation by $T$ yields

$$
\text{Holant}([0, 1, 1] \mid \mathcal{E} \mathcal{Q}) \equiv_T \text{Holant}([0, 1, 1] T^{\otimes 2} \mid T^{-1} \mathcal{E} \mathcal{Q}) \\
\equiv_T \text{Holant}(_=2 \mid T^{-1} \mathcal{E} \mathcal{Q}) \\
\equiv_T \text{Holant}(T^{-1} \mathcal{E} \mathcal{Q}).
$$

The EQUALITY signature $(=_k)$ of arity $k$ in $\mathcal{E} \mathcal{Q}$ is transformed by $T^{-1}$ to

$$
f_{(k)} = (T^{-1})^{\otimes k}(=_k) \\
= \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}^{\otimes k} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes k} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes k} \right\} \\
= \begin{bmatrix} 1 \\ i \end{bmatrix}^{\otimes k} + [0]^{\otimes k} \\
= [2, i, -1, -i, 1, i, -1, -i, 1, i, \ldots]
$$

of arity $k$. By Theorem 7.19, $\text{Holant}(T^{-1} \mathcal{E} \mathcal{Q})$ is $\#\text{P}$-hard. Indeed, even $\text{Holant}(f_{(k)})$, the restriction of this problem to $k$-regular graphs is $\#\text{P}$-hard, for $k \geq 3$. This can also be obtained more directly by Theorem 7.52, a dichotomy for a single signature, which is a special case of Theorem 7.19.

**Problem:** $\#\lambda$-$\text{VertexCover}$

**Input:** An undirected graph $G$.

**Output:** $\sum_{C \in \mathcal{C}(G)} \lambda^{e(C)}$, where $\mathcal{C}(G)$ denotes the set of all vertex covers of $G$, and $e(C)$ is the number of edges with both endpoints in the vertex cover $C$.

Our dichotomy also easily handles this edge-weighted vertex cover problem that is de-
noted by Holant([0, 1, λ] | EΩ). Suppose λ ≠ 0. On regular graphs, this problem is equivalent to the so-called hardcore gas model, which is the vertex-weighted problem denoted by Holant([1, 1, 0] | F), where F consists of signatures of the form [1, 0, ..., 0, µ]. By flipping 0 and 1, this is the same as Holant([0, 1, 1] | F') with F' containing [µ, 0, ..., 0, 1]. For k-regular graphs, the following diagonal transformation $T = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ shows this equivalence, where $\lambda = 1/\mu^{1/k}$;

$$\text{Holant([0, 1, λ] | =k)} \equiv_T \text{Holant([0, 1, λ]|(T^{-1})^{\otimes k}(=k))}$$

$$\equiv_T \text{Holant([1/2, 0, 1, 1] | [1, 0, \ldots, 0, λ^k])}$$

$$\equiv_T \text{Holant([0, 1, 1] | [µ, 0, \ldots, 0, 1]).}$$

This problem, denoted by #$k$-$\lambda$-VERTEXCOVER, is #P-hard for $k \geq 3$. To see this, apply the holographic transformation $T = [0, -i, 1]$ to the edge-weighted form of the problem Holant([0, 1, λ] | =k). Then [0, 1, λ] is transformed to [0, 1, λ] $T^{\otimes 2} = \lambda(=2)$ and (=k) is transformed to $(T^{-1})^{\otimes k}(=k) = g_{(λ,k)} = \frac{1}{\lambda^k}[λ^k + 1, i, -1, -i, 1, \ldots]$, where $T^{-1} = \frac{1}{λ}\begin{bmatrix} 1 & λ \\ 1 & λ \end{bmatrix}$. Since Holant($g_{(λ,k)}$) is #P-hard by Theorem 7.19, or more directly by Theorem 7.52, we conclude that #$k$-$\lambda$-VERTEXCOVER is also #P-hard.

If $λ = 0$, then the above problem is Holant([0, 1, 0] | EΩ), which is tractable. However, the transformation $T$ above is singular in this case. We can in fact apply another transformation $T' = \begin{bmatrix} -λ/2 & -i \\ λ/2 & i \end{bmatrix}$ such that it transforms the problem Holant([0, 1, λ] | =k) into Holant($h_{(λ,k)}$) for some $h_{(λ,k)}$ regardless of whether $λ = 0$ or not. Then by applying Theorem 7.52, we reach the same conclusion that #$\lambda$-VERTEXCOVER is #P-hard on k-regular graphs when $λ ≠ 0$. We note that when $λ = 0$, $T' = \begin{bmatrix} 1 & -i \\ 1 & -i \end{bmatrix} = \sqrt{2}Z^{-1}$, where $Z = \frac{1}{\sqrt{2}}[1 \ 1]$.

We now consider some orientation problems.

**Problem:** #NOSSINKORIENTATION

**Input:** An undirected graph $G$.

**Output:** The number of orientations of $G$ such that each vertex has at least one outgoing edge.

This problem is denoted by Holant([0, 1, 0] | F), where F consists of $f(k) = [0, 1, \ldots, 1, 1]$ for any arity $k$. Each degree two vertex on the left side of the bipartite graph must have its incident edges assigned different values. We associate an oriented edge between the neighbors of such vertices with the head on the side assigned 0 and the tail on the side assigned 1. This problem is #P-hard even over k-regular graphs provided $k \geq 3$. Just as with the bipartite form of the vertex cover problem, we do a holographic transformation to apply Theorem 7.19.
This time, we pick \( T = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} Z^{-1}, \) with \( T^{-1} = \sqrt{2} Z = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \) and get

\[
\text{Holant}([0, 1, 0] | f(k)) \equiv_T \text{Holant}([0, 1, 0]T_{\otimes 2} | (T^{-1})_{\otimes k} f(k))
\equiv_T \text{Holant}(\frac{1}{2}[1, 0, 1] | \hat{f}(k))
\equiv_T \text{Holant}(\hat{f}(k)),
\]

where \( \hat{f}(k) = [2^k - 1, -1, 1, 1, -1, \ldots] \). The signature \(-\hat{f}(k) = [1 - 2^k, i, -1, -i, 1, \ldots] \) is actually a special case of the \#\( k \)-\( \lambda \)-VERTEXCOVER problem with \( \lambda = 2e^{\pi i/k} \). Therefore, this problem is \#P-hard. However, if we consider this problem modulo \( 2^k \), \( \hat{f}(k) \) becomes \([-1, -i, 1, i, -1, \ldots] \), and it belongs to one of the tractable cases in Theorem 7.19. Thus, \#\text{NoSinkOrientation} is tractable modulo \( 2^t \), where \( t \) is the minimum degree of the input graph.

**Problem:** \#\text{NoSinkNoSourceOrientation}

**Input:** An undirected graph \( G \).

**Output:** The number of orientations of \( G \) such that each vertex has at least one incoming and one outgoing edge.

This problem is denoted by \( \text{Holant}([0, 1, 0] | \mathcal{F}) \), where \( \mathcal{F} \) consists of \( f(k) = [0, 1, \ldots, 1, 0] \) for any arity \( k \). This problem is also \#P-hard on \( k \)-regular graphs for \( k \geq 3 \). We pick the same \( T \) as in the previous problem and get

\[
\text{Holant}([0, 1, 0] | f(k)) \equiv_T \text{Holant}([0, 1, 0]T_{\otimes 2} | (T^{-1})_{\otimes k} f(k))
\equiv_T \text{Holant}(\frac{1}{2}[1, 0, 1] | \hat{f}(k))
\equiv_T \text{Holant}(\hat{f}(k)),
\]

where \( \hat{f}(k) = [2^k - 2, 0, 2, 0, -2, \ldots] \). Here we transform from one real-weighted Holant problem to another real-weighted Holant problem via a complex-weighted transformation. The \#P-hardness follows from Theorem 7.52. Like the previous problem, this problem is tractable modulo \( 2^t \), where \( t \) is the minimum degree of the input graph.

Theorem 7.19 is applicable to a set of signatures when different vertices may have different constraints.

**Problem:** \#\text{One-In-Or-One-Out-Orientation}

**Input:** An undirected graph \( G \) with each vertex labeled “One-In” or “One-Out”.

**Output:** The number of orientations of \( G \) such that each vertex has exactly one incoming or exactly one outgoing edge as specified by its label.

This problem is denoted by \( \text{Holant}([0, 1, 0] | \mathcal{F}) \), where the set \( \mathcal{F} \) consists of signatures of the form \( f = [0, 1, 0, \ldots, 0] \) and \( g = [0, \ldots, 0, 1, 0] \). Once again, it is \#P-hard on \( k \)-regular graphs for \( k \geq 3 \). We apply the same transformation as in the above two orientation
problems. The result is Holant(\{\hat{f}, \hat{g}\}), where \(\hat{f} = [k, (k-2)i, -(k-4), \ldots]\) and \(\hat{g} = [k, -(k-2)i, -(k-4), \ldots]\) of arity \(k\). In fact, the entries of \(\hat{f}\) satisfy a second order recurrence relation with characteristic polynomial \((x-i)^2\) while the entries of \(\hat{g}\) satisfy one with characteristic polynomial \((x+i)^2\). Thus \(\hat{f} \in R^+_2\) and \(\hat{g} \in R^-_2\). The hardness follows from Theorem 7.19. See also Lemma 7.30. However, the restriction of this problem to planar graphs is tractable since these signatures are all matchgate signatures, by Theorem 4.11. If we only consider one signature, then both Holant(\(\hat{f}\)) and Holant(\(\hat{g}\)) are tractable. This is also true for a set of signatures with different arities but of one type, either of the type \(\hat{f}\) or of the type \(\hat{g}\). This is case 5 of Theorem 7.19. One can also observe that the problem Holant(\(\hat{f}\)) is equivalent to the problem Holant([0, 1, 0] | [0, 1, 0, \ldots, 0]), which is always 0 provided \(k \geq 3\) by a simple counting argument. Similarly for Holant(\(\hat{g}\)). For \(k \geq 3\), \(\hat{f} \in \mathcal{V}^+\) and \(\hat{g} \in \mathcal{V}^−\) are vanishing signatures. However, combining two such vanishing signatures of the opposite type \(\mathcal{V}^+\) and \(\mathcal{V}^−\), the problem Holant(\{\(\hat{f}, \hat{g}\)\}) is \#P-hard.

One sufficient condition for a signature of arity at least 3 to be vanishing is that its entries satisfy a second order recurrence relation with characteristic polynomial \((x \pm i)^2\). If the entries of a non-degenerate signature \(f\) satisfy a second order recurrence relation with characteristic polynomial \((x - \mu)^2\) for \(\mu \neq \pm i\), then there exists an orthogonal holographic transformation such that \(f\) is transformed into a weighted matching signature.

**Problem:** \#\(\lambda\)-\textsc{WeightedMatching}

**Input:** An undirected graph \(G\).

**Output:** \(\sum_{M \in \mathcal{M}(G)} \lambda^{v(M)}\), where \(\mathcal{M}(G)\) is the set of all matchings in \(G\) and \(v(M)\) is the number of unmatched vertices in the matching \(M\).

The Holant expression of this problem is Holant(\(\mathcal{F}\)), where \(\mathcal{F}\) consists of signatures of the form \([\lambda, 1, 0, \ldots, 0]\). When \(\lambda = 0\), this problem counts perfect matchings, which is \#P-hard even for bipartite graphs [?] but tractable over planar graphs by Kasteleyn’s algorithms (Theorem 4.7). When \(\lambda = 1\), this problem counts general matchings. Vadhan [?] proved that counting general matchings is \#P-hard over \(k\)-regular graphs for \(k \geq 5\), but left open the question for \(k = 4\). Theorem 7.19, or more directly Theorem 7.52, shows that \#\(\lambda\)-\textsc{WeightedMatching} is \#P-hard, for any weight \(\lambda\) and on any \(k\)-regular graphs for \(k \geq 3\).

### 7.4 Outline of Hardness Proof for Theorem 7.19

The hardness proof of Theorem 7.19 will take up the remainder of this Chapter. Before proving Theorem 7.19 which deals with a set of signatures \(\mathcal{F}\), we first prove a special case Theorem 7.52, where \(\mathcal{F}\) consists of a single signature. The proof is by induction on the arity of the signature. The induction is done by taking a self loop, which causes the arity to go down by 2. Thus, we need two base cases, a dichotomy for an arity 3 signature, Theorem 6.3,
and a dichotomy for an arity 4 signature, Theorem 7.29, which is a crucial ingredient in our proof of Theorem 7.19. The arity 4 dichotomy not only is a base case of the single signature dichotomy Theorem 7.52, but also is utilized several times in the inductive step.

After obtaining the dichotomy for an arity 4 signature, the proof continues by revisiting the vanishing signatures to determine what signatures combine with them to give \#P-hardness. When adding unary or binary signatures, the only possible combinations that maintain the tractability of the vanishing signatures are as described in cases 4 and 5 in Theorem 7.19. Moreover, combining two vanishing signatures of the opposite type of arity at least 3 implies \#P-hardness. The proof of this last statement uses techniques that are similar to those in the proof of the arity 4 dichotomy.

Another ingredient of the proof is to understand the signatures that are $\mathcal{A}$-transformable or $\mathcal{P}$-transformable. We obtain explicit characterizations of these signatures. We use these characterizations to prove dichotomy theorems for any signature set containing an $\mathcal{A}$- or $\mathcal{P}$-transformable signature. Unless every signature in the set is $\mathcal{A}$- or $\mathcal{P}$-transformable, the problem is \#P-hard. The proofs of these dichotomy theorems utilize the $\#\text{CSP}^d$ dichotomy, Theorem 6.16.

The main dichotomy, Theorem 7.19, depends on Theorem 7.52 and the results regarding vanishing signatures as well as $\mathcal{A}$- and $\mathcal{P}$-transformable signatures. Figure 7.2 summarizes the dependencies among these results.
7.5 Dichotomy for One Signature of Arity 4

Definition 7.20. A 4-by-4 matrix is redundant if its middle two rows and middle two columns are the same. Denote the set of all redundant 4-by-4 matrices over \( \mathbb{C} \) by \( \text{RM}_4(\mathbb{C}) \).

Recall the notion of a signature matrix in (6.16) of a signature of arity 4. As an example, the signature matrix in (6.17) of a symmetric arity 4 signature is redundant.

Define a function \( \varphi : \mathbb{C}^{4 \times 4} \to \mathbb{C}^{3 \times 3} \) by

\[
\varphi(M) = AMB,
\]

where

\[
A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.
\]

Intuitively, the operation \( \varphi \) replaces the middle two columns of \( M \) with their sum and then the middle two rows of \( M \) with their average. (These two steps commute.) In the other direction, we define the following function \( \psi : \mathbb{C}^{3 \times 3} \to \text{RM}_4(\mathbb{C}) \) by

\[
\psi(N) = BNA.
\]

Intuitively, the operation \( \psi \) duplicates the middle row of \( N \) and then splits the middle column evenly into two columns. Notice that \( \varphi(\psi(N)) = N \). When restricted to \( \text{RM}_4(\mathbb{C}) \), \( \varphi \) is an isomorphism between the semi-group of 4-by-4 redundant matrices and the semi-group of 3-by-3 matrices, under matrix multiplication, and \( \psi \) is its inverse. This can be seen by noticing that

\[
AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad BA = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]

are the identity elements of their respective semi-groups.

Definition 7.21. If \( g \) is a signature of arity 4 with a redundant signature matrix \( M_g \), then we define the compressed signature matrix of \( g \) as \( \bar{M}_g = \varphi(M_g) \).

If all signatures in an \( \mathcal{F} \)-gate have even arity, then the \( \mathcal{F} \)-gate also has even arity. Knowing that binary signatures alone in Holant problems do not produce \#P-hardness (see tractable case 1 of Theorem 7.19), with the above constraint in mind, we would like to interpolate other arity 4 signatures using a given arity 4 signature. We are particularly interested in the following signature \( g \) with signature matrix

\[
M_g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.
\]
This is the identity element in the semi-group \(RM_4(\mathbb{C})\). Thus, \(\widetilde{M}_g = I_3\). Lemma 7.24 will show that the Holant problem with this signature is \#P-hard. In Lemma 7.23, we consider when we can interpolate it. This is a key lemma. There are three cases in Lemma 7.23 and one of them requires the following technical lemma.

**Lemma 7.22.** Let \(M = [B_0 \ B_1 \cdots \ B_t]\) be an \(n\)-by-\(n\) block matrix such that there exists \(\lambda \in \mathbb{C}\), for all integers \(0 \leq k \leq t\), block \(B_k\) is an \(n\)-by-\(c_k\) matrix for some integer \(c_k \geq 0\), and the entry of a nonempty block \(B_k\) for \(c_k \geq 1\) at row \(r\) and column \(c\) is \((B_k)_{rc} = r^{c-1}\lambda^{kr}\), where \(1 \leq r \leq n\), \(1 \leq c \leq c_k\). If \(\lambda\) is nonzero and is not a root of unity, then \(M\) is nonsingular.

**Proof.** We prove by induction on \(n\). If \(n = 1\), then the sole entry is \(\lambda^k\) for some nonnegative integer \(k\). This is nonzero since \(\lambda \neq 0\). Assume \(n > 1\) and let the left-most nonempty block be \(B_j\). We divide row \(r\) by \(\lambda^{jr}\), which is allowed since \(\lambda \neq 0\). This effectively changes block \(B_i\) into a block of the form \(B_{r-j}\). Thus, we have another matrix of the same form as \(M\) but with a nonempty block \(B_0\). To simplify notation, we denote this matrix again by \(M\). The first column of \(B_0\) is all 1’s. We subtract row \(r-1\) from row \(r\), for \(r\) from \(n\) down to 2. This gives us a new matrix \(M' = [B'_0 \ B'_1 \cdots B'_t]\), and \(\det M = \det M'\). Then \(\det M'\) is the determinant of the \((n-1)\)-by-\((n-1)\) submatrix \(M''\) obtained from \(M'\) by removing the first row and column. Now we do column operations on \(M''\) to return the blocks to the proper form so that we can apply induction.

For any nonempty block \(B'_k\) other than \(B'_0\), we prove by induction on the number of columns in \(B'_k\) that \(B'_k\) can be repaired. In the base case, the \(r\)th element of the first column is \((B'_k)_{r1} = \lambda^{kr} - \lambda^{k(r-1)} = \lambda^{k(r-1)}(\lambda^k - 1)\) for \(r \geq 2\). We divide this column by \(\lambda^k - 1\) to obtain \(\lambda^{k(r-1)}\), which is allowed since \(\lambda\) is not a root of unity and \(k \neq 0\). This is now the correct form for the \(r'\)th element of the first column of a block in \(M''\), for the row index \(r' = r - 1 \geq 1\).

Now for the inductive step of \(B'_k\), assume that the first \(d - 1\) columns of the block \(B'_k\) have been repaired to the correct form to be a block in \(M''\). That is, for row index \(r \geq 2\), which denotes the \((r - 1)\)-th row of \(M''\), the \(r\)th element in the first \(d - 1\) columns of \(B'_k\) has the form \((B'_k)_{rc} = (r - 1)^{c-1}\lambda^{k(r-1)}\). The \(r\)th element in column \(d\) of \(B'_k\) currently has the form \((B'_k)_{rd} = r^{d-1}\lambda^{kr} - (r - 1)^{d-1}\lambda^{k(r-1)}\). Then we do column operations

\[
(B'_k)_{rd} - \sum_{c=1}^{d-1} \binom{d-1}{c-1} (B'_k)_{rc} = r^{d-1}\lambda^{kr} - (r - 1)^{d-1}\lambda^{k(r-1)} - \sum_{c=1}^{d-1} \binom{d-1}{c-1} (r - 1)^{c-1}\lambda^{k(r-1)}
\]

\[
= r^{d-1}\lambda^{kr} - r^{d-1}\lambda^{k(r-1)}
\]

\[
= r^{d-1}\lambda^{k(r-1)}(\lambda^k - 1)
\]

and divide by \((\lambda^k - 1)\) to get \(r^{d-1}\lambda^{k(r-1)}\). Once again, this is allowed since \(\lambda\) is not a root of unity and \(k \neq 0\). Then more column operations of the same kind yield

\[
r^{d-1}\lambda^{k(r-1)} - \sum_{c=1}^{d-1} \binom{d-1}{c-1} (r - 1)^{c-1}\lambda^{k(r-1)} = \lambda^{k(r-1)} \left( r^{d-1} + (r - 1)^{d-1} - \sum_{c=1}^{d} \binom{d-1}{c-1} (r - 1)^{c-1} \right)
\]

235
and the term in parentheses is precisely \((r - 1)^{d-1}\). This gives the correct form for the \(r\)th element in column \(d\) of \(B_k'\) in \(M''\).

Finally we repair the columns in \(B_0'\), also by induction on the number of columns. In the base case, if \(B_0'\) only has one column, then there is nothing to prove, since this block has disappeared in \(M''\). Otherwise, \((B_0')_{r2} = r - (r - 1) = 1\), so the second column is already in the correct form to be the first column in \(M''\), and there is still nothing to prove. For the inductive step, assume that columns 2 to \(d - 1\) are in the correct form to be the first block in \(M''\) for \(d \geq 3\). That is, the entry at row \(r \geq 2\) and column \(c\) from 2 through \(d - 1\) has the form \((B_0')_{rc} = (r - 1)^{c-2}\). The \(r\)th element in column \(d\) currently has the form \((B_0')_{rd} = r^{d-1} - (r - 1)^{d-1}\). Then we do the column operations

\[
(B_0')_{rd} - \sum_{c=2}^{d-1} \frac{(d - 1)}{c - 2} (B_0')_{rc} = r^{d-1} - (r - 1)^{d-1} - \sum_{c=2}^{d-1} \frac{(d - 1)}{c - 2} (r - 1)^{c-2}
\]

and divide by \(d - 1\), which is nonzero, to get \((r - 1)^{d-2}\). Recall that the block \(B_0'\) in \(M''\) is indexed by \(r\) and \(d\) starting at 2. This is the correct form for the \(r\)th element in column \(d\) of \(B_0'\) in \(M''\). Therefore, we invoke our original induction hypothesis that the \((n - 1)\)-by-\((n - 1)\) matrix \(M''\) has a nonzero determinant, which completes the proof.

\begin{lemma}
Let \(g\) be the arity 4 signature with \(M_g\) given in Equation (7.5) and let \(f\) be an arity 4 signature with complex weights. If \(M_f\) is redundant and \(\tilde{M}_f\) is nonsingular, then
\end{lemma}
for any set $\mathcal{F}$ containing $f$, we have

$$\text{Holant}(\mathcal{F} \cup \{g\}) \leq_T \text{Holant}(\mathcal{F}).$$

The proof also works for planar graphs, thus $\text{Pl-Holant}(\mathcal{F} \cup \{g\}) \leq_T \text{Pl-Holant}(\mathcal{F})$.

Proof. Consider an instance $\Omega$ of Holant($\mathcal{F} \cup \{g\}$). Suppose that $g$ appears $n$ times in $\Omega$. We construct from $\Omega$ a sequence of instances $\Omega_s$ of Holant($\mathcal{F}$) indexed by $s \geq 1$. We obtain $\Omega_s$ from $\Omega$ by replacing each occurrence of $g$ with the gadget $N_s$ in Figure 7.3 with $f$ assigned to all vertices. In $\Omega_s$, the edge corresponding to the $i$th significant index bit of $N_s$ connects to the same location as the edge corresponding to the $i$th significant index bit of $g$ in $\Omega$. Since $f$ may not be symmetric, in Figure 7.3 we place a diamond on the edge corresponding to the most significant index bit. The remaining index bits are in order of decreasing significance as one travels counterclockwise around the vertex.

Now to determine the relationship between Holant$_\Omega$ and Holant$_{\Omega_s}$, we use the isomorphism between RM$_4(\mathbb{C})$ and $\mathbb{C}^{3 \times 3}$. To compute Holant$_{\Omega_s}$, we effectively replace $M_g$ in Holant$_\Omega$ with $M_{N_s} = (M_f)^s$, the $s$th power of the signature matrix $M_f$. By the Jordan normal form of $\tilde{M}_f$, there exists $T$ and $\Lambda \in \mathbb{C}^{3 \times 3}$ such that

$$\tilde{M}_f = T\Lambda T^{-1} = T \begin{bmatrix} \lambda_1 & b_1 & 0 \\ 0 & \lambda_2 & b_2 \\ 0 & 0 & \lambda_3 \end{bmatrix} T^{-1},$$

where $b_1, b_2 \in \{0, 1\}$. By assumption $\lambda_1\lambda_2\lambda_3 = \text{det}(\tilde{M}_f) \neq 0$. Also since $\tilde{M}_g = \varphi(M_g) = I_3$, and $TI_3T^{-1} = I_3$, we have $\psi(T)M_g\psi(T^{-1}) = M_g$. We can view our construction of $\Omega_s$ as first replacing each $M_g$ by $\psi(T)M_g\psi(T^{-1})$, which does not change the Holant value, and then replacing each new $M_g$, sandwiched between $\psi(T)$ and $\psi(T^{-1})$, by $\psi(\Lambda^s) = \psi(\Lambda)^s$ to obtain $\Omega_s$. Observe that

$$\varphi(\psi(T)\psi(\Lambda^s)\psi(T^{-1})) = T\Lambda^s T^{-1} = (\tilde{M}_f)^s = (\varphi(M_f))^s = \varphi((M_f)^s) = \varphi(M_{N_s}),$$

and both $\psi(T)\psi(\Lambda^s)\psi(T^{-1})$ and $M_{N_s}$ are redundant, hence $\psi(T)\psi(\Lambda^s)\psi(T^{-1}) = M_{N_s}$. Since $M_g = \psi(T)M_g\psi(T^{-1})$ and $M_{N_s} = \psi(T)\psi(\Lambda^s)\psi(T^{-1})$, replacing each $M_g$, sandwiched between $\psi(T)$ and $\psi(T^{-1})$, by $\psi(\Lambda^s)$ indeed transforms $\Omega$ to $\Omega_s$. (We note that, by the isomorphism, $\psi(T^{-1})$ is the multiplicative inverse of $\psi(T)$ within the semi-group RM$_4(\mathbb{C})$; but we prefer not to write it as $\psi(T)^{-1}$ since it is not the usual matrix inverse as a 4-by-4 matrix. Indeed, $\psi(T)$ is not invertible as a 4-by-4 matrix.)

In the analysis below, we stratify assignments in $\Omega_s$ based on assignment values to $\psi(\Lambda^s)$. The inputs to $\psi(\Lambda^s)$ are from $\{0, 1\}^2 \times \{0, 1\}^2$. However, we can combine the inputs 01 and 10, since $\psi(\Lambda^s)$ is redundant. Thus we actually stratify assignments in $\Omega_s$ based on assignment values to $\Lambda^s$, which are from $\{0, 1, 2\} \times \{0, 1, 2\}$. In this compressed form, the row and column assignments to $\Lambda^s$ are the Hamming weight of the two actual inputs from $\{0, 1\}^2$ to the uncompressed form $\psi(\Lambda^s)$. 

237
Now we begin the analysis based on the values of \(b_1\) and \(b_2\), i.e., the shape \(\Lambda\) takes as the Jordan normal form of \(M_f\).

1. \(b_1 = b_2 = 0\). In this case \(\tilde{M}_f\) is diagonalizable, we only need to consider the assignments to \(\Lambda^s\) that assign
   
   - \((0, 0)\) \(i\) many times,
   - \((1, 1)\) \(j\) many times, and
   - \((2, 2)\) \(k\) many times

   such that \(i + j + k = n\), the number of times \(g\) occurs in \(\Omega\), which is also the number of times the gadget \(N_s\) occurs in \(\Omega_s\), for all \(s \geq 1\). Note that any other assignment contributes 0 to the Holant sum. Also note that \(\psi(\Lambda^s) = \begin{bmatrix} \lambda_1^s & 0 & 0 \\ 0 & \lambda_2^s/2 & \lambda_2^s/2 \\ 0 & \lambda_2^s/2 & \lambda_3^s \end{bmatrix}\). Let \(c_{ijk}\) be the sum over all such assignments of the products of evaluations (excluding \(\Lambda^s\) but including the contributions from \(T\) and \(T^{-1}\)) on \(\Omega_s\). Note that \(c_{ijk}\) is independent of \(s\). Then by the form of \(M_g\) in (7.5), we have

   \[\text{Holant}_\Omega = \sum_{i+j+k=n} c_{ijk} \frac{2^j}{2^j}\]

   and the value of the Holant on \(\Omega_s\), for \(s \geq 1\), is

   \[\text{Holant}_{\Omega_s} = \sum_{i+j+k=n} (\lambda_1^i \lambda_2^j \lambda_3^k)^s \left( \frac{c_{ijk}}{2^j} \right),\]

   where the factor \(\frac{1}{2}\) comes from the middle 4 entries in \(M_g\) in (7.5) and in \(\psi(\Lambda^s)\). The coefficient matrix is Vandermonde, but it may not have full rank because it may be that \(\lambda_1^i \lambda_2^j \lambda_3^k = \lambda_1^{i'} \lambda_2^{j'} \lambda_3^{k'}\) for some \((i, j, k) \neq (i', j', k')\). However, this is not a problem since we are only interested in the sum \(\sum c_{ijk}/2^j\). If two coefficients are the same, we replace their corresponding unknowns \(\frac{c_{ijk}}{2^j}\) and \(\frac{c_{ijk'}}{2^{j'}}\) with their sum as a new variable. After all such combinations, we have a Vandermonde system of full rank. In particular, none of the entries are zero since \(\lambda_1 \lambda_2 \lambda_3 = \det(M_f) \neq 0\). Therefore, we can solve the linear system in polynomial time and obtain the value of Holant_{\Omega}.

2. \(b_1 \neq b_2\). We can assume that the Jordan blocks in \(\Lambda\) have been permuted so that \(b_1 = 1\) and \(b_2 = 0\). Then \(\lambda_1 = \lambda_2\), and we denote it by \(\lambda\). Denote \(\lambda_3\) by \(\mu\). We have

   \[\Lambda^s = \begin{bmatrix} \lambda^s & s\lambda^{s-1} & 0 \\ 0 & \lambda^s & 0 \\ 0 & 0 & \mu^s \end{bmatrix}, \quad \text{and} \quad \psi(\Lambda^s) = \begin{bmatrix} \lambda^s & s\lambda^{s-1}/2 & s\lambda^{s-1}/2 & 0 \\ 0 & \lambda^s/2 & \lambda^s/2 & 0 \\ 0 & \lambda^s/2 & \lambda^s/2 & 0 \\ 0 & 0 & 0 & \mu^s \end{bmatrix}.\]

   We only need to consider the assignments to \(\Lambda^s\) that assign
   
   - \((0, 0)\) \(i\) many times,
• (1, 1) \( j \) many times,
• (2, 2) \( k \) many times, and
• (0, 1) \( \ell \) many times

with \( i + j + k + \ell = n \), since any other assignment contributes 0 to the Holant sum.

Let \( c_{ijkl} \) be the sum over all such assignments of the products of evaluations (excluding \( \Lambda^s \) but including the contributions from \( T \) and \( T^{-1} \)) on \( \Omega_s \); \( c_{ijkl} \) is independent of \( s \).

Then

\[
\text{Holant}_\Omega = \sum_{i+j+k=\ell=n} \frac{c_{ijk0}}{2^j} \tag{7.6}
\]

and the value of the Holant on \( \Omega_s \), for \( s \geq 1 \), is

\[
\text{Holant}_{\Omega_s} = \sum_{i+j+k+\ell=n} \lambda^{(i+j)s} \mu^{ks} \left(s \lambda^{s-1}\right)^{\ell} \left(\frac{c_{ijkl}}{2^{j+\ell}}\right) = \lambda^{ns} \sum_{i+j+k+\ell=n} \left(\frac{\mu}{\lambda}\right)^{ks} s^{\ell} \left(\frac{c_{ijkl}}{\lambda^\ell2^{j+\ell}}\right). \tag{7.7}
\]

If \( \mu/\lambda \) is a root of unity, then take a \( t \) such that \( (\mu/\lambda)^t = 1 \). Then

\[
\text{Holant}_{\Omega_{st}} = \lambda^{nst} \sum_{i+j+k+\ell=n} s^{\ell} \left(\frac{t^\ell c_{ijkl}}{\lambda^\ell2^{j+\ell}}\right).
\]

For \( s \geq 1 \), this gives a coefficient matrix \( \left(\left(\frac{st}{2\lambda}\right)^\ell\right) \) (with rows indexed by \( s \) and columns indexed by the tuple \((i, j, k, \ell))\) that is Vandermonde. Although this system is not full rank, we can replace all the unknowns \( c_{ijkl}/2^j \) having \( i+j+k = n-\ell \) by their sum to form a new unknown \( c'_{\ell} = \sum_{i+j+k=n-\ell} c_{ijkl}/2^j \), where \( 0 \leq \ell \leq n \). The unknown \( c'_{0} \) is the Holant of \( \Omega \) that we seek. The resulting Vandermonde system

\[
\text{Holant}_{\Omega_{st}} = \lambda^{nst} \sum_{\ell=0}^{n} \left(\frac{st}{2\lambda}\right)^\ell c'_{\ell}
\]

has full rank, so we can solve for the unknowns in polynomial time and obtain the value of \( c'_{0} = \sum_{i+j+k=n} \frac{c_{ijkl}}{2^j} \).

If \( \mu/\lambda \) is not a root of unity, then we replace all the unknowns \( c_{ijkl}/(\lambda^\ell2^{j+\ell}) \) having \( i+j = m \) with their sum to form new unknowns \( c'_{m\ell \ell} \) for any \( 0 \leq m, k, \ell \) and \( m+k+\ell = n \). Then by (7.6) and (7.7),

\[
\text{Holant}_\Omega = \sum_{m+k=n} c'_{m\ell 0} \quad \text{and} \quad \text{Holant}_{\Omega_s} = \lambda^{ns} \sum_{m+k+\ell=n} \left(\frac{\mu}{\lambda}\right)^{ks} s^{\ell} c'_{m\ell \ell}.
\]

After a suitable ordering of the columns, the matrix of coefficients whose rows are indexed by \( s \) satisfies the hypothesis of Lemma 7.22. Therefore, the linear system has full rank. We can solve for the unknowns in polynomial time and obtain the value of \( \text{Holant}_\Omega \).
3. \( b_1 = b_2 = 1 \). In this case, we have \( \lambda_1 = \lambda_2 = \lambda_3 \), denoted by \( \lambda \). We have

\[
\Lambda^s = \begin{bmatrix}
\lambda^s & s\lambda^{s-1} & \frac{s(s-1)}{2}\lambda^{s-2} \\
0 & \lambda^s & s\lambda^{s-1} \\
0 & 0 & \lambda^s
\end{bmatrix}, \quad \text{and} \quad \psi(\Lambda^s) = \begin{bmatrix}
\lambda^s & \frac{s}{2}\lambda^{s-1} & \frac{s(s-1)}{2}\lambda^{s-2} \\
0 & \frac{s}{2}\lambda^s & \frac{s(s-1)}{2}\lambda^{s-1} \\
0 & 0 & \frac{s}{2}\lambda^s
\end{bmatrix}.
\]

We only need to consider the assignments to \( \Lambda^s \) that assign

- \((0, 0)\) or \((2, 2)\) \( i \) many times,
- \((1, 1)\) \( j \) many times,
- \((0, 1)\) \( k \) many times,
- \((1, 2)\) \( \ell \) many times,
- \((0, 2)\) \( m \) many times

with \( i + j + k + \ell + m = n \), since any other assignment contributes 0 to the Holant sum. Let \( c_{ijklm} \) be the sum over all such assignments of the products of evaluations (excluding \( \Lambda^s \) but including the contributions from \( T \) and \( T^{-1} \)) on \( \Omega_s \); \( c_{ijklm} \) is independent of \( s \). Then

\[
\text{Holant}_{\Omega} = \sum_{i+j=0} c_{ij000} \frac{2^j}{2^i}
\]

and the value of the Holant on \( \Omega_s \), for \( s \geq 1 \), is

\[
\text{Holant}_{\Omega_s} = \sum_{i+j+k+\ell+m=0} \lambda^{i+j+s} \left( s\lambda^{s-1} \right)^{k+\ell} \left( s(s-1)\lambda^{s-2} \right)^m \left( \frac{c_{ijklm}}{2^{i+k+m}} \right)
= \lambda^{ns} \sum_{i+j+k+\ell+m=0} s^{k+\ell+m} (s-1)^m \left( \frac{c_{ijklm}}{\lambda^{k+\ell+2m} 2^{i+k+m}} \right).
\]

We replace all the unknowns \( c_{ijklm}/(\lambda^{k+\ell+2m} 2^{i+k+m}) \) having \( i + j = p \) and \( k + \ell = q \) with their sum to form new unknowns \( c'_{pqm} \), for any \( 0 \leq p, q, m \) and \( p + q + m = n \). The Holant of \( \Omega \) is now \( c'_{000} \). This new linear system is

\[
\text{Holant}_{\Omega_s} = \lambda^{ns} \sum_{p+q+m=0} s^{q+m} (s-1)^m c'_{pqm}.
\]

But this linear system is still rank deficient. We now index the columns by the tuple \((q, m)\), where \( q \geq 0, m \geq 0 \), and \( q + m \leq n \). Correspondingly, we rename the variables \( x_{q,m} = c'_{pqm} \). Note that \( p = n - q - m \) is determined by \((q, m)\). Observe that the column indexed by \((q, m)\) is the sum of the columns indexed by \((q - 1, m)\) and \((q - 2, m + 1)\) provided \( q - 2 \geq 0 \). Namely, \( s^{q+m}(s-1)^m = s^{q-1+m}(s-1)^m + s^{q-2+m+1}(s-1)^{m+1} \). Of course this is only meaningful if \( q \geq 2 \), \( m \geq 0 \) and \( q + m \leq n \). Let \( a_{s,(q,m)} = s^{q+m}(s-1)^m \). Then the linear system is

\[
\sum_{q \geq 0, m \geq 0, q + m \leq n} a_{s,(q,m)} x_{q,m} = \frac{\text{Holant}_{\Omega_s}}{\lambda^{ns}}.
\]
For any fixed $s \geq 1$, we write $\alpha_{q,m} = a_{s,(q,m)}$, then $\alpha_{q,m} x_{q,m} = (\alpha_{q-1,m} + \alpha_{q-2,m+1}) x_{q,m}$. Therefore we can eliminate variable $x_{q,m}$ by adding its value to $x_{q-1,m}$ and to $x_{q-2,m+1}$. More precisely we define new variables

\[
x_{q-1,m} \leftarrow x_{q-1,m} + x_{q,m}
\]
\[
x_{q-2,m+1} \leftarrow x_{q-2,m+1} + x_{q,m}
\]

for every $0 \leq m \leq n - 2$ and for $q$ from $n - m$ down to 2. The order of these updates is important. Once a variable $x_{q,m}$ has been added to $x_{q-1,m}$ and $x_{q-2,m+1}$, there should not be any update involving $x_{q,m}$, which has been eliminated. One valid sequence is a double nested loop, with the outer loop being “For $m = 0$ to $n - 2$” and for each $m$ in the range there is the inner loop “For $q = n - m$ down to 2”. This is illustrated schematically for $n = 6$ in Figure 7.4.

![Figure 7.4](image)

Figure 7.4: A triangular table of variables and their update patterns, for $n = 6$. Rows are indexed by $q$ and columns are indexed by $m$.

Observe that in each update, some variable $x_{q,m}$ is added to a variable $x_{q',m'}$, where $q' < q$ (a strictly lower index value of $q$), and $m' \geq m$ (an index value that is at least $m$). Thus in the order by the double nested loop, once $x_{q,m}$ has been added to $x_{q-1,m}$ and $x_{q-2,m+1}$, there will be no more updates involving $x_{q,m}$.

A more crucial observation is that the column in the linear system (7.8) indexed by the tuple $(q, m) = (0, 0)$ is never changed. This is because, in order to be an updated entry, there must be some $q \geq 2$ and $m \geq 0$ such that $(q - 1, m) = (0, 0)$ or $(q - 2, m + 1) = (0, 0)$, which is clearly impossible. Hence $x_{0,0} = c'_{000}$ is still the Holant value on $\Omega$. The $2n + 1$ unknowns that remain are

$x_{0,0}$, $x_{1,0}$, $x_{0,1}$, $x_{1,1}$, $x_{0,2}$, $x_{1,2}$, $\ldots$, $x_{0,n-1}$, $x_{1,n-1}$, $x_{0,n}$

and their coefficients in row $s$ are

$1, s, s(s - 1), s^2(s - 1), s^2(s - 1)^2, \ldots, s^{n-1}(s - 1)^{n-1}, s^n(s - 1)^{n-1}, s^n(s - 1)^n$.
For $0 \leq \delta \leq 2n$, it is clear that the $\delta$-th entry in this row is a monic polynomial in $s$ of degree $\delta$, and thus $s^\delta$ is a linear combination of the first $\delta$ entries. It follows that the coefficient matrix is a product of the standard Vandermonde matrix multiplied to its right by an upper triangular matrix with all 1's on the diagonal. Hence the matrix is nonsingular, and we can solve the linear system in polynomial time and, in particular, compute $c'_{n00}$.

For an asymmetric signature, we often want to reorder the input bits under a circular permutation. For a single counterclockwise rotation of $90^\circ$, the effect on the entries of the signature matrix of an arity 4 signature is given in Figure 7.5.

Figure 7.5: The movement of the entries in the signature matrix of an arity 4 signature under a counterclockwise rotation of the input edges. The Hamming weight one entries are in the dotted cycle, the Hamming weight two entries are in the two solid cycles (one has length 4 and the other one is a swap), and the entries of Hamming weight three are in the dashed cycle.

We will derive most of our #P-hardness results through Lemma 7.23 and the following Lemma 7.24. This is done by a reduction from the problem of counting Eulerian orientations on 4-regular graphs, which is the Holant problem $\text{Holant}((\neq 2) \mid [0, 0, 1, 0, 0])$. In Theorem 6.29 we have shown that this problem $\#EO$ is #P-hard even for 4-regular planar graphs. Recall that under a holographic transformation by $[\frac{1}{2} \frac{1}{2}]$, this bipartite Holant problem becomes the Holant problem $\text{Holant}([1, 0, 1/3, 0, 1])$, up to a nonzero factor.

**Lemma 7.24.** Let $g$ be the arity 4 signature with $M_g$ given in Equation (7.5) so that $\tilde{M}_g = I_3$. Then $\text{Holant}(g)$ is #P-hard. In fact $\text{PI-Holant}(g)$ is #P-hard.

**Proof.** We reduce from the $\#EO$ problem $\text{Holant}(\theta)$, where $\theta = [1, 0, 1/3, 0, 1]$, which is #P-hard by Theorem 6.29 even for 4-regular planar graphs. We achieve this via an arbitrarily close approximation using the recursive construction in Figure 7.6 with $g$ assigned to every vertex.
Figure 7.6: Recursive construction to approximate $[1, 0, \frac{1}{3}, 0, 1]$. The vertices are assigned $g$.

We claim that the signature matrix $M_{N_k}$ of gadget $N_k$ is

$$M_{N_k} = \begin{bmatrix}
1 & 0 & 0 & a_k \\
0 & a_{k+1} & a_{k+1} & 0 \\
0 & a_{k+1} & a_{k+1} & 0 \\
a_k & 0 & 0 & 1
\end{bmatrix}.$$  \hspace{1cm} (7.9)

where $a_k = \frac{1}{3} - \frac{1}{3} \left(\frac{-1}{2}\right)^k$. This is true for $N_0$. Inductively assume $M_{N_k}$ has this form. Then the rotated form of the signature matrix for $N_k$, as described in Figure 7.5, is

$$\begin{bmatrix}
1 & 0 & 0 & a_{k+1} \\
0 & a_k & a_{k+1} & 0 \\
0 & a_{k+1} & a_k & 0 \\
a_{k+1} & 0 & 0 & 1
\end{bmatrix}.$$  \hspace{1cm} (7.9)

The action of $g$ on the far right side of $N_{k+1}$ in Figure 7.6 is to replace each of the middle two entries in the middle two rows of the matrix in (7.9) with their average, $(a_k + a_{k+1})/2 = a_{k+2}$. This gives $M_{N_{k+1}}$.

Let $G$ be a graph with $n$ vertices and $H_\theta$ (resp. $H_{N_k}$) be the Holant value on $G$ with all vertices assigned $\theta$ (resp. $N_k$). Since each signature entry in $\theta$ can be expressed as a rational number with denominator 3, each term in the sum of $H_\theta$ can be expressed as a
rational number with denominator $3^n$, and $H_\phi$ itself is a sum of $2^{2n}$ such terms, where $2n$ is the number of edges in $G$. If the error $|H_{N_k} - H_\phi|$ is at most $1/3^{n+1}$, then we can recover $H_\phi$ from $H_{N_k}$ by selecting the nearest rational number to $H_{N_k}$ with denominator $3^n$.

For each signature entry $x$ in $M_\phi$, its corresponding entry $\tilde{x}$ in $M_{N_k}$ satisfies $|\tilde{x} - x| \leq x/2^k$. Then for each term $t$ in the Holant sum $H_\phi$, its corresponding term $\tilde{t}$ in the sum $H_{N_k}$ satisfies $t(1 - 1/2^k)^n \leq \tilde{t} \leq t(1 + 1/2^k)^n$, thus $-t(1 - (1 - 1/2^k)^n) \leq \tilde{t} - t \leq t((1 + 1/2^k)^n - 1)$. Since $1 - (1 - 1/2^k)^n \leq (1 + 1/2^k)^n - 1$, which is easy to see by their binomial expansions, we get $|\tilde{t} - t| \leq t((1 + 1/2^k)^n - 1)$. Also each term $t \leq 1$. Hence

$$|H_{N_k} - H_\phi| \leq 2^{2n}((1 + 1/2^k)^n - 1) < 1/3^{n+1},$$

if we take $k = 4n$, as $(1 + x)^n - 1 < xn(1 + x)^n < xe^n$, for $x = 1/2^4n$. \qed

We summarize our progress with the following corollary of Lemmas 7.23 and 7.24.

**Corollary 7.25.** Let $f$ be an arity 4 signature with complex weights. If $M_f$ is redundant and $\tilde{M}_f$ is nonsingular, then Holant($f$) is $\#P$-hard. In fact Pl-Holant($f$) is $\#P$-hard.

**Exercise:** By Lemma 7.24, Pl-Holant($g$) is $\#P$-hard, where $g$ is defined in Equation (7.5). The reduction chain is from the $\#P$-hardness of $T(G; 3, 3)$ of the Tutte polynomial to $\#EO$ (by Theorem 6.29), then to Pl-Holant($g$) (by Lemma 7.24).

Prove this directly from the $\#P$-hardness of $T(G; 3, 3)$, skipping $\#EO$. (Hint: First apply a rotation (Figure 7.5) to $2g$, then interpolate Pl-Holant($\hat{f}$) where $\hat{f}$ is defined in (6.18).)

In order to make Corollary 7.25 more applicable, we show that for an arity 4 signature $f$, the redundancy of $M_f$ and the nonsingularity of $\tilde{M}_f$ are invariant under an invertible holographic transformation.

**Lemma 7.26.** Let $f$ be an arity 4 signature with complex weights, $T \in \mathbb{C}^{2 \times 2}$ a matrix, and $f = T \otimes 4 f$. If $M_f$ is redundant, then $M_{\tilde{f}}$ is also redundant and

$$\det(\varphi(M_f)) = \det(\varphi(M_{\tilde{f}})) \det(T)^6.$$  

**Proof.** Since $\tilde{f} = T \otimes 4 f$, we can express $M_{\tilde{f}}$ in terms of $M_f$ and $T$ as

$$M_{\tilde{f}} = T \otimes 2 M_f (T^T) \otimes 2.$$  

This can be directly checked. Let $T = (t_{ij}^a)$, where row index $i$ and column index $j$ are from $\{0, 1\}$. The entry $f(a, b, c, d)$ appears in $M_f$ at row $(a, b) \in \{0, 1\}^2$ and column $(d, c) \in \{0, 1\}^2$ in lexicographic order. The matrix $T \otimes 2$ has the form $(t_{ij}^a t_{ij}^b)$ at row $(a, b)$ and column $(a', b')$ in lexicographic order. Similarly $(T^T) \otimes 2$ has the form $(t_{ij}^c t_{ij}^d)$ at row $(c, d)$ and column $(c', d')$. Thus row $(a, b)$ and column $(d, c)$ of $T \otimes 2 M_f (T^T) \otimes 2$ has entry

244
\[
\sum_{a',b',c',d' \in \{0,1\}} t_{a'}^{a} t_{b'}^{b} f(a',b',c',d') t_{d'}^{d} t_{c'},
\]
which is the entry \( \hat{f}(a,b,c,d) \) placed at row \((a,b)\) and column \((d,c)\) of \( M_f \).

An alternative, but essentially the same, proof of (7.10) is by linearity, and observing that it is valid for the basis vectors \( f = e_a \otimes e_b \otimes e_c \otimes e_d \), for \( a, b, c, d \in \{0, 1\} \).

Let \( E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \). Then \( X \in \text{RM}_4(\mathbb{C}) \) iff \( EX = X = XE \). Then it follows that \( M_f \in \text{RM}_4(\mathbb{C}) \) if \( M_f \in \text{RM}_4(\mathbb{C}) \), since \( ET^{\otimes 2}E = T^{\otimes 2} \), i.e., a simultaneous row flip \( ab \leftrightarrow ba \) and column flip \( a'b' \leftrightarrow b'a' \) keep \( T^{\otimes 2} \) unchanged. For the two matrices \( A \) and \( B \) in the definition of \( \varphi \), we note that \( BA = M_g \), where \( M_g \) given in Equation (7.5) is the identity element of the semi-group \( \text{RM}_4(\mathbb{C}) \). Since \( M_f \in \text{RM}_4(\mathbb{C}) \), we have \( BAM_f = M_f = M_fBA \). Then we have

\[
\varphi(M_f) = AM_fB = A \left( T^{\otimes 2}M_f (T^T)^{\otimes 2} \right) B
= (AT^{\otimes 2}B)(AM_fB)(A (T^T)^{\otimes 2} B)
= \varphi(T^{\otimes 2})\varphi(M_f)\varphi((T^T)^{\otimes 2}).
\]

(7.11)

Another direct calculation shows that

\[
\det(\varphi(T^{\otimes 2})) = \det(T)^3 = \det(\varphi((T^T)^{\otimes 2})).
\]

(7.12)

To verify (7.12) one can save some calculation by noticing that for any \( T_1 \) and \( T_2 \), \( \varphi((T_1T_2)^{\otimes 2}) = AT^{\otimes 2}_1T^{\otimes 2}_2B \) by definition, and \( AT^{\otimes 2}_1 \) has identical two middle columns, thus \( AT^{\otimes 2}_1 = AT^{\otimes 2}_1BA \), thus \( \varphi \) is multiplicative on \( T^{\otimes 2} \), i.e., \( \varphi((T_1T_2)^{\otimes 2}) = \varphi(T^{\otimes 2}_1)\varphi(T^{\otimes 2}_2) \). Note that (7.12) is easy to check for triangular matrices.

Thus, by applying determinant to both sides of Equation (7.11), we have

\[
\det(\varphi(M_f)) = \det(\varphi(M_f)) \det(T)^6
\]
as claimed.

In particular, for a nonsingular matrix \( T \in \mathbb{C}^{2 \times 2} \), \( M_f \) is redundant and \( \tilde{M}_f \) is nonsingular iff \( M_f \) is redundant and \( \tilde{M}_f \) is nonsingular. From Corollary 7.25 and Lemma 7.26 we have the following corollary.

**Corollary 7.27.** Let \( f \) be an arity 4 signature with complex weights. If there exists a nonsingular matrix \( T \in \mathbb{C}^{2 \times 2} \) such that \( \hat{f} = T^{\otimes 4}f \), where \( M_f \) is redundant and \( \tilde{M}_f \) is nonsingular, then \( \text{Holant}(f) \) is \#P-hard. In fact \( \#\text{P}-\text{Holant}(f) \) is \#P-hard.

The following lemma applies Corollary 7.25.

**Lemma 7.28.** Let \( f_k = ck^k + dk^k \), where \( c \neq 0 \) and \( 0 \leq k \leq 4 \). Then the problem \( \text{Holant}([f_0, f_1, f_2, f_3, f_4]) \) is \#P-hard unless \( \lambda = \pm i \), in which case \([f_0, f_1, f_2, f_3, f_4] \) is a vanishing signature and the Holant vanishes.
Figure 7.7: The tetrahedron gadget. Each vertex is assigned $\hat{f} = [t, 1, 0, 0, 0]$.

Proof. If $\lambda = \pm i$, then $r_d^\pm(f) = 1$, $v_d^\pm(f) = 3$, and so $f = [f_0, f_1, f_2, f_3, f_4]$ is vanishing by Theorem 7.13. Otherwise, a holographic transformation with the orthogonal matrix $T = \frac{1}{\sqrt{1 + \lambda^2}} \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix}$ transforms $f$ to $\hat{f} = [t, 1, 0, 0, 0]$ for some $t \in \mathbb{C}$ after normalizing the second entry.

This orthogonal transformation can be directly verified; but a more informative derivation is as follows: Consider a symmetric signature $f = [f_0, \ldots, f_n]$ of general arity $n$, such that $f_k = c_k \lambda^{k-1} + d_k \lambda^k$, where $c_k \neq 0$, and $\lambda \neq \pm i$. Let $S = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \end{bmatrix}$. Note that $\det S = c \neq 0$.

Then the signature $f$ can be expressed as $f = S^{\otimes n} [1, 1, 0, \ldots, 0]$, where $[1, 1, 0, \ldots, 0]$ should be understood as a column vector of dimension $2^n$, which has a 1 in entries with index weight at most one and 0 elsewhere. This identity can be verified by observing that

$$[1, 1, 0, \ldots, 0] = [1, 0]^{\otimes n} + \frac{1}{(n-1)!} \text{Sym}_{n-1}([1, 0]; [0, 1])$$

Let $T = \frac{1}{\sqrt{1 + \lambda^2}} \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix}$, then $T = T^\top T = T^{-1} \in O_2(\mathbb{C})$ is orthogonal, and $R = TS = [u \ w]$ is upper triangular, where $v, w \in \mathbb{C}$ and $u = \sqrt{1 + \lambda^2} \neq 0$. However, $\det R = \det T \det S = (-1)c \neq 0$, so we also have $v \neq 0$. It follows that

$$T^{\otimes n} f = (TS)^{\otimes n} [1, 1, 0, \ldots, 0] = R^{\otimes n} [1, 1, 0, \ldots, 0] = R^{\otimes n} \left( [1, 0]^{\otimes n} + \frac{1}{(n-1)!} \text{Sym}_{n-1}([1, 0]; [0, 1]) \right) = [u, 0]^{\otimes n} + \frac{1}{(n-1)!} \text{Sym}_{n-1}([u, 0]; [w, v]) = [u^n + nu^{n-1}w, u^{n-1}v, 0, \ldots, 0].$$

Since $u^{n-1}v \neq 0$, we can normalize the entry of Hamming weight one to 1 by a scalar multiplication. Thus, we have $[t, 1, 0, \ldots, 0]$ for some $t \in \mathbb{C}$. 246
Using the tetrahedron gadget in Figure 7.7 with \( \hat{f} \) assigned to each vertex, we have a gadget with signature
\[
h = [t^4 + 6t^2 + 3, t^3 + 3t, t^2 + 1, t, 1].
\]
One way to verify this signature is to treat \([t, 1, 0, 0, 0]\) as a weighted matching signature, and compute the signature of the gadget similarly to a matchgate. Thus, for example, the Hamming weight 0 entry has a contribution \( t^4 \) from the empty matching of the six internal edges, a contribution \( 6t^2 \) from the six matchings of a single edge, and a contribution 3 from the three perfect matchings. The other entries can be computed similarly.

Since the determinant of \( \tilde{M}_h \) is 4, the compressed signature matrix of this gadget is nonsingular, so we are done by Corollary 7.25. \( \square \)

Notice that Lemma 7.28 does not address the complexity of \( \text{Pl-Holant}([f_0, f_1, f_2, f_3, f_4]) \). In particular \( \text{Pl-Holant}([0, 1, 0, 0, 0]) \) is solvable in polynomial time, since \([0, 1, 0, 0, 0]\) is the Perfect Matching signature. Even though Corollary 7.25 applies to Pl-Holant problems as well, the reduction in Lemma 7.28 is nonplanar, as the tetrahedron gadget in Figure 7.7 is not planar. We will show that \( \text{Pl-Holant}([t, 1, 0, 0, 0]) \) remains \#P-hard for all \( t \neq 0 \).

Now we are ready to prove a dichotomy for a single arity 4 signature. For future convenience we state the following theorem for both the case of an arity 3 and arity 4 signature. The arity 3 case is essentially a restatement of Theorem 6.3. Theorem 7.29 does not address the complexity of Pl-Holant problems; Theorem 8.12 in Chapter ?? will be the version of Theorem 7.29 for Pl-Holant problems.

**Theorem 7.29.** If \( f \) is a non-degenerate, symmetric, complex-valued signature of arity 3 or arity 4 in Boolean variables, then \( \text{Holant}(f) \) is \#P-hard unless \( f \) is \( \mathcal{A} \)-transformable, or \( \mathcal{P} \)-transformable, or vanishing, in which case the problem is in \( \mathcal{P} \).

**Proof.** We first consider \( f \) of arity 3. We show that the tractability conditions in Theorem 6.3 are equivalent to the statement that \( f \) is \( \mathcal{A} \)- or \( \mathcal{P} \)-transformable, or vanishing, for any symmetric signature of arity 3. Clearly conditions 1 and 2 in Theorem 6.3 imply that \( f \) is \( \mathcal{A} \)- or \( \mathcal{P} \)-transformable, and condition 3 is equivalent to \( f \) is vanishing by Lemma 7.17.

Suppose \( f \) is \( \mathcal{P} \)-transformable. Then for some \( T, (\equiv_2)T^{\otimes 2} \in \mathcal{P} \), and \((T^{-1})^{\otimes 3}f \in \mathcal{P} \). Since \((T^{-1})^{\otimes 3}f \) is symmetric and has arity 3, either it is degenerate, which would imply \( f \) is degenerate, or \((T^{-1})^{\otimes 3}f = [\begin{smallmatrix} a \\ 0 \\ 0 \end{smallmatrix}]^{\otimes 3} + [\begin{smallmatrix} 0 \\ b \\ 0 \end{smallmatrix}]^{\otimes 3} \), for some \( ab \neq 0 \). Let \( D = [\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}] \), then \( D \in \text{Stab}(\mathcal{P}) \). Then \((TD)^{-1})^{\otimes 3}f = (\equiv_3) \) and \((\equiv_2)(TD)^{\otimes 2} \in \mathcal{P} \).

Suppose \( f \) is \( \mathcal{A} \)-transformable and non-degenerate. Then for some \( T, (\equiv_2)T^{\otimes 2} \in \mathcal{A} \), and \((T^{-1})^{\otimes 3}f \in \mathcal{A} \). Since \((T^{-1})^{\otimes 3}f \) is symmetric, it is in \( \mathcal{F}_{123} \). By the form of \( \mathcal{F}_{123} \), there exists a matrix \( M \in \{ I, [\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}], [\begin{smallmatrix} 1 & -1 \\ 1 & -1 \end{smallmatrix}] \} \) such that \((TM)^{-1})^{\otimes 3}f = \lambda \{ [\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}]^{\otimes 3} + i^r [\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}]^{\otimes 3} \} \), for some \( \lambda \neq 0 \) and \( r \in \{ 0, 1, 2, 3 \} \), and \((\equiv_2)(TM)^{\otimes 2} \in \mathcal{A} \) since \( M \in \text{Stab}(\mathcal{A}) \). We take \( \rho = i^r \), and \( D = \lambda^4 \begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \in \text{Stab}(\mathcal{A}) \). Note that \( \rho^3 = i^r \). We conclude that \((\equiv_2)(TM)D^{\otimes 2} \in \mathcal{A} \) and \((TMD)^{-1})^{\otimes 3}f = (\equiv_3) \). Hence \( f \) satisfies condition 2 of Theorem 6.3.
Now we prove the dichotomy for a single arity 4 signature. Let $f = [f_0, f_1, f_2, f_3, f_4]$. If the compressed signature matrix $\widetilde{M}_f$ is nonsingular, then $\text{Holant}(f)$ is $\#P$-hard by Corollary 7.25, so assume that the rank of $\widetilde{M}_f$ is at most 2. Then we have

$$a \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} + 2b \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} + c \begin{pmatrix} f_2 \\ f_3 \\ f_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

for some $a, b, c \in \mathbb{C}$ that are not all zero. If $a = c = 0$, then $b \neq 0$, so $f_1 = f_2 = f_3 = 0$. In this case, $f$ is a generalized equality Gen-Eq, then $f \in \mathcal{P}$ and in particular $f$ is $\mathcal{P}$-transformable. Now suppose $a$ and $c$ are not both 0. Then $f$ satisfies a second order recurrence relation. If the roots of the characteristic polynomial of the recurrence relation are distinct, then $f_k = \alpha^{4-k} \beta^k + \gamma^{4-k} \delta^k \ (0 \leq k \leq 4)$, for some $\alpha, \beta, \gamma$ and $\delta$, where $\alpha \delta - \beta \gamma \neq 0$. Let $T_1 = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$, then $f = (=4)T_1^{\otimes 4}$, and

$$\text{Holant}(f) \equiv_T \text{Holant}((=4)T_1^{\otimes 4} | (=2)) \equiv_T \text{Holant}((=4) | T_1^{\otimes 2}(=2)).$$

By Theorem 5.60 this problem is $\#P$-hard except when there exists some $T_2$ such that for $\mathcal{C} \in \{\mathcal{P}, \mathcal{A}\}$, $(=4)(T_2^{-1})^{\otimes 4} \in \mathcal{C}$ and $(T_2 T_1)^{\otimes 2}(=2) \in \mathcal{C}$. Let $T = (T_2 T_1)^{-1}$, then $f T^{\otimes 4} \in \mathcal{C}$ and $(T^{-1})^{\otimes 2}(=2) \in \mathcal{C}$. Hence $f$ is either $\mathcal{A}$- or $\mathcal{P}$-transformable. Note that Theorem 5.60 is the transformational form of Theorem 5.3, which is equivalent to Theorem 5.1. The case $k = 4$ of Theorem 5.1 is proved in Chapter 5.

Now suppose the characteristic polynomial has a double root $\lambda$ and there are two cases. In the first, for any $0 \leq k \leq 4$, $f_k = c k \lambda^{k-1} + d \lambda^k$, where $c \neq 0$. In the second, for any $0 \leq k \leq 4$, $f_k = c(4-k)\lambda^3 + d \lambda^{4-k}$, where $c \neq 0$. This case is mapped to the first case by a holographic transformation by $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then we are done by Lemma 7.28.

### 7.6 Vanishing Signatures Revisited

#### 7.6.1 Orientation with a Pair from $\mathcal{V}^+$ and $\mathcal{V}^-$

We prove a lemma on an orientation problem with a pair of vanishing signatures of the opposite type. The proof uses similar techniques to those in Section 7.5. This problem is also called $\#\text{ONE-IN-OR-ONE-OUT-ORIENTATION}$ in Section 7.3 where a set of such signatures are allowed.

**Lemma 7.30.** If $f = [0, 1, 0, \ldots, 0]$ and $g = [0, \ldots, 0, 1, 0]$ are both of arity $n \geq 3$, then the problem $\text{Holant}(|\neq 2| \{f, g\})$ is $\#P$-hard.

**Proof.** Our goal is to obtain a signature that satisfies the hypothesis of Corollary 7.27.
The circle is assigned $f$, the triangle is assigned $g$, and the squares are assigned $(\neq 2)$.

(b) The circle is assigned $h'$, the triangle is assigned $h''$, and the squares are given $(\neq 2)$.

Figure 7.8: Gadget constructions used in Lemma 7.30. Figure 7.8a constructs $h$. Its rotated versions $h'$ and $h''$ are used in Figure 7.8b to construct $r$.

The gadget in Figure 7.8a, with $f$ assigned to the circle vertex, $g$ assigned to the triangle vertex, and $(\neq 2)$ assigned to the square vertices, has signature $h$ with signature matrix

$$M_h = \begin{bmatrix} 0 & 0 & 0 & m \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where $m = n - 2$ is positive since $n \geq 3$. Although this signature matrix is redundant, its compressed form is singular. Rotating this gadget 90° clockwise and 90° counterclockwise yield signatures $h'$ and $h''$ respectively, with signature matrices

$$M_{h'} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & m & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad M_{h''} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & m & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

The gadget in Figure 7.8b, with $h'$ assigned to the circle vertex, $h''$ assigned to the triangle vertex, and $(\neq 2)$ assigned to the square vertices, has a signature $r$ with signature matrix

$$M_r = M_{h'} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} M_{h''} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & m & m^2 + 1 & 0 \\ 0 & 1 & m & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Note that the effect of the $(\neq 2)$ signatures is to reverse all four rows of $M_{h''}$ before multiplying it to the right of $M_{h'}$. Although this signature matrix is not redundant, every entry of Hamming weight two is nonzero since $m$ is positive.

Now we claim that we can use $r$ to interpolate the following signature $R$, for any nonzero value $t \in \mathbb{C}$, via the construction in Figure 7.9. Define $p^\pm = (m \pm \sqrt{m^2 + 4})/2$, $P = \begin{bmatrix} 1 & 1 \\ p^+ & p^- \end{bmatrix}$, and $T = P \begin{bmatrix} 1 & 0 \\ t & t^{-1} \end{bmatrix} P^{-1}$ where $t \in \mathbb{C}$ is any nonzero value. We define the signature $R$ by its
Figure 7.9: Recursive construction to interpolate a signature $R$ that is only a rotation away from having a redundant signature matrix and nonsingular compressed matrix. The circles are assigned $r$ and the squares are assigned $(\not= 2)$.

The signature matrix

$$M_R = \begin{bmatrix} 0 & 0 & 1 \\ 0 & T & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \quad (7.13)$$

Consider an instance $\Omega$ of Holant($\not= 2 | \mathcal{F} \cup \{R\}$). Suppose that $R$ appears $n$ times in $\Omega$. We construct from $\Omega$ a sequence of instances $\Omega_s$ of Holant($\not= 2 | \mathcal{F}$) indexed by $s \geq 1$. We obtain $\Omega_s$ from $\Omega$ by replacing each occurrence of $R$ with the gadget $N_s$ in Figure 7.9 with $r$ assigned to the circle vertices and $(\not= 2)$ assigned to the square vertices. In $\Omega_s$, the edge corresponding to the $i$th significant index bit of $N_s$ connects to the same location as the edge corresponding to the $i$th significant index bit of $R$ in $\Omega$.

The signature matrix of $N_s$ is the $s$th power of the matrix obtained from $M_r$ after reversing all rows, and then switching the first and last rows of the final product, namely

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0^s \\ 0 & 1 & 0 & m \\ 0 & 0 & m & m^2 + 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & m & 0 \\ 0 & m & m^2 + 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}^s \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & m & 0 \\ 0 & m & m^2 + 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{s-1}.$$

The twist of the two input edges on the left side for the first copy of $M_r$ switches the middle two rows, which is equivalent to a total reversal of all rows, followed by the switching of the first and last rows. The total reversals of rows for all subsequent $s - 1$ copies of $M_r$ are due to the presence of $(\not= 2)$ signatures.

After such reversals of rows, it is clear that the matrix is a direct sum of block matrices indexed by $\{00, 11\} \times \{00, 11\}$ and $\{01, 10\} \times \{10, 01\}$. Furthermore, in the final product,
the block indexed by \( \{00, 11\} \times \{00, 11\} \) is \( [0 1] \). Thus in the gadget \( N_s \), the only entries of \( M_{N_s} \) that vary with \( s \) are the four entries in the middle. These middle four entries of \( M_{N_s} \) form the 2-by-2 matrix \( \begin{bmatrix} \frac{1}{m} & m \\ m & m^2 + 1 \end{bmatrix}^s \). Since \( \begin{bmatrix} \frac{1}{m} & m \\ m & m^2 + 1 \end{bmatrix} = P \begin{bmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{bmatrix} P^{-1} \), where \( \lambda_\pm = (m^2 + 2 \pm m\sqrt{m^2 + 4})/2 \) are the eigenvalues, we have

\[
\begin{bmatrix} 1 & m \\ m & m^2 + 1 \end{bmatrix}^s = P \begin{bmatrix} \lambda_+^s & 0 \\ 0 & \lambda_-^s \end{bmatrix} P^{-1}.
\]

As \( \det \begin{bmatrix} \frac{1}{m} & m \\ m & m^2 + 1 \end{bmatrix} = 1 \), \( \lambda_+ \lambda_- = 1 \), so the eigenvalues are nonzero. Since \( m \) is positive, \( \lambda_+ > \lambda_- \) and so neither \( \lambda_+ \) nor \( \lambda_- \) is a root of unity.

Now we determine the relationship between Holant\(_\Omega \) and Holant\(_{\Omega_s} \). We can view our construction of \( \Omega_s \) as first replacing \( M_R \) with the right-hand side of

\[
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},
\]

which does not change the Holant value, and then replacing the 4-by-4 matrix in the middle with the matrix

\[
\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & \lambda_+^s & 0 & 0 \\ 0 & 0 & \lambda_-^s & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.
\]

We stratify the assignments in \( \Omega \) based on the assignments to the \( n \) occurrences of the signature matrix

\[
\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & t & 0 & 0 \\ 0 & 0 & t^{-1} & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.
\] (7.14)

The inputs to this matrix are from \( \{0,1\}^2 \times \{0,1\}^2 \), which correspond to the four input bits. Recall the way rows and columns of a signature matrix are ordered from Definition 1.4. Thus, e.g., the entry \( t \) corresponds to the cyclic input bit pattern 0110 in counterclockwise order. We only need to consider the assignments that assign

- \( i \) many times the bit pattern 0110,
- \( j \) many times the bit pattern 1001, and
- \( k \) many times the bit patterns 0011 or 1100,

such that \( i + j + k = n \), since any other assignment contributes 0 to the Holant sum. Let \( c_{ij} \) be the sum over all such assignments of the product of evaluations (excluding (7.14),
but including the contributions from the block matrices containing $P$ and $P^{-1}$ on $\Omega$. Then

$$\text{Holant}_\Omega = \sum_{i+j+k=n} t^{i-j} c_{ijk}$$

and the value of the Holant on $\Omega_s$, for $s \geq 1$, is

$$\text{Holant}_{\Omega_s} = \sum_{i+j+k=n} \lambda_+^{si} \lambda_+^{sj} c_{ijk} = \sum_{i+j+k=n} \lambda_+^{s(i-j)} c_{ijk}.$$  

This Vandermonde system does not have full rank. However, we can define for $-n \leq \ell \leq n$,

$$c'_\ell = \sum_{i+j=\ell} c_{ijk}.$$  

Then the Holant of $\Omega$ is

$$\text{Holant}_\Omega = \sum_{-n \leq \ell \leq n} t^\ell c'_\ell$$

and the Holant of $\Omega_s$ is

$$\text{Holant}_{\Omega_s} = \sum_{-n \leq \ell \leq n} \lambda_+^{s\ell} c'_\ell.$$  

Now this Vandermonde has full rank because $\lambda_+$ is neither zero nor a root of unity. Therefore, we can solve for the unknowns $\{c'_\ell\}$ in polynomial time and obtain the value of $\text{Holant}_\Omega$. This proves the claim that we can interpolate the signature $R$ in Equation (7.13), for any nonzero $t \in \mathbb{C}$.

Let $t = (\sqrt{m^2 + 8} + \sqrt{m^2 + 4})/2$ so $t^{-1} = (\sqrt{m^2 + 8} - \sqrt{m^2 + 4})/2$. Let $a = (\sqrt{m^2 + 8} - m)/2$ and $b = (\sqrt{m^2 + 8} + m)/2$, then $ab = 2$ and both $a$ and $b$ are nonzero. One can verify that

$$P \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix} P^{-1} = \begin{bmatrix} a & 1 \\ 1 & b \end{bmatrix}.$$  

Thus, the signature matrix for $R$ for this particular choice of $t$ is

$$M_R = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & a & 1 & 0 \\ 0 & 1 & b & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$  

After a counterclockwise rotation of $90^\circ$ on the edges of $R$, we have a signature $R'$ with a redundant signature matrix

$$M_{R'} = \begin{bmatrix} 0 & 0 & 0 & a \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ b & 0 & 0 & 0 \end{bmatrix}.$$  

252
Its compressed signature matrix

\[
\overline{M_{R'}} = \begin{bmatrix} 0 & 0 & a \\ 0 & 2 & 0 \\ b & 0 & 0 \end{bmatrix}
\]

is nonsingular. After a holographic transformation by \( Z^{-1} \), where \( Z = \frac{1}{\sqrt{2}} [1 \; \; 1] \), the binary **DISEQUALITY** (\( \neq 2 \)) is transformed to the binary **EQUALITY** (\( = 2 \)). Thus Holant(\( \neq 2 \mid R' \)) is transformed to Holant(\( = 2 \mid Z^{\otimes 4} R' \)), which is the same as Holant(\( Z^{\otimes 4} R' \)). We conclude that this Holant problem Holant(\( Z^{\otimes 4} R' \)) is \#P-hard by Corollary 7.27, where we choose \( T = Z^{-1} \).

To summarize, Holant(\( Z^{\otimes 4} R' \)) is \#P-hard implies that Holant(\( \neq 2 \mid R' \)) is \#P-hard, which implies that Holant(\( \neq 2 \mid R \)) is \#P-hard as \( R' \) is a rotated form of \( R \), which implies that Holant(\( \neq 2 \mid r \)) is \#P-hard since it can interpolate Holant(\( \neq 2 \mid R \)) for all choices of \( t \), which implies that Holant(\( \neq 2 \mid \{f, g\} \)) is \#P-hard since it can directly realize Holant(\( \neq 2 \mid \emptyset \)).

We note that the planar problem Pl-Holant(\( \neq 2 \mid \{f, g\} \)) is computable in P by matchgates. The twist introduced in Figure 7.9 is unavoidable; otherwise we would have collapsed \#P and P, as all other steps of the reduction chain are planar, all the way back to the \#P-hard planar **Eulerian-Orientations** problem Pl-Holant(\( \neq 2 \mid \emptyset \)).

### 7.6.2 Combining Vanishing Signatures with Others

With Corollary 7.25, Corollary 7.27, and Lemma 7.30 in hand, we consider Holant problems when vanishing signatures are combined with other signatures. We begin with degenerate symmetric signatures which include unary signatures. Note that in the following lemma proving \#P-hardness, to avoid the tractable case 5 in Theorem 7.19, it is necessary to require \( \text{rd}^{\sigma}(f) \geq 2 \), for otherwise \( f \in \mathcal{R}_2^{\sigma} \).

**Lemma 7.31.** Let \( f \in \mathcal{V}^{\sigma} \) be a symmetric signature with \( \text{rd}^{\sigma}(f) \geq 2 \) where \( \sigma \in \{+, -\} \). Let \( v \) be a degenerate symmetric signature, thus \( v = u^{\otimes m} \) for some unary signature \( u \), and some integer \( m \geq 1 \). If \( u \) is not a multiple of \([1, \sigma i]\), then Holant(\( \{f, v\} \)) is \#P-hard.

**Proof.** We consider \( \sigma = + \) since the other case is similar. Since \( f \in \mathcal{V}^+ \), we have \( \text{arity}(f) > 2 \text{rd}^+(f) \geq 4 \), and \( \text{vd}^+(f) = \text{arity}(f) - \text{rd}^+(f) > 0 \). As \( \text{rd}^+(f) \geq 2 \), \( f \) is a nonzero signature. By Lemma 7.15, with zero or more self loops of \( f \), we can construct some \( f' \) with \( \text{rd}^+(f') = 2 \) and \( \text{arity} n \geq 5 \). We can repeatedly apply Lemma 7.15, since in each step we reduce its recurrence degree \( \text{rd}^+ \) and vanishing degree \( \text{vd}^+ \) by exactly one, thus they remain positive and the resulting signature is nonzero. The process can be continued on \( f' \). After two more self loops, we have a nonzero multiple of \([1, i]^{\otimes (n-4)}\).

Let \( t = \gcd(n - 4, m) \). There are integers \( x \) and \( y \) such that \( xm + y(n - 4) = t \). By replacing \( x \) by \( x + z(n - 4) \) and \( y \) by \( y - zm \), for any integer \( z \), we may assume \( x > 0 \) and \( y < 0 \). Then if we connect \( |y| \) copies of \([1, i]^{\otimes (n-4)}\) to \( x \) copies of \( v = u^{\otimes m} \), we can realize \( u^{\otimes t} \). Note that \( u \) is not any multiple (including not the zero multiple) of \([1, i]\) and thus \( \langle u, [1, i] \rangle \)
is a nonzero constant. We can realize \( g = u^{(n-4)} \) by putting \( (n - 4)/t \) many copies of \( u^t \) together.

Now connect this \( g \) back to \( f' \). Since the unary \( u \) is not any multiple of \([1, i]\), we can directly verify that \( g \not\in \mathcal{R}^+_{n-4} \) and thus \( \text{rd}^+(g) = \text{arity}(g) = n - 4 \), and \( \text{vd}^+(g) = 0 \). By Lemma 7.14, we get \( f'' = \langle f', g \rangle \) of arity 4 and \( \text{rd}^+(f'') = 2 \). One can verify that \( \text{Holant}(f'') \) is \#P-hard by Corollary 7.25, by writing \( f''_k = i^k p(k) \) for some polynomial \( p \) of degree exactly 2. A more revealing proof of the \#P-hardness of \( \text{Holant}(f'') \) is by noticing that this is the problem \( \text{Holant}(=^2 \mid f'') \), which is equivalent to \( \text{Holant}(\neq^2 \mid \widehat{f''}) \) under the holographic transformation \( Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \). By \( \text{rd}^+(f'') = 2 \), \( \widehat{f''} \) takes the form \( \begin{bmatrix} f''_0, & f''_1, & f''_2, & 0, & 0 \end{bmatrix} \), with \( f''_2 \neq 0 \), as proved in Subsection 7.1.3. Then \( \text{Holant}(\neq^2 \mid \begin{bmatrix} f''_0, & f''_1, & f''_2, & 0, & 0 \end{bmatrix}) \equiv_T \text{Holant}(\neq^2 \mid \begin{bmatrix} 0, & 0, & 1, & 0, & 0 \end{bmatrix}) \), the Eulerian Orientation problem \#EO (see Subsection 7.1.3), which is \#P-hard by Theorem 6.29.

Next we consider what binary signatures can be combined with a vanishing signature. Before we prove Lemma 7.33, we prove a simple interpolation lemma.

**Lemma 7.32.** Let \( x \in \mathbb{C} \). If \( x \neq 0 \), then for any set \( \mathcal{F} \) containing \([x, 1, 0]\), we have

\[
\text{Holant}(\neq_2 \mid \mathcal{F} \cup \{[v, 1, 0]\}) \leq_T \text{Holant}(\neq_2 \mid \mathcal{F})
\]

for any \( v \in \mathbb{C} \).

**Proof.** Consider an instance \( \Omega \) of \( \text{Holant}(\neq_2 \mid \mathcal{F} \cup \{[v, 1, 0]\}) \). Suppose that \([v, 1, 0]\) appears \( n \) times in \( \Omega \). We stratify the assignments in \( \Omega \) based on the assignments to \([v, 1, 0]\). We only need to consider assignments of Hamming weight at most one since an assignment of Hamming weight two contributes a factor of 0. We have

\[
\text{Holant}_\Omega = \sum_{i=0}^{n} v^i c_i,
\]

where \( c_i \) is the sum, over all assignments with exactly \( i \) many \([v, 1, 0]\)'s in \( \Omega \) assigned with a local assignment 00, and \( n - i \) many \([v, 1, 0]\)'s in \( \Omega \) assigned 01 or 10, of the product of evaluations of all other signatures on \( \Omega \).

We construct from \( \Omega \) a sequence of instances \( \Omega_s \) of \( \text{Holant}(\mathcal{F}) \) indexed by \( s \geq 1 \), by replacing each occurrence of \([v, 1, 0]\) with a gadget \( g_s \) created from \( s \) copies of \([x, 1, 0]\), connected sequentially but with \((\neq_2) = [0, 1, 0] \) between each sequential pair. The signature of \( g_s \) is \([sx, 1, 0]\), which can be verified by the matrix product

\[
\begin{pmatrix}
[x & 1 \\
1 & 0]
\end{pmatrix}
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\begin{pmatrix}
1 & x \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
x & 1 \\
0 & 1
\end{pmatrix}
= 
\begin{pmatrix}
1 & (s-1)x \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
x & 1 \\
1 & 0
\end{pmatrix}
= 
\begin{pmatrix}
sx & 1 \\
1 & 0
\end{pmatrix}.
\]

The Holant on \( \Omega_s \) is

\[
\text{Holant}_{\Omega_s} = \sum_{i=0}^{n} (sx)^i c_i.
\]
Figure 7.10: The circles are assigned $[t, 1, 0, 0]$ and the square is assigned $(\neq_2)$.

For $s \geq 1$, this gives a coefficient matrix that is Vandermonde. Since $x$ is nonzero, $sx$ is distinct for each $s$. Therefore, the Vandermonde system has full rank. We can solve for the unknowns $c_i$ in polynomial time and obtain the value of Holant$_\Omega$.

For binary signatures we have the following lemma. Due to the tractable case 4 in Theorem 7.19, the condition $g \notin \mathcal{R}^*_2$ in Lemma 7.33 is needed.

**Lemma 7.33.** Let $f \in \mathcal{V}^\sigma$ be a symmetric non-degenerate signature where $\sigma \in \{+, -\}$. If $g \notin \mathcal{R}^*_2$ is a symmetric non-degenerate binary signature, then Holant($\{f, g\}$) is $\#P$-hard.

**Proof.** We consider $\sigma = +$ since the other case is similar. A unary signature is degenerate. If a binary symmetric signature $f$ is vanishing, then $vd^+(f) > 1$, and so $vd^+(f) \geq 2$, therefore $f$ is also degenerate. Since we assume $f$ is non-degenerate, $\text{arity}(f) \geq 3$.

We prove the lemma by induction on the arity of $f$. There are two base cases, $\text{arity}(f) = 3$ and $\text{arity}(f) = 4$. However, the arity 3 case is easily reduced to the arity 4 case. We show this first, and then show that the lemma holds for the arity 4 case.

Assume $\text{arity}(f) = 3$. Since $f \in \mathcal{V}^+$, we have $vd^+(f) < 3/2$, thus $f \in \mathcal{R}^+_2$. From $vd^+(f) \leq 1$ we have $vd^+(f) \geq 2$. On the other hand, $f$ is non-degenerate, and so $vd^+(f) < 3$, thus $vd^+(f) = 2$ and $rd^+(f) = 1$.

We connect two copies of $f$ together by one edge to get an arity 4 signature $f'$. By construction, it may not appear that $f'$ is a symmetric signature; but we show that $f'$ is in fact symmetric, non-degenerate and vanishing. It is clearly a vanishing signature, since $f$ is vanishing. Consider the $Z$ transformation, under which $f$ is transformed into $\hat{f} = (Z^{-1})^{\otimes 3}f = [t, 1, 0, 0]$ for some $t$ up to a nonzero constant. The $(=2)$ on the connecting edge between the two copies of $f$ is transformed into $(=2)Z^{\otimes 2} = (\neq 2)$. In the bipartite setting, our construction is the same as the gadget in Figure 7.10. One can verify that the resulting signature is $\hat{f}' = [2t, 1, 0, 0, 0]$. The crucial observation is that it takes the same value 0 on inputs 1010 and 1100, where the left two bits are input to one copy of $f$ and the right two bits are for another. The corresponding signature $f' = Z^{\otimes 4}\hat{f}'$ is non-degenerate, with $rd^+(f') = 1$ and vanishing.

Next we consider the base case of $\text{arity}(f) = 4$. Since $f \in \mathcal{V}^+$, we have $vd^+(f) > 2$ and $rd^+(f) < 2$. Since $f$ is non-degenerate we have $rd^+(f) \neq -1, 0$, hence $rd^+(f) = 1$ and $vd^+(f) = 3$. Also by assumption, the given binary $g \notin \mathcal{R}^+_2$, we have $rd^+(g) = 2$. Once again,
consider the holographic transformation by $Z$. By $\text{rd}^+(g) = 2$, the transformation $(Z^{-1})^{\otimes 2}g$ has the form $[a, b, c]$ with $c \neq 0$. This gives

$$\text{Holant}(=2 \mid \{f, g\}) \equiv_T \text{Holant}((=2)Z^{\otimes 2} \mid \{(Z^{-1})^{\otimes 4}f, (Z^{-1})^{\otimes 2}g\})$$

$$\equiv_T \text{Holant}(\neq 2) \{\hat{f}, \hat{g}\},$$

where up to a nonzero constant, $\hat{f} = [t, 1, 0, 0, 0]$ and $\hat{g} = [a, b, 1]$, for some $t, a, b \in \mathbb{C}$. We have $a - b^2 \neq 0$ since $g$ is non-degenerate.

Our next goal is to show that we can realize a signature of the form $[c, 0, 1]$ where $c \neq 0$. If $b = 0$, then $\hat{g}$ is what we want since in this case $a = a - b^2 \neq 0$. Now we assume $b \neq 0$.

Connecting both edges of $\hat{g}$ to $\hat{f}$ via $(\neq 2)$, we get $[t + 2b, 1, 0]$. If $t \neq -2b$, then by Lemma 7.32, we can interpolate any binary signature of the form $[v, 1, 0]$. If $t = -2b$, then we connect two copies of $\hat{g}$ via $(\neq 2)$, and get $g' = [2ab, a + b^2, 2b]$. Connecting this $g'$ to $\hat{f}$ via $(\neq 2)$, we get $[2(a - b^2), 2b, 0]$, using $t = -2b$. Since $a \neq b^2$ and $b \neq 0$, we can interpolate any $[v, 1, 0]$ again by Lemma 7.32.

Hence, we have any signature $[v, 1, 0]$, where $v \in \mathbb{C}$ is for us to choose. We construct the gadget in Figure 7.11 with the circles assigned $[v, 1, 0]$, the squares assigned $(\neq 2)$, and the triangle assigned $[a, b, 1]$. The resulting gadget has signature $[a + 2bv + v^2, b + v, 1]$, which can be verified by the matrix product

$$\begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ b & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a + 2bv + v^2 & b + v \\ b + v & 1 \end{bmatrix}.$$
is a Gen-Eq signature, and the signature \( \hat{f} \) on the four circle vertices is a weighted version of the matching function At-Most-One. E.g., we may verify that \( \hat{h}_0 = t^4 + 6ct^2 + 3c^2 \) as follows: For the external input pattern 0000, the term \( t^4 \) comes from the empty matching among six internal edges, the term \( 6ct^2 \) comes from the six matchings of size one, each matching edge having weight \( c \), and the term \( 3c^2 \) comes from the three perfect matchings of size two. The other entries can be computed similarly.

The compressed signature matrix of \( \hat{h} \) is

\[
\tilde{M}_\hat{h} = \begin{bmatrix}
3c^2 + 6ct^2 + t^4 & 2(3ct + t^2) & c + t^2 \\
3ct + t^3 & 2(c + t^2) & t \\
c + t^2 & 2t & 1
\end{bmatrix}
\]

and quite amazingly its determinant is \( 4c^3 \neq 0 \) (check it!). Thus \( \tilde{M}_\hat{h} \) is nonsingular. After a holographic transformation by \( Z^{-1} \), where \( Z = \sqrt{3} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \), the binary DisEquality function \( (\neq 2) \) is transformed to the binary Equality \( (= 2) \). Thus Holant(\( \neq 2 \mid \hat{h} \)) is transformed to Holant(\( = 2 \mid Z^{\otimes 4}\hat{h} \)), which is the same as Holant(\( Z^{\otimes 4}\hat{h} \)). Then we are done by Corollary 7.27.

Now we do the induction step. The idea is to produce from \( f \) and \( g \) another non-degenerate signature \( f' \in \mathcal{V}^+ \) that has a smaller arity than \( f \). Assume \( f \) has arity \( n \geq 5 \). Since \( f \) is non-degenerate, \( \text{rd}^+(f) \neq -1, 0 \). First suppose \( \text{rd}^+(f) = 1 \), then connect both edges of the binary \( g \) to \( f \) to get \( f' = \langle f, g \rangle \). We have noted that \( \text{rd}^+(g) = 2 \), then \( \text{vd}^+(g) = 0 \). By Lemma 7.14 we know that \( \text{rd}^+(f') = 1 \) and \( \text{arity}(f') = n - 2 \geq 3 \). Thus \( f' \) is vanishing. Also \( f' \) is non-degenerate, for otherwise let \( f' = [a, b]^{\otimes (n - 2)} \). If \( [a, b] \) is a multiple of \( [1, i] \), then \( \text{rd}^+(f') \leq 0 \), which is false. If \( [a, b] \) is not a multiple of \( [1, i] \), then it can be directly checked that \( f' \notin \mathcal{R}_{n-2}^+ \), and \( \text{rd}^+(f') = n - 2 > 1 \), which is also false. Hence \( f' \) is a non-degenerate vanishing signature of arity \( n - 2 \). By induction hypothesis we are done.
We now suppose $rd^+(f) = t \geq 2$. Since $f$ is non-degenerate it is certainly nonzero. Since it is vanishing, certainly $vd^+(f) > 0$. Hence we may apply Lemma 7.15. Let $f'$ be obtained from $f$ by a self loop, then $rd^+(f') = t - 1 \geq 1$ and $arity(f') = n - 2$. Clearly $f'$ is still vanishing. We claim that $f'$ is non-degenerate. This is proved by the same argument as above. If $f'$ were degenerate, then either $rd^+(f') \leq 0$ which contradicts $rd^+(f') \geq 1$, or $rd^+(f') = arity(f')$ which would contradict $f'$ being a vanishing signature. Therefore, we can apply the induction hypothesis. This finishes our proof.

Finally, we consider a pair of vanishing signatures of the opposite type, both of arity at least 3. We show that opposite types of vanishing signatures cannot mix. More formally, vanishing signatures of opposite types, when put together, lead to $\#P$-hardness.

**Lemma 7.34.** Let $f \in \mathcal{V}^+$ and $g \in \mathcal{V}^-$ be non-degenerate signatures both of arity at least 3. Then Holant($\{f, g\}$) is $\#P$-hard.

**Proof.** Let $rd^+(f) = d$, $rd^-(g) = d'$, $arity(f) = n$ and $arity(g) = n'$, then $2d < n$ and $2d' < n'$. We can apply Lemma 7.15 zero or more times to construct a signature $g'$ obtained from $g$ by adding a certain number of self loops, and the signature is a tensor power of $[1, -1]$. To see this, note that we start with $rd^-(g) < vd^-(g)$ with their sum being $arity(g)$. We are allowed to apply Lemma 7.15, if the signature is nonzero and its $vd^-$ is positive. Each time we apply Lemma 7.15, we reduce $rd^-$ and $vd^-$ by one, and the arity by two. Thus $rd^- < vd^-$ is maintained until $rd^-$ becomes 0, at which point the signature is (a nonzero multiple of) a tensor power of $[1, -1]$. Initially $g$ is non-degenerate by assumption therefore certainly nonzero. During the successive applications of Lemma 7.15 zero or more times, the recursive degree $rd^-$ is positive, and so the signature is nonzero. Thus Lemma 7.15 can be applied. By Lemma 7.31, if $d = rd^+(f) \geq 2$, then Holant($\{f, g'\}$) is $\#P$-hard, and so Holant($\{f, g\}$) is also $\#P$-hard. Similarly, it is $\#P$-hard if $d' \geq 2$. Thus we may assume that $d = d' = 1$.

We perform the $Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ transformation

$$\text{Holant}(\_|\_ \{f, g\}) \equiv_T \text{Holant}(\_|\_ \{Z \otimes 2 \otimes (Z^{-1}) \otimes n \otimes (Z^{-1}) \otimes n' \otimes g\})$$

$$\equiv_T \text{Holant}(\_|\_ \{f', g\}).$$

Since $d = 1$, as shown in Lemma 7.17 in Subsection 7.1.3, $(Z^{-1}) \otimes n f = \hat{f} = [\hat{f}_0, \hat{f}_1, 0, \ldots, 0]$, where $\hat{f}_1 \neq 0$. Similarly, for $g$ with $rd^-(g) = 1$, $(Z^{-1}) \otimes n' g = \hat{g} = [0, \ldots, 0, \hat{g}_{n'-1}, \hat{g}_{n'}]$, where $\hat{g}_{n'-1} \neq 0$.

Therefore, up to a nonzero constant, $\hat{f} = [a, 1, 0, \ldots, 0]$ and $\hat{g} = [0, \ldots, 0, 1, b]$, for some $a, b \in \mathbb{C}$. We show that it is always possible to get two such signatures of the same arity $n^* = \min\{n, n'\}$. Suppose $n > n'$. We form a loop from $\hat{f}$, where the loop is really a path consisting of one vertex and two edges, with the vertex assigned the signature ($\neq 2$). It is easy to see that this signature is the degenerate signature $2[1, 0] \otimes (n-2)$. Similarly, we can form a loop from $\hat{g}$ and can get $2[0, 1] \otimes (n'-2)$. Thus we have both $[1, 0] \otimes (n-2)$ and $[0, 1] \otimes (n'-2)$.
We can connect all $n' - 2$ edges of the second to the first, connected by $(\neq 2)$. This gives $[1, 0]^\otimes(n-n')$. We can continue subtracting the smaller arity from the larger one. We continue this process in a subtractive version of the Euclidean algorithm, and end up with both $[1, 0]^\otimes t$ and $[0, 1]^\otimes t$, where $t = \gcd(n - 2, n' - 2) = \gcd(n - n', n' - 2)$. In particular, $t \mid n - n'$ and by taking $(n - n')/t$ many copies of $[0, 1]^\otimes t$, we can get $[0, 1]^\otimes(n-n')$. Connecting this back to $\hat{f}$ via $(\neq 2)$, we get a symmetric signature of arity $n'$ consisting of the first $n' + 1$ entries of $\hat{f}$. A similar proof works when $n' > n$.

Thus without loss of generality, we may assume we have two signatures $[a, 1, 0, \ldots, 0]$ and $[0, \ldots, 1, 0, b]$, both of arity $n^* \geq 3$. We also have $[1, 0]^\otimes t$ and $[0, 1]^\otimes t$ thus both $[1, 0]^\otimes(n^*-2)$ and $[0, 1]^\otimes(n^*-2)$. Connect $[0, 1]^\otimes(n^*-2)$ to $[a, 1, 0, \ldots, 0]$ of arity $n^*$ via $(\neq 2)$ we get $\hat{h} = [a, 1, 0]$. For $a \neq 0$, translating this back by $Z$, we have a binary signature $h$ together with the given $g \in \mathcal{V}^\to$. We claim that $h \notin \mathcal{R}_2^\to$. Indeed if it were the case that $h \in \mathcal{R}_2^\to$, then $\mathsf{rd}^\to(h) \leq 1$ and thus $\hat{h} = [0, *, *, *]$, a contradiction. By Lemma 7.33, Holant$(\{g, h\})$ is $\#\mathcal{P}$-hard, and thus Holant$(\{f, g\})$ is $\#\mathcal{P}$-hard. A similar proof works for the case $b \neq 0$.

The only case left is when $\hat{f} = [0, 1, 0, \ldots, 0]$, and $\hat{g} = [0, \ldots, 0, 1, 0]$, both of arity at least 3. This is $\#\mathcal{P}$-hard by Lemma 7.30.

\begin{flushright}\[\Box\]\end{flushright}

### 7.7 $\mathcal{A}$-transformable and $\mathcal{P}$-transformable Signatures

In this section, we study $\mathcal{A}$-transformable signatures and $\mathcal{P}$-transformable signatures. We denote by $\alpha = \frac{1+i}{\sqrt{2}} = \sqrt{\frac{1}{2}} = e^{\frac{i\pi}{4}}$ and use $\mathcal{O}_2(\mathbb{C})$ to denote the group of 2-by-2 orthogonal matrices over $\mathbb{C}$. Recall that $\mathcal{R}_{123} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$, where $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_3$ are defined in Chapter 3 Section 3.2.

#### 7.7.1 Characterization of $\mathcal{A}$- and $\mathcal{P}$-transformable Signatures

Recall that by definition, if a set of signatures $\mathcal{F}$ is $\mathcal{A}$-transformable (resp. $\mathcal{P}$-transformable), then the binary EQUALITY $(=\neq 2)$ must be simultaneously transformed into $\mathcal{A}$ (resp. $\mathcal{P}$) along with $\mathcal{F}$. We first characterize all possible matrices for such transformations by considering only the transformation of $(=\neq 2)$. While there are many binary signatures in $\mathcal{A} \cup \mathcal{P}$, it turns out that it is sufficient to consider only three signatures. The following Proposition is best understood in terms of a group action, its stabilizer and orbits.

**Proposition 7.35.** Let $T \in \mathbb{C}^{2 \times 2}$ be a matrix. Then the following hold:

1. $[1, 0, 1]^\otimes 2 = [1, 0, 1]$ iff $T \in \mathcal{O}_2(\mathbb{C})$;
2. $[1, 0, 1]^\otimes 2 = [1, 0, i]$ iff there exists an $H \in \mathcal{O}_2(\mathbb{C})$ such that $T = H [\begin{smallmatrix} 1 & 0 \\ -i & \alpha \end{smallmatrix}]$;
3. $[1, 0, 1]^\otimes 2 = [0, 1, 0]$ iff there exists an $H \in \mathcal{O}_2(\mathbb{C})$ such that $T = \frac{1}{\sqrt{2}} H [\begin{smallmatrix} 1 & 1 \\ -1 & -1 \end{smallmatrix}]$. 


Proof. The first item is clear since

\[ [1, 0, 1]^\otimes 2 = [1, 0, 1] \iff T^T I_2 T = I_2 \iff T^T T = I_2, \]

the definition of a 2-by-2 orthogonal matrix \( T \).

We let \( M_1 = I_2, \ M_2 = \left[ \begin{array}{rr} 1 & 0 \\ 0 & \alpha \end{array} \right] \) and \( M_3 = Z = \frac{1}{\sqrt{2}} \left[ \begin{array}{rr} 1 & 1 \\ 1 & -1 \end{array} \right] \), let \( T_j = HM_j \) (for \( j = 1, 2, 3 \)), where \( H \in \mathbb{O}_2(\mathbb{C}) \). Then

\[ [1, 0, 1]T_j^\otimes 2 = [1, 0, 1](HM_j)^\otimes 2 = [1, 0, 1]M_j^\otimes 2 = f_j, \]

where \( f_1 = [1, 0, 1], \ f_2 = [1, 0, i] \) and \( f_3 = [0, 1, 0] \).

Conversely, suppose that \( [1, 0, 1](T_j)^\otimes 2 = f_j \). Then we have

\[ [1, 0, 1](T_jM_j^{-1})^\otimes 2 = f_j (M_j^{-1})^\otimes 2 = [1, 0, 1], \]

so \( H = T_jM_j^{-1} \in \mathbb{O}_2(\mathbb{C}) \) by case 1. Thus \( T_j = HM_j \) as desired.

We also need the following proposition; the proof is direct.

**Proposition 7.36.** If a symmetric signature \( f = [f_0, f_1, \ldots, f_n] \) can be expressed in the form \( f = a[1, \lambda]^\otimes n + b[1, \mu]^\otimes n \), for some \( a, b, \lambda, \mu \in \mathbb{C} \), then the \( f_k \)'s satisfy the recurrence relation \( f_{k+2} = (\lambda + \mu) f_{k+1} - \lambda \mu f_k \) for \( 0 \leq k \leq n - 2 \).

Recall the stabilizer group \( \text{Stab}({\mathcal{A}}) = \mathbb{C}^* \cdot \langle D, H_2 \rangle \) that was determined in Lemma 6.13, where \( D = \left[ \begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array} \right] \) and \( H_2 = \frac{1}{\sqrt{2}} \left[ \begin{array}{rr} 1 & 1 \\ 1 & -1 \end{array} \right] \).

**Lemma 7.37.** Let \( \mathcal{F} \) be a set of signatures. Then \( \mathcal{F} \) is \( \mathcal{A} \)-transformable under \( T \) iff \( \mathcal{F} \) is \( \mathcal{A} \)-transformable under any \( T' \in T \text{Stab}(\mathcal{A}) \).

**Proof.** Sufficiency is trivial since \( I_2 \in \text{Stab}(\mathcal{A}) \). If \( \mathcal{F} \) is \( \mathcal{A} \)-transformable under \( T \), then by definition, we have \( (=) T^\otimes 2 \in \mathcal{A} \) and \( \mathcal{F}' = T^{-1} \mathcal{F} \subseteq \mathcal{A} \). Let \( T' = TM \in T \text{Stab}(\mathcal{A}) \) for any \( M \in \text{Stab}(\mathcal{A}) \). It then follows that \( (=) T'^\otimes 2 = (=) T^\otimes 2 M^\otimes 2 \in \mathcal{A} M = \mathcal{A} \) and \( T'^{-1} \mathcal{F} = M^{-1} \mathcal{F}' \subseteq M^{-1} \mathcal{A} = \mathcal{A} \). Therefore \( \mathcal{F} \) is \( \mathcal{A} \)-transformable under any \( T' \in T \text{Stab}(\mathcal{A}) \).

After restricting by Proposition 7.35 and normalizing by Lemma 7.37, one only needs to check a small subset of \( \mathbb{GL}_2(\mathbb{C}) \) to determine if \( \mathcal{F} \) is \( \mathcal{A} \)-transformable.

**Lemma 7.38.** Let \( \mathcal{F} \) be a set of signatures. Then \( \mathcal{F} \) is \( \mathcal{A} \)-transformable iff there exists an \( H \in \mathbb{O}_2(\mathbb{C}) \) such that \( \mathcal{F} \subseteq H \mathcal{A} \) or \( \mathcal{F} \subseteq H \left[ \begin{array}{rr} 1 & 0 \\ 0 & \alpha \end{array} \right] \mathcal{A} \).

**Proof.** Sufficiency is easily verified by checking that \( (=) \) is transformed into \( \mathcal{A} \) by \( H \) and by \( H \left[ \begin{array}{rr} 1 & 0 \\ 0 & \alpha \end{array} \right] \) in respective cases.

If \( \mathcal{F} \) is \( \mathcal{A} \)-transformable, then by definition, there exists a matrix \( T \) such that \( (=) T^\otimes 2 \in \mathcal{A} \) and \( T^{-1} \mathcal{F} \subseteq \mathcal{A} \). Since \( (=) \) is non-degenerate and symmetric, \( (=) T^\otimes 2 \in \mathcal{A} \) is equivalent to \( (=) T^\otimes 2 \in \mathcal{F}_{123} \).
Any signature in $\mathcal{F}_{123}$ is expressible as $c(v_1^{\otimes n} + i^t v_2^{\otimes n})$, where $t \in \{0, 1, 2, 3\}$ and $(v_1, v_2)$ is a pair of vectors in the set

$$\left\{ \left([1 \ 0], [0 \ 1]\right), \left([1 \ 1], [1 \ -1]\right), \left([1 \ i], [1 \ -i]\right) \right\}.$$ 

We use $D$ and $H_2 \in \text{Stab}(\mathcal{A})$ to normalize $\mathcal{F}_{123}$ by Lemma 7.37. In particular, $\mathcal{F}_1 = \mathcal{F}_2 H_2$ and $\mathcal{F}_1 = \mathcal{F}_3(D^{-1}H_2)$. Furthermore, the binary signatures in $\mathcal{F}_1$ are just the four signatures $[1, 0, 1]$, $[1, 0, i]$, $[1, 0, -1]$, and $[1, 0, -i]$ up to a scalar. We also normalize these four by $D$, as $[1, 0, 1] = [1, 0, -1]D^{\otimes 2}$ and $[1, 0, i] = [1, 0, -i]D^{\otimes 2}$. Hence $\mathcal{F}$ being $\mathcal{A}$-transformable implies that there exists a matrix $T$ such that $(=)T^{\otimes 2} \in \{[1, 0, 1], [1, 0, i]\}$ and $T^{-1}\mathcal{F} \subseteq \mathcal{A}$. Now we apply Proposition 7.35.

1. If $(=)T^{\otimes 2} = [1, 0, 1]$, then by case 1 of Proposition 7.35, we have $T \in O_2(C)$. Therefore $\mathcal{F} \subseteq H \mathcal{A}$ where $H = T \in O_2(C)$.

2. If $(=)T^{\otimes 2} = [1, 0, i]$, then by case 2 of Proposition 7.35, there exists an $H \in O_2(C)$ such that $T = H [1 \ 0 \ 0 \alpha]$. Therefore $\mathcal{F} \subseteq T \mathcal{A} = H [1 \ 0 \ 0 \alpha] \mathcal{A}$ where $H \in O_2(C)$.

This completes the proof. □

Using these two lemmas, we can characterize all $\mathcal{A}$-transformable signatures. We first define the three sets $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$.

**Definition 7.39.** A symmetric signature $f$ of arity $n$ is in $\mathcal{A}_1$ if there exists an $H \in O_2(C)$ and a nonzero constant $c$ such that $f = cH^{\otimes n} \left([1]^{\otimes n} + \beta [1^{-1}]^{\otimes n}\right)$, where $\beta = \alpha^{tn+2r}$ for some $r \in \{0, 1, 2, 3\}$ and $t \in \{0, 1\}$.

When such an $H$ exists, we say that $f \in \mathcal{A}_1$ with transformation $H$. If $f \in \mathcal{A}_1$ with $I_2$, then we say $f$ is in the canonical form of $\mathcal{A}_1$. If $f$ is in the canonical form of $\mathcal{A}_1$, then by Proposition 7.36, for any $0 \leq k \leq n-2$, we have $f_{k+2} = f_k$ and one of the following holds:

- $f_0 = 0$, or
- $f_1 = 0$, or
- $f_1 = \pm if_0 \neq 0$, or
- $n$ is odd and $f_1 = \pm (1 \pm \sqrt{2})if_0 \neq 0$ (all four sign choices are permissible).

Notice that when $n$ is odd and $t = 1$ in Definition 7.39, it has some complication as described by the factor $\alpha^{n+2r}$, which runs through $\alpha^s$ ($s \in \{0, 1, 2, 3\}$).

**Definition 7.40.** A symmetric signature $f$ of arity $n$ is in $\mathcal{A}_2$ if there exists an $H \in O_2(C)$ and a nonzero constant $c$ such that $f = cH^{\otimes n} \left([1]^{\otimes n} + [1^{-1}]^{\otimes n}\right)$.

Similarly, when such an $H$ exists, we say that $f \in \mathcal{A}_2$ with transformation $H$. If $f \in \mathcal{A}_2$ with $I_2$, then we say $f$ is in the canonical form of $\mathcal{A}_2$. If $f$ is in the canonical form of $\mathcal{A}_2$, then
by Proposition 7.36, for any $0 \leq k \leq n - 2$, we have $f_{k+2} = -f_k$. Since $f$ is non-degenerate, $f_1 \neq \pm if_0$ is implied.

It is worth noting that $H[\mathbb{1}]$ is a nonzero multiple of either $[\mathbb{1}]$ or $[\mathbb{1}]$, for any orthogonal $H \in O_2(\mathbb{C})$. The same is true for $H[\mathbb{1}]$. Using this fact, the following lemma gives a characterization of $A_2$. It says that any signature in $A_2$ is essentially in canonical form.

**Lemma 7.41.** Let $f$ be a non-degenerate symmetric signature. Then $f \in A_2$ if $f$ is of the form $c([\mathbb{1}]^\otimes n + \beta [\mathbb{1}]^\otimes n)$ for some $c, \beta \neq 0$.

**Proof.** Assume that $f = c([\mathbb{1}]^\otimes n + \beta [\mathbb{1}]^\otimes n)$ for some $c, \beta \neq 0$. Consider the orthogonal transformation $H = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$, where $a = \frac{1}{2} \left( \beta^{\frac{1}{2}} + \beta^{-\frac{1}{2}} \right)$ and $b = \frac{1}{2i} \left( \beta^{\frac{1}{2}} - \beta^{-\frac{1}{2}} \right)$. We pick $a$ and $b$ in this way so that $a + bi = \beta^{\frac{1}{2}}$, $a - bi = \beta^{-\frac{1}{2}}$, and $(a + bi)(a - bi) = a^2 + b^2 = 1$. Also $(\frac{a+bi}{a-bi})^n = \beta$. Then

$$H^\otimes n f = c \left( \left[ \begin{array}{c} a + bi \\ -ai + b \end{array} \right] \right)^\otimes n + \beta \left( \left[ \begin{array}{c} a - bi \\ ai + b \end{array} \right] \right)^\otimes n$$

$$= c \left( (a + bi)^n \left[ \begin{array}{c} 1 \\ -i \end{array} \right]^\otimes n + (a - bi)^n \beta \left[ \begin{array}{c} 1 \\ i \end{array} \right]^\otimes n \right)$$

$$= c\sqrt{\beta} \left( \left[ \begin{array}{c} 1 \\ -i \end{array} \right]^\otimes n + \left[ \begin{array}{c} 1 \\ i \end{array} \right]^\otimes n \right),$$

so $f$ can be written as

$$f = c\sqrt{\beta}(H^{-1})^\otimes n \left( \left[ \begin{array}{c} 1 \\ i \end{array} \right]^\otimes n + \left[ \begin{array}{c} 1 \\ -i \end{array} \right]^\otimes n \right).$$

Therefore $f \in A_2$.

On the other hand, the desired form $f = c([\mathbb{1}]^\otimes n + \beta [\mathbb{1}]^\otimes n)$ follows from the fact that $\{[\mathbb{1}]^\otimes n, [\mathbb{1}]^\otimes n\}$ is fixed setwise under any orthogonal transformation up to nonzero constants. □

**Definition 7.42.** A symmetric signature $f$ of arity $n$ is in $A_3$ if there exists an $H \in O_2(\mathbb{C})$ and a nonzero constant $c$ such that $f = cH^\otimes n \left( \left[ \begin{array}{c} 1 \\ a \end{array} \right]^\otimes n + \sqrt{r} \left[ \begin{array}{c} 1 \\ -a \end{array} \right]^\otimes n \right)$ for some $r \in \{0, 1, 2, 3\}$.

Again, when such an $H$ exists, we say that $f \in A_3$ with transformation $H$. If $f \in A_3$ with $I_2$, then we say $f$ is in the canonical form of $A_3$. If $f$ is in the canonical form of $A_3$, then by Proposition 7.36, for any $0 \leq k \leq n - 2$, we have $f_{k+2} = if_k$ and one of the following holds:

- $f_0 = 0$, or
\begin{itemize}
  \item $f_1 = 0$, or
  \item $f_1 = \pm \alpha f_0 \neq 0$.
\end{itemize}

Now we characterize the $\mathcal{A}$-transformable signatures.

**Lemma 7.43.** Let $f$ be a non-degenerate symmetric signature. Then $f$ is $\mathcal{A}$-transformable iff $f \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$.

**Proof.** Assume that $f$ is $\mathcal{A}$-transformable of arity $n$. By applying Lemma 7.38 to $\{f\}$, there exists an $H \in \text{O}_2(\mathbb{C})$ such that $f \in H \mathcal{A}$ or $f \in H \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathcal{A}$. This is equivalent to $(H^{-1})^{\otimes n} f \in \mathcal{A}$ or $(H^{-1})^{\otimes n} f \in \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathcal{A}$. Since $f$ is non-degenerate and symmetric, we can replace $\mathcal{A}$ in the previous expressions with $\mathcal{F}_{123}$. Now we consider the possible cases.

1. If $(H^{-1})^{\otimes n} f \in \mathcal{F}_1$, then a further transformation by $H_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \in \text{O}_2(\mathbb{C})$ puts it into the canonical form of $\mathcal{A}_1$, i.e., $f \in \mathcal{A}_1$ with transformation $HH_2$.
2. If $(H^{-1})^{\otimes n} f \in \mathcal{F}_2$, then it is already in the canonical form of $\mathcal{A}_1$.
3. If $(H^{-1})^{\otimes n} f \in \mathcal{F}_3$, then it is already of the equivalent form of $\mathcal{A}_2$ given by Lemma 7.41.
4. If $(H^{-1})^{\otimes n} f \in \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \mathcal{F}_1$, then a further transformation by $H_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \in \text{O}_2(\mathbb{C})$ puts it into the canonical form of $\mathcal{A}_1$, i.e., $f \in \mathcal{A}_1$ with transformation $HH_2$.
5. If $(H^{-1})^{\otimes n} f \in \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \mathcal{F}_2$, then it is already in the canonical form of $\mathcal{A}_2$.
6. If $(H^{-1})^{\otimes n} f \in \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathcal{F}_3$, then it is of the form $\begin{bmatrix} 1 & \alpha^3 \\ \alpha^3 & 1 \end{bmatrix}^{\otimes n} + i^r \begin{bmatrix} 1 & -\alpha^3 \\ -\alpha^3 & 1 \end{bmatrix}^{\otimes n}$ and a further transformation by $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \in \text{O}_2(\mathbb{C})$ puts it into the canonical form of $\mathcal{A}_3$. To see this,

$$
\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^{\otimes n} \left( \begin{bmatrix} 1 \\ \alpha^3 \end{bmatrix}^{\otimes n} + i^r \begin{bmatrix} 1 \\ -\alpha^3 \end{bmatrix}^{\otimes n} \right) = \begin{bmatrix} -\alpha^3 \\ 1 \end{bmatrix}^{\otimes n} + i^r \begin{bmatrix} \alpha^3 \\ 1 \end{bmatrix}^{\otimes n} = (-\alpha^3)^n \left( \begin{bmatrix} 1 \\ -\alpha^3 \end{bmatrix}^{\otimes n} + (-1)^n i^r \begin{bmatrix} 1 \\ \alpha^3 \end{bmatrix}^{\otimes n} \right) = (-\alpha^3)^n \left( \begin{bmatrix} 1 \\ \alpha \end{bmatrix}^{\otimes n} + i^{2n+r} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix}^{\otimes n} \right).
$$

So $f \in \mathcal{A}_3$ with transformation $H \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. We have shown that if $f$ is $\mathcal{A}$-transformable then $f \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$.

Conversely, if there exists a matrix $H \in \text{O}_2(\mathbb{C})$ such that $(H^{-1})^{\otimes n} f$ is in one of the canonical forms of $\mathcal{A}_1$, $\mathcal{A}_2$, or $\mathcal{A}_3$, then $f$ is $\mathcal{A}$-transformable by applying Lemma 7.38. \Box

We also have a similar characterization for $\mathcal{P}$-transformable signatures. We define the stabilizer group of $\mathcal{P}$ similar to $\text{Stab}(\mathcal{A})$. It is easy to see the left and right stabilizers of $\mathcal{P}$ coincide, which we denote by $\text{Stab}(\mathcal{P})$. Furthermore, $\text{Stab}(\mathcal{P})$ is generated by nonzero scalar multiples of matrices of the form $\begin{bmatrix} 1 & 0 \\ 0 & \nu \end{bmatrix}$ for any nonzero $\nu \in \mathbb{C}$ and $R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.
Lemma 7.44. Let $\mathcal{F}$ be a set of signatures. Then $\mathcal{F}$ is $\mathcal{P}$-transformable iff there exists an $H \in O_2(\mathbb{C})$ such that $\mathcal{F} \subseteq H \mathcal{P}$ or $\mathcal{F} \subseteq H \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \mathcal{P}$.

Proof. Sufficiency is easily verified by checking that $\Rightarrow \mathcal{P}$ is transformed into $\mathcal{P}$ in both cases. In particular, $H$ leaves $\Rightarrow \mathcal{P}$ unchanged.

If $\mathcal{F}$ is $\mathcal{P}$-transformable, then by definition, there exists a matrix $T$ such that $(\Rightarrow \mathcal{P})T \Rightarrow 2 \subseteq \mathcal{P}$ and $T^{-1} \mathcal{F} \subseteq \mathcal{P}$. The non-degenerate binary signatures in $\mathcal{P}$ are either $[0, 1, 0]$ or of the form $[1, 0, \nu]$, up to a scalar. However, notice that $[1, 0, 1] = [1, 0, \nu] \begin{bmatrix} 1 & 0 \\ 0 & \nu^{-\frac{1}{2}} \end{bmatrix} \Rightarrow 2$ and $\begin{bmatrix} 1 & 0 \\ 0 & \nu^{-\frac{1}{2}} \end{bmatrix} \in \text{Stab}(\mathcal{P})$. Thus, we only need to consider $[1, 0, 1]$ and $[0, 1, 0]$. Now we apply Proposition 7.35.

1. If $(\Rightarrow \mathcal{P})T \Rightarrow 2 = [1, 0, 1]$, then by case 1 of Proposition 7.35, we have $T \in O_2(\mathbb{C})$. Therefore $\mathcal{F} \subseteq H \mathcal{P}$ where $H = T \in O(\mathbb{C})$.
2. If $(\Rightarrow \mathcal{P})T \Rightarrow 2 = [0, 1, 0]$, then by case 3 of Proposition 7.35, there exists an $H \in O_2(\mathbb{C})$ such that $T = \frac{1}{\sqrt{2}} H \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Therefore $\mathcal{F} \subseteq H \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathcal{P}$ where $H \in O_2(\mathbb{C})$. □

We also have similar definitions of the sets $\mathcal{P}_1$ and $\mathcal{P}_2$.

Definition 7.45. A symmetric signature $f$ of arity $n$ is in $\mathcal{P}_1$ if there exists $H \in O_2(\mathbb{C})$ and a nonzero constant $c$ such that $f = cH \Rightarrow n \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow n + \beta \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow n$, where $\beta \neq 0$.

We also define $\mathcal{P}_2 = \mathcal{A}_2$.

When such an $H$ exists, we say that $f \in \mathcal{P}_1$ with transformation $H$. If $f \in \mathcal{P}_1$ with $I_2$, then we say $f$ is in the canonical form of $\mathcal{P}_1$. If $f$ is in the canonical form of $\mathcal{P}_1$, then by Proposition 7.36, for any $0 \leq k \leq n - 2$, we have $f_{k+2} = f_k$. Since $f$ is non-degenerate, $f_1 \neq \pm f_0$ is implied.

It is easy to check that $\mathcal{A}_1 \subseteq \mathcal{P}_1$. Now we characterize the $\mathcal{P}$-transformable signatures as we did for the $\mathcal{A}$-transformable signatures in Lemma 7.43.

Lemma 7.46. Let $f$ be a non-degenerate symmetric signature. Then $f$ is $\mathcal{P}$-transformable iff $f \in \mathcal{P}_1 \cup \mathcal{P}_2$.

Proof. Assume that $f$ is $\mathcal{P}$-transformable of arity $n$. By applying Lemma 7.44 to $\{f\}$, there exists an $H \in O_2(\mathbb{C})$ such that $f \in H \mathcal{P}$ or $f \in H \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathcal{P}$. This is equivalent to $(H^{-1}) \Rightarrow n f \in \mathcal{P}$ or $(H^{-1}) \Rightarrow n f \in \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathcal{P}$.

The symmetric signatures in $\mathcal{P}$ take the form $[0, 1, 0]$, or $[a, 0, \ldots, 0, b] = a[1, 0] \Rightarrow n + b[0, 1] \Rightarrow n$, where $ab \neq 0$ since $f$ is non-degenerate. Now we consider the possible cases.

1. If $(H^{-1}) \Rightarrow n f = [0, 1, 0]^{\mathcal{P}} = \frac{1}{2n} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow 2 - \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow 2$, then it is already of the equivalent form of $\mathcal{P}_2 = \mathcal{A}_2$ given by Lemma 7.41.
2. If $(H^{-1})^\otimes n f = a [\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}]^\otimes n + b [\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}]^\otimes n$, then a further transformation by $H_2 = \frac{1}{\sqrt{2}} [\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}] \in O_2(\mathbb{C})$ puts it into the canonical form of $\mathcal{P}_1$, i.e., $f \in \mathcal{P}_1$ with transformation $H H_2$.

3. If $(H^{-1})^\otimes n f = [\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}] \otimes 2 [0, 1, 0]^T = 2[1, 0, 1]^T = \left(\frac{1}{2} \otimes 2 + \frac{1}{2} \otimes 2\right)$, then it is already in the canonical form of $\mathcal{P}_1$.

4. If $(H^{-1})^\otimes n f = [\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}]^\otimes (a[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}]^\otimes n + b[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}]^\otimes n)$, then it is already of the equivalent form of $\mathcal{P}_2 = \mathcal{A}_2$ given by Lemma 7.41.

Conversely, if there exists a matrix $H \in O_2(\mathbb{C})$ such that $(H^{-1})^\otimes n f$ is in one of the canonical forms of $\mathcal{P}_1$ or $\mathcal{P}_2$, then one can directly check that $f$ is $\mathcal{P}$-transformable by applying Lemma 7.44. \qed

Combining Lemma 7.43 and Lemma 7.46, we have a necessary and sufficient condition for a single non-degenerate signature to be $\mathcal{A}$- or $\mathcal{P}$-transformable.

**Corollary 7.47.** Let $f$ be a non-degenerate signature. Then $f$ is $\mathcal{A}$- or $\mathcal{P}$-transformable iff $f \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3$.

Notice that our definitions of $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{A}_3$ each involve an orthogonal transformation. For any single signature $f \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3$, Holant($f$) is tractable. However, this does not imply that Holant($\mathcal{P}_1$), Holant($\mathcal{P}_2$), or Holant($\mathcal{A}_3$) is tractable. One can check, using Theorem 7.19, that Holant($\mathcal{P}_2$) is tractable while Holant($\mathcal{P}_1$) and Holant($\mathcal{A}_3$) are $\#P$-hard.

Given a symmetric signature $f$, whether $f$ is $\mathcal{A}$- or $\mathcal{P}$-transformable is decidable in polynomial time in the size of the symmetric signature $f$ given as $n + 1$ numbers. We will discuss some necessary conditions in Section 7.9.

### 7.7.2 Dichotomies when $\mathcal{A}$- or $\mathcal{P}$-transformable Signatures Appear

Our characterizations of $\mathcal{A}$-transformable signatures in Lemma 7.43 and $\mathcal{P}$-transformable signatures in Lemma 7.46 are up to transformations in $O_2(\mathbb{C})$. Since an orthogonal transformation never changes the complexity of the problem, in the proofs of following lemmas, we assume any signature in $\mathcal{A}_i$ for $i = 1, 2, 3$, or $\mathcal{P}_j$ for $j = 1, 2$, is already in the canonical form. These lemmas state the general principle that if a set $\mathcal{F}$ of symmetric signatures contains a non-degenerate $\mathcal{A}$-transformable or a $\mathcal{P}$-transformable signature of arity $n \geq 3$, then the only way Holant($\mathcal{F}$) is tractable is when the entire set is $\mathcal{P}$-transformable or $\mathcal{A}$-transformable.

We first give a simple lemma to interpolate (=4).

**Lemma 7.48.** Let $a, b \in \mathbb{C}$. If $ab \neq 0$, then for any set $\mathcal{F}$ of complex-weighted signatures containing $[a, 0, \ldots, 0, b]$ of arity $m \geq 3$, 

$$\text{Holant}(\mathcal{F} \cup \mathcal{E}Q_2) \leq T \text{ Holant}(\mathcal{F}),$$

265
where $\mathcal{EQ}_2 = \\{=_2, =_4, \ldots, =_{2k}, \ldots\}$ is the set of all Equalities with an even arity.

The same reduction also works for planar graphs, thus $\text{Pl-Holant}(\mathcal{F} \cup \mathcal{EQ}_2) \leq_T \text{Holant}(\mathcal{F})$.

Proof. We only need to show how to get $(=_4)$. Once we have $(=_4)$, a self loop gives $(=_2)$, and a chain of $m$ copies of $(=_4)$ linked by one edge gives $(=_{2m+2})$.

Since $a \neq 0$, we can normalize the first entry to get $[1, 0, \ldots, 0, x]$, where $x \neq 0$. First, we show how to obtain an arity 4 Gen-Eq signature. If $m = 3$, then we connect two copies together by a single edge to get an arity 4 signature. For larger arities, we form self-loops until realizing a signature of arity 3 or 4. By this process, we have a signature $g = [1, 0, 0, 0, y]$, where $y \neq 0$. If $y$ is a $p$th root of unity, then we can directly realize $(=_4)$ by connecting $p$ copies of $g$ together, two edges at a time as in Figure 7.3. Otherwise, $y$ is not a root of unity and we can interpolate $(=_4)$ as follows.

Consider an instance $\Omega$ of $\text{Holant}(\mathcal{F} \cup \{=_4\})$. Suppose that $(=_4)$ appears $n$ times in $\Omega$. We stratify the assignments in $\Omega$ based on the assignments to $(=_4)$. We only need to consider the all-zero and all-one assignments since any other assignment contributes a factor of 0. We have

$$\text{Holant}_\Omega = \sum_{i=0}^{n} c_i,$$

where $c_i$ is the sum, over all assignments with exactly $i$ many $(=_4)$’s in $\Omega$ assigned 1111, and $n - i$ many $(=_4)$’s in $\Omega$ assigned 0000, of the product of evaluations of all other signatures on $\Omega$.

We construct from $\Omega$ a sequence of instances $\Omega_s$ of $\text{Holant}(\mathcal{F})$ indexed by $s \geq 1$. We obtain $\Omega_s$ from $\Omega$ by replacing each occurrence of $(=_4)$ with a gadget $g_s$ created from $s$ copies of $[1, 0, 0, 0, y]$, connecting two edges together at a time as in Figure 7.3. Then

$$\text{Holant}_{\Omega_s} = \sum_{i=0}^{n} (y^s)^i c_i.$$

For $s \geq 1$, this gives a coefficient matrix that is Vandermonde. Since $y$ is neither zero nor a root of unity, $y^s$ is distinct for each $s$. Therefore, the Vandermonde system $(0 \leq s \leq n)$ has full rank. We can solve for the unknowns $c_i$ in polynomial time and obtain the value of $\text{Holant}_{\Omega}$.

Lemma 7.49. Let $\mathcal{F}$ be a set of symmetric signatures. Suppose $\mathcal{F}$ contains a non-degenerate signature $f \in \mathcal{P}_1$ of arity $n \geq 3$. Then $\text{Holant}(\mathcal{F})$ is $\#P$-hard unless $\mathcal{F}$ is $\mathcal{P}$-transformable or $\mathcal{A}$-transformable.

Proof. By assumption, for any $0 \leq k \leq n - 2$, $f_{k+2} = f_k$ and $f_1 \neq \pm f_0$ since $f$ is not degenerate. We can express $f$ (after an orthogonal transformation) as

$$f = a_0 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes^n + a_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \otimes^n,$$

266
where \( a_0 = (f_0 + f_1)/2, a_1 = (f_0 - f_1)/2, \) and \( a_0 a_1 \neq 0. \) For this \( f, \) we can further perform an orthogonal transformation by \( H_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \) so that \( f \) is transformed into the \textsc{Gen-Eq} signature \([a_0, 0, \ldots, 0, a_1]\) of arity \( n, \) up to a nonzero constant. By Lemma 7.48, we can obtain \((=_4)\). With \((=_4)\), we can realize any equality signature of even arity. Thus, \( \#\text{CSP}^2(H_2F) \leq_T \text{Holant}(F). \)

Now we apply Theorem 6.16, the \( \#\text{CSP}^d \) dichotomy for \( d = 2, \) to the set \( H_2F. \) If \( \#\text{CSP}^2(H_2F) \) is \( \#P \)-hard, then \( \text{Holant}(F) \) is \( \#P \)-hard as well. Otherwise, this problem is tractable. Therefore, there exists some \( T \) of the form \([1 0 \ldots 0 a^k], \) where the integer \( k \in \{0, 1, \ldots, 7\}, \) such that \( TH_2F \) is a subset of \( \mathcal{A} \) or \( \mathcal{P}. \)

If \( TH_2F \subseteq \mathcal{P}, \) then we have \( H_2F \subseteq T^{-1} \mathcal{P}. \) Notice that \( T \in \text{Stab}(\mathcal{P}), \) so \( T^{-1} \mathcal{P} = \mathcal{P}. \) Thus, \( F \) is \( \mathcal{P} \)-transformable under this \( H_2 \) transformation. Otherwise, \( TH_2F \subseteq \mathcal{A}. \) It is easy to verify that \((=_2)((TH_2)^{-1})^{\otimes 2} \) is \([1, 0, i^{-k}] \in \mathcal{A}. \) Thus, \( F \) is \( \mathcal{A} \)-transformable under this \( TH_2 \) transformation. \( \square \)

Note that the hypothesis \( f \in \mathcal{P}_1 \) in Lemma 7.49 may seem to demand that all \( F \) is \( \mathcal{P} \)-transformable in order that the problem \( \text{Holant}(F) \) remain not \( \#P \)-hard. However this is not so. Since \( \mathcal{A}_1 \) is a proper subset of \( \mathcal{P}_1, \) the condition \( f \in \mathcal{P}_1 \) is consistent with possibly \( f \in \mathcal{A}_1. \) In that case \( F \) might be \( \mathcal{A} \)-transformable and not \( \mathcal{P} \)-transformable, and then \( \text{Holant}(F) \) is tractable (thus not \( \#P \)-hard if \( \#P \) does not collapse to \( P). \)

**Lemma 7.50.** Let \( F \) be a set of symmetric signatures. Suppose \( F \) contains a non-degenerate signature \( f \in \mathcal{P}_2 \) of arity \( n \geq 3. \) Then \( \text{Holant}(F) \) is \( \#P \)-hard unless \( F \) is \( \mathcal{P} \)-transformable or \( \mathcal{A} \)-transformable.

**Proof.** By assumption, for any \( 0 \leq k \leq n - 2, \) \( f_{k+2} = -f_k \) and \( f_1 \neq \pm if_0 \) since \( f \) is not degenerate. After an orthogonal transformation we can express \( f \) as

\[
f = a_0 \begin{bmatrix} 1 \\ i \end{bmatrix}^\otimes n + a_1 \begin{bmatrix} 1 \\ -i \end{bmatrix}^\otimes n,
\]

where \( a_0 = (f_0 + if_1)/2 \) and \( a_1 = (f_0 - if_1)/2, \) and \( a_0 a_1 \neq 0. \) Then under the holographic transformation \( Z' = \begin{bmatrix} \psi a_0 & \psi a_1 \\ i \psi a_0 & -i \psi a_1 \end{bmatrix}^{-1}, \) we have

\[
Z'^\otimes n f = (=_n) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\otimes n + \begin{bmatrix} 0 \\ 1 \end{bmatrix}^\otimes n
\]

and

\[
\text{Holant}(=:_2 \mid F \cup \{f\}) \equiv_T \text{Holant}([1, 0, 1](Z'^{-1})^{\otimes 2} \mid Z'F \cup \{Z'^\otimes n f\}) \equiv_T \text{Holant}([0, 1, 0] \mid Z'F \cup \{=_n\}).
\]

Thus, we have a bipartite graph having \((=_n)\) together with \( Z'F \) on the right and \((\neq_2) = [0, 1, 0]\) on the left up to a nonzero scalar. Inductively we can realize \((=_{kn})\) on the right
for every integer \( k > 0 \). Suppose we have already constructed \((=_{(k-1)n})\). Place one copy of \((=_{n})\) and three copies of \((=_{n})\) on the right, and call them \(A\), \(B\) and \(C\). Connect \((n-2)\) edges of \(A\) to \((=_{(k-1)n})\), one edge of \(A\) to one edge of \(B\), and one edge of \(A\) to one edge of \(C\), with all connections via \((\neq_2)\) on the left. This creates a \((=_{kn})\) on the right. Moreover, we can move these EQUALITY signatures to the left side by connecting each edge of \((=_{kn})\) by a copy of \((\neq_2)\). Thus, \(#\text{CSP}^n(Z'F) \leq_T \text{Holant}(F)\).

Now we apply Theorem 6.16, the \(#\text{CSP}^d\) dichotomy with \(d = n\), to the set \(Z'F\). Note that this \(n\) is a constant independent of the input size of any signature grid \(\Omega\). If \(#\text{CSP}^n(Z'F)\) is \#P-hard, then \(\text{Holant}(F)\) is \#P-hard as well. Otherwise, \(#\text{CSP}^n(Z'F)\) is tractable. Let \(\omega\) be a primitive \(4n\)-th root of unity. Then under the holographic transformation \(T = \begin{bmatrix} 1 & 0 \\ 0 & \omega^k \end{bmatrix}\) for some integer \(k\), \(TZ'F \subseteq \mathcal{P}\) or \(\mathcal{A}\). If \(TZ'F \subseteq \mathcal{P}\), then we have \(Z'F \subseteq T^{-1}\mathcal{P}\). Notice that \(T \in \text{Stab}(\mathcal{P})\), so \(T^{-1}\mathcal{P} = \mathcal{P}\). Also recall that \((=_2)(Z'^{-1})^{\otimes 2}\) is a multiple of \((\neq_2)\). Thus, \(F\) is \(\mathcal{P}\)-transformable under this \(Z'\) transformation.

Otherwise, \(TZ'F \subseteq \mathcal{A}\). It is easy to verify that \((=_2)((TZ')^{-1})^{\otimes 2}\) is \([0, 1, 0]\) up to a scalar. Thus, \(F\) is \(\mathcal{A}\)-transformable under this \(TZ'\) transformation.

\[\text{Lemma 7.51.} \text{ Let } F \text{ be a set of symmetric signatures. Suppose } F \text{ contains a non-degenerate signature } f \in \mathcal{A}_3 \text{ of arity } n \geq 3. \text{ Then } \text{Holant}(F) \text{ is } \#P\text{-hard unless } F \text{ is } \mathcal{A}\text{-transformable.}\]

\[\text{Proof.} \text{ After an orthogonal transformation we can express } f \text{ as } f = \lambda \left( \left[ \frac{1}{\alpha} \right]^{\otimes n} + i^r \left[ \frac{1}{-\alpha} \right]^{\otimes n} \right),\]

for some \(\lambda \neq 0\) and integer \(r\). In this canonical form, we have \(f_{k+2} = if_k\) for any \(0 \leq k \leq n-2\).

A self loop on \(f\) yields \(f'\), where \(f_k' = f_k + f_{k+2} = (1+i)f_k\). Thus up to the constant \((1+i)\), \(f'\) is just the first \(n-2\) entries of \(f\). By doing zero or more self loops, we eventually obtain an arity 4 signature when \(n\) is even or a ternary one when \(n\) is odd. There are eight cases depending on the first two entries of \(f\) and the parity of \(n\). However, in each case, we can realize the signature \([1, 0, i]\). We list them here. (In the calculations below, we omit certain nonzero constant factors without explanation.)
\[
M = \begin{bmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0
\end{bmatrix}
\]

- [0, 1, 0, i]: Another self loop gives [0, 1]. Connect it back to the ternary we get [1, 0, i].
- [1, 0, i, 0]: Another self loop gives [1, 0]. Connect it back to the ternary we get [1, 0, i].
- [1, αi, i, −α]: Another self loop gives [1, αi]. Connect two copies of it to the ternary we get [1, −α]. Then connect this back to the ternary we finally get [1, 0, i]. See Figure 7.13a.
- [1, −αi, i, α]: Same construction as in the previous case.
- [0, 1, 0, i, 0]: Another self loop gives [0, 1, 0]. Connect it back to the arity 4 signature we get [1, 0, i].
- [1, 0, i, 0, −1]: Another self loop gives [1, 0, i] directly.
- [1, αi, i, −α, −1]: Another self loop gives [1, αi, i]. Connect two copies of it together we get [1, −α, −i]. Connect this back to the arity 4 signature we get [1, 0, i]. See Figure 7.13b.
- [1, −αi, i, α, −1]: Same construction as in the previous case.

With [1, 0, i] in hand, we can connect three copies to get [1, 0, −i]. Now we construct a bipartite graph, with \(F \cup \{=2\}\) on the right side and [1, 0, −i] on the left. This problem is reducible to Holant(\(F\)). We perform a holographic transformation by \(M^{-1} = \frac{1}{2α} \begin{bmatrix} 1 & 1 & 1 \\ -α & -α & -α \end{bmatrix}\), with \(M = \begin{bmatrix} α & 1 \\ 1 & -1 \end{bmatrix}\), and get

\[
\text{Holant}([1, 0, −i] \mid F \cup \{f, =2\}) \equiv_T \text{Holant}([1, 0, -i](M^{-1})^{\otimes 2} \mid MF \cup \{M^{\otimes n}f, M^{\otimes 2}(=2)\})
\equiv_T \text{Holant}([1, 0, 1] \mid MF \cup \{[1, 0, ..., 0, it^\prime], [1, i, 1]\})
\equiv_T \text{Holant} (MF \cup \{[1, 0, ..., 0, it^\prime], [1, i, 1]\}).
\]

Note that \(M^{\otimes n}f\) is \([1, 0, ..., 0, it^\prime]\) (after normalizing the first entry); similarly \(M^{\otimes 2}(=2)\) is \([1, i, 1]\). On the left side, \([1, 0, −i](M^{-1})^{\otimes 2}\) is a nonzero multiple of \([1, 0, 1] \equiv_T (=2)\). Therefore, we can construct all \(\text{EQUALITY}\) signatures of even arity as follows: Use 4 copies of \([1, 0, ..., 0, it^\prime]\) of arity \(n\) linked as a chain by \(3\) edges we get \((=4n−6)\) with even arity \(4n−6 \geq 6\). Then we can get \((=2)\) and \((=4)\) by self loops. Finally connecting \(k\) copies of \((=4)\) as a chain by \(k−1\) edges gives \((=4k−2(k−1)) \equiv (=2k+2)\) for all \(k \geq 2\). Thus,

\[#\text{CSP}^d(MF \cup \{[1, i, 1]\}) \leq_T \text{Holant}(F)\]

Now we apply Theorem 6.16, the \(\#\text{CSP}^d\) dichotomy for \(d = 2\), to the set \(MF \cup \{[1, i, 1]\}\). If \(#\text{CSP}^2(MF \cup \{[1, i, 1]\})\) is \#P-hard, then \(\text{Holant}(F)\) is \#P-hard as well. Otherwise, \(#\text{CSP}^2(MF \cup \{[1, i, 1]\})\) is tractable. Therefore, there exists some \(T\) of the form \(\begin{bmatrix} 1 & 0 \\ 0 & αd \end{bmatrix}\), where the integer \(d \in \{0, 1, ..., 7\}\), such that \(TMF \cup \{T^{\otimes 2}[1, i, 1]\} \subseteq \mathcal{A}\) or \(\mathcal{P}\).

However, \(T^{\otimes 2}[1, i, 1]\) cannot be in \(\mathcal{P}\) for a diagonal \(T\). Thus \(TMF \cup \{T^{\otimes 2}[1, i, 1]\} \subseteq \mathcal{A}\). Further notice that if \(d \in \{1, 3, 5, 7\}\) in the expression of \(T\), then \(T^{\otimes 2}[1, i, 1]\) is not in \(\mathcal{A}\). Hence, \(T\) must be of the form \(\begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix}\), where the integer \(d \in \{0, 1, 2, 3\}\). For such \(T\), \(T^{\otimes 2}[1, i, 1] \in \mathcal{A}\), and \(T^{-1} \mathcal{A} = \mathcal{A}\) as \(T \in \text{Stab}(\mathcal{A})\). Thus, \(TMF \cup \{T^{\otimes 2}[1, i, 1]\} \subseteq \mathcal{A}\) simply becomes \(MF \subseteq \mathcal{A}\). Moreover, \((=2)(M^{-1})^{\otimes 2}\) is a nonzero multiple of \([1, −i, 1] \in \mathcal{A}\). Therefore, \(F\) is \(\mathcal{A}\)-transformable under the transformation by \(M = \begin{bmatrix} 0 & 1 \\ α & -1 \end{bmatrix}\).
7.8 Proof of Theorem 7.19

In this section, we finish the proof of the main theorem of this Chapter, Theorem 7.19. We begin with a dichotomy for a single signature, which we prove by induction on its arity.

**Theorem 7.52.** If $f$ is a non-degenerate symmetric signature of arity at least 3 with complex weights in Boolean variables, then Holant($f$) is $\#P$-hard unless $f \in \mathcal{P} \cup \mathcal{P}_2 \cup \mathcal{A}_3$ or $f$ is vanishing, in which case the problem is in $P$.

Recall that $\mathcal{A}_1 \subset \mathcal{P}_1$ and $\mathcal{A}_2 = \mathcal{P}_2$. Thus $f \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3$ iff $f$ is $\mathcal{A}$-transformable or $\mathcal{P}$-transformable by Corollary 7.47.

**Proof.** Let the arity of $f$ be $n$. The base cases of $n = 3$ and $n = 4$ are proved in Theorem 7.29. Now assume $n \geq 5$.

Let $f'$ be the signature of arity at least 3 obtained from $f$ by forming a self loop. We consider the cases separately whether $f'$ is degenerate or not.

- Suppose $f' = [a, b]^{\otimes (n-2)}$ is degenerate. There are three cases to consider.
  1. If $a = b = 0$, then $f'$ is the all-zero signature. For $f$, this means $f_{k+2} = -f_k$ for $0 \leq k \leq n - 2$, so $f \in \mathcal{A}_2 = \mathcal{P}_2$ by Lemma 7.41, and therefore Holant($f$) is tractable.
  2. If $a^2 + b^2 \neq 0$, then $f'$ is nonzero and $[a, b]$ is not a constant multiple of either $[1, i]$ or $[1, -i]$. We may normalize so that $a^2 + b^2 = 1$. Then the orthogonal transformation $\begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix}$ transforms the column vector $[a, b]$ to $[1, 0]$. Let $\hat{f}$ be the transformed signature from $f$, and $\hat{f}' = [1, 0]^{\otimes (n-2)}$ the transformed signature from $f'$.

Since an orthogonal transformation keeps $(\_ \_ \_ \_)$ invariant, this transformation commutes with the operation of taking a self loop, i.e., $\hat{f}' = (\hat{f})'$. Here $(\hat{f})'$ is the function obtained from $\hat{f}$ by taking a self loop. So $\hat{f}_0 + \hat{f}_2 = 1$ and for every integer $1 \leq k \leq n - 2$, we have $\hat{f}_k = -\hat{f}_{k+2}$. With one or more self loops, we eventually obtain either $[1, 0]$ when $n$ is odd or $[1, 0, 0]$ when $n$ is even. In either case, we connect an appropriate number of copies of this signature to $\hat{f}$ to get an arity 4 signature $\hat{g} = [f_0, f_1, f_2, -\hat{f}_1, -\hat{f}_2]$. We show that Holant($\hat{g}$) is $\#P$-hard.

To see this, we first compute $\det(\hat{M}_0) = -2(\hat{f}_0 + \hat{f}_2)(\hat{f}_1^2 + \hat{f}_2^2) = -2(\hat{f}_1^2 + \hat{f}_2^2)$, since $\hat{f}_0 + \hat{f}_2 = 1$. Therefore if $\hat{f}_1^2 + \hat{f}_2^2 \neq 0$, Holant($\hat{g}$) is $\#P$-hard by Corollary 7.25. Otherwise $\hat{f}_1^2 + \hat{f}_2^2 = 0$, and we consider $\hat{f}_2 = i\hat{f}_1$ since the other case $\hat{f}_2 = -i\hat{f}_1$ is similar. Since $f$ is non-degenerate, $\hat{f}$ is non-degenerate, which implies $\hat{f}_2 \neq 0$.

We can express $\hat{g}$ as $[1, 0]^{\otimes 4} - \hat{f}_2[1, i]^{\otimes 4}$. Under the holographic transformation by $T = \begin{bmatrix} 1 & (-f_2)^{1/4} \\ 0 & i(-f_2)^{1/4} \end{bmatrix}$, we have

$$\text{Holant}(\_ \_ \_ \_ \_ | \hat{g}) = T \text{ Holant}([1, 0, 1]T^{\otimes 2} | (T^{-1})^{\otimes 4}\hat{g}) = T \text{ Holant}(\_ \_ \_ | =_4),$$

270
where
\[ \hat{h} = [1, 0, 1] T^{\otimes 2} = [1, (-\hat{f}_2)^{1/4}, 0] \]
and \( \hat{g} \) is transformed by \( T^{-1} \) into the arity 4 equality \((=_4)\), since
\[
T^{\otimes 4} \left( \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \otimes 4 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \otimes 4 \right) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \hat{f}_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes 4 = \hat{g}.
\]

By Theorem 5.3 for the case \( k = 4 \) (which is proved in Chapter 5), Holant\( (\hat{h} | =_4) \) is \#P-hard as \( \hat{f}_2 \neq 0 \).

3. If \( a^2 + b^2 = 0 \) but \((a, b) \neq (0, 0)\), then \([a, b]\) is a nonzero multiple of \([1, \pm i]\). We may normalize and have \( f' = [1, i]^{\otimes (n-2)} \) or \([1, -i]^{\otimes (n-2)}\). We consider the first case since the other case is similar.

In the first case, the characteristic polynomial of the recurrence relation of \( f' \) is \( x - i \), so that of \( f \) is \( (x - i)(x^2 + 1) = (x - i)^2(x + i) \). Hence there exist \( a_0, a_1, \) and \( c \) such that
\[
f_k = (a_0 + a_1 k)i^k + c(-i)^k \tag{7.15}
\]
for every integer \( 0 \leq k \leq n \). If \( a_1 = 0 \), then \( f' \) is the all zero signature, a contradiction. If \( c = 0 \), then \( f \) is vanishing, one of the tractable cases. Now we assume \( a_1 c \neq 0 \) and show that Holant\( (f) \) is \#P-hard. Under the holographic transformation \( Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \), we have
\[
\text{Holant}(=_2 | f) \equiv_T \text{Holant}([1, 0, 1] Z^{\otimes 2} | (Z^{-1})^{\otimes n} f)
\equiv_T \text{Holant}([0, 1, 0] | \hat{f}),
\]
where \( \hat{f} = (Z^{-1})^{\otimes n} f \) takes the form \([\hat{f}_0, \hat{f}_1, 0, \ldots, 0, \hat{f}_n] \) with \( \hat{f}_1, \hat{f}_n \neq 0 \). This follows from Lemma 7.17 and linearity since \( \hat{f} \) is the \( Z^{-1} \)-transformation of the sum of two signatures with recurrence degrees \( r^+ \) and \( r^- \) equal to 1 and 0 respectively. Furthermore, by the normalization of \( f' = Z^{\otimes (n-2)} [\hat{f}_1, 0, \ldots, 0] = [1, i]^{\otimes (n-2)} \), we have normalized \( \hat{f}_1 = 1 \). Similarly \( \hat{f}_n = 2^{n/2} \). On the other side, \((=_2) = [1, 0, 1]\) is transformed into \((\neq_2) = [0, 1, 0]\).

Now consider the gadget in Figure 7.14a with \( \hat{f} \) assigned to both vertices, which clearly has a symmetric signature. The Hamming weight 0 entry of this binary signature is 0. This is because the arity of \( \hat{f} \) is \( n \geq 5 \) and so either the left or the right side copy of \( \hat{f} \) has at least two incident edges assigned 1 as they are linked by \((\neq_2)\) signatures. But \( \hat{f} \) takes value 0 for input of Hamming weight \( w \) if \( 2 \leq w < n \). Similarly one can see that for input 01, the only nonzero configuration is all incident edges to the left side copy of \( \hat{f} \) take 0, and those of the right take 1, which has value \( \hat{f}_0 \hat{f}_n \). Finally for input 11, we have two nonzero configurations and get \( 2\hat{f}_n \) as \( \hat{f}_1 = 1 \). Thus the binary signature of this gadget is \([0, \hat{f}_0, \hat{f}_n, 2\hat{f}_n]\), which is equivalent to \([0, \hat{f}_0, 2] \) since \( \hat{f}_n \neq 0 \).

271
Translating back by $Z$ to the original setting,
\[
\text{Holant}((\neq_2) \mid [0, \hat{f}_0, 2], \hat{f}) \equiv_T \text{Holant}((\neq_2)(Z^{-1})^{\otimes 2} \mid Z^{\otimes 2}[0, \hat{f}_0, 2], Z^{\otimes n}\hat{f}) \\
\equiv_T \text{Holant}((=_2) \mid g, f)
\]
where $g = Z^{\otimes 2}[0, \hat{f}_0, 2] = [\hat{f}_0 + 1, -i, \hat{f}_0 - 1]$. This can be verified as
\[
\frac{1}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} 0 & \hat{f}_0 \\ i & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^T = \begin{bmatrix} \hat{f}_0 + 1 & -i \\ -i & \hat{f}_0 - 1 \end{bmatrix}.
\]
It can be directly checked that $g \notin \mathcal{R}_2^+$. If $\hat{f}_0 \neq 0$, then by the matrix product form clearly $g$ is non-degenerate. In this case we construct some function in $\mathcal{V}^+$. First we connect $f'$ back to $f$, getting a binary signature $Z^{\otimes 2}[0, 0, \hat{f}_n]$. This is best seen in the $Z$ basis, where we connect all $n - 2$ edges of $[1, 0]^{\otimes (n-2)}$ via $(\neq_2)$ to $\hat{f}$. So up to a nonzero factor we have $Z^{\otimes 2}[0, 0, 1]$. Then we connect both edges of $Z^{\otimes 2}[0, 0, 1]$ to $f$, the resulting signature is $h = Z^{\otimes n-2}[\hat{f}_0, 1, 0, \ldots, 0]$ of arity $n - 2 \geq 3$, again best seen in the $Z$ basis. Notice that $h$ is non-degenerate and $h \in \mathcal{V}^+$. By Lemma 7.33, Holant($\{h, g\}$) is #P-hard, hence Holant($f$) is also #P-hard.

Otherwise suppose $\hat{f}_0 = 0$. Then we have $g = [1, -i]^{\otimes 2}$. Connecting this degenerate signature to $f$, we get a signature $h = \langle f, g \rangle$. We note that $g$ annihilates the signature $c[1, -i]^{\otimes n}$, and thus $h = \langle f^*, g \rangle$, where $f^*$ is the first summand of $f$ in (7.15), i.e., $f^*_k = (a_0 + a_1k)k^2 (0 \leq k \leq n)$. Then $\text{rd}^+(f^*) = 1$, $\text{vd}^+(g) = 0$, and we can apply Lemma 7.14. It follows that $\text{rd}^+(h) = 1$ and $\text{arity}(h) \geq 3$. This implies that $h$ is non-degenerate and $h \in \mathcal{V}^+$. Moreover, assigning $f$ to both vertices in the gadget of Figure 7.14b, we get a non-degenerate signature $h' \in \mathcal{V}^-$ of arity 4. To see this, consider this gadget after a holographic transformation by $Z$. In this bipartite setting, it is the same as assigning $\hat{f} = [0, 1, 0, \ldots, 0, \hat{f}_n]$ to both the circle and triangle vertices in the gadget of Figure 7.8a. The square vertices there are still assigned $(\neq_2) = [0, 1, 0]$. While it is not apparent from the gadget’s geometry, this signature is in fact symmetric. In particular, its values on inputs 1010 and 1100 are both zero. Note
that as \( n \geq 5 \), the number of parallel paths with \((\neq 2)\) in the middle between the two copies of \( \hat{f} \) is \( n - 2 \geq 3 \), and thus at least one copy of \( \hat{f} \) has two internal incident edges set to 1 due to the \((\neq 2)\) in the middle. For Hamming weight at least 2, the only nonzero value of \( \hat{f} \) is at weight \( n \). From this, one can see that the resulting signature is \( \hat{h'} = (Z^{-1})^{\otimes 4}h' = [0, 0, 0, \hat{f}_n, 0] \). Hence \( \text{rd}^{-}(h') = 1 \) by Lemma 7.17. Therefore \( h' \) is non-degenerate and \( h' \in \mathcal{V}^{-} \).

By Lemma 7.34, \( \text{Holant}([h, h']) \) is \#P-hard, hence \( \text{Holant}(f) \) is also \#P-hard.

- Suppose \( f' \) is non-degenerate. If \( \text{Holant}(f') \) is \#P-hard then so is \( \text{Holant}(f) \). We now assume \( \text{Holant}(f') \) is not \#P-hard. Then, by inductive hypothesis, \( f' \in \mathcal{P}_{1} \cup \mathcal{P}_{2} \cup \mathcal{A}_{3} \) or \( f' \) is vanishing. If \( f' \in \mathcal{P}_{1} \cup \mathcal{P}_{2} \cup \mathcal{A}_{3} \), then applying Lemma 7.49, Lemma 7.50, or Lemma 7.51 to \( f' \) and the set \( \{f, f'\} \), we either have that \( \text{Holant}([f, f']) \) is \#P-hard, so \( \text{Holant}(f) \) is \#P-hard as well, or that \( f \) is \( \mathcal{A} \)- or \( \mathcal{P} \)-transformable, so by Corollary 7.47, \( f \in \mathcal{P}_{1} \cup \mathcal{P}_{2} \cup \mathcal{A}_{3} \).

Otherwise, \( f' \) is vanishing, so \( f' \in \mathcal{V}^{\sigma} \) for \( \sigma \in \{+, -\} \) by Theorem 7.13. We may assume that \( f' \in \mathcal{V}^{+} \); the other case is similar. Let \( \text{rd}^{+}(f') = d - 1 \). Then \( f' \in \mathcal{V}^{+} \) implies that \( 2d < n \) and \( f' \) is non-degenerate implies that \( d \geq 2 \). The entries of \( f' \) can be expressed as

\[
f_{k'} = i^{k} q(k),
\]

where \( q(x) \) is a polynomial of degree exactly \( d - 1 \). Since \( f' \in \mathcal{R}_{d}^{+} \), \( f' \) satisfies a recurrence relation with characteristic polynomial \((x - i)^{d}\). Then \( f \) satisfies a recurrence relation with characteristic polynomial \((x^{2} + 1)(x - i)^{d} = (x - i)^{d+1}(x + i)\), and thus the entries of \( f \) are

\[
f_{k} = i^{k} p(k) + c(-i)^{k}
\]

for some constant \( c \) and a polynomial \( p(x) \) of degree at most \( d \). However, the degree of \( p(x) \) is exactly \( d \), otherwise the polynomial \( q(x) \) for \( f' \) would have degree less than \( d - 1 \). If \( c = 0 \), then \( f \) is vanishing, a tractable case. Now assume \( c \neq 0 \), and we want to show the problem is \#P-hard.

Under the transformation \( Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \), we have

\[
\text{Holant}([=_{2} | f]) \equiv_{T} \text{Holant}(([=_{2}Z^{\otimes 2} | ((Z^{-1})^{\otimes n} f))
\equiv_{T} \text{Holant}(\neq_{2} | \hat{f}),
\]

where \( \hat{f} = (Z^{-1})^{\otimes n} f = [f_{0}, f_{1}, \ldots, f_{d}, 0, \ldots, 0, f_{n}] \), with \( \hat{f}_{d} \neq 0 \) and \( \hat{f}_{n} = 2^{n/2}c \neq 0 \). Note that \( d < n/2 \) and \( \hat{f}_{d} \) occurs strictly in the first half, and there are \( n - d - 1 \) many 0's between \( \hat{f}_{d} \) and \( \hat{f}_{n} \). Taking a self loop in the original setting is equivalent to connecting both edges of \((\neq 2)\) to two inputs of a signature after this transformation. Thus, doing this once on \( \hat{f} \), we can get \( \hat{f}' = 2[\hat{f}_{1}, \ldots, \hat{f}_{d}, 0, \ldots, 0] \) corresponding to \( f' \), namely \((Z^{-1})^{\otimes(n-2)} f' \). Doing this \( d - 2 \) times on \( \hat{f} \), up to a nonzero factor, we get a signature \( \hat{h} = [\hat{f}_{d-2}, \hat{f}_{d-1}, \hat{f}_{d}, 0, \ldots, 0/f_{n}] \) of arity \( n - 2(d - 2) = n - 2d + 4 \). The last entry is \( \hat{f}_{n} \) when \( d = 2 \) and is 0 when \( d > 2 \).
As $n > 2d$, we may do this two more times and get $[\hat{f}_{d}, 0, \ldots, 0]$ of arity $k = n - 2d$. Now connect this signature back to $\hat{f}$ via $(\neq_2)$. We get a signature of arity $n - k$. This signature consists of the last $n - k + 1 = 2d + 1$ signature entries of $\hat{f}$, up to a nonzero factor. We may repeat this operation zero or more times, which reduces the arity of $\hat{f}$ by $k$ each time, until the arity $k'$ of the resulting signature is less than or equal to $k$.

We claim that this signature has the form $\hat{g} = [0, \ldots, 0, \hat{f}_n]$. In other words, the $k' + 1$ entries of $\hat{g}$ consist of the last $\hat{f}_n$ and $k'$ many 0’s in the signature $\hat{f}$, all appearing after $\hat{f}_d$. This is because there are $n - d - 1$ many 0 entries in the signature $\hat{f}$ after $\hat{f}_d$, and $n - d - 1 \geq k \geq k'$.

Translating back by the $Z$ transformation, having both $[\hat{f}_d, 0, \ldots, 0]$ of arity $k$ and $\hat{g} = [0, \ldots, 0, \hat{f}_n]$ of arity $k'$ is equivalent to, in the original setting, having both $[1, i]^{\otimes k}$ and $[1, -i]^{\otimes k'}$, up to a nonzero factor. If $k > k'$, then we can connect $[1, -i]^{\otimes k'}$ to $[1, i]^{\otimes k}$ and get $[1, i]^{\otimes (k - k')}$. Replacing $k$ by $k - k'$ we can repeat this process until the new $k \leq k'$. If the new $k < k'$ we can continue as in the subtractive Euclid algorithm. Keep doing this procedure and eventually we get $[1, i]^{\otimes t}$ and $[1, -i]^{\otimes t}$, where $t = \gcd(k, k')$, and $k' \leq k = n - 2d$ are defined in the previous paragraph. Now putting $k/t$ many copies of $[1, -i]^{\otimes t}$ together, we get $[1, -i]^{\otimes k}$.

In the transformed setting, $[1, -i]^{\otimes k}$ is $[0, \ldots, 0, 1]$ of arity $k$. Then we connect this back to $\hat{h} = [\hat{f}_{d - 2}, \hat{f}_{d - 1}, \hat{f}_d, 0, \ldots, 0, 0/\hat{f}_n]$ via $(\neq_2)$. Doing this is the same as forcing $k$ connected edges of $\hat{h}$ be assigned 0, because $(\neq_2)$ flips the assigned value 1 in $[0, \ldots, 0, 1]$ to 0. Thus we get a signature of arity $n - 2d + 4 = k - 4$, which is $[\hat{f}_{d - 2}, \hat{f}_{d - 1}, \hat{f}_d, 0, 0]$.

Note that the last entry is 0 (and not $\hat{f}_n$), because $k \geq 1$.

However, Holant$(\neq_2 \mid [\hat{f}_{d - 2}, \hat{f}_{d - 1}, \hat{f}_d, 0, 0])$ is equivalent to Holant$(\neq_2 \mid [0, 0, 1, 0, 0])$ when $\hat{f}_d \neq 0$. This is the Eulerian Orientation problem #EO on 4-regular graphs and is #P-hard by Theorem 6.29.

Finally we are ready to give the hardness proof in Theorem 7.19. We prove the contrapositive: If Holant($\mathcal{F}$) is not #P-hard then $\mathcal{F}$ must be in one of the five cases listed in Theorem 7.19. Combined with the tractability proof given in Section 7.2, this completes the proof of Theorem 7.19.

**Proof of hardness in Theorem 7.19.** Assume that Holant($\mathcal{F}$) is not #P-hard. If all of the non-degenerate signatures in $\mathcal{F}$ are of arity at most 2, then it is tractable case 1. Otherwise we have some non-degenerate signatures of arity at least 3. For any such $f$, by Theorem 7.52, $f \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_0$ or $f$ is vanishing. If any of them is in $\mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_0$, then by Lemma 7.49, Lemma 7.50, or Lemma 7.51, we have that $\mathcal{F}$ is $\mathcal{A}$- or $\mathcal{P}$-transformable, which are tractable cases 2 and 3.

Now we assume all non-degenerate signatures of arity at least 3 in $\mathcal{F}$ are vanishing, and there is a nonempty set of such signatures in $\mathcal{F}$. By Lemma 7.34, they must all be in $\mathcal{V}^\sigma$ for the same $\sigma \in \{+, -\}$. By Lemma 7.33, we know that any non-degenerate binary signature in $\mathcal{F}$ must belong to $\mathcal{R}_2^\sigma$. Furthermore, if there is an $f \in \mathcal{V}^\sigma$ from $\mathcal{F}$ such that
rdF(f) ≥ 2, then by Lemma 7.31, the only unary signature that is allowed in F is [1, σ], and all degenerate signatures in F are a tensor power of [1, σ], up to a nonzero factor. Thus, all non-degenerate signatures of arity at least 3 as well as all degenerate signatures belong to \( V^\sigma \), and all non-degenerate binary signatures belong to \( R^\sigma_2 \). This is tractable case 4.

Finally, we have the following: (i) all non-degenerate signatures of arity at least 3 in F belong to \( V^\sigma \); (ii) all signatures \( f \in F \cap V^\sigma \) have \( rdF(f) ≤ 1 \), which implies that \( f \in R^\sigma_2 \); and (iii) all non-degenerate binary signatures in F belong to \( R^\sigma_2 \). Hence all non-degenerate signatures in F belong to \( R^\sigma_2 \). All unary signatures also belong to \( R^\sigma_2 \) by definition. This is indeed tractable case 5. The proof is complete.

7.9 Decidability of the Dichotomy

Theorem 7.19 is a full dichotomy for Holant(F) for a set of symmetric signatures on Boolean variables. It says that if F satisfies one of the tractability conditions then Holant(F) is computable in P, otherwise it is \#P-hard. A natural question is given a finite set F, how can one decide which case it is. This is the decidability question of the dichotomy theorem.

Clearly conditions 1, 4 and 5 are all easily decidable in polynomial time. Hence we may assume F contains at least one non-degenerate signature of arity at least 3, and the real question is how can one decide whether F is \( \mathcal{A} \)- or \( \mathcal{P} \)-transformable. This problem is non-trivial because it involves a potentially infinite set of transformational matrices.

**Theorem 7.53.** Given any finite set F of symmetric signatures with complex weights in Boolean variables containing a non-degenerate signature f of arity at least 3, it is decidable in polynomial time in the input size of F whether F is \( \mathcal{A} \)- or \( \mathcal{P} \)-transformable. Thus the tractability criterion of Theorem 7.19 is decidable in polynomial time.

In this book we will not give the full proof of this theorem; interested readers can find it in [?]. In this section we discuss some necessary conditions of the tractability criterion based on some structural properties, that both shed new light on the dichotomy and also help in many practical cases to test the criterion on particular functions. The decision algorithm of Theorem 7.53 is built on top of these ideas with some additional technical extensions. We will omit some technical proofs in this section.

By Corollary 7.47, f is \( \mathcal{A} \)- or \( \mathcal{P} \)-transformable iff \( f \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \), or \( f \in \mathcal{P}_1 \cup \mathcal{P}_2 \), where \( \mathcal{A}_1 \subseteq \mathcal{P}_1 \) and \( \mathcal{A}_2 = \mathcal{P}_2 \). We say a signature \( f = [f_0, f_1, \ldots, f_n] \) satisfies a second order recurrence relation, if for all \( 0 ≤ k ≤ n − 2 \), there exist \( a, b, c \in \mathbb{C} \) not all zero, such that \( af_k − bf_{k+1} + cf_{k+2} = 0 \). We say the recurrence has type \( \langle a, b, c \rangle \). The characteristic equation of the recurrence relation is \( a − bx + cx^2 = 0 \). The discriminant is \( b^2 − 4ac \).

**Lemma 7.54.** Let f be a non-degenerate symmetric signature of arity \( n ≥ 3 \). If f satisfies a second order recurrence relation then the recurrence relation is unique up to a nonzero multiple.
Under any holographic transformation, the property that a signature satisfies a second order recurrence relation, and if so, the condition $b^2 - 4ac \neq 0$ are both invariant.

**Lemma 7.55.** Let $f$ be a symmetric signature of arity $n$ and $\mathring{f} = T^\otimes n f$ for some $T \in \text{GL}_2(\mathbb{C})$. Then $f$ satisfies a second order recurrence relation iff $\mathring{f}$ does. Furthermore, the discriminant for $f$ is nonzero iff the one for $\mathring{f}$ is nonzero.

**Proof.** If we write the $2^2 \times 2^{n-2}$ signature matrix of $f$ as $M_f$, with rows indexed by the first two input bits and the rest are the other $n - 2$ input bits, as in Definition 1.4, then $(a,-b/2,-b/2,c)$ is the solution of $(x, y, z) M_f = 0$. The corresponding signature matrix of $T^\otimes n f$ is $T^\otimes 2 M_f (T^T)^{(n-2)}$, and thus $(a,-b/2,-b/2,c) (T^{-1})^\otimes 2$ is the solution of $(x', y', z') M_{T^\otimes n f} = 0$. The discriminant of $f$ is $\det \begin{bmatrix} a & -b/2 \\ -b/2 & c \end{bmatrix}$, and that of $T^\otimes n f$ is multiplied by $(\det T)^2$. 

A simple but important observation is that if $f$ is $A$- or $P$-transformable, then $f$ satisfies a second order recurrence relation with a nonzero discriminant. In fact in their canonical forms, $f \in A_1, A_2$ and $A_3$ have types $\langle -1,0,1 \rangle, \langle 1,0,1 \rangle$ and $\langle -i,0,1 \rangle$, with characteristic polynomials $x^2 - 1$, $x^2 + 1$ and $x^2 - i$, respectively. For $f \in P_1$ and $P_2$, the types are $\langle -1,0,1 \rangle$ and $\langle 1,0,1 \rangle$. By Lemma 7.55, any $f \in P_1 \cup A_2 \cup A_3$ satisfies a second order recurrence relation with a nonzero discriminant. Hence,

**Lemma 7.56.** Let $f \in P_1 \cup A_2 \cup A_3$ be a non-degenerate symmetric signature of arity $n \geq 3$. Then $f$ satisfies a second order recurrence relation with nonzero discriminant. It has the form $f = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix}^\otimes n + \begin{bmatrix} \beta \\ \delta \end{bmatrix}^\otimes n$ where $\det \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \neq 0$.

The following definition of the $\theta$ function is crucial.

**Definition 7.57.** For a pair of linearly independent vectors $u = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix}$ and $v = \begin{bmatrix} \beta \\ \delta \end{bmatrix}$, we define

$$\theta(u,v) = \left( \frac{\langle u,v \rangle}{\det[u,v]} \right)^2 = \left( \frac{\alpha \beta + \gamma \delta}{\alpha \delta - \beta \gamma} \right)^2,$$

where $\langle u,v \rangle$ is the dot product, and $\det[u,v]$ is the determinant with column vectors $u$ and $v$. Furthermore, if $f = u^\otimes n + v^\otimes n$ for linearly independent $u$ and $v$ and $n \geq 3$, then we define $\theta(f) = \theta(u,v)$.

By linear independence, $\det[u,v] \neq 0$, so $\theta(u,v)$ is well-defined.

**Lemma 7.58.** For two linearly independent vectors $u, v \in \mathbb{C}^2$, and non-zero $\lambda$ and $\mu \in \mathbb{C}$, $\theta(\lambda u, \mu v) = \theta(u,v)$. If $H \in O_2(\mathbb{C})$, let $\mathring{u} = Hu$ and $\mathring{v} = Hv$. Then $\theta(\mathring{u}, \mathring{v}) = \theta(u,v)$. In particular $\theta(v,u) = \theta(u,v)$.

**Proof.** The first claim is obvious. Because $H$ is orthogonal, the dot product is preserved, $\langle \mathring{u}, \mathring{v} \rangle = \langle u,v \rangle$. Also $\det[\mathring{u}, \mathring{v}] = \pm \det[u, v]$. Hence $\theta(\mathring{u}, \mathring{v}) = \theta(u,v)$. 

276
Intuitively, $\theta(u, v)$ is the square of the cotangent of the angle from $u$ to $v$. (The cotangent function is extended to the complex domain.) Let $f = u^{\otimes n} + v^{\otimes n}$ be a non-degenerate signature of arity $n \geq 3$. Since $f$ is non-degenerate, $u$ and $v$ are linearly independent. The next proposition implies that this expression for $f$ via $u$ and $v$ is unique up to a root of unity. Therefore, $\theta(f)$ from Definition 7.57 is well-defined from $f$; it is independent of the particular expression of $f$.

**Proposition 7.59.** Let $u, v, u', v' \in \mathbb{C}^2$. Suppose $u$ and $v$ are linearly independent. If for some $n \geq 3$, we have

$$u^{\otimes n} + v^{\otimes n} = u'^{\otimes n} + v'^{\otimes n}, \tag{7.16}$$

then with possibly exchanging $u$ and $v$, there exist $\rho$ and $\omega \in \mathbb{C}$, such that $u' = \rho u, v' = \omega v$ where $\rho^n = \omega^n = 1$.

We omit the proof. Now we give some necessary conditions for membership in $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$. Recall that $\mathcal{A}_1 \subseteq \mathcal{P}_1$.

**Lemma 7.60.** Let $f$ be a non-degenerate symmetric signature of arity at least 3. Then

1. $f \in \mathcal{P}_1 \implies \theta(f) = 0$,
2. $f \in \mathcal{A}_2 \implies \theta(f) = -1$, and
3. $f \in \mathcal{A}_3 \implies \theta(f) = -\frac{1}{2}$.

**Proof.** The result clearly holds when $f$ is in the canonical form of each set. This extends to the rest of each set by Lemma 7.58.

These results imply the following corollary.

**Corollary 7.61.** Let $f$ be a non-degenerate symmetric signature $f$ of arity $n \geq 3$. If $f$ is $\mathcal{A}$-transformable, then $f$ satisfies a second order recurrence relation with nonzero discriminant, thus has the form $f = u^{\otimes n} + v^{\otimes n}$, where $u$ and $v$ are linearly independent, and $\theta(f) \in \{0, -1, -\frac{1}{2}\}$.

Note that the condition on $\theta(f)$ is only a necessary condition for $f$ being $\mathcal{A}$-transformable. The condition given in Corollary 7.61 is not sufficient to determine if $f \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$. In fact, it can be directly verified that $\theta(u, v) = -1$ for linearly independent $u$ and $v$ if and only if at least one of $u, v$ is $[\frac{1}{i}]$ or $[\frac{i}{1}]$, up to a nonzero scalar. Thus, if $f = u^{\otimes n} + v^{\otimes n}$ with $u = [\frac{1}{i}]$ and $v$ is not a multiple of $[\frac{1}{\pm i}]$, then $\theta(f) = -1$. But $f$ is not in $\mathcal{A}_2$ by Proposition 7.59, and Lemma 7.41. A full characterization can be achieved with additional work; but we will not pursue it further here.

To obtain the decision algorithm for Theorem 7.19 as stated in Theorem 7.53 will require more careful work concerning a common transformation for the whole set $\mathcal{F}$. We will not present all the details, but it starts with the following.
Lemma 7.38 showed that every \( \mathcal{A} \)-transformable set \( \mathcal{F} \) is contained in either \( \mathcal{H} \mathcal{A} \) or \( \mathcal{H} \left[ \begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right] \mathcal{A} \) for some \( \mathcal{H} \in \text{O}_2(\mathbb{C}) \). For non-degenerate symmetric signatures, we can replace \( \mathcal{A} \) by \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \). We have the following classification.

\[
\theta(f) = \begin{cases} 0 & \text{if } f \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \left[ \begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right] \mathcal{F}_1, \\
-1 & \text{if } f \in \mathcal{F}_3, \\
-\frac{1}{2} & \text{if } f \in \left[ \begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right] (\mathcal{F}_2 \cup \mathcal{F}_3). 
\end{cases}
\] (7.17)

Regarding \( \mathcal{P} \)-transformability, Lemma 7.60 also gives a useful necessary condition for \( f \) being \( \mathcal{P} \)-transformable. As \( \mathcal{P}_2 = \mathcal{A}_2 \), the same discussion for \( \mathcal{A}_2 \) applies to \( \mathcal{P}_2 \).

The next lemma tells us how to decide membership in \( \mathcal{P}_1 \) for signatures of arity at least 3.

**Lemma 7.62.** Let \( f = v_0^\otimes n + v_1^\otimes n \) be a symmetric signature of arity \( n \geq 3 \), where \( v_0 \) and \( v_1 \) are linearly independent. Then \( f \in \mathcal{P}_1 \) iff \( \theta(f) = 0 \).

With some additional work one can obtain the decision algorithm stated in Theorem 7.53.
Chapter 8

Planar \#CSP for Symmetric Constraints

In Chapter 3 we proved a complexity dichotomy Theorem 3.7 for \#CSP(\mathcal{F}), where \mathcal{F} is a set of constraint functions mapping Boolean inputs to complex numbers. The conclusion of Theorem 3.7 is that the problem \#CSP(\mathcal{F}) is \#P-hard, unless \mathcal{F} \subseteq \mathcal{A} or \mathcal{F} \subseteq \mathcal{P}, in which case the problem is computable in polynomial time. In this chapter, we would like to consider this problem when we restrict to planar input instances, i.e., the bipartite graph describing the connection between variables and constraints is a planar graph. We will show that in addition to \mathcal{A} and \mathcal{P}, there is a third tractable class \mathcal{M}, which is the class of matchgate signatures \mathcal{M} under a holographic transformation by the Hadamard matrix \( H_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \). Everything else remains \#P-hard.

8.1 Introduction

In terms of Holant problems, the problem \#CSP(\mathcal{F}) is equivalent to Holant(\mathcal{F} \cup \mathcal{E}Q), where \( \mathcal{E}Q = \{=1, =2, =3, \ldots \} \) and \( (=k) = [1, 0, \ldots, 0, 1] \) is the equality signature of arity \( k \). Considering the edge-vertex incidence graph, one can see that Holant(\mathcal{F}) is equivalent to Holant(\( =2 \mid \mathcal{F} \)). Then we can apply a holographic transformation, and see that for any nonsingular 2-by-2 matrix \( T \), Holant(\( =2 \mid \mathcal{F} \)) is equivalent to Holant(\( (=2)(T^{-1})^\otimes 2 \mid T\mathcal{F} \)). Here \( T\mathcal{F} = \{ T^\otimes \text{arity}(f) f \mid f \in \mathcal{F} \} \).

Using the operator \text{Sym}^{n-1}_n and the expression of a sum of tensor powers, Theorem 4.11 can be equivalently stated as

\textbf{Theorem 8.1.} Let \( f \) be a symmetric signature in \( \mathcal{M} \) of arity \( n \). Then there exist \( a, b, \lambda \in \mathbb{C} \) such that \( f \) takes one of the following forms:

1. \( [a, b]^\otimes n + [a, -b]^\otimes n = \begin{cases} 2[a^n, 0, a^{n-2}b^2, 0, \ldots, 0, b^n] & \text{if } n \text{ is even}, \\ 2[a^n, 0, a^{n-2}b^2, 0, \ldots, 0, ab^{n-1}, 0] & \text{if } n \text{ is odd}; \end{cases} \)

2. \( [a, b]^\otimes n \) \( - [a, -b]^\otimes n = \begin{cases} 2[0, a^{n-1}b, 0, a^{n-3}b^3, 0, \ldots, 0, ab^{n-1}, 0] & \text{if } n \text{ is even}, \\ 2[0, a^{n-1}b, 0, a^{n-3}b^3, 0, \ldots, 0, b^n] & \text{if } n \text{ is odd}; \end{cases} \)
3. \( \lambda \text{Sym}_n^{n-1}([1, 0]; [0, 1]) = [0, \lambda, 0, \ldots, 0]; \)
4. \( \lambda \text{Sym}_n^{n-1}([0, 1]; [1, 0]) = [0, \ldots, 0, \lambda, 0]. \)

In (4.23) and (4.24) of Chapter 3 these expressions are proved to be equivalent to the list in Theorem 4.11.

From the Holant perspective, the signatures in \( \mathcal{E} \mathcal{Q} \) are always available in \( \#\text{CSP}(\mathcal{F}) \). By the signature theory of matchgates and their transformations, the Hadamard matrix \( H_2 \) is essentially the only holographic transformation under which all \( \mathcal{E} \mathcal{Q} \) become realizable as standard signatures of matchgates. This follows from Lemma 4.29, where we take \( \lambda = \mu = 1 \) and \( \alpha = 0 \). Then we see that \( T \) is a transformation so that \((=_k)(T^{-1})^\otimes k\) is a standard signature of a matchgate iff \( T = [\begin{matrix} 1 & -\omega \\ \omega & 1 \end{matrix}] = H_2 [\begin{matrix} 1 & 0 \\ 0 & \omega \end{matrix}] \), where \( \omega^{2k} = 1 \). Since \( \mathcal{E} \mathcal{Q} \) contains \((=_k)\) for all \( k \geq 1 \), the only common transformations \( T \) are \( H_2 \) and \( H_2 [\begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}] \). However \( \Gamma \in \mathcal{M} \) iff \( \Gamma [\begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}]^{\text{arity} (\Gamma)} \in \mathcal{M} \), where we write a signature \( \Gamma \) as a row vector. This can be seen by the following construction: for any matchgate with signature \( \Gamma \), at every external node \( u \) we extend a path \((u, v, w)\) of length 2, making \( w \) the new external node, and the edge \((u, v)\) having weight \(-1\). This new matchgate has the signature \( \Gamma [\begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}]^{\text{arity} (\Gamma)} \). Note that \( [\begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}]^{-1} = [\begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}] \) and thus we have proved both directions. For symmetric signatures this is also obvious by Theorem 8.1. Thus we may consider \( H_2 \) as the only transformation such that \( \mathcal{E} \mathcal{Q} \) becomes matchgates realizable.

Let \( \widehat{\mathcal{F}} \) denote \( H_2 \mathcal{F} \) for any set \( \mathcal{F} \) of signatures. Then \( \widehat{\mathcal{E}} \mathcal{Q} \) is \{[1, 0], [1, 0, 1], [1, 0, 1, 0], \ldots\} while \((=_2)(H_2^{-1})^\otimes 2\) is still \((=_2)\) up to a nonzero constant. Therefore \( \#\text{CSP}(\mathcal{F}) \), which is really \( \text{Holant}(\mathcal{F} \cup \mathcal{E} \mathcal{Q}) \), is equivalent to \( \text{Holant}(\widehat{\mathcal{F}} \cup \widehat{\mathcal{E}} \mathcal{Q}) \). The equivalence of \( \#\text{CSP}(\mathcal{F}) \), \( \text{Holant}(\mathcal{F} \cup \mathcal{E} \mathcal{Q}) \) and \( \text{Holant}(\widehat{\mathcal{F}} \cup \widehat{\mathcal{E}} \mathcal{Q}) \) is also valid over planar instances. We denote by \( \text{Pl-}\#\text{CSP}(\mathcal{F}) \), \( \text{Pl-Holant}(\mathcal{F} \cup \mathcal{E} \mathcal{Q}) \) and \( \text{Pl-Holant}(\widehat{\mathcal{F}} \cup \widehat{\mathcal{E}} \mathcal{Q}) \) for the corresponding problems.

The main dichotomy theorem of this Chapter is stated as follows.

**Theorem 8.2.** Let \( \mathcal{F} \) be any set of symmetric, complex-valued signatures in Boolean variables. Then \( \text{Pl-}\#\text{CSP}(\mathcal{F}) \) is \#P-hard unless \( \mathcal{F} \) satisfies one of the following conditions, in which case it is tractable:

1. \( \#\text{CSP}(\mathcal{F}) \) is tractable, i.e., \( \mathcal{F} \subseteq \mathcal{A} \) or \( \mathcal{F} \subseteq \mathcal{P} \) (by Theorem 3.7);
2. \( \widehat{\mathcal{F}} \) is realizable as standard matchgate signatures, i.e., \( \mathcal{F} \subseteq \widehat{\mathcal{M}} \).

Thus, for symmetric complex-valued signatures in Boolean variables, to be transformable to matchgates under the Hadamard transformation followed by Kasteleyn’s algorithm (a.k.a. the FKT algorithm, see Chapter 4) is a *universal* methodology to solve \( \text{Pl-}\#\text{CSP} \) problems in polynomial time over planar graphs that are \#P-hard over general graphs.

We note that Theorem 8.2 is only proved for sets of *symmetric* signatures, whereas Theorem 3.7 does not have this restriction. This represents a limitation of our knowledge at this time. We also note that the tractability criteria of Theorem 8.2 are decidable in polynomial time in the length of the description of the symmetric signatures in \( \mathcal{F} \). This
follows from the forms of symmetric signatures $\mathcal{F}_{123}$ in $\mathcal{A}$, and those in $\mathcal{P}$ and in $\mathcal{M}$. Note that for $f = [a, b]^{\otimes n} \pm [a, -b]^{\otimes n} \in \mathcal{M}$, $H_2 f \in \mathcal{M}$ has the form $[x, y]^{\otimes n} \pm [y, x]^{\otimes n}$, where $x = a + b$ and $y = a - b$.

In proving dichotomy theorems for Boolean $\#\text{CSP}(\mathcal{F})$, pinning is an important tool. If one can realize the unary Is-Zero $\Delta_0 = [1, 0]$ and Is-One $\Delta_1 = [0, 1]$, then one can obtain all sub-signatures. However the proof in Chapter 3 obtaining $\Delta_0$ and $\Delta_1$ is by a nonplanar reduction. We note that in the nonplanar setting, $\Delta_0$ and $\Delta_1$ are contained in both maximal tractable sets $\mathcal{A}$ and $\mathcal{P}$. Therefore, pinning in this setting would not imply the collapse of $\#\text{P}$ to $\text{P}$, and therefore it is feasible to prove the reduction $\#\text{CSP}(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \leq_T \#\text{CSP}(\mathcal{F})$ as was done in Lemma 3.13. However, $\mathcal{EQ}$ and $\{[1, 0], [0, 1]\}$ are not simultaneously realizable as matchgates, under any holographic transformation. Suppose Theorem 8.2 is true. Then we should not expect to be able to pin for Pl-$\#\text{CSP}(\mathcal{F})$, i.e., to prove a general reduction Pl-$\#\text{CSP}(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \leq_T \text{Pl-$\#\text{CSP}(\mathcal{F})$}$, since otherwise $\#\text{P}$ collapses to $\text{P}$. The collapse is seen by taking $\mathcal{F} = \mathcal{M}$ which is neither a subset of $\mathcal{A}$ nor $\mathcal{P}$, and $\Delta_0 \not\in \mathcal{M}$ since $H_2 \Delta_0 = [1, 1]^T \not\in \mathcal{M}$. Instead, apply the Hadamard transformation and consider Pl-Holant($\mathcal{F} \cup \mathcal{EQ}$). In this Hadamard basis, it is at least possible again that we can pin in a planar setting, since $\Delta_0$ and $\Delta_1$ are contained in $\mathcal{A} \cap \mathcal{P} \cap \mathcal{M}$. Note that $\mathcal{A} = \mathcal{A}$ and $\mathcal{P}$ still contains all unary signatures. If $\mathcal{F} \subseteq \mathcal{A}$, or $\mathcal{P}$, or $\mathcal{M}$, then $\mathcal{F} \cup \mathcal{EQ} \cup \{\Delta_0, \Delta_1\} \subseteq \mathcal{A}$, or $\mathcal{P}$ or $\mathcal{M}$ respectively. Indeed, we will prove our pinning result in this Hadamard basis in Section 8.6, see (8.5).

For Holant problems, it is often important to understand the complexity of the small arity cases first. Theorem 6.3 is a dichotomy for Pl-Holant($f$) where $f$ is a symmetric signature of arity 3. After some strengthening of unary interpolations, we will prove a dichotomy for Pl-Holant($f$) where $f$ is a symmetric signature of arity 4. This uses the idea of a planar pairing. Then we prove some useful lemmas using a technique called domain pairing that essentially realizes an odd arity signature using only signatures of even arity. After that we show some No-Mixing lemmas, which say that the known classes of tractable signatures for Pl-$\#\text{CSP}$ become $\#\text{P}$-hard when mixed. Then we realize pinning in the Hadamard basis. After that we prove the planar CSP dichotomy for a single signature. The main dichotomy for Pl-$\#\text{CSP}(\mathcal{F})$, Theorem 8.2, follows easily from this and the No-Mixing lemmas.

For easy reference we list some results in a convenient form from previous chapters.

**Theorem 8.3.** For any $x, y, z \in \mathbb{C}$, both Pl-Holant($[x, y, z] \mid \mathcal{EQ}$) and Pl-Holant($[x, y, z] \mid \mathcal{EQ}$) are $\#\text{P}$-hard unless one of the following conditions holds, in which case both problems are computable in polynomial time:

1. $xz = y^2$;
2. $y = 0$;
3. $xz = -y^2$ and $x = -z$;
4. $x = z$.

**Proof.** Theorem 5.1 (equivalently Theorem 5.3) was proved for the cases $k = 3$ and 4 in Chapter 5. For Pl-Holant($[x, y, z] \mid \mathcal{EQ}$) both ($=3$) and ($=4$) are present in $\mathcal{EQ}$, and thus
Pl-Holant([x, y, z] | \mathcal{E}Q) is \#P-hard unless both tractability criteria of Theorem 5.3 for \( k = 3 \) and 4 are satisfied. This gives the listed tractability criterion. Conversely, if one of the tractability conditions is satisfied then clearly Pl-Holant([x, y, z] | \mathcal{E}Q) is computable in P.

Under the transformation \( H_2 = [1 \ 1
\begin{array}{c}
1 \\
1
\end{array}
] \), it is easy to see that the conditions \( xz = y^2 \) and \((xz = -y^2) \land (x = -z)\) are invariant while the conditions \( y = 0 \) and \( x = z \) map to each other. Therefore, by an apparent coincidence, the tractability conditions remain the same for \( \text{Pl-Holant}([x, y, z] | \hat{\mathcal{E}}Q) \).

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure.png}
\caption{Unary recursive construction with starter gadget.}
\end{figure}

\textbf{Theorem 8.4.} Let \( \mathcal{F} \) be any set of symmetric, complex-valued signatures in Boolean variables. Then \( \text{Pl-Holant}^c(\mathcal{F} \cup \hat{\mathcal{E}}Q) \) is computable in polynomial time if \( \mathcal{F} \subseteq \mathcal{A} \), \( \mathcal{F} \subseteq \hat{\mathcal{P}} \), or \( \mathcal{F} \subseteq \mathcal{M} \).

\textit{Proof.} We only need to observe that \( \hat{\mathcal{E}}Q \cup \{[1, 0], [0, 1]\} \subseteq \mathcal{A} \cap \hat{\mathcal{P}} \cap \mathcal{M} \). The case \( \mathcal{F} \subseteq \mathcal{A} \) follows from Theorem 6.12. The case \( \mathcal{F} \subseteq \hat{\mathcal{P}} \) follows from Theorem 6.12 and 3.5. The case \( \mathcal{F} \subseteq \mathcal{M} \) follows from Theorem 4.7. \qed

\section{8.2 Unary Interpolation Revisited}

In this section, we discuss \textit{unary recursive construction} and give an exact condition for when it succeeds. The goal of this construction is to interpolate an arbitrary unary signature.

For a unary recursive construction in the Holant framework, there are two gadgets involved: a \textit{starter} gadget of arity 1 and a \textit{recursive} gadget of arity 2. In the Pl-Holant framework the gadgets must also be planar. The signature of the starter gadget is represented by a vector \( s \in \mathbb{C}^2 \). The signature of the recursive gadget is represented by a matrix \( M \in \mathbb{C}^{2 \times 2} \). The construction begins with the starter gadget and proceeds by connecting \( k \geq 0 \) recursive gadgets, one at a time, to the only available edge (see Figure 8.1). The signature matrix of the resulting gadget is \( M^k s \). This construction is denoted by \( (M, s) \).

The most demanding part of the requirement in these constructions, for the purpose of polynomial interpolation, is to construct an infinite set of signatures that are pairwise linearly independent (see Chapter 5). The pairwise linear independence of signatures translates into distinct evaluation points for the polynomial being interpolated.

We say two vectors are orthogonal if their dot product is 0. We first prove a lemma that gives a determinantal condition for the statement that a vector \( s \) is not orthogonal to any row eigenvector of a matrix \( M \).
Lemma 8.5. Suppose $M \in \mathbb{C}^{n \times n}$ and $s \in \mathbb{C}^n$. Then $s$ is not orthogonal to any row eigenvector of $M$ iff $\det([s \ M s \ldots \ M^{n-1}s]) \neq 0$.

Proof. Suppose $\det([s \ M s \ldots \ M^{n-1}s]) = 0$. Then there is a nonzero row vector $v$ such that $v[s \ M s \ldots \ M^{n-1}s] = 0$ is the zero vector. Consider the linear span $S$ by row vectors in the set $\{v, vM, \ldots, vM^{n-1}\}$. By the Cayley-Hamilton theorem, $M$ satisfies its own characteristic polynomial, which is a monic polynomial of degree $n$. Thus, $M^n$ is a linear combination of $I_n, M, \ldots, M^{n-1}$. This shows that $S$ is an invariant subspace of row vectors under the action of multiplication by $M$ from the right. Being an invariant subspace over $\mathbb{C}$, there exists a row eigenvector $u \in S$ of $M$, and the dot product of $s$ with $u$ is 0.

Conversely, suppose $\det([s \ M s \ldots \ M^{n-1}s]) \neq 0$ and assume for a contradiction that $s$ is orthogonal to some row eigenvector $v$ of $M$ with eigenvalue $\lambda$. Then $v[s \ M s \ldots \ M^{n-1}s] = 0$ is the zero vector because $vM^i s = \lambda^i vs = 0$. Since $v \neq 0$, this is a contradiction. \qed

To construct an infinite set of signatures that are pairwise linearly independent, a necessary condition is that $M$ has infinite order modulo a scalar. Otherwise, $M^k = cI_n$ for some $k > 0$ and $c \neq 0$, then any vector of the form $M^\ell s$ for $\ell \geq k$ is a multiple of a vector in the set $\{M^i s\}_{0 \leq i < k}$.

Lemma 8.6. Suppose $M \in \mathbb{C}^{n \times n}$ and $s \in \mathbb{C}^n$. If the following three conditions are satisfied,

1. $\det(M) \neq 0$;
2. $\det([s \ M s \ldots \ M^{n-1}s]) \neq 0$;
3. $M$ has infinite order modulo a scalar;

then the vectors in the set $V = \{M^k s\}_{k \geq 0}$ are pairwise linearly independent.

Proof. Since $\det(M) \neq 0$, $M$ is nonsingular and all eigenvalues are nonzero. Let $M = P^{-1}JP$ be the Jordan decomposition of $M$ and let $v = Ps \in \mathbb{C}^n$. Clearly $v \neq 0$ since $s \neq 0$ by condition 2. Suppose for a contradiction that the vectors in $V$ are not pairwise linearly independent. This means that there exist integers $k > \ell \geq 0$ such that $M^k s = cM^\ell s$ for some nonzero $c \in \mathbb{C}$. Let $m = k - \ell > 0$. Then we have $P^{-1}J^mPs = M^ms = cs$ and thus $J^mv = cv$.

Suppose that $J$ contains some nontrivial Jordan block and consider the 2-by-2 submatrix $[\lambda \ 1; 0 \ 1]$ in the bottom right corner of this block, where $\lambda$ is an eigenvalue of $M$. Without loss of generality we may assume this 2-by-2 submatrix are indexed by $n - 1$ and $n$. From this portion of $J$, the two equations given by $J^mv = cv$ are $\lambda v_{n-1} + m\lambda^{m-1}v_n = cv_{n-1}$ and $\lambda^m v_n = cv_n$. By Lemma 8.5, $s$ is not orthogonal to any row eigenvector of $M$, therefore $v_n \neq 0$ since the last row of $P$ is a row eigenvector of $M$. But then these equations imply that $m\lambda^{m-1}v_n = 0$, a contradiction.

Otherwise, $J$ is a diagonal matrix. From $J^mv = cv$, we get the equations $\lambda^mv_i = cv_i$ for $1 \leq i \leq n$. Since $s$ is not orthogonal to any row eigenvector of $M$, $v_i \neq 0$ for $1 \leq i \leq n$. 283
But then $M^m = cI_n$, which contradicts the assumption that $M$ has infinite order modulo a scalar.

These conditions guarantee success of interpolation by a unary recursive construction.

**Lemma 8.7.** Let $\mathcal{F}$ be a set of signatures. If there exists a planar $\mathcal{F}$-gate with signature matrix $M \in \mathbb{C}^{2 \times 2}$ and a planar $\mathcal{F}$-gate with signature $s \in \mathbb{C}^2$ satisfying the following conditions,

1. $\det(M) \neq 0$;
2. $\det([s Ms]) \neq 0$;
3. $M$ has infinite order modulo a scalar;

then $\text{Pl-Holant}(\mathcal{F} \cup \{[a, b]\}) \leq_T \text{Pl-Holant}(\mathcal{F})$ for any $a, b \in \mathbb{C}$.

**Proof.** Consider an instance $\Omega$ of $\text{Pl-Holant}(\mathcal{F} \cup \{[a, b]\})$. Let $V'$ be the subset of vertices assigned $[a, b]$ and suppose that $|V'| = n$. We construct from $\Omega$ a sequence of instances $\Omega_k$ of $\text{Pl-Holant}(\mathcal{F})$ indexed by $k \geq 0$. We obtain $\Omega_k$ from $\Omega$ by replacing each occurrence of $[a, b]$ with the unary recursive construction $(M, s)$ in Figure 8.1 containing $k$ copies of the recursive gadget. This unary recursive construction has the signature $[x_k, y_k] = M^k s$.

By Lemmas 8.6, the signatures in the set $V = \{[x_k, y_k] \mid 0 \leq k \leq n + 1\}$ are pairwise linearly independent. In particular, at most one $y_k$ can be 0.

We stratify the assignments in $\Omega$ based on the assignment to $[a, b]$. Let $c_\ell$ be the sum over all assignments of products of evaluations at all vertices other than those from $V'$ such that exactly $\ell$ occurrences of $[a, b]$ have their incident edge assigned 0 (and $n - \ell$ have their incident edge assigned 1). Then

$$\text{Holant}(\Omega) = \sum_{0 \leq \ell \leq n} a^\ell b^{n-\ell} c_\ell$$

and the value of the Holant on $\Omega_k$, for $0 \leq k \leq n+1$, is $\text{Holant}(\Omega_k) = \sum_{0 \leq \ell \leq n} x_k^\ell y_k^{n-\ell} c_\ell$, and for at least $n + 1$ of them where $y_k \neq 0$, we have

$$\frac{\text{Holant}(\Omega_k)}{y_k^n} = \sum_{0 \leq \ell \leq n} \left(\frac{x_k}{y_k}\right)^\ell c_\ell.$$ 

This is a linear system with unknowns $c_\ell$ and a Vandermonde coefficient matrix. Since the signatures in $V$ are pairwise linearly independent, the ratios $x_k/y_k$ are all distinct (and well-defined since $y_k \neq 0$), which implies that the Vandermonde matrix has full rank. Therefore, we can solve the linear system for the unknown $c_\ell$'s and obtain the value of $\text{Holant}(\Omega)$.

With this lemma, we obtain a tight characterization for the success of interpolation by a unary recursive construction. For example, the construction using a recursive gadget with
signature matrix \( M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \) and a starter gadget with signature \( s = [0] \) is successful because \( M \) and \( s \) satisfy the conditions of Lemma 8.6 even though the eigenvalues of \( M \) are equal.

The first two conditions of Lemma 8.7 are easy to check. The third condition holds in one of these two cases: either the eigenvalues are the same but \( M \) is not a multiple of the identity matrix, or the eigenvalues are different but their ratio is not a root of unity.

These conditions work well with the anti-gadget technique. The power of this lemma is that when the third condition fails to hold, there exists an integer \( m \) such that \( M^m = cI_2 \), where \( c \neq 0 \) and \( I_2 \) is the 2-by-2 identity matrix. Therefore we can construct \( M^{m-1} = M^{-1} \) and use this in other gadget constructions. We remark that if any such \( m \) exists then the minimum such \( m \) is bounded by a constant. Regardless of the number of internal nodes of the \( F \)-gate, its signature entries in \( M \in \mathbb{C}^{2 \times 2} \) all belong to the fixed algebraic extension field \( \mathbb{F} \) over \( \mathbb{Q} \), from which functions in \( \mathcal{F} \) take values. The ratio of the two eigenvalues \( \lambda \) and \( \mu \) of \( M \) belongs to the splitting field which has a degree of extension at most \( 2[\mathbb{F} : \mathbb{Q}] \). If \( M \) has finite order modulo a scalar, then let \( m \) be the minimum exponent \( \ell > 0 \) such that \( M^\ell = cI_2 \) for some \( c \neq 0 \). Then \( m = \min\{\ell > 0 \mid \left(\frac{\lambda}{\mu}\right)^\ell = 1\} \). Thus the \( m \)-th primitive root of unity \( e^{2\pi i/m} \) belongs to the splitting field. However the \( m \)-th cyclotomic field \( \Phi_m = \mathbb{Q}(e^{2\pi i/m}) \) has degree of extension \( [\Phi_m : \mathbb{Q}] = \phi(m) \), where \( \phi(m) \) is the Euler’s totient function. By an elementary estimate \( \phi(m)^2 \geq m/2 \). Hence \( m \) is upper bounded by a constant only depending on \( \mathbb{F} \).

## 8.3 Planar Pairing

In this section, we prove a dichotomy for \( \text{Pl-Holant}(f) \) where \( f \) is a symmetric signature of arity 4. The key part is to consider the problem \( \text{Pl-Holant}([t, 1, 0, 0, 0]) \) when \( t \) is different from 0. Over the next two lemmas, we prove that this problem is \#P-hard by reducing from \( \text{Pl-Holant}([t, 1, 0, 0]) \). These problems are weighted versions of counting matchings over planar \( k \)-regular graphs for \( k = 4 \) and \( k = 3 \) respectively. We use the idea of a planar pairing to be defined in Definition 8.9, and apply Lemma 8.7.

In the first lemma, we show how to use either the anti-gadget technique or the unary interpolation technique based on Lemma 8.7 from Section 8.2 to effectively obtain \([1, 0, 0, 0]\). The construction in this proof is actually not a unary recursive construction, but a binary recursive construction. However, the particular forms in the starter and recursive gadgets permit analysis equivalent to that of the unary recursive construction.

**Lemma 8.8.** For any \( t \in \mathbb{C} \) and signature set \( \mathcal{F} \) containing \([t, 1, 0, 0, 0]\),

\[
\text{Pl-Holant}(\mathcal{F} \cup \{[1, 0, 0, 0]\}) \preceq_T \text{Pl-Holant}(\mathcal{F}).
\]

**Proof.** Consider the gadget construction in Figure 8.2. For \( k \geq 0 \), the signature of \( N_k \) is of the form \([a_k, b_k, 0]\), and \( N_0 = [t, 1, 0] \). Since \( N_k \) is symmetric and always ends with 0, we can
Figure 8.2: Binary recursive construction with starter gadget used to interpolate $[1,0,0,0]$. The vertices are assigned $[t,1,0,0,0]$.

analyze this construction as though it were a unary recursive construction. Let $s_k = \begin{bmatrix} a_k \\ b_k \end{bmatrix}$, so $s_0 = [t]$. It is clear that $s_k = M^k s_0$, where $M = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$.

Now $M$ is nonsingular since $\det(M) = -2$. If $M$ has finite order modulo a scalar, then $M^m = c I_2$ for some integer $m \geq 1$ and some nonzero $c \in \mathbb{C}$. Then $N_m = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an anti-gadget of $M$ with signature $M^m s_0 = c M^{-1} s_0 = c [t]$. Note that $[t]$ here denotes the signature $[1,0,0]$ as an abbreviated notation. Hence we have $[1,0,0]$ after normalizing.

Now assume that $M$ has infinite order modulo a scalar. Since $\det([s_0 \ M s_0]) = -2$, we can interpolate any signature of the form $[x,y,0]$ by Lemma 8.7, including $[1,0,0]$. \qed

Note that having $[1,0,0] = [1,0]^{\otimes 2}$ is equivalent to be able to pin two variables to 0 simultaneously. To apply $[1,0]^{\otimes 2}$ in a planar setting, one must be able to connect the two variables in a planar way.

**Definition 8.9 (Planar pairing).** A planar pairing $M$ in a planar graph $G = (V,E)$ is a partition of $V$ into pairs such that $(V,E \cup M)$ is a planar multigraph, where $E \cup M$ is the multiset formed by disjoint union.

Lemma 8.9 shows that for any 3-regular planar graph, a planar pairing exists and can be easily found in polynomial time. Now we use the planar pairing technique to show the following.

**Lemma 8.10.** Let $t \in \mathbb{C}$. Then Pl-Holant$([t,1,0,0,0]) \leq_T$ Pl-Holant$([t,1,0,0,0])$.

**Proof.** An instance of Pl-Holant$([t,1,0,0,0])$ is a signature grid $\Omega$ with underlying graph $G = (V,E)$ that is planar and 3-regular. By Lemma 8.9, there exists a planar pairing $M$ in $G$ and it can be found in polynomial time. Then $G' = (V,E \cup M)$ is a planar and 4-regular multigraph. We assign $[t,1,0,0,0]$ to every vertex in $G'$. By Lemma 8.8, we can assume that we have $[1,0,0]$. We replace each edge in $M$ with a path of length 2 to form a graph $G''$ and assign $[1,0,0] = [1,0]^{\otimes 2}$ to the new vertices on the paths. Note that applying $[1,0]$ to $[t,1,0,0,0]$ gives $[t,1,0,0,0]$. Then the signature grid $\Omega''$ with underlying graph $G''$ has the same Holant value as the original signature grid $\Omega$. \qed

**Corollary 8.11.** The problem Pl-Holant$([t,1,0,0,0])$ is $\#P$-hard for all nonzero $t \in \mathbb{C}$. 286
Proof. By Theorem 6.3, the planar Holant dichotomy for a single ternary signature, the problem Pl-Holant([t, 1, 0, 0]) is \#P-hard. Then the Corollary follows from Lemma 8.10. □

One feature of this reduction is worth noticing. Most gadget constructions in hardness proofs for Holant problems are local but the planar pairing technique is a global argument, which permits reductions that are not otherwise possible. Note that from any set of signatures all with even arities, any gadget construction will produce a signature also of even arity. This is by a simple parity argument, because the sum of arities of all signatures used in the construction is twice the number of internal edges plus the arity of the signature produced. Thus, with [t, 1, 0, 0] alone one cannot construct any signature of arity 3 by gadget construction.

Now we are ready to prove our Pl-Holant dichotomy for a symmetric arity 4 signature.

**Theorem 8.12.** If \( f \) is a non-degenerate, symmetric, complex-valued signature of arity 4 in Boolean variables, then Pl-Holant(\( f \)) is \#P-hard unless \( f \) is \( \mathcal{A} \)-transformable or \( \mathcal{P} \)-transformable or vanishing or \( \mathcal{M} \)-transformable, in which case the problem is computable in polynomial time.

**Proof.** Let \( f = [f_0, f_1, f_2, f_3, f_4] \). If there do not exist \( a, b, c \in \mathbb{C} \), not all zero, such that for all \( k \in \{0, 1, 2\} \), \( af_k + bf_{k+1} + cf_{k+2} = 0 \), then Pl-Holant(\( f \)) is \#P-hard by Corollary 7.25. Otherwise, there do exist such \( a, b, c \). If \( a = c = 0 \), then \( b \neq 0 \), so \( f_1 = f_2 = f_3 = 0 \). In this case, \( f \in \mathcal{P} \) is a generalized equality signature, so \( f \) is \( \mathcal{P} \)-transformable. Now suppose \( a \) and \( c \) are not both 0. If \( b^2 - 4ac \neq 0 \), then \( f_k = \alpha^{4-k} \beta^k + \gamma^{4-k} \delta^k \), where \( \det \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \neq 0 \). A holographic transformation by \( \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \) transforms \( f \) to \( \mathcal{A} \)- and we can use Theorem ?? to show that Pl-Holant(\( f \)) is \#P-hard unless \( f \) is either \( \mathcal{A} \)-, \( \mathcal{P} \)-, or \( \mathcal{M} \)-transformable. Otherwise, \( b^2 - 4ac = 0 \) and there are two cases. In the first, for any \( 0 \leq k \leq 4 \), \( f_k = ck \lambda^{k-1} + d \lambda^k \), where \( c \neq 0 \). In the second, for any \( 0 \leq k \leq 4 \), \( f_k = c(4-k) \lambda^{3-k} + d \lambda^{4-k} \), where \( c \neq 0 \). These cases map between each other under a holographic transformation by \( \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \), so assume that we are in the first case. If \( \lambda = \pm i \), then \( f \) is vanishing. Otherwise, a further holographic transformation by \( \frac{1}{\sqrt{1 + k}} \begin{bmatrix} 1 & \lambda_k \\ \lambda_k & 1 \end{bmatrix} \) transforms \( f \) to \( \tilde{f} = [t, 1, 0, 0, 0] \) for some \( t \in \mathbb{C} \) after normalizing the second entry. If \( t = 0 \), then the problem is counting perfect matchings over planar 4-regular graphs, so \( \tilde{f} \in \mathcal{M} \) and \( f \) is \( \mathcal{M} \)-transformable. Otherwise, \( t \neq 0 \) and we are done by Corollary 8.11.

\[ \square \]

### 8.4 Domain Pairing

In this section, we discuss a technique called domain pairing, which pairs input variables to simulate a problem on a domain of size four and then reduces a problem in the Boolean domain to it. As explained in Section 8.1, we work in the Hadamard basis instead of the standard basis. The goal then becomes a dichotomy for Pl-Holant(\( \mathcal{F} \cup \mathcal{E} \mathcal{Q} \)).

287
Figure 8.3: Gadget designed for the paired domain. One vertex is assigned \([1, 0, 1, 0]\) and the other is assigned \([x, 0, y, 0]\).

By a simple parity argument, gadgets constructed with signatures of even arity can only realize other signatures of even arity. In particular, this means that \(=4\) cannot by itself be used to construct \(=3\). Nevertheless, there is an indirect method that can realize \(=3\) using \(=4\). The idea is to change the domain from individual elements to pairs of elements. Thus, we call this reduction technique \textit{domain pairing}.

The following is a lemma where the reduction is carried out by this method.

**Lemma 8.13** (Domain pairing). Let \(a, b, x, y \in \mathbb{C}\). Suppose \(f = [x, 0, y, 0] \text{ and } g = [a, 0, \ldots, 0, b] \) with arity at least 3. If \(aby \neq 0\) and \(x^2 \neq y^2\), then

\[
\text{Pl-Holant}([x, y, y] | \mathcal{E}Q) \lesssim \text{Pl-Holant}([f, g] \cup \hat{\mathcal{E}}Q).
\]

In particular, \(\text{Pl-Holant}([f, g] \cup \hat{\mathcal{E}}Q) \) is \#P-hard.

**Proof.** We prove the reduction. The problem \(\text{Pl-Holant}([x, y, y] | \mathcal{E}Q)\) is \#P-hard by Theorem 8.3 since \(y \neq 0\) and \(x^2 \neq y^2\). By Lemma 7.48, we have \(\mathcal{E}Q_2\), the set of all \textit{Equalities} of even arity.

Notice that if one were able to obtain \((=3)\) on the Boolean domain, then we could have constructed \([x, y, y]\) from \([x, 0, y, 0]\) easily. Simply take a self-loop on \((=3)\) gives \([1, 1]\), and connect \([1, 1]\) to \([x, 0, y, 0]\) gives \([x, y, y]\). However as we remarked earlier, there is an obstacle to obtaining \((=3)\) due to parity. Instead we use the domain pairing technique.

An instance of \(\text{Pl-Holant}([x, y, y] | \mathcal{E}Q)\) is a planar bipartite graph \(G = (U, V, E)\) in which every vertex in \(U\) has degree 2. We replace every vertex in \(V\) of degree \(k\) (which is assigned \((=k)\) \(\in \mathcal{E}Q\)) with a vertex of degree \(2k\) and assign \((=2k)\) \(\in \mathcal{E}Q_2\). Furthermore, we bundle two adjacent variables to form \(k\) bundles of 2 edges each. The \(k\) bundles correspond to the \(k\) incident edges of the original vertex with degree \(k\).

If the inputs to these \textit{Equality} signatures are restricted to \(\{(0, 0), (1, 1)\}\) on each bundle, then these \textit{Equality} signatures take value 1 on \((0, 0), (1, 1)\) and take value 0 elsewhere. Thus, if we restrict the domain to \(\{(0, 0), (1, 1)\}\), it is the \textit{Equality} signature \((=k)\).

To simulate \([x, y, y]\), we connect \(f = [x, 0, y, 0]\) to \(e = [1, 0, 1, 0] \in \hat{\mathcal{E}}Q\) by a single edge as shown in Figure 8.3 to form a gadget with signature

\[
h(u_1, u_2, v_1, v_2) = \sum_{w=0,1} f(u_1, v_1, w) \cdot e(u_2, v_2, w).
\]
We replace every (degree 2) vertex in \( U \) (which is assigned \([x, y, y]\)) by a degree 4 vertex assigned \( h \), where the variables of \( h \) are bundled as \((u_1, u_2)\) and \((v_1, v_2)\).

The vertices in this new graph \( G' \) are connected as in the original graph \( G \), except that every original edge is replaced by two edges that connect to the same side of the gadget in Figure 8.3, either the left or the right side. Clearly \( G' \) is still a planar graph. Notice that \( h \) is only connected by \((u_1, u_2)\) and \((v_1, v_2)\) to some bundle of two incident edges of an \textsc{Equality} signature. Since this \textsc{Equality} signature enforces that the value on each bundle is either \((0, 0)\) or \((1, 1)\), we only need to consider the restriction of \( h \) to the domain \( \{(0, 0), (1, 1)\} \).

On this domain, \( h \) is exactly the symmetric binary signature \([x, y, y]\). Therefore, the Holant of \( G' \) has the same Holant value as the original graph \( G \).

Lemma 8.13 is mainly used to prove that generally speaking, \textsc{Gen-Eq} that do not belong to \( \mathcal{A} \) cannot mix with \( \mathcal{EQ} \). As \( \mathcal{EQ} \subseteq \mathcal{A} \) and \( \mathcal{A} \) is a tractable class, only having \textsc{Gen-Eq} that do belong to \( \mathcal{A} \) together with \( \mathcal{EQ} \) cannot lead to \#P-hardness (if \#P does not collapse to P). However we show that having \textsc{Gen-Eq} of arity at least 3 that do not belong to \( \mathcal{A} \) together with \( \mathcal{EQ} \) will lead to \#P-hardness. There are two scenarios, Corollary 8.14 and Lemma 8.16, that will let us apply Lemma 8.13. Lemma 8.16 says that a non-degenerate \textsc{Gen-Eq} of arity at least 3 that does not belong to \( \mathcal{A} \) together with \( \mathcal{EQ} \) will lead to \#P-hardness. Corollary 8.14 says that even if this \textsc{Gen-Eq} of arity at least 3 may belong to \( \mathcal{A} \), if there is another binary non-degenerate \textsc{Gen-Eq} that does not belong to \( \mathcal{A} \), then together with \( \mathcal{EQ} \) we still have \#P-hardness.

Corollary 8.14 is an immediate corollary of Lemma 8.13, and is also used in Lemma 8.16, so we prove that first.

**Corollary 8.14.** Let \( a, b, x, y \in \mathbb{C} \). Suppose \( f = [x, 0, y] \) and \( g = [a, 0, \ldots, 0, b] \) with arity at least 3. If \( abxy \neq 0 \) and \( x^4 \neq y^4 \), then \( \text{Pl-Holant}\{f, g\} \cup \mathcal{EQ} \) is \#P-hard.

**Proof.** Connect three copies of \( f = [x, 0, y] \) to \([1, 0, 1, 0]\), with one on each edge, to get \( x[x^2, 0, y^2, 0] \) and we apply Lemma 8.13.

The proof of Lemma 8.16 applies Corollary 8.14 and an interpolation of a unary signature. We prove this interpolation first; it will be used again later.

**Lemma 8.15.** Let \( x \in \mathbb{C} \). Suppose \( \mathcal{F} \) is a set of signatures containing \( f = [1, x, 1] \). If \( x \notin \{0, \pm 1\} \) and \( M_f \) has infinite order modulo a scalar, then

\[
\text{Pl-Holant}(\mathcal{F} \cup \{a, b\} \cup \mathcal{EQ}) \leq_T \text{Pl-Holant}(\mathcal{F} \cup \mathcal{EQ})
\]

for any \( a, b \in \mathbb{C} \).

**Proof.** Consider the unary recursive construction \((M_f, s)\) in Figure 8.1, where \( s = [\frac{1}{x}] \). The determinant of \( M_f \) is \( 1 - x^2 \neq 0 \). The determinant of \([s M_f s]\) is \( x \neq 0 \). By assumption, \( M_f \) has infinite order modulo a scalar. Therefore, we can interpolate any unary signature by Lemma 8.7.

289
Lemma 8.16. Let $a, b \in \mathbb{C}$. Suppose $f = [a, 0, \ldots, 0, b]$ with arity at least 3. If $ab \neq 0$ and $a^4 \neq b^4$, then Pl-Holant($\{f\} \cup \mathcal{E}_Q$) is $\#P$-hard.

Proof. Since $a \neq 0$, we normalize $f$ to $[1, 0, \ldots, 0, x]$, where $x \neq 0$ and $x^4 \neq 1$. If the arity of $f$ is even, then after some number of self-loops, we have $[1, 0, x]$ and are done by Corollary 8.14. Otherwise, the arity of $f$ is odd. After zero or more self-loops, we have $g = [1, 0, 0, x]$. If we have the signature $[1, 1]$, then we can connect this to $g$ to get $[1, 0, x]$ and be done by Corollary 8.14.

We use the signature $[1, x]$, which we obtain via a self-loop on $g$. Suppose $\Re(x)$, the real part of $x$, is nonzero. Connecting $[1, x]$ to $[1, 0, 1, 0] \in \mathcal{E}_Q$ gives $h = [1, x, 1]$. The eigenvalues of $M_h$ are $\lambda_{\pm} = 1 \pm x$. They are nonzero because $x \neq \pm 1$. Note that $|x - 1| = |x + 1|$ iff $x$ lies on the pure imaginary line $\Re(x) = 0$, so we have $|\lambda_{\pm}| \neq 1$. Hence the ratio of the eigenvalues is not a root of unity, and so $M_h$ has infinite order modulo a scalar. Therefore, we can interpolate $[1, 1]$ by Lemma 8.15.

Otherwise, $\Re(x) = 0$ but $x$ is not a root of unity since $x \neq \pm i$. Connecting $[1, x]$ to $g$ gives $h = [1, 0, x^2]$. Clearly $(x^2)^4 \neq 1$. Hence we apply Corollary 8.14 on $h$ and $f$, implying that Pl-Holant($\mathcal{F} \cup \mathcal{E}_Q$) is $\#P$-hard. \qed

8.5 No-Mixing of Tractable Signatures

Perhaps a natural goal at this point would be to attempt a dichotomy Pl-$\#$CSP($f$) for a single signature in the form of Pl-Holant($\{f\} \cup \mathcal{E}_Q$), which should say that in order not to be $\#P$-hard, $\check{f}$ must belong to $\mathcal{A} \cup \mathcal{P} \cup \mathcal{M}$, the union of the three tractable classes. This will indeed be proved (Theorem 8.32) before we prove the main dichotomy (Theorem 8.34). However, seemingly taking a detour, we will actually prove a No-Mixing theorem (Theorem 8.22) first. This No-Mixing theorem assumes that a given signature set $\mathcal{F}$ already satisfies $\mathcal{F} \subseteq \mathcal{A} \cup \mathcal{P} \cup \mathcal{M}$ (a necessary condition implied by the single signature dichotomy), and states that the problem is still $\#P$-hard unless $\mathcal{F}$ is contained in a single tractable class. In other words, no mixing of distinct tractable classes can occur. The reason we organize the proof in this way is because the No-Mixing theorem will be useful in the proof of the single signature dichotomy.

Thus in this section, we prove that various tractable signatures, if they do not belong to a single tractable class, will produce $\#P$-hardness in combination. There are three tractable classes. To help understand these classes, Figure 8.4 contains a Venn diagram of the signatures in $\mathcal{A}$, $\mathcal{P}$, and $\mathcal{M}$. Note that non-trivial intersections occur; but crucially $\mathcal{M} \cap \mathcal{P} \subset \mathcal{A}$, thus there is an empty cell in the Venn diagram

$$\mathcal{M} \cap \mathcal{P} = \emptyset.$$  \hspace{1cm} (8.1)

We will prove five lemmas giving pairwise No-Mixing statements. In the first two lemmas
Figure 8.4: Venn diagram of the tractable Pl-#CSP signature sets in the Hadamard basis. Each signature has been normalized for simplicity of presentation. For a signature $f$, the notation $f_{\geq k}$ is short for “arity($f$) $\geq k$”. Notice that $\mathcal{M} \cap \mathcal{P} - \mathcal{A}$ is empty.
one of the signatures is unary. The last three lemmas consider the general case of the three pairwise no-mixing statements between three tractable classes, respectively.

Lemma 8.17 is a no-mixing lemma of $\mathcal{A}$ versus $\overline{\mathcal{P}}$, where the unary is not in $\mathcal{A}$. Note that the condition $ab \neq 0$ and $a^4 \neq b^4$ is equivalent to $[a, b] \notin \mathcal{A}$. Also note that since $\mathcal{P}$ contains all unary signatures, so does $\overline{\mathcal{P}}$.

**Lemma 8.17.** Let $a, b \in \mathbb{C}$. Suppose $f \in \mathcal{A} - \overline{\mathcal{P}}$. If $ab \neq 0$ and $a^4 \neq b^4$, then the problem Pl-Holant($\{f, [a, b]\} \cup \mathcal{EQ}$) is $\#P$-hard.

**Proof.** To understand what forms a signature $f \in \mathcal{A} - \overline{\mathcal{P}}$ may take, we may consider a holographic transformation $H_2 = [\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}]$. Note that $H_2 \in \text{Stab}(\mathcal{A})$, then $H_2 f \in \mathcal{A} - \overline{\mathcal{P}}$. Since $\mathcal{F}_1 \subset \mathcal{P}$, we have $H_2 f \in \mathcal{F}_2 \cup \mathcal{F}_3 - \mathcal{P}$. By analyzing the forms in $\mathcal{F}_2 \cup \mathcal{F}_3 - \mathcal{P}$ we can obtain the following possibilities for $f$, up to a nonzero scalar.

- $[1, 0, \pm i]$;
- $[1, 0, \ldots, 0, x]$ of arity at least 3 with $x^4 = 1$;
- $[1, \pm 1, -1, \mp 1, 1, \pm 1, -1, \mp 1, \ldots]$ of arity at least 2;
- $[1, 0, -1, 0, 1, 0, -1, 0, \ldots, 0]$ or $[1, 0, (-1), 0, \ldots]$ of arity at least 3;
- $[0, 1, 0, -1, 0, 1, 0, -1, \ldots, 0]$ or $[0, 1, (-1), 0, \ldots]$ of arity at least 3.

We handle these cases below. Our general technique is to produce a binary signature that does not fit the tractability criterion of Theorem 8.3 for spin systems. In one case we also use Corollary 8.14 and Lemma 8.16 based on the domain pairing technique.

1. Suppose $f = [1, 0, \pm i]$. Connecting $[a, b]$ to $[1, 0, 1, 0]$ gives $[a, b, a]$, and connecting two copies of $[1, 0, \pm i]$ to $[a, b, a]$, one on each edge, gives $g = [a, \pm ib, -a]$. Then Pl-Holant($g \mid \mathcal{EQ}$) is $\#P$-hard by Theorem 8.3.

2. Suppose $f = [1, 0, \ldots, 0, x]$ of arity at least 3 with $x^4 = 1$. Connecting $[a, b]$ to $f$ gives $g = [a, 0, \ldots, 0, bx]$ of arity at least 2. Note that $(bx)^4 = b^4 \neq a^4$. If the arity of $g$ is exactly 2, then Pl-Holant($\{f, g\} \cup \mathcal{EQ}$) is $\#P$-hard by Corollary 8.14, so we are done. Otherwise, the arity of $g$ is at least 3 and Pl-Holant($\{g\} \cup \mathcal{EQ}$) is $\#P$-hard by Lemma 8.16.

3. Suppose $f = [1, \pm 1, -1, \ldots]$ of arity at least 2. Connecting some number of $[1, 0]$ gives $[1, \pm 1, -1]$ of arity exactly 2. Connecting $[a, b]$ to $[1, 0, 1, 0]$ gives $[a, b, a]$ and connecting two copies of $[a, b, a]$ to $[1, \pm 1, -1]$, one on each edge, gives $g = [a^2 \pm 2ab - b^2, \pm(a^2 + b^2), -a^2 \pm 2ab + b^2]$. This is easily verified by

$$
\begin{bmatrix}
    a & b \\
    b & a
\end{bmatrix}
\begin{bmatrix}
    1 & \pm 1 \\
    \pm 1 & -1
\end{bmatrix}
\begin{bmatrix}
    a & b \\
    b & a
\end{bmatrix}
= 
\begin{bmatrix}
    a^2 \pm 2ab - b^2 & \pm(a^2 + b^2) \\
    \pm(a^2 + b^2) & -a^2 \pm 2ab + b^2
\end{bmatrix}.
$$

Then Pl-Holant($g \mid \mathcal{EQ}$) is $\#P$-hard by Theorem 8.3.
4. Suppose \( f = [1, 0, -1, 0, \ldots] \) of arity at least 3. Connecting some number of \([1, 0]\) gives \( g = [1, 0, -1, 0] \) of arity exactly 3. Connecting \([a, b]\) to \( g \) gives \( h = [a, -b, -a] \). Then \( \text{Pl-Holant}(h \mid \widehat{\mathcal{Q}}) \) is \#P-hard by Theorem 8.3.

5. The argument for \( f = [0, 1, 0, -1, \ldots] \) is similar to the previous case. \( \square \)

Before we move to the next lemma, we observe that in Lemma 8.17 we did not explicitly require that \( \{f, [a, b]\} \not\subseteq \mathcal{M} \). In general, a pairwise No-Mixing statement should say that for some \( f \) and \( g \), if \( \{f, g\} \) is not contained in a single tractable class, then their joint presence leads to \#P-hardness. The condition in Lemma 8.17 ensures that \( \{f, [a, b]\} \not\subseteq \mathcal{A} \) and \( \{f, [a, b]\} \not\subseteq \mathcal{P} \). However, even though it was not explicitly stated that \( \{f, [a, b]\} \not\subseteq \mathcal{M} \), this is implied since \( ab \neq 0 \) violates the parity requirement for \([a, b] \in \mathcal{M} \).

The next lemma is a no-mixing statement of \( \mathcal{M} \) versus \( \mathcal{A} \) involving a unary signature \([a, b]\). Note that the condition \( ab \neq 0 \) is equivalent to \([a, b] \not\subseteq \mathcal{M} \). We explicitly require that \( f \in \mathcal{M} - \mathcal{A} \), but we do not require that \([a, b] \in \mathcal{A} \). This gives us a slightly wider applicability. Also note that we do not explicitly require that \( \{f, [a, b]\} \not\subseteq \mathcal{P} \). This is because \( f \in \mathcal{M} - \mathcal{A} \) implies that \( f \not\subseteq \mathcal{P} \) by the observation (8.1).

**Lemma 8.18.** Let \( a, b \in \mathbb{C} \). If \( f \in \mathcal{M} - \mathcal{A} \) and \( ab \neq 0 \), then \( \text{Pl-Holant}(\{f, [a, b]\} \cup \widehat{\mathcal{Q}}) \) is \#P-hard.

**Proof.** By considering the forms in Theorem 4.11 for \( f \in \mathcal{M} \), it is easy to see that, up to a nonzero scalar, the possibilities for \( f \) are as follows:

- \([1, 0, r] \) with \( r \neq 0 \) and \( r^4 \neq 1 \);
- \([1, 0, r, 0, r^2, 0, \ldots] \) of arity at least 3 with \( r \neq 0 \) and \( r^2 \neq 1 \);
- \([0, 1, 0, r, 0, r^2, \ldots] \) of arity at least 3 with \( r \neq 0 \) and \( r^2 \neq 1 \);
- \([0, 1, 0, \ldots, 0] \) of arity at least 3;
- \([0, \ldots, 0, 1, 0] \) of arity at least 3.

We handle these cases below. Again our technique is to produce a binary signature that does not fit the tractability criterion of Theorem 8.3 for spin systems.

1. Suppose \( f = [1, 0, r] \) with \( r^4 \neq 1 \) and \( r \neq 0 \). Connecting \([a, b]\) to \([1, 0, 1, 0]\) gives \([a, b, a]\) and connecting two copies of \([1, 0, r]\) to \([a, b, a]\), one on each edge, gives \( g = [a, br, ar^2] \).

   If \( a^2 \neq b^2 \), then \( \text{Pl-Holant}(g \mid \widehat{\mathcal{Q}}) \) is \#P-hard by Theorem 8.3.

   Otherwise, \( a^2 = b^2 \) and we first connect \([a, b]\) to \([1, 0, r]\) to get \([a, br]\). Then by the same construction, we have \( g' = [a, br^2, ar^2] \) and \( \text{Pl-Holant}(g' \mid \widehat{\mathcal{Q}}) \) is \#P-hard by Theorem 8.3.

2. Suppose \( f = [1, 0, r, 0, \ldots] \) of arity at least 3 with \( r^2 \neq 1 \) and \( r \neq 0 \). Connecting some number of \([1, 0]\) gives \( g = [1, 0, r, 0] \) of arity exactly 3. Connecting \([a, b]\) to \( g \) gives \( h = [a, br, a] \). If \( a^2 \neq b^2r \), then \( \text{Pl-Holant}(h \mid \widehat{\mathcal{Q}}) \) is \#P-hard by Theorem 8.3.
Otherwise, \( a^2 = b^2r \) and we first connect \([1,0] \) and \([a,b] \) to \([1,0,r,0] \) to get \([a,br] \). Then by the same construction, we have \( g' = [a,br^2,ar] \) and \( \text{Pl-Holant}(g' \mid \hat{\mathcal{E}}\mathcal{Q}) \) is \#P-hard by Theorem 8.3.

3. The argument for \( f = [0,1,0,r,\ldots] \) is similar to the previous case.

4. Suppose \( f = [0,1,0,\ldots,0] \) of arity \( k \geq 3 \). Connecting \( k - 2 \) copies of \([a,b] \) to \( f \) gives \( g = a^{k-3}[(k-2)b,a,0] \). Then \( \text{Pl-Holant}(g \mid \hat{\mathcal{E}}\mathcal{Q}) \) is \#P-hard by Theorem 8.3.

5. The argument for \( f = [0,\ldots,0,1,0] \) is similar to the previous case. \( \Box \)

Now we consider the general case of two signatures from two different tractable sets. The three tractable classes \( \mathcal{A}, \hat{\mathcal{P}} \) and \( \mathcal{M} \) give rise to three pairs of pairwise cases to consider, each of which is covered in one of the next three lemmas.

**Lemma 8.19 ((\( \mathcal{A}, \hat{\mathcal{P}} \))-No-Mixing).** If \( f \in \mathcal{A} - \hat{\mathcal{P}} \) and \( g \in \hat{\mathcal{P}} - \mathcal{A} \), then \( \text{Pl-Holant}(\{f,g\} \cup \hat{\mathcal{E}}\mathcal{Q}) \) is \#P-hard.

**Proof.** The only possibility for \( g \) is \([a,b,a,b,\ldots] \), where \( ab \neq 0 \) and \( a^4 \neq b^4 \). Connecting some number of \([1,0] \) to \( g \) gives \([a,b] \) and we are done by Lemma 8.17. \( \Box \)

**Lemma 8.20 ((\( \mathcal{M}, \mathcal{A} \))-No-Mixing).** If \( f \in \mathcal{M} - \mathcal{A} \) and \( g \in \mathcal{A} - \mathcal{M} \), then \( \text{Pl-Holant}(\{f,g\} \cup \hat{\mathcal{E}}\mathcal{Q}) \) is \#P-hard.

**Proof.** Suppose \( g \) does not contain a 0 entry. Then after connecting some number of \([1,0] \) to \( g \), we have a unary signature \([a,b] \) with \( ab \neq 0 \), and are done by Lemma 8.18.

Otherwise, \( g \) contains a 0 entry. Then, by the forms of \( \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \) and \( \mathcal{M} \), it must be a \text{GEN-EQ} (see Figure 8.4), \( g = [x,0,\ldots,0,y] \) of arity at least 3 with \( xy \neq 0 \) (and \( x^4 = y^4 \)). Up to a nonzero scalar, the possibilities for \( f \) are as follows:

- \([1,0,r] \) with \( r \neq 0 \) and \( r^4 \neq 1 \);
- \([1,0,r,0,r^2,0,\ldots] \) of arity at least 3 with \( r \neq 0 \) and \( r^2 \neq 1 \);
- \([0,1,0,r,0,r^2,\ldots] \) of arity at least 3 with \( r \neq 0 \) and \( r^2 \neq 1 \);
- \([0,1,0,0,\ldots,0] \) of arity at least 3;
- \([0,\ldots,0,1,0] \) of arity at least 3.

We handle these cases below.

1. Suppose \( f = [1,0,r] \) with \( r \neq 0 \) and \( r^4 \neq 1 \). Then we are done by Corollary 8.14.
2. Suppose \( f = [1,0,r,0,\ldots] \) of arity at least 3 with \( r \neq 0 \) and \( r^2 \neq 1 \). After connecting some number of \([1,0] \) to \( f \), we have \( h = [1,0,r,0] \) of arity exactly 3. Then \( \text{Pl-Holant}(\{g,h\} \cup \hat{\mathcal{E}}\mathcal{Q}) \) is \#P-hard by Lemma 8.13.

294
3. Suppose $f = [0, 1, 0, r, \ldots]$ of arity at least 3 with $r \neq 0$ and $r^2 \neq 1$. After connecting some number of $[1, 0]$ to $f$, we have $h = [0, 1, 0, r]$ of arity exactly 3. Connecting two more copies of $[1, 0]$ to $f$ gives $h = [0, 1]$. Then we apply a holographic transformation by $T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, so $g$ is transformed to $\hat{g} = [y, 0, \ldots, 0, x]$ and $h$ is transformed to $\hat{h} = [r, 0, 1, 0]$. Since $T$ is orthogonal, Pl-Holant($\{g, h\} \cup \tilde{\mathcal{Q}}$) is transformed to Pl-Holant($\{\hat{g}, \hat{h}\} \cup T\tilde{\mathcal{Q}}$). Every even arity signature in $\mathcal{Q}$ remains unchanged after a holographic transformation by $T$. By attaching $[0, 1]$ to $\{\hat{g}, \hat{h}\} \cup T\tilde{\mathcal{Q}}$, we obtain all of the odd arity signatures in $\mathcal{Q}$ again. Then Pl-Holant($\{\hat{g}, \hat{h}\} \cup \tilde{\mathcal{Q}}$) is #P-hard by Lemma 8.13.

4. Suppose $f = [0, 1, 0, \ldots, 0]$ of arity $k \geq 3$. The gadget in Figure 8.5 with $f$ assigned to both vertices has signature $h = [k - 1, 0, 1]$. Then Pl-Holant($\{g, h\} \cup \tilde{\mathcal{Q}}$) is #P-hard by Corollary 8.14.

5. The argument for $f = [0, \ldots, 0, 1, 0]$ is similar to the previous case.

Note that in Lemma 8.19 and 8.20 we do not explicitly require that $\{f, g\}$ is not contained in the third tractable class. But that is implied (as it should). In Lemma 8.19, $g \not\in \mathcal{M} - \mathcal{A}$ implies that $g \not\in \mathcal{M}$. In Lemma 8.20, $f \not\in \mathcal{M} - \mathcal{A}$ implies that $f \not\in \mathcal{P}$. Both these statements follow from the observation (8.1). However in Lemma 8.21 we do have to explicitly require that $\{f, g\} \not\subseteq \mathcal{A}$.

**Lemma 8.21** ((\(\mathcal{P}, \mathcal{M}\))-No-Mixing). If $f \in \mathcal{M} - \mathcal{P}$ and $g \in \mathcal{P} - \mathcal{M}$ and $\{f, g\} \not\subseteq \mathcal{A}$, then Pl-Holant($\{f, g\} \cup \tilde{\mathcal{Q}}$) is #P-hard.

**Proof.** The only possibility for $g$ is $[a, b, a, b, \ldots]$ with $ab \neq 0$ (see Figure 8.4). Connecting some number of $[1, 0]$ to $g$ gives $h = [a, b]$. If $f \not\subseteq \mathcal{A}$, then Pl-Holant($\{f, h\} \cup \tilde{\mathcal{Q}}$) is #P-hard by Lemma 8.18. Otherwise, $f \in \mathcal{A}$, so $g \not\subseteq \mathcal{A}$ and Pl-Holant($\{f, g\} \cup \tilde{\mathcal{Q}}$) is #P-hard by Lemma 8.19.

Observe that in the proof of Lemma 8.21 when we apply Lemma 8.18 to $\{f, h\}$, we have $f \in \mathcal{M} - \mathcal{A}$, but we do not necessarily have the condition $h \in \mathcal{A} - \mathcal{M}$. We only have $h \not\in \mathcal{M}$. However Lemma 8.18 does not require that $h \in \mathcal{A}$. Also note that we choose to prove the three lemmas, Lemma 8.19, 8.20, and 8.21, in that order so that the proof of Lemma 8.21 can apply Lemma 8.19.
We summarize this section with the following theorem, which says that the tractable signature sets cannot mix. Signatures from different tractable sets, when put together, lead to \#P-hardness.

**Theorem 8.22** (No-Mixing Theorem). Let \( F \) be any set of symmetric, complex-valued signatures in Boolean variables. If \( F \subseteq A \cup \widehat{P} \cup \mathcal{M} \), then \( \text{Pl-Holant}(F \cup \widehat{E}Q) \) is \#P-hard unless \( F \subseteq A \), or \( F \subseteq \widehat{P} \), or \( F \subseteq \mathcal{M} \), in which case \( \text{Pl-Holant}(F \cup \widehat{E}Q) \) is tractable.

**Proof.** If \( F \) is a subset of \( A \), \( \widehat{P} \), or \( \mathcal{M} \), then the tractability is given in Theorem 8.4. Otherwise \( F \) is not a subset of \( A \), \( \widehat{P} \), or \( \mathcal{M} \). Then \( F \) contains a signature \( g \in (\widehat{P} \cup \mathcal{M}) - A \) since \( F \not\subseteq A \). As \( F \subseteq A \cup \widehat{P} \cup \mathcal{M} \), either \( F \) contains a signature \( f \in A - (\widehat{P} \cup \mathcal{M}) \), or \( F \subseteq \widehat{P} \cup \mathcal{M} \). Suppose it is the first case. If \( g \in \widehat{P} - A \), then \( \text{Pl-Holant}(F \cup \widehat{E}Q) \) is \#P-hard by Lemma 8.19. Otherwise, \( g \in A - \widehat{P} \) and \( \text{Pl-Holant}(F \cup \widehat{E}Q) \) is \#P-hard by Lemma 8.20.

Now suppose it is the second case that \( F \subseteq \widehat{P} \cup \mathcal{M} \). Since \((\widehat{P} \cap \mathcal{M}) - A \) is empty by the observation (8.1) (see Figure 8.4), either \( g \in (\mathcal{M} - \widehat{P}) - A \) or \( g \in (\mathcal{M} - \widehat{P}) - A \). If \( g \in (\mathcal{M} - \widehat{P}) - A \), then there exists a signature \( f \in \mathcal{M} - \widehat{P} \) since \( F \not\subseteq \widehat{P} \). In which case, \( \text{Pl-Holant}(F \cup \widehat{E}Q) \) is \#P-hard by Lemma 8.21. Otherwise, \( g \in (\mathcal{M} - \widehat{P}) - A \) and there exists a signature \( f \in \mathcal{M} - \widehat{P} \) since \( F \not\subseteq \widehat{P} \). In which case, \( \text{Pl-Holant}(F \cup \widehat{E}Q) \) is \#P-hard again by Lemma 8.21.

**8.6 Pinning for Planar Graphs**

The idea of pinning is a common reduction technique between counting problems. For the Boolean \#CSP framework, pinning fixes some variables to constants 0 or 1, and is represented by the unary signatures \([1; 0]\) and \([0; 1]\) respectively.

Dyer, Goldberg, and Jerrum [?] showed that one can pin both to 0 and 1 in the Pl-\#CSP framework; this is Lemma 3.13. However this reduction is not planar. As it was explained in Section 8.1, it is not possible to prove a general pinning lemma as Lemma 3.13 for the Pl-\#CSP framework, assuming \#P does not collapse to P. However, in [?] it is shown that for real-weighted Pl-\#CSP this difficulty can be overcome if we first perform a holographic transformation by the Hadamard matrix \( H_2 = [\begin{array}{ll} 1 & 1 \\ 1 & 1 \end{array}] \). Thus in this Hadamard basis, pinning can be done in a planar way. This holographic transformation is necessary.

The expression of \( \text{Pl-\#CSP}(F) \) in the Hadamard basis is \( \text{Pl-Holant}(H_2F \cup \widehat{E}Q) \). In \( \widehat{E}Q = \{[1, 0], [1, 0, 1], [1, 0, 1, 0], \ldots\} \) we already have “half” of the pinning pair \([1, 0]\). So pinning in \( \text{Pl-Holant}(H_2F \cup \widehat{E}Q) \) amounts to obtaining the missing signature \([0, 1]\). We remark that there is no intrinsic precedence of \([1, 0]\) over \([0, 1]\). This is merely an artifact of our choice of \( H_2 \). We could have made an equally valid choice \( H'_2 = [\begin{array}{ll} -1 & 1 \\ 1 & -1 \end{array}] \), then \([0, 1] \in H'_2\widehat{E}Q \).
but \([1,0] \not\in H_2^sEQ\). Getting the other “half” of the pair \([[1,0],[0,1]]\) will be a significant step toward proving the dichotomy theorem Theorem 8.34.

**The Road to Pinning**

We will try to prove a dichotomy for \(\text{Pl-Holant}(C \cup \overline{EQ})\), for a general \(C\). When it is translated back to \(\text{Pl-}\#\text{CSP}(F)\) for any \(F\), we will apply the dichotomy on \(\text{Pl-Holant}(\tilde{F} \cup \overline{EQ})\). An important step for the dichotomy of \(\text{Pl-Holant}(C \cup \overline{EQ})\) is to prove that it is equivalent to \(\text{Pl-Holant}^c(C \cup \overline{EQ})\).

We begin the road to pinning with a lemma that assumes the presence of \([0,0,1] = [0,1]^{\otimes 2}\), which is the tensor product of two copies of \([0,1]\). In our pursuit to realize \([0,1]\), this may be as close as we can get to directly. For example suppose every signature in the set has even parity. Recall that a signature has even parity if its support is on entries of even Hamming weight. By a simple parity argument, gadgets constructed with signatures of even parity can only realize signatures of even parity. But obviously \([0,1]\) does not have even parity.

However, if every signature has even parity and \([0,0,1]\) is present, then we can already prove a dichotomy. To prove this, we first prove the following simple lemma.

**Lemma 8.23.** Let \(a, b, c \in \mathbb{C}\). If \(ab \neq 0\), then \(\text{Pl-Holant}([a,0,0,0,b,c])\) is \(\#P\)-hard.

**Proof.** Let \(f\) be the signature of the gadget in Figure 8.6 with \([a,0,0,0,b,c]\) assigned to both vertices. The signature matrix of \(f\) is

\[
\begin{bmatrix}
a^2 & 0 & 0 & 0 \\
0 & b^2 & b^2 & bc \\
0 & b^2 & b^2 & bc \\
0 & bc & bc & 3b^2 + c^2
\end{bmatrix},
\]

which is redundant. Its compressed form is nonsingular since its determinant is \(6a^2b^4 \neq 0\). Thus, we are done by Corollary 7.25. \(\square\)

**Lemma 8.24.** Suppose \(F\) is a set of symmetric signatures with complex weights containing \([0,0,1]\). If every signature in \(F\) has even parity, then either \(\text{Pl-Holant}(F \cup \overline{EQ})\) is \(\#P\)-hard or \(F\) is a subset of \(A\), \(\tilde{P}\), or \(M\), in which case \(\text{Pl-Holant}^c(F \cup \overline{EQ})\) is tractable.
Proof. The tractability is given in Theorem 8.4.

Any degenerate symmetric signature having even parity has the form \( \lambda[1, 0]^{\otimes n} \), or \( \lambda[0, 1]^{\otimes n} \) if the arity \( n \) is even. Both belong to \( \mathcal{A} \cap \mathcal{P} \cap \mathcal{M} \). In the following we consider non-degenerate signatures of \( \mathcal{F} \).

Every signature of arity at most 3 satisfying the parity condition is in \( \mathcal{M} \). If \( \mathcal{F} \) has no signature of arity greater than 3, then \( \mathcal{F} \subseteq \mathcal{M} \) and we are done.

So we may suppose \( \mathcal{F} \) contains some non-degenerate signature of arity at least 4. For every signature \( f \in \mathcal{F} \) with \( f = [f_0, f_1, \ldots, f_m] \) and \( m \geq 4 \), using \([0, 0, 1]\) and \([1, 0]\), we can obtain all subsignatures of the form \( [f_{k-2}, 0, f_k, 0, f_{k+2}] \) for any even \( k \) such that \( 2 \leq k \leq m - 2 \). If there is any subsignature \( g \) of this form satisfies \( f_{k-2} f_{k+2} \neq f_k^2 \) and \( f_k \neq 0 \), then we are done by Corollary 7.25.

Otherwise all subsignatures of the above form satisfy

\[
f_{k-2} f_{k+2} = f_k^2 \quad \text{or} \quad f_k = 0
\]

for all even \( k \) such that \( 2 \leq k \leq m - 2 \). We classify all non-degenerate signatures \( f \in \mathcal{F} \) with arity \( m \geq 4 \) into two types as follows. If there exists an even index \( 2 \leq k \leq m - 2 \) such that \( f_k \neq 0 \), then by the equation \( f_{k-2} f_{k+2} = f_k^2 \neq 0 \) we conclude that \( f_k 
eq 0 \) for all even index \( 0 \leq \ell \leq m \), and the signature entries of even Hamming weight form a geometric progression. These signatures have the form

\[
[\alpha^n, 0, \alpha^{n-1} \beta, 0, \ldots, 0, \alpha \beta^{n-1}, 0, \beta^n] \quad \text{or} \quad [\alpha^n, 0, \alpha^{n-1} \beta, 0, \ldots, 0, \alpha \beta^{n-1}, 0, \beta^n, 0] \quad (8.2)
\]

for some \( \alpha, \beta \in \mathbb{C} \), which are in \( \mathcal{M} \). Otherwise, \( f_k = 0 \) for all even index \( 2 \leq k \leq m - 2 \). Then the signature has the form

\[
[x, 0, \ldots, 0, y] \quad (8.3)
\]

of even arity \( m \geq 4 \), where \( xy \neq 0 \) by being non-degenerate, or

\[
[x, 0, \ldots, 0, y, 0] \quad (8.4)
\]

of odd arity \( m \geq 5 \), where \( y \neq 0 \) by being non-degenerate. We can group the subcase of (8.4) with \( x = 0 \) as part of the second case of (8.2) with \( \alpha = 0 \). Then we call signatures in (8.2) type 1 signatures. Type 2 signatures are those in (8.3) and (8.4) with both \( x \neq 0 \) and \( y \neq 0 \). Note that there are an odd number of 0’s between \( x \) and \( y \) in type 2 signatures (since they have even parity). If all non-degenerate signatures of arity at least 4 in \( \mathcal{F} \) are of type 1, then including all signatures of arity at most 3 in \( \mathcal{F} \), we have \( \mathcal{F} \subseteq \mathcal{M} \).

Otherwise \( \mathcal{F} \) contains a signature of type 2. Suppose there is some signature \( f = [x, 0, \ldots, 0, y, 0] \) of arity at least 5 with \( xy \neq 0 \) in (8.4). After zero or more self-loops, we have \( g = [x, 0, 0, 0, y, 0] \) of arity exactly 5. Then we are done by Lemma 8.23.

Otherwise all type 2 signatures in \( \mathcal{F} \) are from (8.3). If there is any such \( f = [x, 0, \ldots, 0, y] \) of arity at least 4 with \( xy \neq 0 \) and \( x^4 \neq y^4 \), then we are done by Lemma 8.16. Otherwise all type 2 signatures in \( \mathcal{F} \) are from (8.3) and satisfy \( x^4 = y^4 \). This puts every signature in \( \mathcal{F} \) of type 2 in \( \mathcal{A} \). Therefore \( \mathcal{F} \subseteq \mathcal{A} \cup \mathcal{M} \) and we are done by Theorem 8.22. \qed
The conclusion of every result in the remainder of Section 8.6 states that we are able to pin (under various assumptions on \( F \)). Formally speaking, we repeatedly prove that Pl-Holant\(^c\)(\( F \cup \widehat{\mathcal{Q}} \)) is \#P-hard (or in P) if and only if Pl-Holant(\( F \cup \widehat{\mathcal{Q}} \)) is \#P-hard (or in P). The difference between these two counting problems is the presence of \([0, 1]\) in Pl-Holant\(^c\)(\( F \cup \widehat{\mathcal{Q}} \)). We always prove this statement in one of three ways:

1. either we show that Pl-Holant\(^c\)(\( F \cup \widehat{\mathcal{Q}} \)) is tractable (so Pl-Holant(\( F \cup \widehat{\mathcal{Q}} \)) is as well);
2. or we show that Pl-Holant(\( F \cup \widehat{\mathcal{Q}} \)) is \#P-hard (so Pl-Holant\(^c\)(\( F \cup \widehat{\mathcal{Q}} \)) is as well);
3. or we show how to reduce Pl-Holant\(^c\)(\( F \cup \widehat{\mathcal{Q}} \)) to Pl-Holant(\( F \cup \widehat{\mathcal{Q}} \)) by realizing \([0, 1]\) using signatures in \( F \cup \widehat{\mathcal{Q}} \).

Note that these 3 cases together logically imply that

\[
\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{Q}}) \leq_T \text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{Q}}).
\]  

(8.5)

However this reduction statement is not proved by directly obtaining the signature \([0, 1]\) with a gadget construction or interpolation in all cases.

**Lemma 8.25.** Let \( F \) be any set of complex-weighted symmetric signatures containing \([0, 0, 1]\). Then Pl-Holant\(^c\)(\( F \cup \widehat{\mathcal{Q}} \)) is \#P-hard (or in P) iff Pl-Holant(\( F \cup \widehat{\mathcal{Q}} \)) is \#P-hard (or in P).

**Proof.** If we had a unary signature \([a, b]\) where \( b \neq 0 \), then connecting \([a, b]\) to \([0, 0, 1]\) gives the signature \([0, b]\), which is \([0, 1]\) after normalizing. Thus, in order to reduce Pl-Holant\(^c\)(\( F \cup \widehat{\mathcal{Q}} \)) to Pl-Holant(\( F \cup \widehat{\mathcal{Q}} \)) by constructing \([0, 1]\), it suffices to construct a unary signature \([a, b]\) with \( b \neq 0 \).

For every signature \( f \in \mathcal{F} \) with \( f = [f_0, f_1, \ldots, f_m] \), using \([0, 0, 1]\) and \([1, 0]\), we can obtain all subsignatures of the form \([f_{k-1}, f_k]\) for any odd \( k \) such that \( 1 \leq k \leq m \). If any subsignature satisfies \( f_k \neq 0 \), where \( k \) is odd, then we can construct \([0, 1]\).

Otherwise all signatures in \( \mathcal{F} \) have even parity and we are done by Lemma 8.24.

There are two scenarios that lead to Lemma 8.25, which are the focus of the next two lemmas.
Lemma 8.26. For $x \in \mathbb{C}$, let $F$ be any set of complex-weighted symmetric signatures containing $[1,0,x]$ such that $x \notin \{0, \pm 1\}$. Then $\text{Pl-Holant}^c(F \cup \mathcal{EQ})$ is #P-hard (or in P) iff $\text{Pl-Holant}(F \cup \mathcal{EQ})$ is #P-hard (or in P).

Proof. There are two cases. In either case, we realize $[0,0,1]$ and finish by applying Lemma 8.25.

First we claim that the conclusion holds provided the norm $|x| \notin \{0,1\}$. Combining $k$ copies of $[1,0,x]$ gives $[1,0,x^k]$. Since $|x| \notin \{0,1\}$, $x$ is neither zero nor a root of unity, so we can use polynomial interpolation to realize $[a,0,b]$ for any $a,b \in \mathbb{C}$, including $[0,0,1]$.

Otherwise $|x| = 1$. The gadget in Figure 8.7 has signature $[f_0,f_1,f_2] = [1+x^2,0,2x]$. If $x = \pm i$, then we have $[0,0,\pm 2i]$, which is $[0,0,1]$ after normalizing.

Otherwise $x \neq \pm i$, so $f_0 \neq 0$. Since $x \neq 0$, we have $f_2 \neq 0$. Since $x \neq \pm 1$ and the norm $|x| = 1$, we have $|f_0| < 2$. However, $|f_2| = 2$. Therefore, after normalizing, the signature $[1,0,y]$ with $y = \frac{2x}{1+x^2}$ has $|y| > 1$, so it can interpolate $[0,0,1]$ by our initial claim since $|y| \notin \{0,1\}$. 

Lemma 8.27. Let $F$ be any set of complex-weighted symmetric signatures containing a signature $[f_0,f_1,\ldots,f_n]$ that is not identically zero but has $f_0 = 0$. Then $\text{Pl-Holant}^c(F \cup \mathcal{EQ})$ is #P-hard (or in P) iff $\text{Pl-Holant}(F \cup \mathcal{EQ})$ is #P-hard (or in P).

Proof. If $f_1 \neq 0$, then we connect $n-1$ copies of $[1,0]$ to $f$ to get $[0,f_1]$, which is $[0,1]$ after normalizing. If $f_1 = 0$, then $n \geq 2$. If $f_2 \neq 0$, then we connect $n-2$ copies of $[1,0]$ to $f$ to get $[0,0,f_2]$, which is $[0,0,1]$ after normalizing. Then we are done by Lemma 8.25. If $f_1 = f_2 = 0$, then $n \geq 3$, and $f = [0,0,0,f_3,\ldots,f_n]$. After zero or more self-loops, we get a signature with exactly one or two initial 0’s, which is one of the above scenarios.

As a significant step toward pinning for any signature set $F$, we show how to pin given any binary signature. Some cases resist pinning and are excluded.

Lemma 8.28. Let $F$ be any set of complex-weighted symmetric signatures containing a binary signature $f$. Then $\text{Pl-Holant}^c(F \cup \mathcal{EQ})$ is #P-hard (or in P) iff $\text{Pl-Holant}(F \cup \mathcal{EQ})$ is #P-hard (or in P) unless $f \in \{[0,0,0],[1,0,-1],[1,r,r^2],[1,b,1]\}$, up to a nonzero scalar, for any $b,r \in \mathbb{C}$.

Proof. Let $f = [f_0,f_1,f_2] \neq [0,0,0]$. If $f_0 = 0$ then we are done by Lemma 8.27. Otherwise $f_0 \neq 0$, and we normalize $f_0$ to 1. If $\text{Pl-Holant}(f \mid \mathcal{EQ})$ is #P-hard by Theorem 8.3, then $\text{Pl-Holant}(F \cup \mathcal{EQ})$ is also #P-hard. Otherwise, $f$ is one of the tractable cases, which implies that

$$f \in \{[1,r,r^2],[1,0,x],[1,\pm 1,-1],[1,b,1]\}.$$  

If $f = [1,\pm 1,-1]$, then we connect $f$ to $[1,0,1,0] \in \mathcal{EQ}$ to get $[0,\pm 2]$, which is $[0,1]$ after normalizing. If $f = [1,0,x]$, then we are done by Lemma 8.26 unless $x \in \{0,\pm 1\}$. The remaining cases are all excluded by assumption, so we are done.
Pinning in the Hadamard Basis

Before we show how to pin in the Hadamard basis, we handle two simple cases.

**Lemma 8.29.** If $F$ is a set of signatures containing $[1, \pm i]$, then we have $\text{Pl-Holant}^c(F \cup \widehat{\mathcal{Q}}) \leq \text{Pl-Holant}(F \cup \widehat{\mathcal{Q}})$.

**Proof.** Connect two copies of $[1, \pm i]$ to $[1, 0, 1, 0] \in \widehat{\mathcal{Q}}$ to get $[0, \pm 2i]$, which is $[0, 1]$ after normalizing. \hfill \square

The next lemma considers the signature $[1, b, 1, b^{-1}]$, which we will also use in Theorem 8.32, the single signature dichotomy.

**Lemma 8.30.** Let $b \in \mathbb{C}$. If $b \notin \{0, \pm 1\}$, then $\text{Pl-Holant}^c([1, b, 1, b^{-1}] \cup \widehat{\mathcal{Q}})$ is $\#P$-hard.

**Proof.** Connect two copies of $[1, 0]$ to $f = [1, b, 1, b^{-1}]$ to get $[1, b]$. Connecting this back to $f$ gives $g = [1 + b^2, 2b, 2]$. Then $\text{Pl-Holant}(g | \widehat{\mathcal{Q}})$ is $\#P$-hard by Theorem 8.3. \hfill \square

Now we are ready to prove our pinning result.

**Theorem 8.31 (Pinning).** Let $F$ be any set of complex-weighted symmetric signatures. Then $\text{Pl-Holant}^c(F \cup \widehat{\mathcal{Q}})$ is $\#P$-hard (or in $P$) iff $\text{Pl-Holant}(F \cup \widehat{\mathcal{Q}})$ is $\#P$-hard (or in $P$).

This theorem does not exclude the possibility that either framework contains a problem of intermediate complexity, i.e., a problem that is neither in $P$ nor $\#P$-hard. It only says that if one framework does not contain a problem of intermediate complexity, then the other framework also does not. We will indeed prove neither happens. Our goal is to prove a dichotomy for $\text{Pl-Holant}(F \cup \widehat{\mathcal{Q}})$. By Theorem 8.31, this is equivalent to proving a dichotomy for $\text{Pl-Holant}^c(F \cup \widehat{\mathcal{Q}})$.

**Proof of Theorem 8.31.** If $F$ contains any identically 0 signature, we can remove it. This does not change the complexity of the problem. For simplicity, we normalize the first nonzero entry of every signature in $F$ to 1. If $F$ contains the degenerate signature $[0, 1]^\otimes n$ for some $n \geq 1$, then we take self-loops on this signature until we have either $[0, 1]$ or $[0, 0, 1]$ (depending on the parity of $n$). If we have $[0, 1]$, we are done. Otherwise, we have $[0, 0, 1]$ and are done by Lemma 8.25.

So we may assume that any degenerate signature in $F$ has the form $[1, b]^\otimes n$. Then we can replace all degenerate signatures in $F$ by their unary versions $[1, b]$ using $[1, 0]$. This does not change the complexity of the problem. If $F$ contains only degenerate signatures, then $F \subseteq \widehat{\mathcal{P}}$ and $\text{Pl-Holant}^c(F \cup \widehat{\mathcal{Q}})$ is tractable by Theorem 8.4.

Otherwise $F$ contains a non-degenerate signature $f$ of arity at least two. We connect some number of $[1, 0]$ to $f$ until we obtain a signature with arity exactly two. We call the
resulting signature the binary prefix of $f$. If this binary prefix is not one of the exceptional forms in Lemma 8.28, then we are done, so assume that it is one of the exceptional forms.

Now we perform case analysis according to the exceptional forms in Lemma 8.28. There are five cases below because we consider $[1, r, r^2]$ as $[1, 0, 0]$ and $[1, r, r^2]$ with $r \neq 0$ as separate cases. In each case, we either show that the conclusion of the theorem holds or that $f \in \mathcal{A} \cup \mathcal{F} \cup \mathcal{M}$. After the case analysis, we then handle all of these tractable $f$ together.

1. Suppose the binary prefix of $f$ is $[0, 0, 0]$. Since $f$ is not identically zero, and $f_0 = 0$, we are done by Lemma 8.27.

2. Suppose the binary prefix of $f$ is $[1, 0, -1]$. If $f$ is not of the form

$$[1, 0, -1, 0, 1, 0, -1, 0, \ldots, 0 \text{ or } 1 \text{ or } (-1)],$$ (8.6)

then in particular $f$ has arity at least 3 and after one self-loop, we have a signature of arity at least one with 0 as its first entry but is not identically zero, so we are done by Lemma 8.27.

Thus, in this case, we may assume $f$ has the form given in (8.6).

3. Suppose the binary prefix of $f$ is $[1, 0, 0]$. Since $f$ is assumed to be non-degenerate, $f \neq [1, 0, \ldots, 0]$, then after connecting $[1, 0]$ for some $k \geq 0$ times, we have $[1, 0, \ldots, 0, x]$ of arity at least 3, where $x \neq 0$. If $x^4 \neq 1$, then Pl-Holant($\{f\} \cup \mathcal{E} \mathcal{Q}$) is #P-hard by Lemma 8.16, so Pl-Holant($\mathcal{F} \cup \mathcal{E} \mathcal{Q}$) is also #P-hard. Otherwise, $x^4 = 1$. Suppose that $x$ is not the last entry in $f$. Then $k \geq 1$. By connecting one fewer $[1, 0]$ than before, we have $g = [1, 0, \ldots, 0, x, y]$ and there are two cases to consider. If the index of $x$ in $g$ is odd, then after zero or more self-loops, we have $h = [1, 0, 0, x, y]$. The determinant of the compressed signature matrix of $h$ is $-2x^2 \neq 0$. Thus, Pl-Holant($h$) is #P-hard by Corollary 7.25, so Pl-Holant($\mathcal{F} \cup \mathcal{E} \mathcal{Q}$) is also #P-hard.

Otherwise, the index of $x$ in $g$ is even. After zero or more self-loops, we have $h = [1, 0, 0, 0, x, y]$. Then by Lemma 8.23, Pl-Holant($h$) is #P-hard, so Pl-Holant($\mathcal{F} \cup \mathcal{E} \mathcal{Q}$) is also #P-hard.

Thus, in this case, we may assume $f = [1, 0, \ldots, 0, x] \text{ with } x^4 = 1$.

4. Suppose the binary prefix of $f$ is $[1, r, r^2]$, where $r \neq 0$. Since $f$ is non-degenerate, $f \neq [1, r, \ldots, r^n]$, then after connecting some number of $[1, 0]$, we have $[1, r, \ldots, r^m, y]$, where $y \neq r^{m+1}$ and $m \geq 2$. Using $[1, 0]$, we can get $[1, r]$. If $r = \pm i$, then we are done by Lemma 8.29, so assume that $r \neq \pm i$. Then we can attach $[1, r]$ back to the initial signature some number of times to get $g = [1, r, r^2, x]$ after factoring out the normalizing constant $1 + r^2 \neq 0$, where $x \neq r^3$. We connect $[1, r]$ once more to get $h = [1 + r^2, r(1 + r^2), r^2 + rx]$. If $h$ does not have one of the exceptional forms in Lemma 8.28, then we are done, so assume that it does.

Since the second entry of $h$ is not 0 and $x \neq r^3$, the only possibility is that $h$ has the form $[1, b, 1]$ up to a scalar. This gives $x = r^{-1}$. Note that $r \neq \pm 1$ since $x \neq r^3$. A
Suppose the binary prefix of $c$. Also every unary signature is in Theorem 8.22, the No-Mixing theorem.

5. Suppose the binary prefix of $f$ is $[1, b, 1]$. If $b = \pm 1$, then this binary prefix is degenerate and thus removed. So we assume $b \neq \pm 1$. If $f$ is not of the form $[1, b, 1, b, \ldots]$, then first suppose that the index of the first entry in $f$ to break the pattern is even. Then after connecting some number of $[1, 0]$, we have $[1, b, 1, \ldots, b, y]$ of arity $\geq 4$, where $y \neq 1$. Then after some number of self-loops and normalizing, we have $g = [1, b, 1, b, x]$, where $x \neq 1$. The determinant of its compressed signature matrix is $2(b^2 - 1)(1 - x) \neq 0$. Thus, Pl-Holant$(g)$ is #P-hard by Cor 7.25, so Pl-Holant$(\mathcal{F} \cup \mathcal{E} \mathcal{Q})$ is also #P-hard.

Otherwise, the index of the first entry in $f$ to break the pattern is odd. Then after connecting some number of $[1, 0]$, we have $[1, b, 1, \ldots, 1, y]$ of arity $\geq 3$, where $y \neq b$. Then after some number of self-loops and normalizing, we have $[1, b, 1, x]$, where $x \neq b$. We do a self-loop to get $g = [2, b + x]$. If $b = 0$, then connecting $g$ to $[1, 0, 1, x]$ gives $h = [2, x, 2 + x^2]$. We assume that $h$ has one of the exceptional forms in Lemma 8.28 since we are done otherwise. Because $x \neq 0$, the only possibility is that $h$ has the form $[1, r, r^2]$ up to a scalar. Then we get $x^2 = -4$, so $g = [2, x] = 2[1, \pm i]$ and we are done by Lemma 8.29. We use the signature $g$ again below.

Otherwise, $b \neq 0$. Using $[1, 0]$, we can get $h = [1, b, 1]$. As $b \neq \pm 1$, $h$ is non-degenerate. If $M_h$ has infinite order modulo a scalar, then we can interpolate $[0, 1]$ by Lemma 8.15 since $b \notin \{0, \pm 1\}$. If the signature matrix $M_h$ of $h$ has finite order modulo a scalar, then $M_h^k = \beta I_2$ for some positive integer $k$ and some nonzero complex value $\beta$. Thus after normalizing, we can construct the anti-gadget $[1, -b, 1]$ by connecting $k - 1$ copies of $h$ together. Connecting $[1, 0]$ to $[1, -b, 1]$ gives $[1, -b]$ and connecting this to $[1, b, 1, x]$ gives $[1 - b^2, 0, 1 - bx]$. If $1 - bx \notin \{0, \pm 1\}$, then we are done by Lemma 8.26.

Otherwise, $y = \frac{1 - bx}{1 - b^2} \notin \{0, \pm 1\}$. For $y = 0$, we get $x = b^{-1}$ and are done by Lemma 8.30 since we have $[1, b, 1, b^{-1}]$ and $b \notin \{0, \pm 1\}$. For $y = 1$, we get $b = x$, a contradiction. For $y = -1$, we get $2 - b^2 - bx = 0$. Then connecting $g = [2, b + x]$ to $[1, -b, 1]$ gives $[2 - b^2 - bx, x - b] = [0, x - b]$, which is $[0, 1]$ after normalizing since $x \neq b$.

Thus, in this case, we may assume $f = [1, b, 1, b, \ldots]$.

At this point, every non-degenerate signature in $\mathcal{F}$ must be of one of the following forms:

- $[1, 0, -1, 0, 1, 0, -1, 0, \ldots, 0 \text{ or } 1 \text{ or } (-1)]$, which is in $\mathcal{A} \cap \mathcal{M}$;
- $[1, 0, \ldots, 0, x]$, where $x^4 = 1$, which is in $\mathcal{A}$;
- $[1, b, 1, b, \ldots, 1 \text{ or } b]$, which is in $\mathcal{P}$.

Also every unary signature is in $\mathcal{P}$. Therefore $\mathcal{F} \subseteq \mathcal{A} \cup \mathcal{P} \cup \mathcal{M}$ and we are done by Theorem 8.22, the No-Mixing theorem. \qed

303
8.7 Planar CSP Dichotomy

In this section, we prove our main dichotomy theorem. We begin with a dichotomy for a single signature.

Recall that Pl-Holant(\{f\} \cup \widehat{\mathcal{Q}}) is the equivalent expression of Pl-\#CSP(\{f\}), where \( \hat{f} = H_2 f \). Theorem 8.32 is expressed in the Pl-Holant language.

**Theorem 8.32.** If \( f \) is a non-degenerate symmetric signature of arity at least 2 with complex weights in Boolean variables, then Pl-Holant(\{f\} \cup \widehat{\mathcal{Q}}) is \#P-hard unless \( f \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M} \), in which case the problem is computable in polynomial time.

**Proof.** When \( f \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M} \), the problem is tractable by Theorem 8.4. When \( f \notin \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M} \), we prove that Pl-Holant(\{f\} \cup \widehat{\mathcal{Q}}) is \#P-hard, which is equivalent because of pinning (Theorem 8.31). The advantage of Pl-Holant is that we can use \([1,0]\) and \([0,1]\) to obtain any subsignature of \( f \).

Notice that once we have \([0,1]\) and \( \widehat{\mathcal{Q}} \), we can realize every signature in \( T\widehat{\mathcal{Q}} \), where \( T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \). In fact, every even arity signature in \( \widehat{\mathcal{Q}} \) is also in \( T\widehat{\mathcal{Q}} \), and we obtain all the odd arity signatures in \( T\widehat{\mathcal{Q}} \) by attaching \([0,1]\) to all the even arity signatures in \( \widehat{\mathcal{Q}} \). Also note that \( T[0,1] = [1,0] \) and \( T[1,0] = [0,1] \). Therefore, Pl-Holant(\{f\} \cup \widehat{\mathcal{Q}}) \equiv_T \text{Pl-Holant}(\{Tf\} \cup \widehat{\mathcal{Q}}). \) Furthermore, \( \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M} \) is closed under \( T \). We use these facts later.

The possibilities for \( f \) can be divided into three cases:

- \( f \) satisfies the parity condition;
- \( f \) does not satisfy the parity condition but does contain a 0 entry;
- \( f \) does not contain a 0 entry.

We handle these cases below.

1. Suppose that \( f \) satisfies the parity condition. If \( f \) has even parity, then we are done by Lemma 8.24.

Otherwise, \( f \) has odd parity. If \( f \) has odd arity, \( f = [0, f_1, 0, \ldots, 0, f_{2k+1}] \), then under a holographic transformation by \( T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \), \( f \) is transformed to \( f' = [f_{2k+1}, 0, \ldots, 0, f_1, 0] \), which has even parity. Then either Pl-Holant(\{f'\} \cup \widehat{\mathcal{Q}}) is \#P-hard by Lemma 8.24 (and thus Pl-Holant(\{f\} \cup \widehat{\mathcal{Q}}) is also \#P-hard), or \( f' \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M} \) (and thus \( f \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M} \)).

Otherwise, \( f \) has odd parity and even arity, \( f = [0, f_1, 0, \ldots, 0, f_{2k-1}, 0] \). Connect \([0,1]\) to \( f \) to get a signature \( g = [f_1, 0, \ldots, 0, f_{2k-1}, 0] \) with even parity and odd arity. Then either Pl-Holant(\{g\} \cup \widehat{\mathcal{Q}}) is \#P-hard by Lemma 8.24 (and thus Pl-Holant(\{f\} \cup \widehat{\mathcal{Q}}) is also \#P-hard), or \( g \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M} \). In the latter case, the last entry of \( g \) is 0, since
$g$ has odd arity and even parity. Then inspection of Figure 8.4 shows that $g \in \mathcal{M}$. In particular, the even parity entries of $g$ form a geometric progression, by Theorem 4.11 (case 2, this includes the form $[0, \ldots, 0, *, 0]$). Therefore $f \in \mathcal{M}$ since $f$ has odd parity and the same geometric progression among its odd parity entries.

2. Suppose that $f$ contains a 0 entry but does not satisfy the parity condition. Then there are two nonzero entries $f_i$ and $f_j$, where $i < j$ and $i$ and $j$ are of the opposite parity. Take two such entries with minimum $j - i$, which is odd. If $j - i > 1$ then any entry $f_\ell$ with $i < \ell < j$ must be 0, since $j - i = (j - \ell) + (\ell - i)$ and one of them must be odd. Thus $f_i$ and $f_j$ are separated by an even number $j - i - 1 \geq 0$ of 0 entries. And so $f$ contains a subsignature $g = [a, 0, \ldots, 0, b]$ of odd arity $n = 2^k + 1 \geq 1$, where $ab \neq 0$.

If $k = 0$, then $n = 1$ and $g = [a, b]$. But since $f$ does contain a 0 entry, $f$ has arity greater than 1. Then we can shift from the positions of $g$ either to the right or to the left and find the 0 entry in $f$ and obtain a binary subsignature $h$ of the form $[c, d, 0]$ or $[0, c, d]$, where $cd \neq 0$. Then Pl-Holant($h | \overline{\mathcal{Q}}$) is #P-hard by Theorem 8.3, so Pl-Holant($\{f\} \cup \overline{\mathcal{Q}}$) is also #P-hard.

Otherwise $k \geq 1$, so $n \geq 3$. If $a^4 \neq b^4$, then Pl-Holant($\{g\} \cup \overline{\mathcal{Q}}$) is #P-hard by Lemma 8.16, so Pl-Holant($\{f\} \cup \overline{\mathcal{Q}}$) is also #P-hard.

Otherwise, $a^4 = b^4$, so $g \in \mathcal{A}$. If $f = g$, then we are done, so assume that $f \neq g$, which implies that there is another entry just before $a$ or just after $b$. If this entry is nonzero, then $f$ has a subsignature $h$ of the form $[c, a, 0]$ or $[0, b, d]$, where $abcd \neq 0$. Then Pl-Holant($h | \overline{\mathcal{Q}}$) is #P-hard by Theorem 8.3, so Pl-Holant($\{f\} \cup \overline{\mathcal{Q}}$) is also #P-hard.

Otherwise, this entry is 0 and $f$ has a subsignature $h$ of the form

$$[0, a, 0, \ldots, 0, b] \quad \text{or} \quad [a, 0, \ldots, 0, b, 0]$$

of arity at least 4. If the arity of $h$ is even, then after some number of self-loops, we have a signature $h'$ of the form $[0, a, 0, 0, b]$ or $[a, 0, 0, b, 0]$ of arity exactly 4. Then Pl-Holant($h'$) is #P-hard by Corollary 7.25 since $ab \neq 0$, so Pl-Holant($\{f\} \cup \overline{\mathcal{Q}}$) is also #P-hard.

Otherwise, the arity of $h$ is odd. After some number of self-loops, we have a signature $h'$ of the form $[0, a, 0, 0, b]$ or $[a, 0, 0, b, 0]$ of arity exactly 5. Then Pl-Holant($h'$) is #P-hard by Lemma 8.23 since $ab \neq 0$, so Pl-Holant($\{f\} \cup \overline{\mathcal{Q}}$) is also #P-hard.

3. Suppose $f$ contains no 0 entry. If $f$ has a binary subsignature $g$ such that Pl-Holant($g | \overline{\mathcal{Q}}$) is #P-hard by Theorem 8.3, then Pl-Holant($\{f\} \cup \overline{\mathcal{Q}}$) is also #P-hard.

Otherwise every binary subsignature $[a, b, c]$ of $f$ satisfies the conditions of some tractable case in Theorem 8.3. Since there are no 0 entries, there are three possible tractable cases for $[a, b, c]$:

(D) it is degenerate with condition $ac = b^2$ (case 1);
(A) it has the affine type $A$ with condition $ac = -b^2 \wedge a = -c$ (case 3); or
(P) it has a Hadamard-transformed product type $\widehat{P}$ with condition $a = c$ (case 4).

If every binary subsignature $[a, b, c]$ of $f$ satisfies $ac = b^2$, then $f$ is degenerate, a contradiction. If every binary subsignature $[a, b, c]$ of $f$ satisfies $ac = -b^2 \wedge a = -c$, then $f = [1, \pm 1, -1, \mp 1, 1, \pm 1, -1, \mp 1, \ldots] \in A$ (up to a nonzero scalar) and we are done. If every binary subsignature $[a, b, c]$ of $f$ satisfies $a = c$, then $f \in \widehat{P}$ and we are done.

Otherwise, no single one of the three tractable conditions (D), (A) or (P) is satisfied by all binary subsignatures of $f$. Hence $f$ has arity $n \geq 3$. Let $h_i = [f_i, f_{i+1}, f_{i+2}]$ for all $0 \leq i \leq n - 2$ be binary subsignatures of $f$. Suppose there exists an $i$ such that $h_i$ satisfies the affine condition (A) (case 3). We claim that there must exist two successive subsignatures $h = [a, b, c]$ that is affine and $h' = [b, c, d]$ satisfying either the degenerate (D) or the product-type (P) condition, up to a transformation of $[\frac{1}{i} | \frac{1}{0}]$. This is because we can start from $h_i$ and search in both directions $h_{i-1}$ and $h_{i+1}$ until we found such $h$ and $h'$. It is always successful because not all $h_j$ satisfies the affine condition (A). Let $g = [a, b, c, d]$ be the ternary subsignature of $f$. Then for either case (D) or (P) of $h'$, we have $g = [1, \varepsilon, -1, \varepsilon]$ after normalization, where $\varepsilon^2 = 1$. Connecting two copies of $[0, 1]$ to $g$ gives $[-1, \varepsilon]$. Connecting this back to $g$ gives $g' = [0, -2\varepsilon, 2]$. Then Pl-Holant($g' | \widehat{EQ}$) is #P-hard by Theorem 8.3, so Pl-Holant($\{f\} \cup \widehat{EQ}$) is also #P-hard.

Now we suppose no $h_i$ satisfies the affine condition (A). Then every $h_i$ satisfies either (D) or (P), but no single condition (D) or (P) is satisfied by all. Therefore some $h_i$ must not satisfy (D), which must satisfy (P), and some other $h_j$ must satisfy (D), otherwise all would satisfy (P). Start from $h_i$ we can shift either to the right or to the left to find the first $h_j$ satisfying (D). Up to a transformation $[\frac{1}{i} | \frac{1}{0}]$, we may assume $j > i$. Then $h_{j-1}$ does not satisfy (D), therefore $h_{j-1}$ must satisfy (P). Thus there exists a ternary subsignature $g = [a, b, c, d]$ of $f$ such that $h = [a, b, c]$ satisfies (P) but not (D), and $h' = [b, c, d]$ satisfies (D). Then $g = [1, b, 1, b^{-1}]$ after normalization, where $b^2 \neq 1$. Then Pl-Holant($g | \widehat{EQ}$) is #P-hard by Lemma 8.30, so Pl-Holant($\{f\} \cup \widehat{EQ}$) is also #P-hard.

Now we are ready to prove our main dichotomy theorem.

**Theorem 8.33.** Let $F$ be any set of symmetric, complex-valued signatures in Boolean variables. Then Pl-Holant($F \cup \widehat{EQ}$) is #P-hard unless $F \subseteq A$, $F \subseteq \widehat{P}$, or $F \subseteq M$, in which case the problem is computable in polynomial time.

**Proof.** The tractability is given in Theorem 8.4. When $F$ is not a subset of $A$, $\widehat{P}$, or $M$, we prove that Pl-Holant($F \cup \widehat{EQ}$) is #P-hard, which is sufficient because of pinning (Theorem 8.31).
For any degenerate signature $f \in \mathcal{F}$, we connect some number of $[1,0]$ or $[0,1]$ to $f$ to get its corresponding unary signature. We replace $f$ by this unary signature, which does not change the complexity. Thus, assume that the only degenerate signatures in $\mathcal{F}$ are unary signatures. Recall that all unary signatures belong to $\mathcal{P}$. 

If $\mathcal{F} \not\subseteq \mathcal{A} \cup \mathcal{P} \cup \mathcal{M}$, then there exists some non-degenerate signature $f \in \mathcal{F} - (\mathcal{A} \cup \mathcal{P} \cup \mathcal{M})$ of arity at least 2. Then the problem is $\#P$-hard by Theorem 8.32. Otherwise, $\mathcal{F} \subseteq \mathcal{A} \cup \mathcal{P} \cup \mathcal{M}$ and we are done by Theorem 8.22.

Since $\text{Pl-}\#\text{CSP}(\mathcal{F}) \equiv_{\tau} \text{Pl-Holant}(\hat{\mathcal{F}} \cup \hat{\mathcal{Q}})$, where $\hat{\mathcal{F}} = H_2 \mathcal{F}$, and $\hat{\mathcal{Q}} = H_2 \mathcal{Q}$, we can translate Theorem 8.33 to obtain the main theorem of this Chapter, Theorem 8.34, the Pl-\#CSP dichotomy. Recall that $H_2 \mathcal{A} = \mathcal{A}$, $H_2 \mathcal{P} = \mathcal{P}$ and $H_2 \mathcal{M} = \mathcal{M}$. Theorem 8.34 is Theorem 8.33 stated in the standard basis.

**Theorem 8.34.** Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then $\text{Pl-}\#\text{CSP}(\mathcal{F})$ is $\#P$-hard unless $\mathcal{F} \subseteq \mathcal{A}$, $\mathcal{F} \subseteq \mathcal{P}$, or $\mathcal{F} \subseteq \mathcal{M}$, in which case the problem is computable in polynomial time.
Bibliography