Valiant's Holant Theorem and Matchgate Tensors

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

Abstract

We propose *matchgate tensors* as a natural and proper language to develop Valiant's new theory of Holographic Algorithms. We give a treatment of the central theorem in this theory—the Holant Theorem—in terms of matchgate tensors. Some generalizations are presented.

1 Background

In a remarkable paper, Valiant [9] in 2004 has proposed a completely new theory of Holographic Algorithms or Holographic Reductions. In this framework, Valiant has developed a most novel methodology of designing polynomial time (indeed NC^2) algorithms, a methodology by which one can design a custom made process capable of carrying out a seemingly exponential computation with exponentially many cancellations so that the computation can actually be done in polynomial time.

The simplest analogy is perhaps with Strassen's matrix multiplication algorithm [5]. Here the algorithm computes some extraneous quantities in terms of the submatrices, which do not directly appear in the answer yet only to be canceled later, but the purpose of which is to speedup computation by introducing cancelations. In the several cases such clever algorithms had been found, they tend to work in a linear algebraic setting, in particular the computation of the determinant figures prominently [8, 2, 6]. Valiant's new theory manages to create a process of custom made cancelation which gives polynomial time algorithms for combinatorial problems which do not appear to be linear algebraic.

In terms of its broader impact in complexity theory, one can view Valiant's new theory as another algorithmic design paradigm which pushes back the frontier of what is solvable by polynomial time. Admittedly, at this early stage, it is still premature to say what drastic consequence it might have on the landscape of the big questions of complexity theory, such as P vs. NP. But the new theory has already been used by Valiant to devise polynomial time algorithms for a number of problems for which no polynomial time algorithms were known before.

Unless and until a proof of P $\neq$ NP is found, one should regard this as an open problem. We can ask ourselves on what basis we derive confidence on the truth of this conjecture. In our view this confidence is not based on any partial lower bound which are either for very restricted models of computation or are still very weak. Fundamentally this source of confidence in P $\neq$ NP comes from the fact that all existing algorithmic approaches do not seem to tackle a myriad of NP-complete

*Supported by NSF CCR-0208013 and CCR-0511679.

†Supported by NSF CCR-0208013.

problems. Valiant's new theory of holographic algorithms challenges us to re-examine this belief critically.

The theory is quite unlike anything before, and it is a delicate theory that will be difficult to explain without all the definitions. The central theorem in this theory is the beautiful *Holant Theorem*, which is the linchpin that holds everything together and makes it all possible. But, at least to this author, the actual proof of the theorem in [9] was a little mysterious and somewhat difficult to understand. We believe the source of this difficulty lies in the way how one defines the main concepts of the theory.

The main purpose of this paper is to give a development of the theory based on the concept of *tensors*. While *tensor product* as an *operation* was already used by Valiant in [9], here our viewpoint is different in that we start off with the concepts of *covariant* and *contravariant* tensors, and, as it is customary in modern geometry, we strive to give it a *coordinate free* framework. Then various transformations of these tensors follow from general principles in tensor space. We then give a tensor theoretic proof of Valiant's Holant Theorem. It is suggested that once we have properly defined all the concepts based on *covariant* and *contravariant* tensors, Valiant's beautiful Holant Theorem can be understood as a *natural* expression of tensors.

Given the conceptual clarity afforded by the tensor perspective, we can easily see some generalizations of the *Holant Theorem* which follow from this framework.

2 Valiant's definitions

In this section we give a brief account of the key definitions of Valiant's theory, starting with the matching problem. More details can be found in [9].

Given a graph G, a matching of G is a set of edges no two of which share a vertex. A perfect matching M is a matching such that every vertex of G is incident to one edge of M. The decision problem of whether there is a perfect matching in G is computable in P, one of the notable achievements in the study of Algorithms. However, it is known that counting the number of perfect matchings in G is #P-complete.

We assign to every edge $e = (i, j)$ a variable x_{ij}, where $i < j$. Then we define the following polynomial

$$\mathrm{PerfMatch}(G) = \sum_M \prod_{(i,j) \in M} x_{ij},$$

where the sum is over all perfect matchings M. $\mathrm{PerfMatch}(G)$ is a polynomial on $\binom{n}{2}$ many variables $x_{ij}, 1 \leq i < j \leq n$. If the graph is a weighted graph with weights w_{ij}, we can also evaluate $\mathrm{PerfMatch}(G)$ at $x_{ij} = w_{ij}$. Note that if all the weights are 1, then $\mathrm{PerfMatch}(G)$ just counts the number of perfect matchings in the graph.

A most remarkable result due to Fisher, Kasteleyn and Temperley (FKT), ([7], [3], and [4]) from statistical physics is that for planar graphs, this Perfect Matching polynomial $\mathrm{PerfMatch}(G)$ can be evaluated in polynomial time. In fact it can be evaluated as a Pfaffian of a skew-symmetric matrix which is constructible from a planar embedding of G in polynomial time.

In effect, Valiant's theory allows the expression of a desired computation as an exponential sum, called the *Holant*, and via the Holant Theorem, reduces to the problem of computing the number of perfect matchings on planar graphs. This is done via the evaluation of $\mathrm{PerfMatch}(G)$ by the FKT method, for a suitably constructed *Macthgrid*, composed of *matchgates*, which we proceed to define. These reductions are called holographic reductions, because they carry out exponentially many cancellations analogous to a pattern of interference in quantum computing.

Define a *planar matchgate* Γ as a triple (G, X, Y) where G is a planar embedding of a weighted planar graph (V, E, W), $X \subseteq V$ is a set of input nodes, $Y \subseteq V$ is a set of output nodes, and $X \cap Y = \emptyset$. Furthermore in the planar embedding of G, counter-clock wise one encounters vertices of X, labeled $1, \ldots, |X|$ and then vertices of Y, labeled $|Y|, \ldots, 1$.

Valiant defines the standard signature, $u(\Gamma)$, of Γ to be a $2^{|X|} \times 2^{|Y|}$ matrix whose entries are indexed by subsets $X' \subseteq X$ and $Y' \subseteq Y$, and the entry indexed by (X', Y') is $\mathrm{PerfMatch}(G - Z)$, where $Z = X' \cup Y'$. Here $G - Z$ denotes the subgraph of G obtained by removing the subset of nodes in Z (and all their incident edges). We will make one slight (harmless) change here. We take the transpose of this matrix to be the standard signature. This is to conform to (one standard) notation in view of later development in terms of covariant and contravariant tensors [1]. Thus the standard signature for us is a $2^{|Y|} \times 2^{|X|}$ matrix.

Matchgates with only output nodes are called *generators*. Matchgates with only input nodes are called *recognizers*. More generally, with both input and output nodes a matchgate is called a *transducer*. We note that the standard signature of a generator is a column vector and the standard signature of a recognizer is a row vector.

Let $\mathbf{b}$ denote the standard basis for two dimensional space, $\mathbf{b} = [e_0, e_1] = \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right]$. Consider another basis $\boldsymbol{\beta} = [n, p] = \left[\begin{pmatrix} n_0 \\ n_1 \end{pmatrix}, \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}\right]$.

Let Γ be a generator with m output nodes. Then by definition its standard signature $u(\Gamma)$ is a 2^m-vector. Valiant then defines the *signature of this generator with respect to the basis* $\boldsymbol{\beta}$ as the coefficients of $u(\Gamma)$ when expressed in the new basis $\boldsymbol{\beta}$. More precisely, for an m-tuple tensor product $x = x_1 \otimes x_2 \otimes \cdots \otimes x_m$, where $x_i = n$ or p, Valiant defines $\mathrm{valG}(\Gamma, x)$, "the signature element corresponding to x" [9], to be the coefficient of x when $u(\Gamma)$ is expressed as a sum over $\{n, p\} \otimes \{n, p\} \otimes \cdots \otimes \{n, p\}$. (Technically, Valiant's theory also allows a basis to be a set of dependent vectors; but in order that $u(\Gamma)$ be expressible in the new basis, it is implicitly required that the standard signature be in the linear span of the tensor products of the new basis. In this case, any such linear expression gives arise to a set of values $\mathrm{valG}(\Gamma, x)$. We will see that this slight complication can be easily handled (see the discussion at the end of the Section 3 and 4); but for simplicity of development, we will assume for now that the basis $\boldsymbol{\beta} = [n, p]$ consists of independent vectors as a basis ordinarily does.)

Turning to recognizers, let Γ' be a recognizer with m input nodes. Let $x = x_1 \otimes x_2 \otimes \cdots \otimes x_m$ range over 2^m possible values, where each $x_i = n$ or p. Now Valiant treats x as a 2^m-vector in the standard basis, and defines $\mathrm{valR}(\Gamma', x)$, "the recognizer matchgate Γ' 'evaluated at input' x" [9], to be the inner product of the standard signature $u(\Gamma)$ with x.

Next Valiant defines a *matchgrid* $\Omega = (A, B, C)$ to be a weighted planar graph consisting of a disjoint union of: a set of g generators $A = (A_1, \ldots, A_g)$, a set of r recognizers $B = (B_1, \ldots, B_r)$, and a set of f connecting edges $C = (C_1, \ldots, C_f)$, where each C_i edge has weight 1 and joins an output node of a generator with a input node of a recognizer, so that every input and output node in every constituent matchgate has exactly one such incident connecting edge.

Now we come to the central definition of Valiant's theory—the Holant.

$$\mathrm{Holant}(\Omega) = \sum_{x \in \boldsymbol{\beta}^{\otimes f}} \left\{ [\Pi_{1 \le i \le g} \mathrm{valG}(A_i, x|_{A_i})] \cdot [\Pi_{1 \le j \le r} \mathrm{valR}(B_j, x|_{B_j})] \right\}.$$

The following is the beautiful Holant Theorem

Theorem 2.1 (Valiant) *For any matchgrid Ω over any basis $\boldsymbol{\beta}$, let G be its underlying weighted graph, then*

$$\mathrm{Holant}(\Omega) = \mathrm{PerfMatch}(G).$$

3 A treatment in terms of vectors

In this section we rephrase Valiant's definitions in terms of vectors, this serves as a transition to the ultimate tensor framework.

Let Γ be a generator with m output nodes. We now consider the object called $\mathrm{valG}(\Gamma)$ as a (column) vector, whose entries are indexed by $x \in \boldsymbol{\beta}^{\otimes m} = \{n, p\}^{\otimes m}$. Let T be the transformation matrix from $\mathbf{b}$ to $\boldsymbol{\beta}$, namely

$$[n, p] = [e_0, e_1]T,$$

where

$$T = \begin{pmatrix} n_0 & p_0 \\ n_1 & p_1 \end{pmatrix}.$$

We form the tensor product matrix $T^{\otimes m}$ which transforms the basis $\mathbf{b}^{\otimes m}$ to $\boldsymbol{\beta}^{\otimes m}$. This follows because tensor product "distributes" over matrix product, from $\boldsymbol{\beta} = \mathbf{b}T$ we get,

$$\boldsymbol{\beta}^{\otimes m} = (\mathbf{b}T)^{\otimes m} = \mathbf{b}^{\otimes m} T^{\otimes m}.$$

Then we claim that the vector $\mathrm{valG}(\Gamma)$ is obtained from the standard signature $u(\Gamma)$ by multiplying the tensor product matrix $(T^{\otimes m})^{-1} = (T^{-1})^{\otimes m}$:

$$\mathrm{valG}(\Gamma) = (T^{-1})^{\otimes m} u(\Gamma),$$

where for a generator Γ, we recall that the standard signature $u(\Gamma)$ is a column vector of dimension 2^m. This agrees with Valiant's definition since

$$\mathbf{b}^{\otimes m} = \boldsymbol{\beta}^{\otimes m} (T^{\otimes m})^{-1} = \boldsymbol{\beta}^{\otimes m} (T^{-1})^{\otimes m},$$

and therefore

$$(\mathbf{b})^{\otimes m} u(\Gamma) = (\boldsymbol{\beta})^{\otimes m} (T^{-1})^{\otimes m} u(\Gamma)$$

is the expression of the standard signature expressed in the new basis $\boldsymbol{\beta}$, i.e., the entry of the vector $(T^{-1})^{\otimes m} u(\Gamma)$ indexed by $x \in \{n, p\}^{\otimes m}$ is what was called $\mathrm{valG}(\Gamma, x)$ in Section 2.

We next consider recognizers. Let Γ' be a recognizer with m input nodes. We will define $\mathrm{valR}(\Gamma')$ as a (row) vector. But more precisely we will consider $\mathrm{valR}(\Gamma')$ as a vector belonging to the dual space X^*, where X is the linear span of $\boldsymbol{\beta}^{\otimes m}$.

Let $\boldsymbol{\beta}^* = \begin{pmatrix} n^* \\ p^* \end{pmatrix}$ denote the dual basis to $\boldsymbol{\beta}$, namely n^*, p^* are linear functions on the linear space spanned by $\boldsymbol{\beta}$, such that $n^*(n) = 1, n^*(p) = 0, p^*(n) = 0, p^*(p) = 1$. Then the dual basis to $\boldsymbol{\beta}^{\otimes m}$ is simply $(\boldsymbol{\beta}^*)^{\otimes m}$.

When we have a basis transformation $\boldsymbol{\beta} = \mathbf{b}T$ from $\mathbf{b}$ to $\boldsymbol{\beta}$, the dual basis transforms as follows

$$\boldsymbol{\beta}^* = T^{-1} \mathbf{b}^*.$$

This follows from general principles. (See Section 4.)

Now we claim that what was defined by Valiant as $\mathrm{valR}(\Gamma', x)$, as x ranges over $\boldsymbol{\beta}^{\otimes m}$, amounts to a dual vector $\mathrm{valR}(\Gamma')$ in X^*, whose entries are indexed by $x^* \in (\boldsymbol{\beta}^*)^{\otimes m}$, i.e., we claim

$$\mathrm{valR}(\Gamma') = u(\Gamma') T^{\otimes m},$$

under the basis $(\boldsymbol{\beta}^*)^{\otimes m}$ in X^*.

The standard signature $u(\Gamma')$ is really a dual vector in X^*,

$$u(\Gamma')(\mathbf{b}^*)^{\otimes m}.$$

Since the dual basis transforms as

$$\mathbf{b}^* = T\boldsymbol{\beta}^*,$$

we get

$$(\mathbf{b}^*)^{\otimes m} = T^{\otimes m}(\boldsymbol{\beta}^*)^{\otimes m},$$

and therefore $u(\Gamma')(\mathbf{b}^*)^{\otimes m}$ takes the form

$$u(\Gamma')(\mathbf{b}^*)^{\otimes m} = u(\Gamma')T^{\otimes m}(\boldsymbol{\beta}^*)^{\otimes m},$$

in the new basis. Notice that the entry of this vector indexed by $x^* \in (\boldsymbol{\beta}^*)^{\otimes m}$ is precisely the inner product of $u(\Gamma')$ with the column of $T^{\otimes m}$ indexed by x^*, and the latter is nothing but the vector of coefficients when $x \in \boldsymbol{\beta}^{\otimes m}$ is expressed in terms of the standard basis $\mathbf{b}^{\otimes m}$. Thus we have reconciled this formulation with Valiant's definition.

Now consider the definition of the Holant. We assume $\Omega = (A, B, C)$ is a matchgrid where each generator A_i has m_i output nodes and each recognizer B_j has ℓ_j input nodes.

The definition of valG and valR in our linear algebra formulation makes the following observation transparent. The Holant in fact is an evaluation of an inner product of two vectors, one of which is the tensor product of all the $\mathrm{valG}(A_i)$ over the generators, and the other is the tensor product of all the $\mathrm{valR}(B_j)$ over the recognizers. More precisely (and it gives the same numerical result) this quantity $\mathrm{Holant}(\Omega)$ is the result of applying a dual vector in X^*, which is the tensor product $\bigotimes_j[\mathrm{valR}(B_j)(\boldsymbol{\beta}^*)^{\otimes \ell_j}]$, on a primal vector in X, which is also a tensor product $\bigotimes_i[(\boldsymbol{\beta})^{\otimes m_i}\mathrm{valG}(A_i)]$, where the f copies of the basis vectors in $\boldsymbol{\beta}$ are in 1-1 correspondence as given by the f connecting edges in C.

Thus

$$\mathrm{Holant}(\Omega) = \sum_{x \in \boldsymbol{\beta}^{\otimes f}} \left[\bigotimes_i \mathrm{valG}(A_i)\right]_x \cdot \left[\bigotimes_j \mathrm{valR}(B_j)\right]_{x^*} = \langle \bigotimes_j \mathrm{valR}(B_j), \bigotimes_i \mathrm{valG}(A_i)\rangle.$$

Note that the sum $\sum_{x \in \boldsymbol{\beta}^{\otimes f}}$ is precisely over all the entries in the two tensor product vectors which are indexed by $x \in \boldsymbol{\beta}^{\otimes f}$, and by the corresponding $x^* \in (\boldsymbol{\beta}^*)^{\otimes f}$, respectively. Here we have adopted the conventional notation $\langle \cdot, \cdot \rangle$ for the inner product. For a row vector Y and a column vector Z of the same dimension, the inner product $\langle Y, Z\rangle$ is just $Y \cdot Z = \sum_i Y_i Z_i$.

The total number of output nodes of all A_i is the same as the total number of input nodes of all B_j, i.e., $\sum_i m_i = \sum_j \ell_j = f$, the total number of interconnecting wires between the generators and the recognizers. Note that, according to an appropriate ordering of the indices, $\bigotimes_i \mathrm{valG}(A_i)$ can be expressed by the matrix-vector product form

$$[\otimes_i (T^{\otimes m_i})^{-1}][\otimes_i u(A_i)],$$

which is just $(T^{\otimes f})^{-1}[\otimes_i u(A_i)]$.

Similarly the tensor product $\bigotimes_j \mathrm{valR}(B_j, \cdot)$ can be expressed by

$$[\otimes_j u(B_j)]T^{\otimes f}.$$

Now the beautiful thing is that the adjacent $T^{\otimes f}$ and $(T^{\otimes f})^{-1}$ cancel in the inner product, and finally we get

$$\text{Holant}(\Omega) = \langle \otimes_j u(B_j), \otimes_i u(A_i) \rangle.$$

What we have now is the definition of the Holant under the standard basis $\mathbf{b}$.

Stripped away of all its linear algebraic layers, we have come now to the combinatorial reason why the Holant Theorem holds: The set of all perfect matchings on G can be partitioned according to exactly the subset of edges S among the f connecting edges $C_1, C_2, \ldots, C_f$ that is part of the matching. And summed over this partition is precisely what the Holant Theorem states in the standard basis $\mathbf{b}$:

$$\text{Holant}(\Omega) = \text{PerfMatch}(G).$$

As a postcript, we note that the transformation matrix T need not be invertible or even square, as long as the standard signature of the generators can be expressed in the linear span of $\boldsymbol{\beta}^{\otimes}$.

Assume the standard signature $\mathbf{b}^{\otimes m} u(\Gamma)$ is in the linear span of $\boldsymbol{\beta}^{\otimes m}$, where Γ has m output nodes. Then there exists a (column) vector v such that

$$\mathbf{b}^{\otimes m} u(\Gamma) = \boldsymbol{\beta}^{\otimes m} v.$$

We can then define $\text{valG}(\Gamma)$ to be this v,

$$\text{valG}(\Gamma) = v.$$

(This v may not be unique, but any such v will do.) It follows that

$$\mathbf{b}^{\otimes m} u(\Gamma) = \boldsymbol{\beta}^{\otimes m} v = \mathbf{b}^{\otimes m} T^{\otimes m} \text{valG}(\Gamma).$$

So

$$u(\Gamma) = T^{\otimes m} \text{valG}(\Gamma).$$

Then the proof of the Holant Theorem above still holds, as

$$\begin{aligned}
\text{Holant}(\Omega) &= \langle \bigotimes_j \text{valR}(B_j), \bigotimes_i \text{valG}(A_i) \rangle \\
&= [\otimes_j u(B_j) T^{\otimes f}] \cdot [\otimes_i \text{valG}(A_i)] \\
&= [\otimes_j u(B_j)] \cdot [\otimes_i u(A_i)] \\
&= \langle \otimes_j u(B_j), \otimes_i u(A_i) \rangle.
\end{aligned}$$

4 Valiant's theory based on tensors

In this section we will give a tensor theoretic treatment of Valiant's Holant Theorem.

4.1 Covariant and contravariant tensors

First we briefly recall some notations regarding *covariant* and *contravariant* tensors. We will avoid any overly abstract framework of these concepts, but will appeal to the notion of a coordinate-free definition of a tensor, which exists in a certain tensor space *a priori*. Such a tensor has various expressions according to the basis of the tensor space chosen, and these expressions transform according to simple transformation rules when one changes from one basis to another.

At an operational level, one can think of it as follows: Let V be a vector space of dimension d over some field $\mathbf{F}$. Let $\mathbf{b} = \{\mathbf{b}_1, \ldots, \mathbf{b}_d\}$ be a basis. Now with respect to this basis $\mathbf{b}$, every vector of V has a unique expression as $\sum_{i=1}^{d} x^i \mathbf{b}_i$, (which is usually abbreviated in this area of mathematics as just $x^i \mathbf{b}_i$, with a matching upper and lower index i automatically being summed.) One has a dual space V^* and a dual basis to $\mathbf{b}$, denoted as $\mathbf{b}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^d\}$, where $\mathbf{b}^i(\mathbf{b}_j) = \delta^i_j$.

From V and V^* one can form tensor product space of any arity. So, e.g., the space V_2^3 of type $\binom{3}{2}$ is a tensor product space of dimension d^5, and has a basis $\{\mathbf{b}_i \otimes \mathbf{b}_j \otimes \mathbf{b}_k \otimes \mathbf{b}^\ell \otimes \mathbf{b}^m\}$, where all indices run from 1 to d. Any element $\mathbf{x} \in V_2^3$ is called a tensor, and has the expression $\sum_{ijk\ell m} x^{ijk}_{\ell m} \mathbf{b}_i \otimes \mathbf{b}_j \otimes \mathbf{b}_k \otimes \mathbf{b}^\ell \otimes \mathbf{b}^m$, or simply $(x^{ijk}_{\ell m})$, and is called *covariant* on ℓ, m and *contravariant* on i, j, k. In particular vectors in V are contravariant and dual vector in V^* are covariant. They are called as such because the way they transform under a basis transformation.

Let $\boldsymbol{\beta} = \mathbf{b}T$ be a new basis. In coordinates,

$$\boldsymbol{\beta}_j = \sum_i \mathbf{b}_i t^i_j,$$

where the (i, j) entry of T is t^i_j. (Upper index is for row, lower index is for column.) Then it can be easily verified for the dual basis that

$$\boldsymbol{\beta}^* = T^{-1} \mathbf{b}^*,$$

where $\boldsymbol{\beta}^* = \{\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^d\}$ is dual to $\boldsymbol{\beta}$. Indeed, by denoting $T^{-1} = (\tilde{t}^i_j)$, we have $\boldsymbol{\beta}^i(\boldsymbol{\beta}_j) = \sum_k \tilde{t}^i_k \mathbf{b}^k (\sum_l \mathbf{b}_l t^l_j) = \sum_{k,l} \tilde{t}^i_k t^l_j \delta^k_l = \delta^i_j$.

In coordinates, if $\mathbf{x} = \sum x^i \mathbf{b}_i \in V$, then under a basis transformation, $\mathbf{x} = \sum (x')^{i'} \boldsymbol{\beta}_{i'}$ where

$$(x')^{i'} = \sum_i \tilde{t}^{i'}_i x^i.$$

If $\mathbf{x}^* = \sum x_i \mathbf{b}^i \in V^*$ in the dual space, then under the same basis transformation, $\mathbf{x}^* = \sum (x')_{i'} \boldsymbol{\beta}^{i'}$ where

$$(x')_{i'} = \sum_i t^i_{i'} x_i.$$

Thus, vectors in V are contravariant and vectors in V^* are covariant.

This extends to tensors of any "type". E.g., a tensor in V_2^3, $\mathbf{x} = (x^{ijk}_{\ell m})$ is contravariant on the three upper indices and covariant on the two lower indices. And it transforms as

$$(x')^{i'j'k'}_{\ell' m'} = \sum_{i,j,k,\ell,m} \tilde{t}^{i'}_i \tilde{t}^{j'}_j \tilde{t}^{k'}_k t^\ell_{\ell'} t^m_{m'} x^{ijk}_{\ell m}.$$

Finally a *contraction* on an index i for a pair of tensors $(x^{i\cdots}_{j\ldots})$ and $(y^{k\cdots}_{i\ldots})$ is simply an application of the dual on the primal; in terms of coordinates

$$\sum_i x^{i\ldots}_{j\ldots} y^{k\ldots}_{i\ldots}.$$

The reader is referred to [1] for more details.

4.2 Holant Theorem based on tensors

In this section we will give a tensor theoretic treatment of Valiant's Holant Theorem.

In Section 2, we defined the objects valG and valR as vectors. However, an even more appropriate home for these objects are in tensor spaces of type $\binom{m}{0}$ for the generators and $\binom{0}{m}$ for the recognizers.

Thus, consider a generator matchgate Γ whose underlying weighted graph G has m output nodes. We consider a vector space V of dimension 2 over some field $\mathbf{F}$ has already been fixed. We may choose some basis $\mathbf{b}$ of V and consider it the standard basis. We assign to this matchgate a tensor $\mathbf{G} \in V_0^m$ of type $\binom{m}{0}$. This tensor under the standard basis has the form

$$\sum G^{i_1 i_2 \ldots i_m} \mathbf{b}_{i_1} \otimes \mathbf{b}_{i_2} \otimes \cdots \otimes \mathbf{b}_{i_m},$$

where

$$G^{i_1 i_2 \ldots i_m} = \mathrm{PerfMatch}(G - Z),$$

where Z is the subset of the output nodes having the characteristic sequence $\chi_Z = i_1 i_2 \ldots i_m$. Note that we are putting the matchgate tensor $\mathbf{G}$ in a tensor space *a priori*, and the expression of $\mathbf{G}$ under a particular basis is subordinate to that. In particular, $\mathbf{G}$ transforms as a contravariant tensor under a basis transformation $\boldsymbol{\beta} = \mathbf{b}T$, as

$$(G')^{i'_1 i'_2 \ldots i'_m} = \sum G^{i_1 i_2 \ldots i_m} \tilde{t}^{i'_1}_{i_1} \tilde{t}^{i'_2}_{i_2} \ldots \tilde{t}^{i'_m}_{i_m}.$$

This tensor is what we have been calling $\mathrm{valG}(\Gamma)$. As a tensor of type $\binom{m}{0}$, it is usually abbreviated as simply $G^{i_1 i_2 \ldots i_m}$.

Now consider a recognizer Γ^* whose underlying weighted graph G^* has m input nodes. To Γ^* we will assign a tensor $\mathbf{R} \in V_m^0$ of type $\binom{0}{m}$. This tensor under the standard (dual) basis has the form

$$\sum R_{i_1 i_2 \ldots i_m} \mathbf{b}^{i_1} \otimes \mathbf{b}^{i_2} \otimes \cdots \otimes \mathbf{b}^{i_m},$$

where

$$R_{i_1 i_2 \ldots i_m} = \mathrm{PerfMatch}(G^* - Z),$$

where Z is the subset of the input nodes having $\chi_Z = i_1 i_2 \ldots i_m$. Again we put the matchgate tensor $\mathbf{R}$ in a tensor space *a priori*. In particular, when changing a basis $\boldsymbol{\beta}_j = \sum_i \mathbf{b}_i t^i_j$, $\mathbf{R}$ transforms as a covariant tensor should, namely

$$(R')_{i'_1 i'_2 \ldots i'_m} = \sum R_{i_1 i_2 \ldots i_m} t^{i_1}_{i'_1} t^{i_2}_{i'_2} \ldots t^{i_m}_{i'_m}.$$

This tensor is what we have been calling $\mathrm{valR}(\Gamma^*)$.

In a *matchgrid* $\Omega = (A, B, C)$ the indices of various generators and recognizers are matched up in a 1-1 correspondence by the f connecting edges. Then, in the language of tensors, the definition of the Holant is just a *contraction* on all pairs of corresponding indices.

We denote by $\mathbf{G}$ the tensor product of all the generator tensors over A_i and $\mathbf{R}$ the tensor product of all the recogizer tensors over B_j. Then $\mathbf{G} \in V_0^f$ and $\mathbf{R} \in V_f^0$, and the Holant is the contraction of $\mathbf{R}$ with $\mathbf{G}$ by contracting on all the corresponding indices, which we can denote simply as

$$\langle \mathbf{R}, \mathbf{G} \rangle.$$

Note that the coordinate-free definition of valG and valR as tensors immediately implies that the Holant is independent of any basis. In terms of coordinates we can verify that pairwise $\sum_{i'} t^i_{i'} \tilde{t}^{i'}_j = \delta^i_j$. One can say that the corresponding pair of (t^i_k) and $(\tilde{t}^k_j)$ cancels out. Thus we can use the standard basis with $\mathrm{PerfMatch}(A_i - Z)$ and $\mathrm{PerfMatch}(B_j - Z)$, where A_i and B_j are the constituent generators and recognizers respectively. Then combinatorially we see clearly,

$$\mathrm{Holant}(\Omega) = \mathrm{PerfMatch}(G),$$

as all perfect matchings M of G are partitioned according to the subset $M \cap C$.

We can now consider a generalization of the Holant Theorem. We will consider a more general matchgrid having transducers, in addition to generators and recognizers.

Let Γ be a transducer matchgate with ℓ input nodes and k output nodes. We attach to Γ a tensor $\mathbf{T}$ in V_ℓ^k, contravariant on k upper indices and covariant on ℓ lower indices. Under basis $\mathbf{b}$ it has the expression

$$\sum T_{i_1 i_2 \ldots i_\ell}^{j_1 j_2 \ldots j_k} \mathbf{b}^{i_1} \otimes \mathbf{b}^{i_2} \otimes \cdots \otimes \mathbf{b}^{i_\ell} \otimes \mathbf{b}_{j_1} \otimes \mathbf{b}_{j_2} \otimes \cdots \otimes \mathbf{b}_{j_k},$$

where

$$T_{i_1 i_2 \ldots i_\ell}^{j_1 j_2 \ldots j_k} = \text{PerfMatch}(G - Z),$$

and $G - Z$ is the graph of Γ obtained by removing the subset of the input/output vertices with $\chi_Z = i_1 i_2 \ldots i_\ell j_1 j_2 \ldots j_k$. This agrees with the definition of the standard signature $u(\Gamma)$, except now we have a tensor in V_ℓ^k. In short $\mathbf{T} = (T_{i_1 i_2 \ldots i_\ell}^{j_1 j_2 \ldots j_k})$.

Then it follows from general principles that under a basis transformation $\boldsymbol{\beta}_j = \sum_i \mathbf{b}_i t_j^i$, $\mathbf{T}$ transforms as

$$(T')_{a_1 a_2 \ldots a_\ell}^{b_1 b_2 \ldots b_k} = \sum T_{i_1 i_2 \ldots i_\ell}^{j_1 j_2 \ldots j_k} t_{a_1}^{i_1} t_{a_2}^{i_2} \ldots t_{a_\ell}^{i_\ell} \tilde{t}_{j_1}^{b_1} \tilde{t}_{j_2}^{b_2} \ldots \tilde{t}_{j_k}^{b_k}.$$

In Valiant's notation [9], under a basis $\boldsymbol{\beta}$, this could have been denoted as $\text{valT}(\Gamma, \cdot)$.

We define a generalized matchgrid $\Omega = (A, B, C, D)$ to be a weighted planar graph G which consists of a disjoint set of g generators $A_1, \ldots, A_g$, r recognizers $B_1, \ldots, B_r$, t transducers $C_1, \ldots, C_t$, and f connecting edges $D_1, \ldots, D_f$, where each D_i has weight 1 and they connect output nodes of some A_α or C_γ to input nodes of some B_β or $C_{\gamma'}$ in a 1-1 fashion.

Then we can define the extended Holant in the notation in [9]:

$$\text{Holant}(\Omega) = \sum_{x \in \boldsymbol{\beta}^{\otimes f}} \left\{ [\Pi_{1 \le \alpha \le g} \text{valG}(A_\alpha, x|_{A_\alpha})] \cdot [\Pi_{1 \le \beta \le r} \text{valR}(B_\beta, x|_{B_\beta})] \cdot [\Pi_{1 \le \gamma \le t} \text{valT}(C_\gamma, x|_{C_\gamma})] \right\}.$$

In terms of the tensors, we simply compute a contraction on all the matching pairs of upper and lower indices, indicated by the f connecting edges.

Since all the corresponding pairs of (t_k^i) and $(\tilde{t}_j^k)$ cancel out, the extended Holant also reduces to the expression under the standard basis. Then it follows from the same combinatorial reason that

Theorem 4.1 *For matchgrid $\Omega = (A, B, C, D)$,*

$$\text{Holant}(\Omega) = \text{PerfMatch}(G).$$

Finally, we briefly discuss what happens when the new "basis" $\boldsymbol{\beta}$ is only a set of vectors (and not necessarily a basis in the linear algebra sense.) This allows for the possibility that the transformation matrix $T = (t_j^i)$ is not a square matrix. This flexibility was shown to be useful for one problem solved by Valiant in [9].

Consider a generator Γ with m output nodes, and its tensor $\mathbf{G} \in V_0^m$. Even though $\boldsymbol{\beta}$ may be linearly dependent, we *will* assume that $\mathbf{G} = \sum G^{i_1 i_2 \ldots i_m} \mathbf{b}_{i_1} \otimes \mathbf{b}_{i_2} \otimes \cdots \otimes \mathbf{b}_{i_m}$ belong to the linear span of $\{\boldsymbol{\beta}_{j_1} \otimes \boldsymbol{\beta}_{j_2} \otimes \cdots \otimes \boldsymbol{\beta}_{j_m}\}$,

$$\sum_i G^{i_1 i_2 \ldots i_m} \mathbf{b}_{i_1} \otimes \mathbf{b}_{i_2} \otimes \cdots \otimes \mathbf{b}_{i_m} = \sum_j G'^{j_1 j_2 \ldots j_m} \boldsymbol{\beta}_{j_1} \otimes \boldsymbol{\beta}_{j_2} \otimes \cdots \otimes \boldsymbol{\beta}_{j_m},$$

for some numbers $G'^{j_1 j_2 \ldots j_m}$. By a slight abuse of notation, we also say the tensor $\mathbf{G}$ takes the form $(G'^{j_1 j_2 \ldots j_m})$ in the new basis. These numbers are not unique, when $\boldsymbol{\beta}$ is not linearly independent, i.e., the columns of T are not linearly independent. But any such set of numbers will do. This will be called valG(Γ).

Now consider a recognizer Γ' with m input nodes, to which we have already assigned a covariant tensor $\mathbf{R} \in V_m^0$. When T is not invertible, there will not be a set of corresponding dual basis $\{\boldsymbol{\beta}_j\}$ as before. However, the covariant tensor $\mathbf{R} = \sum R_{i_1 i_2 \ldots i_m} \mathbf{b}^{i_1} \otimes \mathbf{b}^{i_2} \otimes \cdots \otimes \mathbf{b}^{i_m}$ has the following evaluations: It sends $\boldsymbol{\beta}_{j_1} \otimes \boldsymbol{\beta}_{j_2} \otimes \cdots \otimes \boldsymbol{\beta}_{j_m} \mapsto \sum_i R_{i_1 i_2 \ldots i_m} t_{j_1}^{i_1} t_{j_2}^{i_2} \ldots t_{j_m}^{i_m}$. Thus, when we consider the transformation $\boldsymbol{\beta}_j = \sum_i \mathbf{b}_i t_j^i$, we will denote this tensor as $({R'}_{j_1 j_2 \ldots j_m})$, simply as a notation, where the values ${R'}_{j_1 j_2 \ldots j_m} = \sum_i R_{i_1 i_2 \ldots i_m} t_{j_1}^{i_1} t_{j_2}^{i_2} \ldots t_{j_m}^{i_m}$.

The simple proof above for the Holant Theorem is still valid. Note that in the tensor framework, we did not change the intrinsic meanings of $\mathbf{G}$, $\mathbf{R}$ and the Holant as a contraction $\langle \mathbf{R}, \mathbf{G} \rangle$. Under a change of vectors, from $\mathbf{b}$ to $\boldsymbol{\beta}$, we merely changed the expression of the tensors. This change of expression is only useful in expressing a desired computation by the matchgrid. It has no effect on the validity and proof of the Holant Theorem.

5 Performance and Defect Problem

In [9] Valiant showed how to solve several combinatorial problems in polynomial time using holographic algorithms.

To his list of problems, we add the following problem.

A Boolean formula F consists of a set of clauses $\{C_j\}$, each of which is a set of literals x_i or $\overline{x_i}$. F is called a planar formula if it can be drawn as a planar graph where vertices correspond to variables x_i and clauses C_j, and an edge exists between x_i and C_j iff x_i or $\overline{x_i}$ appear in C_j.

We will consider a planar formula F where each clause has three literals. Each clause is labeled as either compulsory or non-compulsory. For a clause C and any assignment σ, let $w(\sigma\mid_C) = \#$ of 1's that σ assigns to the literals in C. σ is called *exacting* on C if $w(\sigma\mid_C) = 0$ or 3. σ is k-*exacting* on F if σ is exacting on all the compulsory clauses and precisely k non-compulsory clauses. Let

$$\mathrm{perf}(\sigma) = (-1)^{|\{C \mid w(\sigma|_C) \geq 2\}|},$$

and

$$\mathrm{defect}(\sigma) = (-1)^{|\{C \mid w(\sigma|_C) \leq 1\}|}.$$

PERFORMANCE

Input A planar formula F where each clause has three literals, and is labeled as either compulsory or non-compulsory; integer k.
Output $\sum_{\sigma:k\text{-exacting}} \mathrm{perf}(\sigma)$.

DEFECT

Input A planar formula F as above.
Output $\sum_{\sigma:k\text{-exacting}} \mathrm{defect}(\sigma)$.

Comment: The two problems are $\#P$-hard if the -1 is replaced by 1. For the PERFORMANCE problem if we call an assignment σ *Even* if the number of clauses C for which $w(\sigma\mid_C) \geq 2$ is even, and *Odd* otherwise, then $\sum_{\sigma:k\text{-exacting}} \mathrm{perf}(\sigma)$ is clearly the number of Even k-exacting assignments minus the number of Odd ones. Similarly for the DEFECT problem. Viewed in this way, one can easily see that the two problems are essentially the same problem.

To describe the holographic polynomial time solution to the PERFORMANCE problem, we use the basis $\mathbf{b2} = [n, p]$, where $n = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $p = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. It can be shown that the following *symmetric* signature $[x, y, -y, -x]$ is achievable by a matchgate under basis $\mathbf{b2}$, for any real values x and y. Here the notation $[x, y, -y, -x]$ is a short hand for the 8-dimensional tensor, with coefficients x on $n \otimes n \otimes n$ (for the bit pattern 000), $-x$ on $p \otimes p \otimes p$ (for the bit pattern 111), y on $n \otimes n \otimes p$, $n \otimes p \otimes n$ and $p \otimes n \otimes n$ (for bit patterns of Hamming weight 1), and $-y$ on bit patterns of Hamming weight 2. In particular we have $[1, 0, 0, -1]$ and $[1, y, -y, -1]$.

We also note that the matrix $T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ has inverse $T^{-1} = \frac{1}{2}T$, so that for this basis $\mathbf{b2}$ what is achievable as a generator tensor $\mathbf{G}$ is also achievable as a recognizer tensor $\mathbf{T}$.

One can also realize the symmetric signature $[1, 0, 0, 1]$, $[1, 0, 1]$, and $[0, 1, 0]$. $[1, 0, 0, 1]$ has the effect of 3 equal bits. $[1, 0, 1]$ has the effect of 2 equal bits. $[0, 1, 0]$ has the effect of 2 unequal bits. Using a planar generator matchgate with its matchgate tensor corresponding to $[1, 0, 0, 1]$ for a variable x has the effect of setting the possible truth assignments of n (for 0) or p (for 1) with 3 output nodes. If a variable appears more than 3 times as a literal in clauses, then we can "chain" together two such generator matchgates above, with a recognizer having the symmetric signature $[1, 0, 1]$. This effectively produces a generator matchgate with 4 output nodes, which sets truth assignments to x. "Chaining" k such generators together gives a "truth-setting" matchgate with $k + 2$ output nodes.

For each clause C, if it is compulsory, we use a clause matchgate with symmetric signature $[1, 0, 0, -1]$. If it is non-compulsory we use $[1, y, -y, -1]$. If a variable appears positively in a clause we can use the "equal" matchgate with the symmetric signature $[1, 0, 1]$ to connect to this clause matchgate. If x appears negatively in a clause we can use the "unequal" matchgate with the symmetric signature $[0, 1, 0]$.

Then, in the Holant evaluation, for each assignment σ, for every exacting clause (either compulsory or non-compulsory) we get a value 1 for $w(\sigma \mid_C) = 0$ and a value -1 for $w(\sigma \mid_C) = 3$. For a non-exacting clause (which must be non-compulsory) we get a value y for $w(\sigma \mid_C) = 1$ and a value $-y$ for $w(\sigma \mid_C) = 2$. Overall, we get a polynomial in y, where the coefficient of y^d is a sum over all assignments σ, which are exacting on all the compulsory clauses (and perhaps some non-compulsory clauses) and non-exacting on precisely d non-compulsory clauses; and for each such a σ, the contribution to the coefficient is the value $(-1)^{|\{C | w(\sigma|_C)=2 \text{ or } 3\}|} = \mathrm{perf}(\sigma)$.

Now if one evaluates the Holant at $m + 1$ many distinct values of y, where m is the number of clauses, we can find all the coefficients of this polynomial.

For the DEFECT problem we can use the symmetric signatures $[-x, -y, y, x]$ and $[-1, 0, 0, 1]$ instead.

Acknowledgments

We would like to thank Leslie Valiant for very interesting discussions. We also thank Andrew Yao, and his group of students in Tsinghua University, for listening to the lectures by the first author on this material and for offering many constructive comments. We also thank in particular Rakesh Kumar and Anand Sinha for many interesting discussions on this and related topics.

References

[1] C. T. J. Dodson and T. Poston. *Tensor Geometry*, Graduate Texts in Mathematics 130, Second edition, Springer-Verlag, New York, 1991.

[2] M. Jerrum and M. Snir: Some Exact Complexity Results for Straight-Line Computations over Semirings. J. ACM 29(3): 874-897 (1982).

[3] P. W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27: 1209-1225 (1961).

[4] P. W. Kasteleyn. Graph Theory and Crystal Physics. In *Graph Theory and Theoretical Physics*, (F. Harary, ed.), Academic Press, London, 43-110 (1967).

[5] V. Strassen, "Gaussian Elimination is Not Optimal." Numerische Mathematik 13, 35 4-356, 1969.

[6] É. Tardos: The gap between monotone and non-monotone circuit complexity is exponential. Combinatorica 8(1): 141-142 (1988).

[7] H. N. V. Temperley and M. E. Fisher. Dimer problem in statistical mechanics – an exact result. *Philosophical Magazine* 6: 1061– 1063 (1961).

[8] L. G. Valiant: Negation can be Exponentially Powerful. Theor. Comput. Sci. 12: 303-314 (1980).

[9] L. G. Valiant. Holographic Algorithms (Extended Abstract). In *Proc. 45th IEEE Symposium on Foundations of Computer Science*, 2004, 306–315. A more detailed version appeared in Electronic Colloquium on Computational Complexity Report TR05-099.

[10] L. G. Valiant. Holographic circuits. In *Proc. 32nd International Colloquium on Automata, Languages and Programming*, 1–15, 2005.

[11] L. G. Valiant. Completeness for parity problems. In *Proc. 11th International Computing and Combinatorics Conference*, 2005.