# Essentially every unimodular matrix defines an expander

Jin-Yi Cai [*]
Department of Computer Science and Engineering
State University of New York
Buffalo, NY 14260
and
Computer Sciences Department
University of Wisconsin
Madison, WI 53706
jyc@cs.wisc.edu

**Abstract**

We generalize the construction of Gabber and Galil to essentially every unimodular matrix in $SL_2(\mathbf{Z})$. It is shown that every parabolic or hyperbolic fractional linear transformation explicitly defines an expander of bounded degree and constant expansion. Thus all but a vanishingly small fraction of unimodular matrices define expanders.

## 1 Introduction

It has been recognized in the last 25 years that certain combinatorial objects called expanders are extremely useful in a number of computer science applications. These include sorting networks, superconcentrators and sparse connection networks in general, pseudorandom generators and amplifications and deterministic simulations, to name just a few.

An $(n, k, d)$ *expander* is a bipartite graph $G = (L, R, E)$, with $|L| = |R| = n$ and at most $kn$ edges, such that for every subset $X$ of $L$, the neighbor set in $R$ has $|\Gamma(X)| \geq [1 + d(1 - |X|/n)]|X|$. Thus, for every subset of input vertices of cardinality at most, say, $n/2$, its neighbor set *expands*, having cardinality at least a constant multiple more than $|X|$. It is generally desired to have $k$ and $d$ fixed and $n$ growing to infinity.

The first existence theorems on expanders were provided by probabilistic counting argument [11][34]. Roughly speaking, such a proof starts by defining a certain probability space of graphs, and then one shows that the probability of such graphs is non-zero. In fact it is usually shown that such probability tends to 1. Thus not only such graphs exist, but they exist in *abundance*. The weakness of such a proof is that it is not explicit.

Margulis [30] was the first to give an explicit construction of a sequence of graphs $\{G_n\}$. This major achievement uses group representation theory. However, while his construction is explicit, the constant of expansion was not explicitly known. Gabber and Galil [21] in a beautiful paper

[*]Research supported in part by NSF grant CCR-9634665, and by a Guggenheim Fellowship. A preliminary version appeared in *The 11th International Symposium on Algorithm and Computation* (ISAAC) 2000, Taipei, Taiwan. Springer-Verlag Lecture Notes in Computer Science, D. T. Lee and Shang-Hua Teng (Eds.) Vol **1969** (2000) 2–22.

gave an explicit construction of graphs $\{G_n\}$ with an explicitly stated constant of expansion. The Gabber-Galil proof also has the added advantage of being relatively elementary. We will follow the proofs of [21] closely. There is an extensive literature on expanders and their applications to the theory of computing, the reference section contains an incomplete list of important works. It was realized that expansion properties are closely related to the second largest eigenvalues of the graph $\lambda(G)$ (see [36, 7]), and for $d$-regular graphs the gap between $d$ and $\lambda(G)$ provides estimates for both upper and lower bound for the expansion constant. The best construction was given by Lubotsky, Phillip and Sarnak [29] and by Margulis [31], where asymptotically optimal $\lambda(G)$ was achieved. The proofs in [29] use deep results from number theory, especially results of Eichler and Igusa concerning the Ramanujan conjecture.

We also mention the interesting construction of Ajtai, Komlós and Szemerédi [4], where they showed a randomly chosen transposition and a full cycle over the group $S_n$ also supply an expander. If the original probabilistic constructions are one extreme of showing the "abundance" of expander graphs, the proof in [4] can be viewed as a construction with reduced randomness. The other extreme is of course the explicit constructions mentioned above. Recently, Reingold et. al. [35] considered a new construction technique called zig-zag graph product.

In this paper, we generalize the construction of Gabber and Galil [21] to essentially every unimodular matrix in $SL_2(\mathbf{Z})$. Our proofs are relatively elementary. They do provide a certain "abundance" as well as being explicit, with the same expansion constant $1 - \sqrt{3}/2$ as in [21]. It is shown that *every* parabolic or hyperbolic fractional linear transformation explicitly defines an expander of bounded degree and constant expansion.

Regarding the complexity of deciding for a graph whether it is an expander, Blum et. al. [14] have shown that it is coNP-complete. We will not discuss the complexity aspect in this paper.

Here is an outline of the paper. In Section 2 we give some preliminary remarks on the matrices we will use in the construction. In Section 3 we prove Theorem 1 which summarizes the first key property of the matrices we use. In Section 4 we further develop these properties of the matrices which are summarized in Theorem 3. This is our main combinatorial handle on the matrices we use in our construction. Then we turn to analytic techniques in Section 5, where we use Fourier analysis to derive the key estimate in Lemma 20. Here the proof uses our combinatorial properties of Theorem 3 and some basic properties of Fourier analysis, including Parseval's equality. In Section 6 we give the construction of the family of graphs and prove that they are indeed expanders. Finally in Section 7 we give some further geometric descriptions of the expanders we constructed.

## 2 Preliminary Remarks

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an integral unimodular matrix, i.e., $A \in SL_2(\mathbf{Z})$, where $a, b, c, d \in \mathbf{Z}$ and $\det A = ad - bc = 1$.

We define a companion matrix $\tilde{A}$ to be $\begin{pmatrix} d & c \\ b & a \end{pmatrix}$. Note that in terms of the mappings they define on $\mathbf{R}^2$, $\tilde{A}$ is merely an exchange of the $x$ and $y$ coordinates. More formally, let $R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then $R = R^{-1}$ is the matrix form of the permutation (12). Thus $\tilde{A} = RAR$.

We are going to consider the set $\Sigma = \{A, \tilde{A}, A^{-1}, \tilde{A}^{-1}\}$. We will use this set to define a constant degree expander. To this end we want all 4 matrices in $\Sigma$ to be distinct.

**Lemma 1** *$A = \tilde{A}$ iff $A = \pm I$.*
*$A = A^{-1}$ iff $A = \pm I$.*

$A = \tilde{A}^{-1}$ *iff* $b + c = 0$.

For the other $\binom{4}{2}$ possibilities, we note that $\tilde{A} = RAR$, and thus

**Lemma 2** $\tilde{A} = A^{-1}$ *iff* $A = \tilde{A}^{-1}$ *iff* $b + c = 0$.
$\tilde{A} = \tilde{A}^{-1}$ *iff* $A = A^{-1}$ *iff* $A = \pm I$.
$A^{-1} = \tilde{A}^{-1}$ *iff* $A = \tilde{A}$ *iff* $A = \pm I$.

The goal in the next two sections is to prove the necessary combinatorial properties of our matrices in $\Sigma$, as summarized in Theorem 3. Regarding the choice of the matrices in $\Sigma$ starting from $A$, especially the definition of the companion matrix $\tilde{A}$, one may think of the possibility of choosing the transpose $A^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ instead as the companion matrix. However there are examples where Theorem 3 is not valid for this choice.

We will henceforth assume $A \neq \pm I$ and $b + c \neq 0$.

## 3 One less, three more

Our goal regarding the combinatorial properties of the set $\Sigma$ is to prove Theorem 3. In this section we first prove a partial result in Theorem 1.

We will assume none of $a, b, c, d$ is zero, and deal with the case where $abcd = 0$ just prior to Theorem 3.

Let $p = (x, y)$. Define the max (or $\infty$-) norm $||p|| = \max\{|x|, |y|\}$. The goal in this section is to show that, under a mild condition, if one of the norms

$$\{||Ap||, ||\tilde{A}p||, ||A^{-1}p||, ||\tilde{A}^{-1}p||\}$$

is strictly less than the corresponding norm $||p||$, then the three other norms are all strictly greater than $||p||$. The proof involves an examination of all the cases with reductions using suitable symmetries.

Let us start with the following Lemma:

**Lemma 3** $||Ap|| < ||p|| \Longrightarrow ||\tilde{A}p|| > ||p||$.

Given $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, for a contradiction assume $||Ap|| < ||p||$ and $||\tilde{A}p|| \leq ||p||$, where $p = (x, y)$. First let's assume $|y| \geq |x|$, thus $||p|| = |y|$. We have

$$\begin{aligned} |ax + by| &< |y| \\ |cx + dy| &< |y| \\ |dx + cy| &\leq |y| \\ |bx + ay| &\leq |y|. \end{aligned}$$

Let $\xi = -\frac{x}{y}$. We note that since the strict inequality $||Ap|| < ||p||$ holds, $y \neq 0$. Dividing through by $y$ and $a, b, c, d$ respectively, we get the rational approximations of $\xi$

$$\begin{aligned} |\xi - \frac{b}{a}| &< \frac{1}{|a|} \\ |\xi - \frac{d}{c}| &< \frac{1}{|c|} \end{aligned}$$

$$
\begin{aligned}
|\xi - \frac{a}{b}| &\leq \frac{1}{|b|} \\
|\xi - \frac{c}{d}| &\leq \frac{1}{|d|}.
\end{aligned}
$$

(We recall that none of $a, b, c, d$ is zero as assumed.) It follows that

$$
\begin{aligned}
\left| |\xi| - |\frac{b}{a}| \right| &< \frac{1}{|a|} \\
\left| |\xi| - |\frac{d}{c}| \right| &< \frac{1}{|c|} \\
\left| |\xi| - |\frac{a}{b}| \right| &\leq \frac{1}{|b|} \\
\left| |\xi| - |\frac{c}{d}| \right| &\leq \frac{1}{|d|}.
\end{aligned}
$$

Then

$$
\frac{|b|-1}{|a|} < |\xi| < \frac{|b|+1}{|a|}
$$
$$
\frac{|a|-1}{|b|} \leq |\xi| \leq \frac{|a|+1}{|b|}.
$$

Thus,

$$
\frac{|a|-1}{|b|} < \frac{|b|+1}{|a|},
$$

and

$$
\frac{|b|-1}{|a|} < \frac{|a|+1}{|b|}.
$$

If $|b| < |a|$ then, being integral, we get $|b| + 1 \leq |a|$ and $|b| \leq |a| - 1$, and so the following contradiction follows

$$
1 \leq \frac{|a|-1}{|b|} < \frac{|b|+1}{|a|} \leq 1.
$$

If $|a| < |b|$ then $|a| + 1 \leq |b|$, $|a| \leq |b| - 1$, and the following contradiction arises

$$
1 \leq \frac{|b|-1}{|a|} < \frac{|a|+1}{|b|} \leq 1.
$$

Hence it follows that $|a| = |b|$. Being a row of a unimodular matrix $A$, the gcd of $(a, b)$ is 1. Thus $|a| = |b| = 1$.

The exact same argument can be made for the pair $(c, d)$. We conclude that $|c| = |d| = 1$ as well. Hence

$$
a, b, c, d = 1 \pmod 2.
$$

It follows that

$$
ad - bc = 0 \pmod 2,
$$

which is a contradiction to $\det A = 1$.

Next we consider the case $|x| \geq |y|$. This is essentially symmetric. We have

$$\begin{aligned}
|ax+by| &< |x| \\
|cx+dy| &< |x| \\
|dx+cy| &\leq |x| \\
|bx+ay| &\leq |x|.
\end{aligned}$$

Let $\eta = -\frac{y}{x}$. Since $x \neq 0$ in this case, $\eta$ is well defined. Dividing through by $x$ and $a, b, c, d$ respectively, we get the rational approximations of $\eta$

$$\begin{aligned}
|\eta - \frac{a}{b}| &< \frac{1}{|b|} \\
|\eta - \frac{c}{d}| &< \frac{1}{|d|} \\
|\eta - \frac{b}{a}| &\leq \frac{1}{|a|} \\
|\eta - \frac{d}{c}| &\leq \frac{1}{|c|}.
\end{aligned}$$

Then

$$\begin{aligned}
\frac{|a|-1}{|b|} < |\eta| < \frac{|a|+1}{|b|} \\
\frac{|b|-1}{|a|} \leq |\eta| \leq \frac{|b|+1}{|a|},
\end{aligned}$$

and thus

$$\begin{aligned}
\frac{|b|-1}{|a|} < \frac{|a|+1}{|b|} \\
\frac{|a|-1}{|b|} < \frac{|b|+1}{|a|}.
\end{aligned}$$

The rest is the same.

This concludes the proof of Lemma 3.

By the symmetry of $a \leftrightarrow d$ and $b \leftrightarrow c$, which effects $A \leftrightarrow \tilde{A}$ we also have the following Lemma,

**Lemma 4** $||\tilde{A}p|| < ||p|| \Longrightarrow ||Ap|| > ||p||$.

We next consider the pair $(||Ap||, ||A^{-1}p||)$.

**Lemma 5** *Suppose* $|\mathrm{tr}(A)| = |a+d| \geq 2$*, then*

$$||Ap|| < ||p|| \Longrightarrow ||A^{-1}p|| > ||p||.$$

Before we give the proof of this lemma, we shall discuss briefly the condition on the trace.

The elements in $SL_2(\mathbf{Z})$ with trace $|a+d| < 2$ are called elliptic elements, $|a+d| = 2$ parabolic elements, and $|a+d| > 2$ hyperbolic elements. (A final class called loxodromic elements for complex linear fractional transformations $z \mapsto \frac{az+b}{cz+d}$ do not occur here since our matrix $A$ is real.) We note that for an integral matrix $A$, these classes are more simply stated as

- Elliptic elements: $a+d=0,\pm 1$.
- Parabolic elements: $|a+d|=2$.
- Hyperbolic elements: $|a+d|>2$.

In view of the mapping properties of these classes, it is not surprising that we needed, for the construction of expanders, the condition that the mappings be parabolic or hyperbolic, and not elliptic. Using Cayley-Hamilton Theorem, it is easy to verify that for every elliptic $A \in SL_2(\mathbf{Z})$, $A^{12}=I$. We also note that except for a vanishingly small fraction, virtually all elements are hyperbolic.

We now turn to the proof of Lemma 5.

Assume for a contradiction that

$$||Ap|| < ||p|| \text{ and yet } ||A^{-1}p|| \leq ||p||.$$

First let's assume that $|y| \geq |x|$. Then we have the inequalities

$$\begin{array}{rcl} |ax+by| & < & |y| \\ |cx+dy| & < & |y| \\ |dx-by| & \leq & |y| \\ |-cx+ay| & \leq & |y|. \end{array}$$

With the second and the fourth inequalities we get

$$|(a+d)y| \leq |cx+dy| + |-cx+ay| < 2|y|,$$

and thus

$$|a+d| < 2,$$

where we have also used the fact that $y \neq 0$ as implied by the strict inequality $||Ap|| < ||p|| = |y|$. This is a contradiction to the assumption that $A$ is not elliptic.

The remaining case for Lemma 5 is when $|x| \geq |y|$. Then

$$\begin{array}{rcl} |ax+by| & < & |x| \\ |cx+dy| & < & |x| \\ |dx-by| & \leq & |x| \\ |-cx+ay| & \leq & |x|. \end{array}$$

This time with the first and the third inequalities we again get

$$|a+d| < 2.$$

The proof of Lemma 5 is complete.

Exactly the same argument gives us the following

**Lemma 6** *Suppose* $|\mathrm{tr}(A)| = |a+d| \geq 2$, *then* $||A^{-1}p|| < ||p|| \Longrightarrow ||Ap|| > ||p||$.

We next consider the pair $(||Ap||, ||\tilde{A}^{-1}p||)$. We now require the condition $|b+c| \geq 2$. This condition is the same as requiring the trace of the permuted matrix $RA$ to be at least 2 in absolute value: $|\mathrm{tr}(RA)| = |b+c| \geq 2$. In terms of the symmetry involved for $x$ and $y$, this is quite natural.

**Lemma 7** *Suppose* $|\mathrm{tr}(RA)| = |b + c| \geq 2$, *then*

$$||Ap|| < ||p|| \Longrightarrow ||\tilde{A}^{-1}p|| > ||p||.$$

For the proof of Lemma 7, again we assume for a contradiction that

$$||Ap|| < ||p|| \text{ and yet } ||\tilde{A}^{-1}p|| \leq ||p||.$$

First assume that $|y| \geq |x|$. Then

$$\begin{aligned} |ax + by| &< |y| \\ |cx + dy| &< |y| \\ |ax - cy| &\leq |y| \\ |-bx + dy| &\leq |y|. \end{aligned}$$

With the first and the third inequalities we get

$$|(b + c)y| = |(ax + by) - (ax - cy)| \leq |ax + by| + |ax - cy| < 2|y|,$$

and thus

$$|b + c| < 2,$$

just as before.

Similarly if $|x| \geq |y|$, then we use the second and the fourth inequalities to get the same contradiction

$$|b + c| < 2.$$

This completes the proof of Lemma 7.

Exactly the same argument gives us the following

**Lemma 8** *Suppose* $|\mathrm{tr}(RA)| = |b + c| \geq 2$, *then*

$$||\tilde{A}^{-1}p|| < ||p|| \Longrightarrow ||Ap|| > ||p||.$$

Combining the 6 Lemmata above (Lemma 3 to Lemma 8), we conclude that under the condition $|\mathrm{tr}(A)| = |a + d| \geq 2$ and $|\mathrm{tr}(RA)| = |b + c| \geq 2$, for each of the 3 pairs

$$(||Ap||, ||\tilde{A}p||), (||Ap||, ||A^{-1}p||), (||Ap||, ||\tilde{A}^{-1}p||),$$

involving $||Ap||$ from the following set

$$\{||Ap||, ||\tilde{A}p||, ||A^{-1}p||, ||\tilde{A}^{-1}p||\}$$

there can be at most one of the entry to be strictly less than $||p||$, and in that case the other entry of the pair is strictly greater than $||p||$.

This is not quite enough for the goal of this section as stated, which includes the remaining 3 pairs not involving $||Ap||$ (and corresponding 6 Lemmata above). However we will handle the remaining proofs by symmetry.

For the pair $(||\tilde{A}p||, ||A^{-1}p||)$ we apply the symmetry $a \leftrightarrow d$, $b \leftrightarrow c$, thus $A \leftrightarrow \tilde{A}$. This reduces the pair $(||\tilde{A}p||, ||A^{-1}p||)$ to the pair $(||Ap||, ||\tilde{A}^{-1}p||)$ and Lemma 7, Lemma 8 give us respectively

**Lemma 9** *Suppose* $|\mathrm{tr}(RA)| = |b + c| \geq 2$, *then* $||\tilde{A}p|| < ||p|| \Longrightarrow ||A^{-1}p|| > ||p||$.

and

**Lemma 10** *Suppose* $|\mathrm{tr}(RA)| = |b+c| \geq 2$*, then* $||A^{-1}p|| < ||p|| \Longrightarrow ||\tilde{A}p|| > ||p||$.

For the pair $(||\tilde{A}p||, ||\tilde{A}^{-1}p||)$ we apply the symmetry $b \leftrightarrow -c$, (and $c \leftrightarrow -b$, $a \leftrightarrow a$, and $d \leftrightarrow d$), thus, $A \leftrightarrow \tilde{A}^{-1}$ and $\tilde{A} \leftrightarrow A^{-1}$. Thus this reduces the pair $(||\tilde{A}p||, ||\tilde{A}^{-1}p||)$ to the pair $(||A^{-1}p||, ||Ap||)$. Now Lemma 6, Lemma 5 give us respectively

**Lemma 11** *Suppose* $|\mathrm{tr}(A)| = |a+d| \geq 2$*, then* $||\tilde{A}p|| < ||p|| \Longrightarrow ||\tilde{A}^{-1}p|| > ||p||$.

and

**Lemma 12** *Suppose* $|\mathrm{tr}(A)| = |a+d| \geq 2$*, then* $||\tilde{A}^{-1}p|| < ||p|| \Longrightarrow ||\tilde{A}p|| > ||p||$.

Finally for the pair $(||A^{-1}p||, ||\tilde{A}^{-1}p||)$ we apply the same symmetry $b \leftrightarrow -c$ as above, which transforms it to the pair $(||\tilde{A}p||, ||Ap||)$. Then we apply Lemma 4, Lemma 3 respectively,

**Lemma 13** $||A^{-1}p|| < ||p|| \Longrightarrow ||\tilde{A}^{-1}p|| > ||p||$.

and

**Lemma 14** $||\tilde{A}^{-1}p|| < ||p|| \Longrightarrow ||A^{-1}p|| > ||p||$.

Combining Lemma 3 to Lemma 14 we have

**Theorem 1** *For any* $A \in SL_2(\mathbf{Z})$*, where* $abcd \neq 0$ *and* $A$*,* $RA$ *not elliptic, then if any one of the following 4 entries*

$$\{||Ap||, ||\tilde{A}p||, ||A^{-1}p||, ||\tilde{A}^{-1}p||\}$$

*is strictly less than the corresponding norm* $||p||$*, then the three other norms are all strictly greater than* $||p||$.

We note that the condition that none of $a, b, c, d$ is zero is only technical, and will be handled later. Only the conditions on the trace are real restrictions.

# 4 At most two equalities

As shown in Section 3 if there is any one among

$$\{||Ap||, ||\tilde{A}p||, ||A^{-1}p||, ||\tilde{A}^{-1}p||\}$$

to be strictly less than $||p||$, then the three other norms are all strictly greater than $||p||$. In particular there are no equalities in this case. Suppose now, for this section, that there are no one among the four to be strictly less than $||p||$, i.e.,

$$\begin{aligned} ||Ap|| &\geq ||p|| && (1) \\ ||\tilde{A}p|| &\geq ||p|| && (2) \\ ||A^{-1}p|| &\geq ||p|| && (3) \\ ||\tilde{A}^{-1}p|| &\geq ||p||. && (4) \end{aligned}$$

We count the number of equalities among these four inequalities. The goal in this section is to show that, for $p \neq 0$, there can be at most two among the four to be equalities. It follows that the other terms, at least two among four, are all strictly greater than $||p||$. Clearly the condition that $p \neq 0$ is necessary for handling the equalities.

We prove this by contradiction. Suppose there are at least three among the four inequalities in (1), (2), (3) and (4) are in fact equalities. Then there are the following *two Alternatives*.

*Alternative (1):*

$$\begin{aligned} ||Ap|| &= ||p|| && (5) \\ ||\tilde{A}p|| &= ||p|| && (6) \end{aligned}$$

both hold and at least one of the following holds

$$\begin{aligned} ||A^{-1}p|| &= ||p|| && (7) \\ ||\tilde{A}^{-1}p|| &= ||p||; && (8) \end{aligned}$$

*Alternative (2):*

Both (7) and (8) hold and at least one of (5) and (6) holds.

In either *Alternatives*, without loss of generality we may assume that $|y| \geq |x|$. We note that the symmetry $x \leftrightarrow y$ exchanges and permutes the equalities

$$\begin{aligned} ||Ap|| = ||p|| &\quad\leftrightarrow\quad ||\tilde{A}p|| = ||p|| \\ ||A^{-1}p|| = ||p|| &\quad\leftrightarrow\quad ||\tilde{A}^{-1}p|| = ||p|| \end{aligned}$$

respectively, and thus the assumption $|y| \geq |x|$ is indeed without loss of generality.

Let us first assume *Alternative (1)*.

Since $p \neq 0$, and $|y| \geq |x|$, we have $y \neq 0$. The first alternative leads to

$$\begin{aligned} |ax + by| &\leq |y| \\ |cx + dy| &\leq |y| \\ |dx + cy| &\leq |y| \\ |bx + ay| &\leq |y|, \end{aligned}$$

and at least one of the following holds

$$\begin{aligned} |dx - by| &\leq |y| \\ |-cx + ay| &\leq |y| \end{aligned}$$

or

$$\begin{aligned} |ax - cy| &\leq |y| \\ |-bx + dy| &\leq |y|. \end{aligned}$$

As in the proof of Lemma 3, denoting $\xi = -\frac{x}{y}$, and dividing through by $y$ and $a, b, c, d$ respectively, we get the rational approximations of $\xi$

$$\begin{aligned} |\xi - \frac{b}{a}| &\leq \frac{1}{|a|} && (9) \\ |\xi - \frac{d}{c}| &\leq \frac{1}{|c|} && (10) \\ |\xi - \frac{a}{b}| &\leq \frac{1}{|b|} && (11) \\ |\xi - \frac{c}{d}| &\leq \frac{1}{|d|}. && (12) \end{aligned}$$

**Lemma 15** *For any unimodular matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ either $|a| \neq |b|$ or $|c| \neq |d|$.*

Assume instead both equalities hold $|a| = |b|$ and $|c| = |d|$. Since they form the rows of a unimodular matrix, the gcd of both $(a, b)$ and $(c, d)$ are 1. Thus

$$|a| = |b| = |c| = |d| = 1,$$

and taking modulo 2

$$a = b = c = d = 1 \pmod 2.$$

However this leads to

$$\det(A) = ad - bc = 0 \pmod 2$$

which contradicts the unimodularity of $A$. Lemma 15 is proved.

Hence we have two possibilities from Lemma 15: 1. $|a| \neq |b|$ or 2. $|c| \neq |d|$. We show that in either cases it leads to $|x| = |y|$.

1. $|a| \neq |b|$

   Suppose $ab > 0$, i.e., they are of the same sign, then $\frac{b}{a} = \frac{|b|}{|a|}$, and from the rational approximations to $\xi$,

   $$\frac{|b| - 1}{|a|} \leq \xi \leq \frac{|b| + 1}{|a|},$$

   and also

   $$\frac{|a| - 1}{|b|} \leq \xi \leq \frac{|a| + 1}{|b|}.$$

   Note that these two bounds on $\xi$ are symmetric for $a$ and $b$. Thus, without loss of generality $|a| > |b|$. Then, by being integral, $|a| \geq |b| + 1$, it follows that

   $$1 \leq \frac{|a| - 1}{|b|} \leq \xi \leq \frac{|b| + 1}{|a|} \leq 1,$$

   which means that these inequalities are in fact all equalities, and $\xi = 1$. By definition of $\xi$, $x = -y$. This is true regardless $|a| > |b|$ or $|a| < |b|$, as long as $ab > 0$.

   The case where $a$ and $b$ are of opposite signs, i.e., $ab < 0$, is handled similarly with $\frac{b}{a} = -\frac{|b|}{|a|}$, and the corresponding rational approximations of $-\xi$. So we obtain $-\xi = 1$. Hence $x = y$.

   We conclude in this case that $|x| = |y|$.

2. $|c| \neq |d|$

   This case is handled by the symmetry $a \leftrightarrow d$ and $b \leftrightarrow c$. Note that the set of rational approximations in Eqn. (9) to Eqn. (12) is invariant under this substitution. Hence we also get $|x| = |y|$.

We now proceed to deal with the possibility $|x| = |y|$, which is $\neq 0$, under the assumption that

$$\begin{aligned} |ax + by| &\leq |y| \\ |cx + dy| &\leq |y| \\ |dx + cy| &\leq |y| \\ |bx + ay| &\leq |y| \end{aligned}$$

and at least one of the following holds

$$\begin{aligned} |dx - by| &\leq |y| \\ |-cx + ay| &\leq |y| \end{aligned}$$

or

$$\begin{aligned} |ax - cy| &\leq |y| \\ |-bx + dy| &\leq |y|. \end{aligned}$$

Consider two cases $x = -y$ and $x = y$:

Case 1. $x = -y$

Dividing through by $|y|$ we have

$$\begin{aligned} |a - b| &\leq 1 \\ |c - d| &\leq 1 \end{aligned}$$

and

$$\begin{aligned} |d + b| &\leq 1 \\ |c + a| &\leq 1. \end{aligned}$$

From these we obtain

$$\begin{aligned} |a + d| &\leq 2 \\ |b + c| &\leq 2. \end{aligned}$$

By our condition on the trace of $A$ and $RA$, i.e., they are not elliptic, we get

$$|a + d| = |b + c| = 2.$$

Hence we get

$$\begin{aligned} |a - b| &= 1 \\ |c - d| &= 1 \\ |d + b| &= 1 \\ |c + a| &= 1. \end{aligned}$$

Thus we can write

$$\begin{pmatrix} b & b \\ c & c \end{pmatrix} = \begin{pmatrix} a & -d \\ d & -a \end{pmatrix} + \mathcal{E}, \tag{13}$$

where we let

$$\mathcal{E} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix},$$

and $\epsilon_{ij} = \pm 1$ for $i, j = 1, 2$.

In $\mathcal{E}$ the top row cannot be of the same sign, otherwise $a + d = 0$. Similarly the bottom row cannot be of the same sign, otherwise $a + d = 0$ as well.

Furthermore, we observe that

$$a + d + (\epsilon_{11} - \epsilon_{12}) = 0$$

and

$$a + d + (\epsilon_{21} - \epsilon_{22}) = 0.$$

Thus the trace $a + d = -2$ iff

$$\mathcal{E} = \begin{pmatrix} +1 & -1 \\ +1 & -1 \end{pmatrix},$$

and the trace $a + d = +2$ iff

$$\mathcal{E} = \begin{pmatrix} -1 & +1 \\ -1 & +1 \end{pmatrix}.$$

However in either way we obtain

$$b + c = 0,$$

by adding the diagonal entries in the matrix equation Eqn. (13).

So under $|a + d| \geq 2, |b + c| \geq 2$ we conclude that $x = -y$ is impossible.

Case 2. $x = y$

This case is handled by the symmetry $b \leftrightarrow -b$ and $c \leftrightarrow -c$ in the above argument for Case 1. with $x = -y$. Thus $x = y$ is also impossible.

We have proved that *Alternative (1)* is in fact impossible.

Finally we consider the second alternative.
Assume *Alternative (2)*, we have: $|y| \geq |x|$ and,

$$\begin{aligned} |dx - by| &\leq |y| \\ |-cx + ay| &\leq |y| \\ |ax - cy| &\leq |y| \\ |-bx + dy| &\leq |y| \end{aligned}$$

and at least one of the following holds

$$\begin{aligned} |ax + by| &\leq |y| \\ |cx + dy| &\leq |y| \end{aligned}$$

or

$$\begin{aligned} |dx + cy| &\leq |y| \\ |bx + ay| &\leq |y|. \end{aligned}$$

Use $\eta = -\xi = \frac{x}{y}$, and the symmetry $a \leftrightarrow d$, and $b \leftrightarrow -b$, $c \leftrightarrow -c$, we conclude that the second alternative is also impossible.

**Theorem 2** *For any $A \in SL_2(\mathbf{Z})$, where $abcd \neq 0$ and $A$, $RA$ not elliptic, then for $p \neq 0$, among*

$$\{||Ap||, ||\tilde{A}p||, ||A^{-1}p||, ||\tilde{A}^{-1}p||\}$$

*there cannot be more than two of them equal to $||p||$.*

We now briefly handle the case with $abcd = 0$. Suppose $a = 0$ or $d = 0$. Then $bc = -1$ by unimodularity. Being both integral, $b = -c = \pm 1$. Then $b + c = 0$. This is excluded.

Suppose $b = 0$, then $ad = 1$ and being integral, $a = d = \pm 1$. Thus the matrix we are dealing with is $A = \pm \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$.

The case of $c = \pm 1$ with $b = 0$ is *the* matrix dealt with by Gabber and Galil [21]. They showed that the matrix $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ does satisfy our statement in Theorem 3, and from that it defines an expander. In fact, based on the properties of $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, with some care, they further showed that $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ also defines an expander with a smaller expansion constant. We note however, that as far as Theorem 3 is concerned, the condition of $RA$ being non-elliptic technically excludes the matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

For the general $c$, it is not difficult to see that the properties stated in Theorem 3 are valid for $|c| \geq 2$ and $b = 0$. The best way to see this is to consider directly the mappings by matrices in $\Sigma$, which are "shears" along the $x$-axis or $y$-axis, respectively, on the set of lattice points with $||(x, y)||_\infty = r$ for any $r \geq 1$. (This *is* the Gabber-Galil proof for $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ in [21].)

By symmetry, the same is true for the case $|b| \geq 2$ and $c = 0$.

Combining Theorem 1, Theorem 2, and the above discussion regarding $abcd = 0$, we have

**Theorem 3** *For any $A \in SL_2(\mathbf{Z})$, where $A$, $RA$ are not elliptic, i.e., $|a + d| \geq 2$ and $|b + c| \geq 2$, and $p \neq 0$, then among*

$$\{||Ap||, ||\tilde{A}p||, ||A^{-1}p||, ||\tilde{A}^{-1}p||\},$$

- *Either one is less than $||p||$ and three others are greater than $||p||$,*
- *Or no more than two are equal to $||p||$ and the rest are all greater than $||p||$.*

# 5 Analytic proof of expansion

In this section we prove some explicit estimates using Fourier analysis. We will follow [21] and adapt their proof for special matrices to general matrices.

Let $B = A$ or $\tilde{A}$ and let $U = [0, 1)^2$. $B$ defines a measure preserving automorphism $\beta = \beta_B$ of $U$ as follows:

$$\beta : (x, y) \mapsto (x, y)B \bmod 1,$$

where mod 1 is taken component-wise in the two-dimensional row vector $(x, y)B$. We will denote $\alpha = \beta_A$ and $\tilde{\alpha} = \beta_{\tilde{A}}$. It is easy to check that $\beta$ is a bijection on $U$ with inverse map $\beta^{-1}(x, y) = (x, y)B^{-1} \bmod 1$. That it is measure preserving follows from the fact that the Jacobi of the map is $\det B = 1$.

For any function $\phi$ on $U$, we can define the function

$$B^*(\phi)(x, y) = \phi(\beta^{-1}(x, y)).$$

We will restrict our discussion to square integrable functions $\phi$ on $U$. For such $\phi$ the Fourier coefficients are defined as follows

$$a_{\binom{m}{n}}(\phi) = \int_U \phi(x, y) e^{-2\pi i(mx+ny)} d\mu(x, y),$$

where $m, n \in \mathbf{Z}$, and $\mu$ is the Lebesgue measure on $U$. The next lemma relates the Fourier coefficients of $\phi$ with that of $B^*(\phi)$.

**Lemma 16**

$$a_{\binom{m}{n}}(B^*(\phi)) = a_{B\binom{m}{n}}(\phi).$$

**Proof:**

$$\begin{aligned} a_{\binom{m}{n}}(B^*(\phi)) &= \int_U \phi(\beta^{-1}(x,y))e^{-2\pi i(x,y)\cdot\binom{m}{n}}d\mu(x,y) \\ &= \int_U \phi(\beta^{-1}(x,y))e^{-2\pi i(x,y)B^{-1}B\binom{m}{n}}d\mu(x,y) \end{aligned}$$

We can replace $(x,y)B^{-1}$ by $\beta^{-1}(x,y)$ in the exponent since the function $\exp[-2\pi i X]$ has integral period 1. Hence, by a substitution of variables $(x', y') = \beta^{-1}(x,y)$, and note that the Jacobi is 1, we get

$$\begin{aligned} a_{\binom{m}{n}}(B^*(\phi)) &= \int_U \phi(x', y')e^{-2\pi i(x',y')B\binom{m}{n}}d\mu(x', y') \\ &= a_{B\binom{m}{n}}(\phi). \end{aligned}$$

Our goal is to obtain a non-trivial estimate for

$$\sum_q \left[|a_{Aq} - a_q|^2 + |a_{\tilde{A}q} - a_q|^2\right],$$

where $q$ ranges over $\mathbf{Z}^2$, and $\{a_q\}$ is square summable $\sum_q |a_q|^2 < \infty$. Note that $A$ and $\tilde{A}$ define permutations on $\mathbf{Z}^2 - \{0\}$ while $A0 = \tilde{A}0 = 0$. Thus the above sum can also range over $\mathbf{Z}^2 - \{0\}$.

Let $f, g$ be any complex square summable functions on $\mathbf{Z}^2 - \{0\}$. The inner product is defined as

$$\langle f, g\rangle = \sum_{q\neq 0} f(q)\cdot \overline{g(q)},$$

and the norm is

$$||f|| = \langle f, f\rangle^{1/2} = \sum_{q\neq 0} |f(q)|^2.$$

It follows that

$$||f - f\circ A||^2 + ||f - f\circ \tilde{A}||^2 = 4||f||^2 - C,$$

where the *cross terms*

$$C = \langle f, f\circ A\rangle + \langle f\circ A, f\rangle + \langle f, f\circ \tilde{A}\rangle + \langle f\circ \tilde{A}, f\rangle,$$

thus $|C| \leq 2\ [\langle |f|, |f\circ A|\rangle + \langle |f|, |f\circ \tilde{A}|\rangle]$.

**Lemma 17**

$$||f - f\circ A||^2 + ||f - f\circ \tilde{A}||^2 \geq (4 - 2\sqrt{3})\ ||f||^2.$$

**Proof:** We only need to show an upper bound $|C| \leq 2\sqrt{3}\ ||f||^2$. Define

$$\lambda(p,q) = \begin{cases} \sqrt{3} & \text{if } ||q|| < ||p|| \\ 1 & \text{if } ||q|| = ||p|| \\ 1/\sqrt{3} & \text{if } ||q|| > ||p|| \end{cases}$$

By Cauchy-Schwarz, $2|XY| \leq \lambda |X|^2 + \frac{1}{\lambda}|Y|^2$. Note that $\lambda(p,q) = \lambda(q,p)^{-1}$, and thus for $\sigma = A$ or $\tilde{A}$,

$$\begin{aligned} 2\sum_{q\neq 0} |f(q)||f(\sigma(q))| &\leq \sum_{q\neq 0} \left[\lambda(q,\sigma(q))|f(q)|^2 + \lambda(\sigma(q),q)|f(\sigma(q))|^2\right] \\ &= \sum_{q\neq 0} |f(q)|^2 \left[\lambda(q,\sigma(q)) + \lambda(q,\sigma^{-1}(q))\right]. \end{aligned}$$

Hence

$$|C| \leq \sum_{q\neq 0} |f(q)|^2 \left[\sum_{\sigma\in\Sigma} \lambda(q,\sigma(q))\right].$$

(Recall that $\Sigma = \{A, \tilde{A}, A^{-1}, \tilde{A}^{-1}\}$.) By Theorem 3, the sum of four terms $\sum_{\sigma\in\Sigma} \lambda(q,\sigma(q)) \leq 2\sqrt{3}$ in all cases (being either $\leq \sqrt{3} + 3/\sqrt{3}$, or $\leq 4/\sqrt{3}$, or $\leq 1 + 3/\sqrt{3}$, or $\leq 2 + 2/\sqrt{3}$.) It follows that $|C| \leq 2\sqrt{3}\ ||f||^2$.

Stated for $\{a_q\}$ we have

**Lemma 18** *If $a_0 = 0$ and $\sum_{q\neq 0} |a_q|^2 < \infty$, then*

$$\sum_q \left[|a_{Aq} - a_q|^2 + |a_{\tilde{A}q} - a_q|^2\right] \geq (4 - 2\sqrt{3}) \sum_q |a_q|^2.$$

We next translate this lemma to integrals via Parseval's equality.

**Lemma 19** *For square integrable function $\phi$ on $U$ with $\int_U \phi = 0$,*

$$\int_U |A^*(\phi) - \phi|^2 + \int_U |\tilde{A}^*(\phi) - \phi|^2 \geq (4 - 2\sqrt{3}) \int_U |\phi|^2.$$

**Proof:** By Parseval's equality, for square integrable $\psi$,

$$\int_U |\psi|^2 = \sum_q |a_q(\psi)|^2,$$

where $a_q(\psi)$ are the Fourier coefficients. Note that $a_0(\phi) = \int_U \phi = 0$. By linearity and Lemma 16, $a_q(A^*(\phi) - \phi) = a_q(A^*(\phi)) - a_q(\phi) = a_{Aq}(\phi) - a_q(\phi)$. Lemma 19 follows from Lemma 18.

Recall the definition of $\beta = \beta_B$ for $B \in \Sigma$, as a mapping from $U$ to $U$: for $\xi \in U$, $\xi \mapsto \beta_B(\xi) = \xi B \bmod 1$.

**Lemma 20** *For measurable set $Z \subseteq U$,*

$$\sum_{B=A,\tilde{A}} \mu[Z - \beta_B^{-1}(Z)] \geq (2 - \sqrt{3})\ \mu(Z)\mu(Z^c).$$

**Proof:** Define $\phi = \chi_Z - \mu(Z) = \begin{cases} \mu(Z^c) & \text{on } Z \\ -\mu(Z) & \text{on } Z^c \end{cases}$, where $\chi_Z$ is the characteristic function on $Z$. Then $\int_U \phi = 0$, and

$$\int_U |\phi|^2 = \mu(Z)\mu(Z^c) < \infty.$$

Let $\xi \in U$, and denote $\beta_A$ by $\alpha$, i.e., $\alpha(\xi) = \xi A \bmod 1$. We observe that

$$\begin{aligned} A^*(\phi)(\xi) &= \phi(\alpha^{-1}(\xi)) \\ &= \begin{cases} \mu(Z^c) & \text{for } \xi \in \alpha(Z) \\ -\mu(Z) & \text{for } \xi \notin \alpha(Z) \end{cases} \\ &= \chi_{\alpha(Z)} - \mu(Z) \end{aligned}$$

It follows that

$$A^*(\phi) - \phi = \chi_{\alpha(Z)} - \chi_Z.$$

Hence for $\int_U |A^*(\phi) - \phi|^2$, the integrand is 1 on the symmetric difference $\alpha(Z)\Delta Z$, and 0 elsewhere. So

$$\int_U |A^*(\phi) - \phi|^2 = \mu[\alpha(Z)\Delta Z].$$

However, $\alpha(Z)\Delta Z = [\alpha(Z) - Z] \cup [Z - \alpha(Z)]$. Since $\alpha$ is bijective and measure preserving,

$$\begin{aligned} \mu[Z - \alpha(Z)] &= \mu[Z] - \mu[Z \cap \alpha(Z)] \\ &= \mu[\alpha(Z)] - \mu[Z \cap \alpha(Z)] \\ &= \mu[\alpha(Z) - Z] \\ &= \mu[Z - \alpha^{-1}(Z)] \end{aligned}$$

Thus

$$\int_U |A^*(\phi) - \phi|^2 = 2\mu[Z - \alpha^{-1}(Z)].$$

Similarly, denote $\tilde{\alpha} = \beta_{\tilde{A}}$, we have

$$\int_U |\tilde{A}^*(\phi) - \phi|^2 = 2\mu[Z - \tilde{\alpha}^{-1}(Z)].$$

Then by Lemma 19,

$$\sum_{B=A,\tilde{A}} \mu[Z - \beta_B^{-1}(Z)] = \frac{1}{2} \sum_{B=A,\tilde{A}} \int_U |B^*(\phi) - \phi|^2 \geq (2 - \sqrt{3}) \int_U |\phi|^2 = (2 - \sqrt{3})\ \mu(Z)\mu(Z^c).$$

# 6 The graph

In this section we give the construction of a family of bipartite graphs, constructed from every matrix $A$ considered in Theorem 3, and prove an explicit expansion constant for the graphs.

We will first define a notion of neighborhood. Denote the unit square by $U = [0,1)^2$. For $p = (i,j) \in \mathbf{Z}^2$, the translated square by $p$ is denoted by $U_p = p + U$. We define a set of "neighborhood" points as follows: For $B = A, \tilde{A}$,

$$N_B = \{q \in \mathbf{Z}^2 \mid \mu[UB \cap U_q] \neq 0\}, \tag{14}$$

where $\mu$ denotes the Lebesgue measure, and $UB = \{\xi B \mid \xi \in U\}$ is the image of $U$ under $B$.

For $k \geq 1$, let the "mod $k$ neighborhood" $N_{B,k}$ be

$$N_{B,k} = N_B \bmod k. \tag{15}$$

Note that the cardinality of $N_{B,k}$ is at most that of $N_B$ for every $k$. In particular since $|N_B|$ is independent of $k$, $|N_{B,k}|$ is bounded in $k$. For any measurable set $V \subseteq \mathbf{R}^2$, denote its mod $k$ fold in the torus $(\mathbf{R}/k\mathbf{Z})^2$ by $(V)_k = V \bmod k$. We will identify $(\mathbf{Z}/k\mathbf{Z})^2$ with the subset $\{0, 1, \ldots, k-1\}^2 \subset \mathbf{Z}^2$, and also identify $(\mathbf{R}/k\mathbf{Z})^2$ with $[0, k)^2 \subset \mathbf{R}^2$. Note that for $q \in (\mathbf{Z}/k\mathbf{Z})^2$, $U_q \subseteq [0, k)^2$, so that $(U_q)_k = U_q$.

We claim that

$$N_{B,k} = \{q \in (\mathbf{Z}/k\mathbf{Z})^2 \mid \mu_k[(UB)_k \cap U_q] \neq 0\}, \tag{16}$$

where $\mu_k$ is the Lebesgue measure on the torus $(\mathbf{R}/k\mathbf{Z})^2$. Also

$$N_{B,k} = (\{q \in \mathbf{Z} \mid \mu_k[(UB)_k \cap (U_q)_k] \neq 0\}) \bmod k. \tag{17}$$

It is easy to see (16) and (17) are equivalent. Indeed let $q \in (\mathbf{Z}/k\mathbf{Z})^2$. For any integral vector $v$, $(U_{q+kv})_k = (U_q)_k = U_q$. Now if $q$ belongs to the set defined in (16), then clearly $q$ also belongs to the set defined in (17). Conversely, suppose $q$ belongs to (17). Then there is some integral vector $v$, such that $\mu_k[(UB)_k \cap (U_{q+kv})_k] \neq 0$. Then as $(U_{q+kv})_k = U_q$, $\mu_k[(UB)_k \cap U_q] \neq 0$. So $q$ also belongs to the set defined in (16).

To show that both (16) and (17) indeed define the set $N_{B,k}$, consider any $q \in (\mathbf{Z}/k\mathbf{Z})^2$.

If $q \in N_{B,k}$, then there exists some integral vector $v$, such that $q + kv \in N_B$, i.e., $\mu[UB \cap U_{q+kv}] \neq 0$. Then $\mu_k[(UB)_k \cap (U_{q+kv})_k] \neq 0$. Since $(U_{q+kv})_k = U_q$ for $q \in (\mathbf{Z}/k\mathbf{Z})^2$, it follows that $\mu_k[(UB)_k \cap U_q] \neq 0$.

Conversely suppose $\mu_k[(UB)_k \cap U_q] \neq 0$. Then there exists some integral vector $v$, such that $\mu[UB \cap U_{q+kv}] \neq 0$. Hence $q + kv \in N_B$, and $q \in N_{B,k}$.

We also note that $|N_A| = |N_{\tilde{A}}|$ and $|N_{A,k}| = |N_{\tilde{A},k}|$. This is because $\tilde{A} = RAR$ simply exchanges the $x$ and $y$ coordinates from $A$, where $R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $R$ is $\mu$-invariant, and $UR = U$ is invariant under $R$.

We now define the family of graphs. For every $k \geq 1$, the bipartite graph $G_k = (L, R, E)$ has $n = k^2$ vertices on both sides, $L = R = (\mathbf{Z}/k\mathbf{Z})^2$. We will arbitrarily order the vertices in $N_{A,k}$ and $N_{\tilde{A},k}$ as $\{q_\ell\}_{\ell=1}^{|N_{A,k}|}$ and $\{\tilde{q}_\ell\}_{\ell=1}^{|N_{\tilde{A},k}|}$, then a vertex $p \in L$ is connected to $p \in R$ and every $p_\ell = pA + q_\ell \bmod k$, and $\tilde{p}_\ell = p\tilde{A} + \tilde{q}_\ell \bmod k$, for every $q_\ell \in N_{A,k}$, and $\tilde{q}_\ell \in N_{\tilde{A},k}$. (The set $\{p_\ell\}_{\ell=1}^{|N_{A,k}|}$ consists of $|N_{A,k}|$ distinct elements, since $\{q_\ell\}_{\ell=1}^{|N_{A,k}|}$ consists of $|N_{A,k}|$ distinct elements; similarly $\{\tilde{p}_\ell\}_{\ell=1}^{|N_{\tilde{A},k}|}$ consists of the same number of distinct elements. But the union of these two sets together with $p \in R$ may have repeated elements. In case there are repeated elements, we do not place multiple edges.) It is clear that the maximum degree of $G_k$ is bounded, being at most $d = 1 + 2|N_A|$.

We denote by $\sigma_0$ the identity map on $(\mathbf{Z}/k\mathbf{Z})^2$, and by $\sigma_\ell$ the permutation $p \mapsto p_\ell = pA + q_\ell \bmod k$. Similarly denote by $\tilde{\sigma}_\ell$ the permutation $p \mapsto \tilde{p}_\ell = p\tilde{A} + \tilde{q}_\ell \bmod k$. Thus for $p \in L$ the neighbor set of $p$ in $R$ is $\Gamma(p) = \{p\} \cup \{\sigma_i(p) \mid 1 \leq i \leq |N_{A,k}|\} \cup \{\tilde{\sigma}_j(p) \mid 1 \leq j \leq |N_{A,k}|\}$. (There is only one edge for every distinct element in the set $\Gamma(p)$.)

For $B = A, \tilde{A}$, we claim that the neighbors of $p$ of the form $p' = pB + q \bmod k$ are precisely those satisfying

$$\mu_k[(U_pB)_k \cap U_{p'}] \neq 0. \tag{18}$$

Moreover, if $p' = pA + q_\ell \bmod k$, the measure

$$\mu_k[(U_pA)_k \cap U_{p'}] = \mu_k[(UA)_k \cap (U_{q_\ell})_k] \tag{19}$$

is independent of $p$ (and depends on $\ell$ only). We note that as $q_\ell \in (\mathbf{Z}/k\mathbf{Z})^2$, $(U_{q_\ell})_k = U_{q_\ell}$, thus this quantity is also the same as $\mu_k[(UA)_k \cap U_{q_\ell}]$.

A similar statement holds true for $p' = p\tilde{A} + \tilde{q}_\ell \bmod k$.

In fact, for any subsets $S, T \subseteq \mathbf{R}^2$, and any point $x \in \mathbf{R}^2$, we claim

$$\mu_k[(x+S)_k \cap (x+T)_k] = \mu_k[(S)_k \cap (T)_k], \tag{20}$$

is independent of $x$.

Assuming (20) for now, we prove the assertions in (18) and (19). We note that $U_pA = pA + UA$. And for $p' \in (\mathbf{Z}/k\mathbf{Z})^2$, let $q = (p' - pA) \bmod k \in (\mathbf{Z}/k\mathbf{Z})^2$. Then $U_{p'} = (U_{p'})_k = (U_{pA+q})_k = (pA + U_q)_k$, and $(U_q)_k = U_q$. Thus, by (20)

$$\mu_k[(U_pA)_k \cap U_{p'}] = \mu_k[(pA + UA)_k \cap (pA + U_q)_k] = \mu_k[(UA)_k \cap (U_q)_k] = \mu_k[(UA)_k \cap U_q].$$

Hence it is non-zero iff $q \in N_{A,k}$ by (16), i.e., $p'$ is of the form $pA + q_\ell \bmod k \in (\mathbf{Z}/k\mathbf{Z})^2$, a neighbor of $p$. This proves (18) and (19).

To prove (20), we may visualize it as follows: Imagine replicating $S$ and $T$ through out the plane $\mathbf{R}^2$ by a lattice vector of $k\mathbf{Z}^2$, i.e., consider $S + k\mathbf{Z}^2$ and $T + k\mathbf{Z}^2$. These subsets are invariant under shifts by $k\mathbf{Z}^2$. Now form the intersection $\Delta = [S + k\mathbf{Z}^2] \cap [T + k\mathbf{Z}^2]$. As an intersection of subsets invariant under shifts by $k\mathbf{Z}^2$, it is easy to see that

$$(\Delta)_k = \Delta \cap [0, k)^2$$

and also

$$(\Delta)_k = (S)_k \cap (T)_k.$$

Now for any point $x$, consider the shift $x + \Delta$. Clearly $x + \Delta = [x + S + k\mathbf{Z}^2] \cap [x + T + k\mathbf{Z}^2]$, also an intersection of subsets invariant under shifts by $k\mathbf{Z}^2$. Then replacing $x + \Delta$ for $\Delta$ above we get

$$(x + \Delta)_k = [x + \Delta] \cap [0, k)^2$$

and also

$$(x + \Delta)_k = (x + S)_k \cap (x + T)_k.$$

Furthermore the measure of $\Delta$ in any shifted square of $[0, k)^2$ is the same, i.e., for all $x$,

$$\mu[\Delta \cap [-x + [0, k)^2]] = \mu[\Delta \cap [0, k)^2].$$

Then

$$\begin{aligned}
\mu_k[(x+S)_k \cap (x+T)_k] &= \mu_k[(x+\Delta)_k] \\
&= \mu_k[[x+\Delta] \cap [0,k)^2] \\
&= \mu[[x+\Delta] \cap [0,k)^2]
\end{aligned}$$

$$
\begin{aligned}
&= \mu[\Delta \cap [-x + [0,k)^2]] \\
&= \mu[\Delta \cap [0,k)^2] \\
&= \mu_k[\Delta \cap [0,k)^2] \\
&= \mu_k[(\Delta)_k] \\
&= \mu_k[(S)_k \cap (T)_k].
\end{aligned}
$$

This proves (20), and thus also (18) and (19). This concludes the definition of the graph $G_k$.

The next Lemma discretizes Lemma 20.

**Lemma 21** *Let $X \subseteq L$. There exists $\tau = \sigma_\ell$, or $\tau = \tilde{\sigma}_\ell$, for some $1 \le \ell \le |N_{A,k}|$, such that*

$$|\tau(X) - X| \ge (1 - \sqrt{3}/2)|X||X^c|/n,$$

*where $n = k^2$.*

**Proof:** For $X$, define a subset of the torus $(\mathbf{R}/k\mathbf{Z})^2$ by $Y = \bigcup_{p \in X} U_p$. Thus each point $p = (i,j) \in X$ is replaced by the translated square $U_p$. Clearly $\mu_k(Y) = |X|$ and $\mu_k(Y^c) = |X^c|$. If we shrink $Y$ by a factor of $k$, we may consider $Z = \frac{1}{k}Y \subseteq U$, in which we can identify $U$ with the unit torus $(\mathbf{R}/\mathbf{Z})^2$. Clearly $\mu(Z) = \frac{|X|}{n}$ and $\mu(Z^c) = \frac{|X^c|}{n}$.

We next consider where does the small square $\frac{1}{k}U_p$ get mapped to under $\alpha$; more specifically, for $p' \in (\mathbf{Z}/k\mathbf{Z})^2$, we ask which $\frac{1}{k}U_{p'}$ contains a subset of an image of $\frac{1}{k}U_p$ with non-zero measure. For $\xi = [(i,j) + (u,v)]/k \in \frac{1}{k}U_p$, where $p = (i,j)$ and $(u,v) \in U$,

$$
\begin{aligned}
\alpha(\xi) &= \xi A \bmod 1 \\
&= \frac{(i+u, j+v)A \bmod k}{k}.
\end{aligned}
$$

So $\alpha(\xi) \in \frac{1}{k}U_{p'}$ iff $(i+u, j+v)A \bmod k \in U_{p'}$. Thus

$$\alpha(\frac{1}{k}U_p) = \frac{1}{k}(U_pA)_k,$$

and

$$\alpha(\frac{1}{k}U_p) \cap \frac{1}{k}U_{p'} = \frac{1}{k}[(U_pA)_k \cap U_{p'}].$$

It follows that

$$\mu[\alpha(\frac{1}{k}U_p) \cap \frac{1}{k}U_{p'}] = \frac{1}{n}\mu_k[(U_pA)_k \cap U_{p'}], \tag{21}$$

and, by (19), this is non-zero iff $p'$ is a neighbor of $p$ of the form $p' = \sigma_\ell(p) = pA + q_\ell \bmod k$. A similar statement is true for $\tilde{\alpha}(\xi)$.

Let $w_\ell = \mu[\alpha(\frac{1}{k}U_p) \cap \frac{1}{k}U_{\sigma_\ell(p)}] > 0$ be the weight of intersection. Note that these weights correspond to disjoint slices of $\alpha(\frac{1}{k}U_p)$ that together make up $\alpha(\frac{1}{k}U_p)$. Since $\alpha$ is measure preserving, $\sum_{1 \le \ell \le |N_{A,k}|} w_\ell = 1/n$. One also observes that $w_\ell$ is independent of $p$, by (21) and (19).

Similarly one can define $\tilde{w}_\ell = \mu[\tilde{\alpha}(\frac{1}{k}U_p) \cap \frac{1}{k}U_{\tilde{\sigma}_\ell(p)}] > 0$, for $\tilde{A}$, and they also sum to $1/n$.

By definition, $Z = \bigcup_{p \in X} \frac{1}{k}U_p$. Within each $\frac{1}{k}U_p$, divide it according to

$$\left[\frac{1}{k}U_p\right] \cap \alpha^{-1}\left(\frac{1}{k}U_{\sigma_\ell(p)}\right),$$

each with weight $w_\ell$.

$\xi \in Z - \alpha^{-1}(Z)$ iff $[\xi \in Z \ \& \ \alpha(\xi) \notin Z]$. For $\xi \in Z$, $\xi \in \frac{1}{k}U_p$ for a unique $p \in X$, and within $\frac{1}{k}U_p$ those $\xi \in (\frac{1}{k}U_p) \cap \alpha^{-1}(\frac{1}{k}U_{\sigma_\ell(p)})$ are mapped to $\frac{1}{k}U_{\sigma_\ell(p)}$. For those $\xi$, $\alpha(\xi) \notin Z$ iff $\sigma_\ell(p) \notin X$. It follows that

$$\begin{aligned}\mu[Z - \alpha^{-1}(Z)] &= \sum_{p \in X} \sum_{1 \le \ell \le |N_{A,k}|} w_\ell \mathbf{1}_{[\sigma_\ell(p) \notin X]} \\ &= \sum_{1 \le \ell \le |N_{A,k}|} w_\ell \sum_{p \in L} \mathbf{1}_{[p \in X \text{ and } \sigma_\ell(p) \notin X]} \\ &= \sum_{1 \le \ell \le |N_{A,k}|} w_\ell |X - \sigma_\ell^{-1}(X)|,\end{aligned}$$

where $\mathbf{1}_{[\cdot]}$ is the indicator function.

Similarly

$$\mu[Z - \tilde{\alpha}^{-1}(Z)] = \sum_{1 \le \ell \le |N_{A,k}|} \tilde{w}_\ell |X - \tilde{\sigma}_\ell^{-1}(X)|.$$

By Lemma 20,

$$\mu[Z - \alpha^{-1}(Z)] + \mu[Z - \tilde{\alpha}^{-1}(Z)] \ge (2 - \sqrt{3}) \frac{|X|}{n} \frac{|X^c|}{n}.$$

Hence,

$$\sum_{1 \le \ell \le |N_{A,k}|} w_\ell |X - \sigma_\ell^{-1}(X)| + \sum_{1 \le \ell \le |N_{A,k}|} \tilde{w}_\ell |X - \tilde{\sigma}_\ell^{-1}(X)| \ge (2 - \sqrt{3})|X||X^c|/n^2.$$

It follows that there exists $\ell_0$, such that either

$$|X - \sigma_{\ell_0}^{-1}(X)| \ge (1 - \sqrt{3}/2)|X||X^c|/n,$$

or

$$|X - \tilde{\sigma}_{\ell_0}^{-1}(X)| \ge (1 - \sqrt{3}/2)|X||X^c|/n,$$

as $\sum_\ell w_\ell = \sum_\ell \tilde{w}_\ell = 1/n$.

In either cases, since $\tau = \sigma_{\ell_0}$ or $\tilde{\sigma}_{\ell_0}$ is a permutation, $|X - \tau^{-1}(X)| = |\tau(X) - X|$, and thus

$$|\tau(X) - X| \ge (1 - \sqrt{3}/2)|X||X^c|/n.$$

Lemma 21 is proved.

Now the neighbor set $\Gamma(X) \supseteq X \cup \tau(X)$, it follows that

$$\begin{aligned}|\Gamma(X)| &= |X| + |\Gamma(X) - X| \\ &\ge |X| + |\tau(X) - X| \\ &\ge \left[1 + (1 - \sqrt{3}/2)\left(1 - \frac{|X|}{n}\right)\right] |X|.\end{aligned}$$

This completes our construction and proof of the expander graphs.

**Theorem 4** *For any $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z})$, where $A$, $RA$ are not elliptic, i.e., $|a + d| \ge 2$ and $|b + c| \ge 2$, we can explicitly construct a family of expander graphs with bounded degree and expansion coefficient $1 - \sqrt{3}/2$.*

# 7 Some geometric descriptions

In this section we give some further concrete geometric descriptions of the neighbor set

$$N_B = \{q \in \mathbf{Z}^2 \mid \mu[UB \cap U_q] \neq 0\}$$

used to define the expander graph. In particular we derive an explicit upper bound on the degree $d$ of the expander graphs in terms of the matrix $B$.

For a subset $S$ in $\mathbf{R}^2$, we denote its set of boundary points by $\partial S$, its closure by $\overline{S} = S \cup \partial S$, and its interior points by $S^o = S - \partial S$ respectively. Consider the unit square $U$ and the parallelogram $UB$. Clearly $\partial U = \{(x,y) \in [0,1]^2 \mid$ either: $[x = 0$ or $1]$, or: $[y = 0$ or $1]\}$, $\overline{U} = [0,1]^2$, and $U^o = (0,1)^2$. Also $\partial(UB) = (\partial U)B$, $\overline{UB} = (\overline{U})B$, and $(UB)^o = (U^o)B$.

Consider the parallelogram $UB$. We observe that the condition on $q \in \mathbf{Z}^2$

$$\mu[UB \cap U_q] \neq 0,$$

is equivalent when $UB$ is replaced by either $\overline{UB}$ or $(UB)^o$, since the measure of the boundary set $\partial(UB)$ is 0. Similarly we can replace $U_q$ by either $\overline{U_q} = q + \overline{U}$ or $U_q^o = q + U^o$. In particular we have

$$N_B = \{q \in \mathbf{Z}^2 \mid \mu[\overline{UB} \cap U_q^o] \neq 0\}. \tag{22}$$

Next we claim that $\mu[\overline{UB} \cap U_q^o] \neq 0$ is equivalent to $\overline{UB} \cap U_q^o \neq \emptyset$. Thus

$$N_B = \{q \in \mathbf{Z}^2 \mid \overline{UB} \cap U_q^o \neq \emptyset\}. \tag{23}$$

To see this, let $\xi \in \overline{UB} \cap U_q^o$. If $\xi \in (UB)^o$, then clearly $\mu[\overline{UB} \cap U_q^o] \neq 0$. Suppose $\xi \in \partial(UB)$. Then at least for a sufficiently small $\epsilon > 0$, there exists one quarter of a small disk

$$D_\epsilon = \{(x,y) \in \mathbf{R}^2 \mid x > 0, y > 0, x^2 + y^2 < \epsilon^2\},$$

whose image under a certain measure-preserving affine linear map (which maps $(0,0)$ to $\xi$) is contained in $(UB)^o \cap U_q^o$. Hence $\mu[\overline{UB} \cap U_q^o] \neq 0$ as well.

It follows that, to identify $N_B$, we should collect all lattice points $q \in \mathbf{Z}^2$ such that for some $\eta \in U^o$, $q + \eta \in \overline{UB}$.

We can reverse this process. Start with an arbitrary $z \in \overline{UB}$, and "place" an open unit square $-U^o$ at $z$. As $z$ runs through $\overline{UB}$, we get a region as the union

$$\overline{UB} + (-U^o) = \bigcup_{z \in \overline{UB}} (z + (-U^o)) = \{z - \xi \mid z \in \overline{UB}, \xi \in U^o\}. \tag{24}$$

We look for all lattice points in this region. i.e.,

$$N_B = \{q \in \mathbf{Z}^2 \mid q \in \overline{UB} + (-U^o)\}. \tag{25}$$

The more interesting claim in this section is the following, which is valid for those $B$ considered in Theorem 4: In (24), it suffices to trace the point $z$ along the boundary of $UB$ only, i.e., there is no need to place $z$ in the interior of $UB$.

**Lemma 22** *For any $A$ considered in Theorem 4, and $B = A, \tilde{A}$,*

$$N_B = \{q \in \mathbf{Z}^2 \mid q \in \partial(UB) + (-U^o)\}. \tag{26}$$

Let $\rho_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ denote the mapping $(x,y) \mapsto (x,y)\rho_\theta$, which is a counter clockwise rotation of $\theta$ on $\mathbf{R}^2$. Let $\rho = \rho_{\frac{\pi}{2}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then $\rho^2 = \rho_\pi = -I$, $\rho^3 = \rho_{\frac{3\pi}{2}} = -\rho$, and $\rho^4 = I$.

The equality (26) in Lemma 22 is in fact *not* valid for matrices in $\{I, \rho, \rho^2, \rho^3\}$. For example for $B = I$, it can be verified that $\partial(UB) + (-U^o) = (-1,1)^2 - \{(0,0)\}$, while $\overline{UB} + (-U^o) = (-1,1)^2$. Simliarly for the other three matrices, the region $\partial(UB) + (-U^o)$ is $\overline{UB} + (-U^o)$ with a center point deleted.

However these matrices do not satisfy the requirements in Theorem 4 and hence are excluded. We now prove that for any integral unimodular matrix $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ other than the 4 powers of $\rho$ above, the equality in (26) holds.

**Proof:** (of Lemma 22) First we consider whether there are any sides of the parallelogram $UB$ which are parallel to the $x$-axis or the $y$-axis.

Suppose for both $x$-axis and $y$-axis, there are sides of the parallelogram $UB$ parallel to it. Then a moment reflection shows that $B$ is one of the four powers of $\rho$ above, which are excluded in Theorem 4.

Now suppose for either the $x$-axis or the $y$-axis, there is no side of the parallelogram $UB$ parallel to it. By symmetry we assume there is no side of $UB$ parallel to the $y$-axis. Label the four vertices of the parallelogram $UB$ as $v_i = (x_i, y_i)$, where $1 \leq i \leq 4$, such that $x_1 = \min\{x_i | 1 \leq i \leq 4\}$, and starting with $v_1$, the four vertices counter clockwise are $v_1, v_2, v_3, v_4$ in that order. Since no side of the parallelogram $UB$ is parallel to the $y$-axis, both $x_2$ and $x_4$ are $> x_1$. Moreover being a parallelogram, $x_3 - x_1 = (x_2 - x_1) + (x_4 - x_1)$, it follows that

$$x_1 < x_2, x_4 < x_3.$$

Suppose there is a lattice point in $z + (-U^o)$, where $z = (x,y) \in \overline{UB}$. If $z \in \partial(UB)$, then we are done. Suppose $z \in (UB)^o$. Consider the vertical line $X = x$ where $x_1 \leq x \leq x_3$. It intersects the parallelogram $\overline{UB}$ in a line segment, possibly degenerate (i.e., a single point). At $x = x_1$ or $x = x_3$ the intersection *is* the single point $z$, which is the vertex $v_1$ or $v_3$ of $\overline{UB}$, and thus not in $(UB)^o$. (Besides, in this case $z$ is a lattice point, and therefore no point in $z + (-U^o)$ is a lattice point.)

Therefore we need only to consider $x_1 < x < x_3$. Then the intersection is a (proper) line segment bounded by the upper end $\overline{z} = (x, \overline{y})$ and the lower end $\underline{z} = (x, \underline{y})$. Note that $\underline{y} < y < \overline{y}$, where the inequalities being strict because $z \in (UB)^o$.

We observe the following, which follows from the fact that $UB$ is a parallelogram: As $x$ varies from $x_1$ to $\min\{x_2, x_4\}$, $\overline{y} - \underline{y}$ is strictly monotonic increasing; it is constant when $x$ varies from $\min\{x_2, x_4\}$ to $\max\{x_2, x_4\}$; finally it is strictly monotonic decreasing when $x$ varies from $\max\{x_2, x_4\}$ to $x_3$. Of course if $x_2 = x_4$ then the middle interval $[\min\{x_2, x_4\}, \max\{x_2, x_4\}]$ is just a single point.

We claim that $0 \leq \overline{y} - \underline{y} \leq 1$ throughout. For, if at any $x$, $\overline{y} - \underline{y} > 1$, then it is so at $x = x_2$. Then there must be a lattice point of the form $(x_2, y^*)$ in the interior of $UB$, where $\underline{y} < y^* < \overline{y}$, which contradicts the unimodularity of $B$.

Moreover, if $x_2 \neq x_4$, then we claim that $0 \leq \overline{y} - \underline{y} < 1$ throughout. Suppose otherwise, then the maximum value at $\min\{x_2, x_4\}$ is $\overline{y} - \underline{y} = 1$. Consider the vertical line at $X = \min\{x_2, x_4\}$. If $x_2 < x_4$ then $\underline{y} = y_2$ is integral, and if $x_2 > x_4$ then $\overline{y} = y_4$ is integral. Then since $\overline{y} - \underline{y} = 1$, in either case, both $\overline{y}$ and $\underline{y}$ are integral at $\min\{x_2, x_4\}$. If $x_2 < x_4$ then $(x_2, \overline{y})$ would be a lattice point on the boundary $\partial(UB)$ of the parallelogram but distinct from the vertices. This is a contradiction to the unimodularity of $B$. Similarly if $x_2 > x_4$ then $(x_4, \underline{y})$ would be such a lattice point.

To cover a lattice point $p \in z + (-U^o)$, we consider the vertical slides of $z$.

If $x_2 \neq x_4$, then as $\overline{y} - \underline{y} < 1$, clearly either $\overline{z} + (-U^o)$ or $\underline{z} + (-U^o)$ covers $p$.

Suppose $x_2 = x_4$. Since $y_4 - y_2 > 0$, being integral, and as $\overline{y} - \underline{y} \leq 1$ always, we get $y_4 - y_2 = 1$. So $\overline{y} - \underline{y}$ has the unique maximum value 1 at $x_2 = x_4$. For any other $x$, $\overline{y} - \underline{y} < 1$. But if $z$ is on the vertical line $X = x_2 (= x_4)$, the $x$-coordinate is integral, thus there are no lattice points in $z + (-U^o)$. For $z$ not on this vertical line, $\overline{y} - \underline{y} < 1$ implies that the sliding argument above works.

This completes the proof of Lemma 22.

The rest of this section is devoted to giving an exact bound for $|N_B|$, which is the number of lattice points in $\partial(UB) + (-U^o)$.

We first note that this bound will be exactly the same, if we used one of the four rotated/reflected copies of $-U^o$, i.e., we could have used $\partial(UB) + S$, for $S \in \{U^o, U^o\rho, -U^o, -U^o\rho\}$. This follows from a more general principle, namely, for any subsets $S$ and $T$ in $\mathbf{R}^2$, when we use a "shifted" copy $x+S$ for $S$, there is an isometry between $T+S$ and $T+(x+S)$, which is a translation by the vector $x$. Now note that every $S \in \{U^o, U^o\rho, -U^o, -U^o\rho\}$ is in fact a shifted copy $x + (-U^o)$ for some lattice vector $x$. In fact, $U^o = (1,1) + (-U^o)$, $U^o\rho = (0,1) + (-U^o)$, and $-U^o\rho = (1,0) + (-U^o)$. Therefore there are exactly the same number of lattice points in $\partial(UB) + S$, for $S \in \{U^o, U^o\rho, -U^o, -U^o\rho\}$.

Now we compute $|N_B|$. We first deal with the case where there are sides of the parallelogram $UB$ which are parallel to the $x$-axis or the $y$-axis. As we saw in the proof of Lemma 22 that for the matrices considered in Theorem 4, not both $x$-axis and $y$-axis can have sides of $UB$ parallel to it. By symmetry, suppose some side of $UB$ is parallel to the $y$-axis (but no side is parallel to the $x$-axis). Then one side of the parallelogram $\overline{UB}$ must be either the segment $\{(0,y) \mid 0 \leq y \leq 1\}$ or $\{(0,y) \mid -1 \leq y \leq 0\}$. After a suitable reflection with respect to $x$-axis and/or $y$-axis we may assume that $UB$ is located in the first quadrant $(x, y \geq 0)$, i.e., $B$ is of the form $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$, for $b \geq 1$. The reflections may have changed $-U^o$ to one of its four reflected copies $\{U^o, U^o\rho, -U^o, -U^o\rho\}$. But as we noted earlier, for counting the lattice points, we may continue to assume it is $-U^o$.

For $B = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$, and $b \geq 1$, it is quite straightforward to compute that the open set $\partial(UB) + (-U^o)$ is a convex polygon with 6 vertices (see Fig 1.).

Counter clockwise starting from the lower left most corner, these are: $(-1,-1), (0,-1), (1, b-1), (1, b+1), (0, b+1), (-1, 1)$. There are 8 lattice points on the boundary of this convex polygon, which are the above 6 points together with $(-1, 0)$ and $(1, b)$. But none of these 8 lattice points should be included since they are not in the *open interior* of the convex polygon $\partial(UB) + (-U^o)$. This interior is bounded by $-1 < x < 1$. Thus the only lattice points in the interior has $x = 0$. Then it is clear that there are exactly $b + 1$ lattice points in $\partial(UB) + (-U^o)$.

If we denote $\left| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right| = |a| + |b| + |c| + |d|$, then we have $|N_B| = |B| - 1$ in this case.

Next we consider a general $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where no side of the parallelogram $UB$ is parallel to either the $x$-axis or the $y$-axis. As before, label the four vertices of the parallelogram $UB$ as $v_i = (x_i, y_i)$, where $1 \leq i \leq 4$, such that $x_1 = \min\{x_i | 1 \leq i \leq 4\}$, and starting with $v_1$, the four vertices counter clockwise are $v_1, v_2, v_3, v_4$ in that order. Since no side of the parallelogram $UB$ is parallel to the $y$-axis, both $x_2$ and $x_4$ are $> x_1$. As observed before, $x_1 < x_2, x_4 < x_3$. Since no side of $UB$ is parallel to the $x$-axis either, $y_2 \neq y_1$ and $y_4 \neq y_1$.

We claim that either both $y_2, y_4 > y_1$ or both $y_2, y_4 < y_1$. Otherwise, by drawing a horizontal line

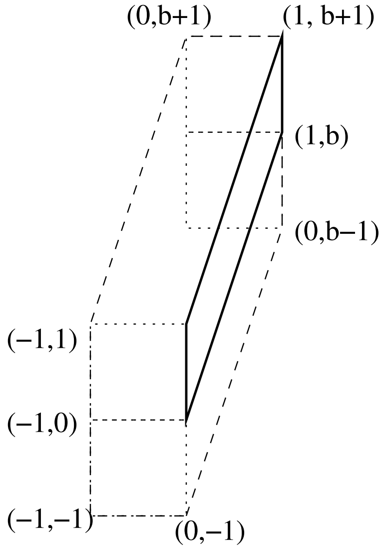


Fig. 1.

at $Y = y_1$, we see that $(\min\{x_2, x_4\}, y_1)$ would be a lattice point in the *interior* of the parallelogram $UB$, which is a contradiction to $B$ being unimodular. After a reflection we may assume both $y_2, y_4 > y_1$.

Now depending on which vertex among $\{v_1, v_2, v_3, v_4\}$ is the origin 0, we have the following four cases. If $v_1 = 0$, then all entries $a, b, c, d > 0$, and $(UB)^o$ is properly contained in the first quadrant $(x, y > 0)$. If $v_2 = 0$, then $a, b > 0$ and $c, d < 0$. In this case by shifting $UB$ with vector $(-c, -d)$, we may assume our matrix is $B = \begin{pmatrix} -c & -d \\ a & b \end{pmatrix}$, and then $(UB)^o$ is properly contained in the first quadrant again. If $v_3 = 0$, then both $a, b < 0$ and $c, d < 0$. We can shift by $(-(a+c), -(b+d))$ and get $B = \begin{pmatrix} -c & -d \\ -a & -b \end{pmatrix}$ in the first quadrant. Finally if $v_4 = 0$, then $a, b < 0$ and $c, d > 0$. By shifting with $(-a, -b)$, we may assume our matrix is $B = \begin{pmatrix} c & d \\ -a & -b \end{pmatrix}$, back to the first quadrant. For all these shifts, $-U^o$ and $|B|$ remain unchanged.

In terms of the matrix $B$, we have arrived at the form $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with all entries $> 0$.

Then the vertices of $UB$ are $(0,0)$, $(a,b)$, $(c,d)$ and $(a+c, b+d)$ respectively. And it can be computed that $\partial(UB) + (-U^o)$ is an open convex polygon with 8 vertices (see Fig 2.):

$$(-1,-1), (0,-1), (a, b-1), (a+c, b+d-1), (a+c, b+d), (a+c-1, b+d), (c-1, d), (-1, 0).$$

Note that while all these 8 vertices are lattice points, none of which is counted in $|N_B|$ since they are all not in the *interior*, and thus does not belong to the open set $\partial(UB) + (-U^o)$.

We can subdivide this convex polygon into 6 parallelograms as in Fig 2. We now count lattice points in $\partial(UB) + (-U^o)$ in each of these 6 parallelograms.

On the boundary of region $IV$ ( $= (UB)^o$) there are 3 lattice points in $\partial(UB) + (-U^o)$, namely $(0,0), (a,b), (c,d)$. Note that the fourth lattice point $(a+c, b+d)$ of region $IV$ is not in $\partial(UB) + (-U^o)$. There are no other lattice points in $\partial(UB) + (-U^o)$ which are on the boundaries of these 6 parallelograms.

Now in the interior of these 6 parallelograms:

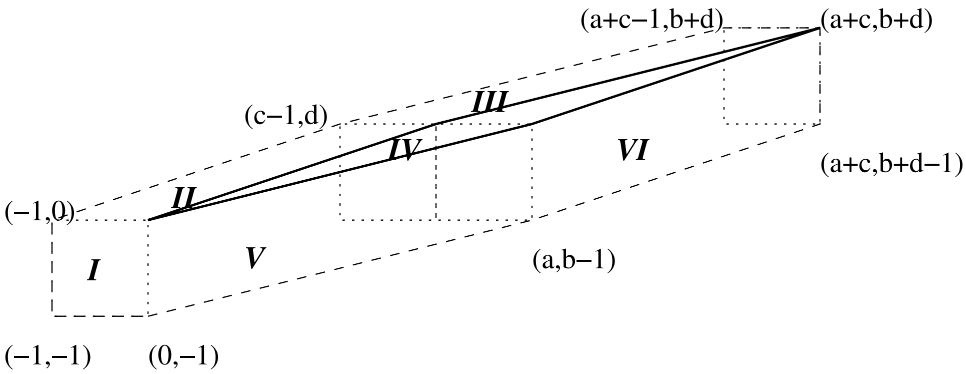


Fig. 2.

1. There are none in regions *I* and *IV*.

2. For region *II*, consider horizontal lines $Y = y$, where $y$ takes an integral values strictly between 0 and $d$. Since each such horizontal line intersects region *II* in an interval of length exact 1, there is exactly one lattice point in the interior. (The lattice point cannot land on the boundary for $0 < y < d$, since otherwise we will have a lattice point on $\partial(UB)$ which is distinct from the vertices of $UB$.) Thus region *II* contributes $d - 1$ lattice points in its interior.

   Similarly, for region *III*, we also consider horizontal lines $Y = y$, where $y$ is integral and $d < y < b + d$. Then the same proof gives $b - 1$ lattice points in the interior of region *III*.

3. For region *V*, we consider vertical lines $X = x$, where $x$ is integral and $0 < x < a$. The same proof idea as above shows that there are $a - 1$ lattice points in the interior of region *V*.

   Similarly, region *VI* has $c - 1$ interior lattice points.

Summarizing, there are exactly $3 + (a - 1) + (b - 1) + (c - 1) + (d - 1) = |B| - 1$ lattice points in $\partial(UB) + (-U^o)$.

**Theorem 5** *The degree $d$ of the expander family constructed in Theorem 4 starting from matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is at most $2|A| - 1 = 2(|a| + |b| + |c| + |d|) - 1$.*

## Acknowledgements

I thank Pavan Aduri, Venkat Chakaravarthy, Bernard Chazelle, Charles Denis, Avi Wigderson and especially Samit Sengupta for valuable discussions, comments and feedbacks. I also thank two anonymous referees for helpful comments. In particular I thank one referee's comment for asking the explicit bound on the degree $d$ of the expander, as now stated in Theorem 5. I also thank Venkat Chakaravarthy and Charles Denis for their help with the xfig program.

## References

[1] M. Ajtai. Recursive construction for Γ-regular expanders. Combinatorica, 14(4):379-416, 1994.

[2] M. Ajtai, J. Komlos and E. Szemeredi, Sorting in $c \log n$ parallel steps, *Combinatorica*, **3**, (1983) 1–19.

[3] M. Ajtai, J. Komlos and E. Szemeredi, Deterministic simulation in LOGSPACE, Proc. of the *19th ACM STOC*, 132–140, 1987.

[4] M. Ajtai, J. Komlós and E. Szemerédi, Generating expanders from two permutations. In *A tribute to Paul Erdös*, edited by A. Baker, B. Bollobás & A. Hajnal. pp. 1–12. Cambridge University Press, 1990.

[5] N. Alon, Eigenvalues, geometric expanders, sorting in rounds and Ramsey Theory, *Combinatorica* **6**, 207–219.

[6] N. Alon, Eigenvalues and Expanders, *Combinatorica* **6** (2), 83–96. 1986.

[7] N. Alon and V. D. Milman, Eigenvalues, expanders and superconcentrators. Proc of *the 25th ACM STOC*, 320–322. 1984.

[8] N. Alon and V. D. Milman. $\Omega$ isoperimetric inequalities for graphs, and superconcentrators. J. Combin. Theory Ser. B, 38(1):73-88, 1985.

[9] N. Alon, Z. Galil, and V. D. Milman. Better expanders an d superconcentrators. J. Algorithms, 8(3):337-347, 1987.

[10] N. Alon, O. Goldreich, J. Hasted and R. Peralta, Simple construction of almost $k$-wise independent random variables, The *31st FOCS*, 544–553.

[11] N. Alon and J. Spencer, with an appendix by P. Erdös, The Probabilistic Method. John Wiley and Sons, Inc.1992.

[12] D. Angulin, A note on a construction of Margulis, *Information Processing Letters*, **8**, pp 17–19, (1979).

[13] M. Blum. Independent unbiased coin flips from a correlated biased source–a finite state Markov chain. *Combinatorica*, 6(2):97-108, 1986.

[14] M. Blum, R. M. Karp, O. Vornberger, C. H. Papadimitriou, M. Yannakakis. The Complexity of Testing Whether a Graph is a Superconcentrator. *Information Processing Letters* 13(3): 164-167 (1981).

[15] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. FOCS 1987. 286-294.

[16] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230-261, 1988.

[17] F. R. K. Chung, On Concentrators, superconcentrators, generalized, and non-blocking networks, *Bell Sys. Tech J.* **58**, pp 1765–1777, (1978).

[18] A. Cohen and A. Wigderson. Dispersers, deterministic amplification, and weak random sources. In Proceedings of the *30th Annual Symposium on Foundations of Computer Science*, pages 14-19, 1989.

[19] J. Friedman. On the second eigenvalue and random walks in random regular graphs. *Combinatorica*, 11(4):331-362, 1991.

[20] J. Friedman, J. Kahn, and E. Szemer'edi. On the second eigenvalue in random regular graphs. STOC 1989. 587-598.

[21] O. Gabber and Z. Galil, Explicit construction of linear size superconcentrators. *JCSS* **22**, pp 407–420 (1981).

[22] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. Security preserving amplification of hardness. FOCS 1990. 318-326.

[23] O. Goldreich and A. Wigderson. Tiny families of functions with random properties: A quality-size trade-off for hashing. *Random Structures and Algorithms*, 11(4):315-343, 1997.

[24] J. Hastad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364-1396, 1999.

[25] R. Impagliazzo, N. Nisan, and A. Wigderson. Pseudorandomness for network algorithms. STOC 1994. 356-364.

[26] R. Impagliazzo, R. Shaltiel, and A. Wigderson. Extractors and pseudo-random generators with optimal seed length. STOC 2000. 1-10.

[27] R. Impagliazzo and D. Zuckerman. How to recycle random bits. FOCS 1989. 248-253.

[28] S. Jimbo and A. Maruoka, Expanders obtained from affine transformations. *Combinatorica*, 7 (4): 343-355, 1987.

[29] A. Lubotsky, R. Phillip and P. Sarnak, Explicit expanders and the Ramanujan conjectures. Proceedings of the *18th ACM STOC*, 1986, 240–246. *Combinatorica*, **8**, 1988, 261–277.

[30] G. A. Margulis, Explicit construction of concentrators. *Problems Inform. Transmission* **9**, 1973, 325–332.

[31] G. A. Margulis, Explicit group-theoretic constructions for combinatorial designs with applications expanders and concentrators. *Problems Inform. Transmission* **24**, 1988, 39–46.

[32] J. Naor and M. Naor, Small bias probability spaces: efficient constructions and applications. Proc. of *22nd ACM STOC*, 1990. 213–223.

[33] N. Pippenger, Superconcentrators. *SIAM J. Computing* **6**, pp 298–304, (1972).

[34] M. Pinsker, On the complexity of a concentrator, The *7th International Teletraffic Conference*, Stockholm, 318/1–318/4, 1973.

[35] O. Reingold, S. Vadhan and A. Wigderson, Entropy Waves, The Zig-Zag Graph Product, and New Constant-Degree Expanders and Extractors, Proc. of *41st IEEE FOCS*, 2000, 3–13.

[36] M. Tanner, Explicit concentrators from generalized $n$-gons. *SIAM J. on Algebraic Discrete Methods*, 5 (3): 287–293, 1984.

[37] L. Valiant, Graph-theoretic properties in computational complexity, *JCSS*, **13**, 1976, 278–285.