Sum–product estimates for well-conditioned matrices

J. Solymosi and V. Vu

Dedicated to the memory of György Elekes

Abstract

We show that if \( A \) is a finite set of \( d \times d \) well-conditioned matrices with complex entries, then the following sum–product estimate holds
\[
|A + A| \times |A \cdot A| = \Omega \left( |A|^{5/2} \right).
\]

1. Introduction

Let \( A \) be a finite subset of a ring \( Z \). The sum–product phenomenon, first investigated by Erdős and Szemerédi [4], suggests that either \( A \cdot A \) or \( A + A \) is much larger than \( A \). This was first proved for \( Z \), the ring of integers, in [4]. Recently, many researchers have studied (with considerable success) other rings. Several of these results have important applications in various fields of mathematics. The interested readers are referred to Bourgain’s survey [1].

In this paper we consider \( Z \) being the ring of \( d \times d \) matrices with complex entries. (We are going to use the notation ‘matrix of size \( d \)’ for \( d \times d \) matrices.) It is well known that one cannot generalize the sum–product phenomenon, at least in the straightforward manner, in this case. The archetypal counterexample is the following:

Example 1.1. Let \( I \) denote the identity matrix and let \( E_{ij} \) be the matrix with only one non-zero entry at position \( ij \) and this entry is one. Let \( M_a := I + aE_{1d} \) and let \( A = \{M_1, \ldots, M_n\} \). It is easy to check that
\[
|A + A| = |A \cdot A| = 2n - 1.
\]

This example suggests that one needs to make some additional assumptions in order to obtain a non-trivial sum–product estimate. Chang [2] proved the following

Theorem 1.2. There is a function \( f = f(n) \) tending to infinity with \( n \) such that the following holds. Let \( A \) be a finite set of matrices of size \( d \) over the reals such that for any \( M \neq M' \in A \), we have \( \det(M - M') \neq 0 \). Then we have
\[
|A + A| + |A \cdot A| \geq f(|A|)|A|.
\]

The function \( f \) in Chang’s proof tends to infinity slowly. In most applications, it is desirable to have a bound of the form \( |A|^{1+c} \) for some positive constant \( c \). In this paper, we show that this is indeed the case (and in fact \( c \) can be set to be \( \frac{1}{4} \)) if we assume that the matrices are far from being singular. Furthermore, this result provides a new insight into the above counterexample (see the discussion following Theorem 2.2).

Received 12 February 2008; revised 9 April 2009; published online 19 July 2009.

2000 Mathematics Subject Classification 11B75 (primary), 15A45, 11C20 (secondary).

The research was conducted while both researchers were members of the Institute for Advanced Study. Funding provided by The Charles Simonyi Endowment. The first author was supported by NSERC and OTKA grants and by Sloan Research Fellowship. The second author was supported by an NSF Career Grant.
NOTATION. We use asymptotic notation under the assumption that $|A| = n$ tends to infinity. Notation such as $f(n) = \Omega_\xi(m)$ means that there is a constant $c > 0$, which depends on $\xi$ only, such that $f(n) \geq c m$ for every large enough $n$. Throughout the paper letter $\xi$ might be a number like $d$ or a vector like $\kappa, d$ or $\alpha, r$. The notation $f(n) = O_\xi(m)$ means that there is a constant $c$, which depends on $\xi$ only, such that $f(n) \leq c m$ for every large enough $n$.

Throughout the paper symbol $\mathbb{C}$ denotes the field of complex numbers.

2. New results

The classical way to measure how close a matrix is to being singular is to consider its condition number. For a matrix $M$ of size $d$, let $\sigma_{\text{max}}(M)$ and $\sigma_{\text{min}}(M)$ be the largest and smallest singular values of $M$. The quantity $\kappa(M) = \sigma_{\text{max}}(M)\sigma_{\text{min}}(M)^{-1}$ is the condition number of $M$. (If $M$ is singular, then $\sigma_{\text{min}}(M) = 0$ and $\kappa(M) = \infty$.)

Our main result shows that if the matrices in $A$ are well conditioned (that is, their condition numbers are small, or equivalently they are far from being singular), then $|A + A| + |A \cdot A|$ is large.

**Definition 2.1.** Let $\kappa$ be a positive number at least one. A set $A$ of matrices is called $\kappa$-well conditioned if the following conditions hold.

(i) For any $M \in A$, we have $\kappa(M) \leq \kappa$.

(ii) For any $M, M' \in A$, we have $\det(M - M') \neq 0$, unless $M = M'$.

**Theorem 2.2.** Let $A$ be a finite $\kappa$-well-conditioned set of size $d$ matrices with complex entries. Then we have

$$|A + A| \times |A \cdot A| \geq \Omega_{\kappa, d}(|A|^{5/2}).$$

Consequently, we have

$$|A + A| + |A \cdot A| \geq \Omega_{\kappa, d}(|A|^{5/4}).$$

Theorem 2.2 is a generalization of the first author’s sum–product bound on complex numbers [7]. Some elements in the proof of Theorem 2.2 were inspired by techniques applied in [7]. The idea of using geometry for sum–product problems was introduced by Elekes [3].

**Remark 2.3.** By following the proof closely, one can set the hidden constant in $\Omega$ as $(\frac{c}{\kappa})^d$, where $c$ is an absolute constant ($\frac{1}{100}$, say, would be sufficient).

**Remark 2.4.** We reconsider the set in the counterexample. It is easy to show that both $\sigma_{\text{max}}(M_a)$ and $\sigma_{\text{min}}(M_a)^{-1}$ are $\Omega_d(a)$. Thus $\kappa(M_a) = \Omega_d(a^2)$, which, for a typical $a$, is $\Omega_d(|A|^2)$. Hence, the matrices in the counterexample have very large condition numbers.

**Remark 2.5.** Note that if the entries of a matrix $M$ of size $d$ are random integers from $\{-n, \ldots, n\}$, then, with probability tending to one as $n$ tends to infinity, $\kappa(M) = O_d(1)$. (In order to see this, note that by Hadamard’s bound, $\sigma_{\text{max}}(M) \leq dn$ with probability one. Moreover, it is easy to show that with high probability $|\det M| = \Omega_d(n^d)$, which implies that $\sigma_{\text{min}}(M) = \Omega_d(n)$.)
The proof of Theorem 2.2 is presented in Sections 3–6.

3. Neighborhoods

Consider a matrix $M$ of size $d$. We can view $M$ as a vector in $\mathbb{C}^{d^2}$ by writing its entries (from left to right, row to row) as the co-ordinates. From now on we consider $A$ as a subset of $\mathbb{C}^{d^2}$.

The matrix operations act as follows:

(i) **addition**: this will be viewed as vector addition;

(ii) **multiplication**: this is a bit more tricky. Take a matrix $M$ of size $d$ and a $d^2$-vector $M'$. To obtain the vector $M'M$, we first rewrite $M'$ as a matrix, then do the matrix multiplication $M'M$, and finally rewrite the result as a vector. This multiplying by $M$ is a linear operator on $\mathbb{C}^{d^2}$.

Next, we need a series of definitions. Note that here we are considering $M$ as a vector in $\mathbb{C}^{d^2}$. The norm $\|M\|$ indicates the length of this vector in $\mathbb{C}^{d^2}$. Then we have the following.

(i) **Radius of** $M$, that is, $r(M) := \min_{M' \in A \setminus \{M\}} \|M - M'\|$.

(ii) **Nearest neighbor** of $M$, that is, $n(M)$ is an $M'$ such that $\|M - M'\| = r(M)$ (if there is more than one $M'$ then choose one arbitrarily).

(iii) **Ball of** $M$, that is, $B(M)$ is the ball in $\mathbb{C}^{d^2}$ around $M$ with radius $r(M)$.

The following lemma will be used frequently in the proof. Let $x, y, z$ be three different points in $\mathbb{C}^r$. The angle $\angle xyz$ is the angle between the rays $yx$ and $yz$. We understand that this angle is at most $\pi$. In $\mathbb{C}^r$ there are various ways of defining the angle between two vectors $x$ and $y$.

(See [6] for a survey of some possible choices.) We are using the notation, where $\text{Re}(y^*x)$ is the real part of the Hermitian product, $(y^*x) = \sum_{i=1}^r \bar{y}_i x_i$. It is important to us that with this definition the law of cosines remains valid, and we have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\|x\|\|y\|\cos(\angle(x, y)). \quad \text{(3.1)}$$

**Lemma 3.1.** For any positive integer $r$ and any constant $0 < \alpha < \pi$, there is a constant $C(\alpha, r)$ such that the following holds. There are at most $C(\alpha, r)$ points on the unit sphere in $\mathbb{C}^r$ such that for any two points $z, z'$, the angle $\angle zoz'$ is at least $\alpha$. (Here $o$ denotes the origin.)

This lemma is equivalent to the statement that a unit sphere in $\mathbb{C}^r$ has at most $C(\delta, r)$ points such that any two has distance at least $\delta$. It can be proved using a simple volume argument. (See [5] for a more advanced approach.) The optimal estimate for $C(\alpha, r)$ is unknown for most pairs $(\alpha, r)$, but this value is not important in our argument.

**Lemma 3.2.** For any positive integer $r$ there is a positive constant $C_1(r)$ such that the following holds. Let $A$ be a set of points in $\mathbb{C}^r$. Then for $z \in \mathbb{C}^r$ there are at most $C_1(r)$ elements $M$ of $A$ such that $z \in B(M)$.

**Proof.** Let $M_1, \ldots, M_k$ be elements of $A$ such that $z \in B(M_i)$ for all $i$. By the definition of $B(M)$ the distance between two distinct elements, $M_i$ and $M_j$, is at least as large as their distances from $z$. Then, by (3.1), the angle $M_i z M_j$ is at least $\pi/3$ for any $i \neq j$. The claim follows from Lemma 3.1.
4. K-normal pairs

Let $K$ be a large constant to be determined. We call an ordered pair $(M, M')$ product $K$-normal if the ellipsoid $B(M)M'$ contains at most $K(|A \cdot A|/|A|)$ points from $A \cdot A$. (Recall that multiplying by $M'$ is a linear operator on $\mathbb{C}^d$, and thus it maps a ball into an ellipsoid.)

**Lemma 4.1.** There is a constant $C_2 = C_2(d)$ such that the following holds. For any fixed $M'$ and $K \geq C_2$, the number of $M$ such that the pair $(M, M')$ is product $K$-normal is at least $(1 - C_2/K)|A|$.

**Proof.** Let $M_1, \ldots, M_m$ be the elements of $A$, where $(M, M)$ is not product $K$-normal. By definition, we have

$$\sum_{i=1}^m |B(M_i)M \cap A \cdot A| \geq Km\frac{|A \cdot A|}{|A|}.$$ 

Set $\varepsilon := m/|A|$. By the pigeon hole principle, there is a point $z$ in $A \cdot A$ belonging to at least $K\varepsilon$ ellipsoids $B(M_i)M$. By applying the map $M^{-1}$, it follows that $zM^{-1}$ belongs to at least $K\varepsilon$ balls $B(M_i)$. By Lemma 3.2, $K\varepsilon = O(d^2) = O(d)$. Thus, $\varepsilon = O(d/K)$, proving the claim.

By the same argument, we can prove the sum version of this lemma. An ordered pair $(M, M')$ is sum $K$-normal if the ball $B(M) + M'$ contains at most $K(|A + A|/|A|)$ points from $A + A$.

**Lemma 4.2.** For any fixed $M'$, the number of $M$ such that the pair $(M, M')$ is sum $K$-normal is at least $(1 - C_2/K)|A|$.

5. Cones

For a ball $B$ in $\mathbb{C}^r$ and a point $x \notin B$, define the cone $\text{Cone}(x, B)$ as

$$\text{Cone}(x, B) := \{tx + (1-t)B | 0 \leq t \leq 1\}.$$ 

Now let $\alpha$ be a positive constant at most $\pi$. For two different points $x$ and $y$, we define the cone $\text{Cone}_\alpha(x, y)$ as $\text{Cone}(x, B_\alpha(y))$, where $B_\alpha(y)$ is the unique ball around $y$ such that the angle of $\text{Cone}(x, B_\alpha(y))$ is exactly $\alpha$. (The angle of $\text{Cone}(x, B_\alpha(y))$ is given by $\max_{s,t \in B_\alpha(y)} \angle sxt$.)

**Lemma 5.1.** For any positive integer $r$ and any constant $0 < \alpha \leq \pi$, there is a constant $C(\alpha, r)$ such that the following holds. Let $A$ be a finite set of points in $\mathbb{C}^r$ and let $L$ be any positive integer. Then for any point $x \in \mathbb{C}^r$, there are at most $C(\alpha, r)L$ points $y$ in $A$ such that the cone $\text{Cone}_\alpha(x, y)$ contains at most $L$ points from $A$.

**Proof.** Case 1: We first prove the case $L = 1$. In this case, if $y \in A$ and $\text{Cone}_\alpha(x, y)$ contains at most one point from $A$, then it contains exactly one point which is $y$. For any two points $y_1, y_2 \in A$ such that both $\text{Cone}_\alpha(x, y_1)$ and $\text{Cone}_\alpha(x, y_2)$ contain exactly one point from $A$, the angle $y_1xy_2$ is at least $\alpha$, by the definition of the cones. Thus, the claim follows from Lemma 3.1.

Case 2: We reduce the case of general $L$ to the case $L = 1$ by a random sparsifying argument. Let $Y = \{y_1, \ldots, y_m\}$ be a set of points in $A$ such that $\text{Cone}_\alpha(x, y_i)$ contains at
most $L$ points from $\mathcal{A}$ for all $1 \leq i \leq m$. We create a random subset $\mathcal{A}'$ of $\mathcal{A}$ by picking each point with probability $p$ (for some $0 < p \leq 1$ to be determined), randomly and independently. We say that $y_i$ survives if it is chosen and no other points in $\mathcal{A} \cap \text{Cone}_\alpha(x, y_i)$ are chosen. For each $y_i \in \mathcal{Y}$, the probability that it survives is at least $p(1-p)^{L-1}$. By linearity of expectations, the expected number of points that survive is at least $mp(1-p)^{L}$. Thus, there are sets $\mathcal{Y}' \subset \mathcal{A}' \subset \mathcal{A}$, where $|\mathcal{Y}'| \geq mp(1-p)^{L}$ with the property that each point $y_i \in \mathcal{Y}'$ is the only point in $\mathcal{A}'$ that appears in $\text{Cone}(x, y_i) \cap \mathcal{A}'$. By the special case $L = 1$, we conclude that $mp(1-p)^{L-1} \leq |\mathcal{Y}'| = O_{\alpha, r}(1)$. The claim of the lemma follows by setting $p = 1/L$.

6. Proof of the main theorem

Consider a point $M$ and its nearest neighbor $n(M)$. Let $M_1$ be another point, viewed as a matrix. We consider the multiplication with $M_1$. This maps the ball $B(M)$ to the ellipsoid $B(M)M_1$ and $n(M)$ to the point $n(M)M_1$.

Since the condition number $\kappa(M_1)$ is not too large, it follows that $B(M)M_1$ is not degenerate. In other words, the ratio between the maximum and minimum distance from $MM_1$ to a point on the boundary of $B(M)M_1$ is bounded from above by $O_{\alpha}(1)$.

Let $b(M, M_1)$ be the largest ball contained in $B(M)M_1$ and $\text{Cone}(M, M_1)$ be the cone with its tip at $n(M)M_1$ defined by

$$\text{Cone}(M, M_1) := \{tn(M)M_1 + (1-t)b(M, M_1)|0 \leq t \leq 1\}.$$ 

The assumption that $M_1$ is well conditioned implies that the angle of this cone is bounded from below by a positive constant $\alpha$ depending only on $\kappa$ and $d$. Thus, we can apply Lemma 5.1 to this system of cones.

Let $T$ be the number of ordered triples $(M_0, M_1, M_2)$ such that $(M_0, M_1)$ is product $K$-normal and $(M_0, M_2)$ is sum $K$-normal.

We choose $K$ sufficiently large so that the constant $(1-C_2/K)$ in Lemmas 4.1 and 4.2 is at least $\frac{4}{5}$. It follows that for any fixed $M_1$ and $M_2$, there are at least $\frac{1}{5}|A|$ matrices $M_0$ such that $(M_0, M_1)$ is product $K$-normal and $(M_0, M_2)$ is sum $K$-normal. This implies that

$$T \geq \frac{4}{5}|A|^3. \quad (6.1)$$

Now we bound $T$ from above. First we embed the triple $(M_0, M_1, M_2)$ into the quadruple $(M_0, n(M_0), M_1, M_2)$. Next, we bound the number of $(M_0, n(M_0), M_1, M_2)$ from above.

The $\kappa$-well-conditioned assumption of Theorem 2.2 guarantees that the quadruple $(M_0, n(M_0), M_1, M_2)$ is uniquely determined by the quadruple

$$(M_0M_1, n(M_0)M_1, M_0 + M_2, n(M_0) + M_2).$$

In order to see this, set $A = M_0M_1, B = n(M_0)M_1, C = M_0 + M_2$ and $D = n(M_0) + M_2$. Then $(M_0 - n(M_0))M_1 = A - B$ and $M_0 - n(M_0) = C - D$. Since $M - M'$ is invertible for any $M \neq M' \in \mathcal{A}$, we have $M_1 = (C - D)^{-1}(A - B)$. (This is the only place where we use this condition.) Since $M_1$ is also invertible (as it has a bounded condition number), it follows that $M_0 = AM_1^{-1}$, $n(M_0) = BM_1^{-1}$ and $M_2 = C - M_0$.

It suffices to bound the number of $(M_0M_1, n(M_0)M_1, M_0 + M_2, n(M_0) + M_2)$.

We first choose $n(M_0)M_1$ from $\mathcal{A} \cdot \mathcal{A}$. There are, of course, $|\mathcal{A} \cdot \mathcal{A}|$ choices. After fixing this point, by Lemma 5.1 and the definition of product $K$-normality, we have $O_{\kappa, d}(K(|\mathcal{A} \cdot \mathcal{A}|/|\mathcal{A}|))$ choices for $M_0M_1$. Similarly, we have $|\mathcal{A} + \mathcal{A}|$ choices for $n(M_0) + M_2$ and for each such choice, we have $O_{\kappa, d}(K(|\mathcal{A} + \mathcal{A}|/|\mathcal{A}|))$ choices for $M_0 + M_2$. It follows that

$$T \leq |\mathcal{A} \cdot \mathcal{A}| \cdot O_{\kappa, d}\left(K\frac{|\mathcal{A} \cdot \mathcal{A}|}{|\mathcal{A}|}\right) \cdot |\mathcal{A} + \mathcal{A}| \cdot O_{\kappa, d}\left(K\frac{|\mathcal{A} + \mathcal{A}|}{|\mathcal{A}|}\right), \quad (6.2)$$

Recall that \( K \) is also a constant depending only on \( \kappa \) and \( d \). Putting (6.1) and (6.2) together, we obtain

\[
\frac{4}{5} |A|^3 \leq O_{\kappa,d} \left( \frac{|A \cdot A||A + A|}{|A|^2} \right),
\]

concluding the proof.

\textbf{Acknowledgements.} The authors thank an anonymous referee for useful comments on a previous draft.

\textbf{References}


\textit{J. Solymosi} \\
Department of Mathematics \\
University of British Columbia \\
1984 Mathematics Road \\
Vancouver, BC \\
Canada V6T 1Z2 \\
solymosi@math.ubc.ca

\textit{V. Vu} \\
Department of Mathematics \\
Rutgers University \\
110 Frelinghuysen Road \\
Piscataway, NJ 08854 \\
USA \\
vantu@math.rutgers.edu