Bull. London Math. Soc. 41 (2009) 817–822
doi:10.1112/blms/bdp054

Sum–product estimates for well-conditioned matrices

J. Solymosi and V. Vu

Dedicated to the memory of György Elekes

Abstract

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

1. *Introduction*

Let $\mathcal{A}$ be a finite subset of a ring Z. The sum–product phenomenon, first investigated by Erdős and Szemerédi [4], suggests that either $\mathcal{A}\cdot\mathcal{A}$ or $\mathcal{A}+\mathcal{A}$ is much larger than $\mathcal{A}$. This was first proved for $\mathbb{Z}$, 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 $\mathcal{A} = \{M_1, \dots, M_n\}$. It is easy to check that $|\mathcal{A}+\mathcal{A}| = |\mathcal{A}\cdot\mathcal{A}| = 2n-1$.

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 $\mathcal{A}$ be a finite set of matrices of size d over the reals such that for any $M \neq M' \in \mathcal{A}$, we have $\det(M-M') \neq 0$. Then we have*

$$|\mathcal{A}+\mathcal{A}| + |\mathcal{A}\cdot\mathcal{A}| \geqslant f(|\mathcal{A}|)|\mathcal{A}|.$$

The function f in Chang's proof tends to infinity slowly. In most applications, it is desirable to have a bound of the form $|\mathcal{A}|^{1+c}$ 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 $|\mathcal{A}| = n$ tends to infinity. Notation such as $f(n) = \Omega_\xi(m)$ means that there is a constant $c > 0$, which depends on ξ only, such that $f(n) \geqslant cm$ for every large enough n. Throughout the paper letter ξ might be a number like d or a vector like κ, d or α, r. The notation $f(n) = \mathrm{O}_\xi(m)$ means that there is a constant c, which depends on ξ only, such that $f(n) \leqslant cm$ for every large enough n. In both cases m is a function of n or it is the constant one function, $m = 1$, in which case we write $\Omega_\xi(1)$ or $\mathrm{O}_\xi(1)$. 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 $\sigma_{\max}(M)$ and $\sigma_{\min}(M)$ be the largest and smallest singular values of M. The quantity $\kappa(M) = \sigma_{\max}(M)\sigma_{\min}(M)^{-1}$ is the *condition* number of M. (If M is singular, then $\sigma_{\min}(M) = 0$ and $\kappa(M) = \infty$.)

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

DEFINITION 2.1. Let κ be a positive number at least one. A set $\mathcal{A}$ of matrices is called κ-well conditioned if the following conditions hold.

(i) For any $M \in \mathcal{A}$, we have $\kappa(M) \leqslant \kappa$.

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

THEOREM 2.2. *Let $\mathcal{A}$ be a finite κ-well-conditioned set of size d matrices with complex entries. Then we have*

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

Consequently, we have

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

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 Ω as $(\frac{c}{\kappa})^{d^2}$, 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 $\sigma_{\max}(M_a)$ and $\sigma_{\min}(M_a)^{-1}$ are $\Omega_d(a)$. Thus $\kappa(M_a) = \Omega_d(a^2)$, which, for a typical a, is $\Omega_d(|\mathcal{A}|^2)$. 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 $\{-n, \dots, n\}$, then, with probability tending to one as n tends to infinity, $\kappa(M) = \mathrm{O}_d(1)$. (In order to see this, note that by Hadamard's bound, $\sigma_{\max}(M) \leqslant dn$ with probability one. Moreover, it is easy to show that with high probability $|\det M| = \Omega_d(n^d)$, which implies that $\sigma_{\min}(M) = \Omega_d(n)$.)

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 $\mathcal{A}$ 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, $r(M) := \min_{M' \in \mathcal{A} \setminus \{M\}} \|M - M'\|$.

(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 xyz is the angle between the rays yx and yz. We understand that this angle is at most π. 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

$$\angle(x, y) = \arccos \frac{\mathrm{Re}(y^* x)}{\|x\|\|y\|}$$

notation, where $\mathrm{Re}(y^* x)$ 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

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

LEMMA 3.1. *For any positive integer r and any constant $0 < \alpha \leqslant \pi$, 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 zoz' is at least α. (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 δ. 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 (α, 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 $\mathcal{A}$ 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 $\mathcal{A}$ such that $z \in B(M)$.*

Proof. Let $M_1, \ldots, M_k$ be elements of $\mathcal{A}$ 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 $K(|\mathcal{A}\cdot\mathcal{A}|/|\mathcal{A}|)$ points from $\mathcal{A}\cdot\mathcal{A}$. (Recall that multiplying by M' is a linear operator on $\mathbb{C}^{d^2}$, 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 $K \geqslant C_2$, the number of M such that the pair (M, M') is product K-normal is at least $(1 - C_2/K)|\mathcal{A}|$.*

Proof. Let $M_1, \dots, M_m$ be the elements of $\mathcal{A}$, where (M_i, M) is not product K-normal. By definition, we have

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

Set $\varepsilon := m/|\mathcal{A}|$. By the pigeon hole principle, there is a point z in $\mathcal{A}\cdot\mathcal{A}$ 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, $K\varepsilon = \mathrm{O}(d^2) = \mathrm{O}(d)$. Thus, $\varepsilon = \mathrm{O}(d)/K$, 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 $K(|\mathcal{A}+\mathcal{A}|/|\mathcal{A}|)$ points from $\mathcal{A}+\mathcal{A}$.

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

5. *Cones*

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

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

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

LEMMA 5.1. *For any positive integer r and any constant $0 < \alpha \leqslant \pi$, there is a constant $C(\alpha, r)$ such that the following holds. Let $\mathcal{A}$ 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 $\mathcal{A}$ such that the cone $\mathrm{Cone}_\alpha(x, y)$ contains at most L points from $\mathcal{A}$.*

Proof. *Case* 1: We first prove the case $L = 1$. In this case, if $y \in \mathcal{A}$ and $\mathrm{Cone}_\alpha(x, y)$ contains at most one point from $\mathcal{A}$, then it contains exactly one point which is y. For any two points $y_1, y_2 \in \mathcal{A}$ such that both $\mathrm{Cone}_\alpha(x, y_1)$ and $\mathrm{Cone}_\alpha(x, y_2)$ contain exactly one point from $\mathcal{A}$, the angle y_1xy_2 is at least α, 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 $\mathcal{Y} = \{y_1, \dots, y_m\}$ be a set of points in $\mathcal{A}$ such that $\mathrm{Cone}_\alpha(x, y_i)$ contains at

most L points from $\mathcal{A}$ for all $1 \leqslant i \leqslant m$. We create a random subset $\mathcal{A}'$ of $\mathcal{A}$ by picking each point with probability p (for some $0 < p \leqslant 1$ to be determined), randomly and independently. We say that y_i *survives* if it is chosen and no other points in $\mathcal{A} \cap \mathrm{Cone}_\alpha(x, y_i)$ 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 $|\mathcal{Y}'| \geqslant mp(1-p)^L$ with the property that each point $y_i \in \mathcal{Y}'$ is the only point in $\mathcal{A}'$ that appears in $\mathrm{Cone}(x, y_i) \cap \mathcal{A}'$. By the special case $L = 1$, we conclude that $mp(1-p)^{L-1} \leqslant |\mathcal{Y}'| = \mathrm{O}_{\alpha,r}(1)$. 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 $\mathrm{O}_\kappa(1)$.

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

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

The assumption that M_1 is well conditioned implies that the angle of this cone is bounded from below by a positive constant α depending only on κ 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 $\frac{9}{10}$. It follows that for any fixed M_1 and M_2, there are at least $\frac{4}{5}|\mathcal{A}|$ 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

$$T \geqslant \frac{4}{5}|\mathcal{A}|^3. \tag{6.1}$$

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 κ-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 $\mathrm{O}_{\kappa,d}(K(|\mathcal{A} \cdot \mathcal{A}|/|\mathcal{A}|))$ 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

$$T \leqslant |\mathcal{A} \cdot \mathcal{A}| \cdot \mathrm{O}_{\kappa,d}\left(K\frac{|\mathcal{A} \cdot \mathcal{A}|}{|\mathcal{A}|}\right) \cdot |\mathcal{A} + \mathcal{A}| \cdot \mathrm{O}_{\kappa,d}\left(K\frac{|\mathcal{A} + \mathcal{A}|}{|\mathcal{A}|}\right). \tag{6.2}$$

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

$$\frac{4}{5}|\mathcal{A}|^3 \leqslant \mathrm{O}_{\kappa,d}\left(\frac{|\mathcal{A}\cdot\mathcal{A}||\mathcal{A}+\mathcal{A}|}{|\mathcal{A}|^2}\right),$$

concluding the proof. □

Acknowledgements. The authors thank an anonymous referee for useful comments on a previous draft.

References

1. J. Bourgain, 'More on the sum–product phenomenon in prime fields and its applications', *Int. J. Number Theory* 1 (2005) 1–32.
2. M.-C. Chang, 'Additive and multiplicative structure in matrix spaces', *Comb. Probab. Comput.* 16 (2007) 219–238.
3. Gy. Elekes, 'On the number of sums and products', *Acta Arith.* 81 (1997) 365–367.
4. P. Erdős and E. Szemerédi, 'On sums and products of integers', *Studies in pure mathematics* (Birkhauser, Basel, 1983) 213–218.
5. O. Henkel, 'Sphere-packing bounds in the Grassmann and Stiefel manifolds', *IEEE Trans. Inf. Theory* 51 (2005) 3445–3456.
6. K. Scharnhorst, 'Angles in complex vector spaces', *Acta Appl. Math.* 69 (2001) 95–103.
7. J. Solymosi, 'On sum-sets and product-sets of complex numbers', *J. Théor. Nombres Bordeaux* 17 (2005) 921–924.

J. Solymosi
Department of Mathematics
University of British Columbia
1984 Mathematics Road
Vancouver, BC
Canada V6T 1Z2

solymosi@math.ubc.ca

V. Vu
Department of Mathematics
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08554
USA

vanvu@math.rutgers.edu