Approximating sparse binary matrices in the cut-norm

Noga Alon *

Abstract

The cut-norm $||A||_C$ of a real matrix $A = (a_{ij})_{i \in R, j \in S}$ is the maximum, over all $I \subset R,\ J \subset S$ of the quantity $|\sum_{i \in I, j \in J} a_{ij}|$. We show that there is an absolute positive constant c so that if A is the n by n identity matrix and B is a real n by n matrix satisfying $||A - B||_C \leq \frac{1}{16}||A||_C$, then $rank(B) \geq cn$. Extensions to denser binary matrices are considered as well.
Keywords: cutnorm, low rank approximation.
AMS Subject Classification: 15A60

1 The main results

The *cut-norm* $||A||_C$ of a real matrix $A = (a_{ij})_{i \in R, j \in S}$ is the maximum, over all $I \subset R,\ J \subset S$ of the quantity $|\sum_{i \in I, j \in J} a_{ij}|$. This concept plays a major role in the work of Frieze and Kannan [7] on efficient approximation algorithms for dense graph and matrix problems.

Consider matrices with a set of rows indexed by R and a set of columns indexed by S. For $I \subset R$ and $J \subset S$, and for a real d, the *cut matrix* $D = CUT(I, J, d)$ is the matrix $(d_{ij})_{i \in R, j \in S}$ defined by $d_{ij} = d$ if $i \in I, j \in J$ and $d_{ij} = 0$ otherwise. A *cut-decomposition* of A expresses it in the form

$$A = D^{(1)} + \ldots + D^{(k)} + W,$$

where the matrices $D^{(i)}$ are cut matrices, and the matrix $W = (w_{ij})$ has a relatively small cut-norm.

The authors of [7] proved that any given n by n matrix A with entries in $[-1, 1]$ admits a cut-decomposition in which the number of cut matrices is $O(1/\epsilon^2)$, and the cut-norm of the matrix W is at most ϵn^2. More generally, it is at most $\epsilon n ||A||_F$, where

$$||A||_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

is the Frobenuis norm of A. The fact that $O(1/\epsilon^2)$ is tight is proved in [3]. Suppose we wish to approximate sparse $\{0,1\}$-matrices, say, n by n binary matrices with m 1's, and our objective is to

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540. Email: nogaa@tau.ac.il. Research supported in part by a USA-Israeli BSF grant 2012/107, by an ISF grant 620/13, by the Israeli I-Core program and by the Fund for Mathematics.

get a cut decomposition so that the cut norm of the matrix W is at most ϵm. How large should k be in this case ? The case of binary matrices arises naturally when considering adjacency matrices of bipartite or general graphs, and the sparse case in which $m = o(n^2)$ is thus interesting.

The first case to consider, which will turn out to be helpful in the study of the general case too, is when A is the n by n identity matrix. Note that in this case the all 0 matrix B satisfies $||A - B||_C = n$, and the constant matrix B' in which each entry is $\frac{1}{n}$ gives $||A - B'||_C \leq n/4$. Therefore, an approximation up to cut norm $\frac{1}{4} \cdot n$ is trivial in this case, and can be done by one cut matrix. It turns out that for smaller values of ϵ, e.g., for $\epsilon = \frac{1}{20}$, the required number k of cut matrices jumps to $\Omega(n)$.

This is proved in the next theorem. In fact, we prove a stronger result: not only does the number of cut matrices in such a cut decomposition have to be linear in n, the rank of any good approximation of the identity matrix in the cut norm has to be $\Omega(n)$.

Theorem 1.1 *There is an absolute positive constant c so that the following holds. Let A be the n by n identity matrix, and let B be an arbitrary real n by n matrix so that $||A-B||_C \leq \frac{n}{16}$. Then the rank of B is at least cn.*

Note that if we replace the cut norm $||A||_C$ of $A = (a_{ij})$ by the ℓ_∞-norm $||A||_\infty = max_{ij}|a_{ij}|$ then it is known (see [1]) that the minimum possible required rank of an ϵ-approximating matrix in this norm (that is, a matrix B so that $||A - B||_\infty \leq \epsilon ||A||_\infty$ $(= \epsilon)$) is between $\Omega(\frac{1}{\epsilon^2 \log(1/\epsilon)} \log n)$ and $O(\frac{1}{\epsilon^2} \log n)$.

The above can be extended to denser binary matrices, yielding the following more general result.

Theorem 1.2 *(i) There is an absolute positive constant c so that the following holds. For every m satisfying $n \leq m \leq n^2$ there is an n by n binary matrix A with m entries equal to 1, so that for any real n by n matrix B satisfying $||A - B||_C \leq \frac{m}{16}$, $rank(B) \geq c\frac{n^2}{m}$.*
(ii) The above is tight in the following sense: any n by n matrix A satisfying $||A||_F^2 \leq m$ admits a cut decomposition

$$A = D^{(1)} + \ldots + D^{(k)} + W,$$

where D_i are cut matrices, $||W||_C \leq \epsilon m$ and $k \leq \frac{n^2}{\epsilon^2 m}$.

It is worth noting that if $n^2/2 \geq m > \epsilon m \geq n$ then there is an n by n binary matrix with m entries equal to 1 so that for any real n by n matrix B satisfying $||A - B||_C \leq \epsilon n$, the rank of B is at least $\Omega(\frac{n^2}{\epsilon m})$. This can be shown by a simple reduction to the result in part (i) of the theorem.

Note also that the constant $\frac{1}{16}$ in both theorems can be improved, we make no attempt to optimize it and the other absolute constants in our estimates. To simplify the presentation we also assume, throughout this note, that n is sufficiently large whenever this is needed, and omit all floor and ceiling signs when these are not crucial.

2 Proofs

Besides the cut-norm $||A||_C$ and the norm $||A||_\infty$, it is convenient to define several other norms of matrices. For a (not-necessarily square) matrix $A = (a_{ij})$ define

$$||A||_{\infty \mapsto 1} = max \sum_{i,j} a_{ij} x_i y_j,$$

where the maximum is taken over all $x_i, y_j \in \{-1, 1\}$. It is easy and well known that for any real matrix A

$$||A||_C \leq ||A||_{\infty \mapsto 1} \leq 4||A||_C. \tag{1}$$

Define also

$$||A||_1 = \sum_{ij} |a_{ij}|$$

and as already mentioned

$$||A||_F^2 = \sum_{i,j} a_{ij}^2.$$

We start with the proof of Theorem 1.1. We need the following two known results. The first is a result of Szarek establishing a tight constant in Khinchin's Inequality.

Lemma 2.1 (Szarek, [9]) *Let $c_1, c_2, \ldots, c_n$ be a set of n reals, let $y_1, \ldots, y_n$ be independent, identically distributed random variables, each distributed uniformly on $\{-1, 1\}$, and define $Y = \sum_i c_i y_i$. Then*

$$E(|Y|) \geq \frac{1}{\sqrt{2}} (c_1^2 + \ldots + c_n^2)^{1/2}.$$

The second result is the following folklore inequality.

Lemma 2.2 (c.f., e.g., [2], Lemma 2.1) *For any real symmetric matrix M,*

$$rank(M) \geq \frac{[trace(M)]^2}{||M||_F^2}.$$

Note that if T is an arbitrary n by n matrix, $(T + T^t)/2$ is symmetric, its rank is at most twice that of T, its trace is equal to the trace of T, and the sum of squares of its entries is at most the sum of squares of the entries of T. Thus the last lemma implies.

Corollary 2.3 *For any real n by n matrix T,*

$$rank(T) \geq \frac{[trace(T)]^2}{2||T||_F^2}.$$

We will also use the following simple corollary of Lemma 2.1.

Corollary 2.4 *For any n by n matrix $B = (b_{ij})$*

$$||B||_{\infty \mapsto 1} \geq \frac{1}{\sqrt{2}} \sum_{i=1}^{n} (\sum_{j=1}^{n} b_{ij}^2)^{1/2}.$$

Proof: Let $y_1, y_2 \ldots, y_n$ be independent and uniform in $\{-1,1\}$. By Lemma 2.1 for each fixed i, $1 \le i \le n$:

$$E[|\sum_{j=1}^n b_{ij} y_j|] \ge \frac{1}{\sqrt{2}} (\sum_{j=1}^n b_{ij}^2)^{1/2}.$$

By linearity of expectation,

$$E[\sum_{i=1}^n |\sum_{j=1}^n b_{ij} y_j|] \ge \frac{1}{\sqrt{2}} \sum_{i=1}^n (\sum_{j=1}^n b_{ij}^2)^{1/2}.$$

Fix values $y_j \in \{-1,1\}$ so that

$$\sum_{i=1}^n |\sum_{j=1}^n b_{ij} y_j| \ge \frac{1}{\sqrt{2}} \sum_{i=1}^n (\sum_{j=1}^n b_{ij}^2)^{1/2}.$$

For each i, let $x_i \in \{-1,1\}$ be the sign of $\sum_{j=1}^n b_{ij} y_j$ (+1, say, if this sum is 0), then

$$||B||_{\infty \mapsto 1} \ge \sum_{i,j} b_{ij} x_i y_j = \sum_{i=1}^n |\sum_{j=1}^n b_{ij} y_j|] \ge \frac{1}{\sqrt{2}} \sum_{i=1}^n (\sum_{j=1}^n b_{ij}^2)^{1/2},$$

completing the proof. □

Proof of Theorem 1.1: Here is an outline of the proof. Let A be the n by n identity matrix, and suppose B is an n by n matrix so that $||A - B||_C \le \frac{n}{16}$. We first show that B does not contain too many rows of Euclidean norm exceeding some absolute constant, and omit these rows, if any, and the corresponding columns. The resulting submatrix B' now approximates a smaller identity matrix. Next we show that the trace of B' is at least $\Omega(n)$ and that the square of its Frobenuis norm is at most $O(n)$. By Corollary 2.3 this implies the desired result. We proceed with the details.

Clearly $||B||_C \le ||A||_C + \frac{n}{16} = \frac{17n}{16}$ and hence by (1), $||B||_{\infty \mapsto 1} \le \frac{17n}{4}$. By Corollary 2.4 this implies

$$\sum_{i=1}^n (\sum_{j=1}^n b_{ij}^2)^{1/2} \le \sqrt{2} ||B||_{\infty \mapsto 1} \le \frac{17\sqrt{2} n}{4}. \tag{2}$$

Define

$$I = \{i : (\sum_{j=1}^n b_{ij}^2)^{1/2} > 34\sqrt{2}.\}$$

By (2), $|I| < n/8$. We can thus delete a set of $n/8$ rows and the same columns from B, getting a matrix $B' = (b'_{ij})$ with $\ell = 7n/8$ rows and columns so that

$$(\sum_{j=1}^{\ell} (b'_{ij})^2)^{1/2} \le 34\sqrt{2}$$

for all $1 \le i \le \ell$. If A' is the ℓ by ℓ identity matrix, then clearly $||A' - B'||_C \le \frac{n}{16}$, as $A' - B'$ is a submatrix of $A - B$. We further assume that ℓ is even, otherwise we omit another row and column to ensure this is the case.

As before, $||B'||_C \le ||A'||_C + \frac{n}{16} = \frac{15n}{16}$ and hence $||B'||_{\infty\mapsto 1} \le \frac{15n}{4}$.

Next we prove that

$$\sum_{1\le i\neq j\le \ell} b'_{ij} \ge -3n. \qquad (3)$$

Indeed, since $||B'||_{\infty\mapsto 1} \le \frac{15n}{4}$, for any $x_i \in \{-1,1\}$

$$-\frac{15n}{4} \le \sum_{i,j} b'_{ij} x_i x_j \le \frac{15n}{4}.$$

Taking expectation over all random choices of $x_i \in \{-1,1\}$ we conclude that

$$-\frac{15n}{4} \le trace(B') \le \frac{15n}{4}.$$

Taking, now, $x_i = y_j = 1$ for all i, j we have

$$trace(B') + \sum_{1\le i\neq j\le \ell} b'_{ij} = \sum_{ij} b'_{ij} \ge \ell - \frac{n}{16} = \frac{13n}{16},$$

implying that

$$\sum_{1\le i\neq j\le \ell} b'_{ij} \ge \frac{13n}{16} - trace(B') \ge -3n,$$

as needed.

Let σ be a random involution of S_ℓ, that is, a random partition of the members of $[\ell] = \{1,2,\ldots,\ell\}$ into $\ell/2$ disjoint pairs $\{i,\sigma(i)\}$. Then the expected value of the random variable $\sum_{i=1}^{\ell} b'_{i,\sigma(i)}$ is exactly

$$\frac{1}{\ell-1} \sum_{1\le i\neq j\le \ell} b'_{ij}$$

since any entry b'_{ij} with $i \neq j$ contributes its value to this sum with probability $\frac{1}{\ell-1}$. By (3) this expectation is at least $-\frac{3n}{\ell-1} > -4$. Fix an involution σ for which $\sum_{i=1}^{\ell} b'_{i,\sigma(i)} \ge -4$.

With this permutation σ, put $S = \sum_{i=1}^{\ell} b'_{i,\sigma(i)}$. Thus $S > -4$. Define also

$$T = \sum_{i,j: i\neq j,\sigma(j)} b'_{i,j}.$$

Consider, now, two random variables X and Y defined as follows. Let $I' \subset [\ell]$ be a random set of indices containing exactly one random member of each pair $\{i,\sigma(i)\}$. Let $J' = [\ell] - I$. Then

$$X = \sum_{i\in I', j\in J'} b'_{ij},$$

whereas

$$Y = \sum_{i\in I', j\in I'} b'_{ij}$$

Note that each diagonal element b'_{ii} of B' contributes to the sum Y with probability 1/2 and does not contribute to X. Each element $b'_{i,\sigma(i)}$ contributes to X with probability 1/2, and does not contribute to Y. Each other element of B' contributes to X with probability 1/4 and to Y with probability 1/4. Thus, by linearity of expectation,

$$E(Y) = \frac{1}{2}trace(B') + \frac{1}{4}T$$

whereas

$$E(X) = \frac{1}{2}S + \frac{1}{4}T$$

showing that

$$E(Y-X) = \frac{1}{2}(trace(B') - S).$$

However, since $||A' - B'||_C \leq \frac{n}{16}$, the variable X is always at most $\frac{n}{16}$ and the variable Y is always at least $\frac{\ell}{2} - \frac{n}{16}$. (This follows by considering the submatrix of $A' - B'$ with rows I' and columns J', and the submatrix with rows I' and columns I'.) Thus the expectation of $Y - X$ is at least $\frac{\ell}{2} - \frac{n}{16} - \frac{n}{16} = \frac{5n}{16}$ and we have

$$E(Y-X) = \frac{1}{2}[trace(B') - S] \geq \frac{5n}{16}$$

implying that

$$trace(B') \geq \frac{10n}{16} + S \geq \frac{5n}{8} - 4 > n/2. \tag{4}$$

Since by the definition of B' for every i

$$(\sum_{j=1}^{\ell}(b'_{ij})^2)^{1/2} \leq 34\sqrt{2}$$

we conclude that

$$\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}(b'_{ij})^2 \leq 34\sqrt{2}\sum_{i=1}^{\ell}(\sum_{j=1}^{\ell}(b'_{ij})^2)^{1/2}$$

$$\leq 34\sqrt{2}\sum_{i=1}^{n}(\sum_{j=1}^{n}b_{ij}^2)^{1/2} \leq 34\sqrt{2}\frac{17\sqrt{2}n}{4} = 289n$$

where here we used (2). We have thus proved that

$$\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}(b'_{ij})^2 \leq 289n.$$

Plugging this and (4) in Corollary 2.3 we conclude that

$$rank(B) \geq rank(B') \geq \frac{n}{8 \cdot 289},$$

completing the proof. □

Proof of Theorem 1.2: The proof of part (i) is by a simple reduction from Theorem 1.1. Given $n \leq m \leq n^2$ let A be the n by n block diagonal matrix consisting of n^2/m blocks, each being a square

m/n by m/n matrix of 1's (we ignore here divisibility issues that are clearly not essential). Let B be an n by n matrix and suppose that $||A-B||_C \leq \frac{m}{16}$. Put $r = \frac{n^2}{m}$.

Let A' be the r by n matrix obtained from A by replacing each block of rows of A by a row which is the sum of them. Similarly, let B' be the r by n matrix obtained from B by the same operation. It is easy to see that $||A'-B'||_C \leq ||A-B||_C$. Next apply the same operation to columns, replacing each block of columns by their sum. We get two r by r matrices, A'' and B'', so that $||A''-B''||_C \leq ||A-B||_C \leq \frac{m}{16}$, with the matrix A'' being exactly $\frac{m^2}{n^2}I$, where I is the r by r identity matrix. Also, clearly $rank(B'') \leq rank(B)$. By Theorem 1.1 and obvious scaling we conclude that if

$$||A''-B''||_C \leq \frac{r}{16}\frac{m^2}{n^2} = \frac{m}{16}$$

then $rank(B'') \geq \Omega(r) = \Omega(\frac{n^2}{m})$, establishing part (i) of the theorem.

The proof of part (ii) follows from the one of Frieze and Kannan in [7]. Starting with $W = A$, as long as $||W||_C \geq \epsilon m$ take a set I of rows and a set J of columns so that $|\sum_{i\in I, j\in J} w_{ij}| \geq \epsilon m$. Define a cut matrix $D = cut(I,J,d)$ where d is the average of the entries of the submatrix of W on the rows I and columns J and define $W' = W - D$. It is easy to check that

$$||W'||_F^2 \leq ||W||_F^2 - \frac{\epsilon^2 m^2}{|I||J|} \leq ||W||_F^2 - \frac{\epsilon^2 m^2}{n^2},$$

and as we started with $||A||_F^2 \leq m$ the process must terminate after at most

$$\frac{m}{(\epsilon^2 m^2/n^2)} = \frac{n^2}{\epsilon^2 m}$$

steps. This completes the proof. □

3 Concluding remarks

- It is interesting to note the striking difference between the cut-approximation considered here and the ℓ_∞-approximation considered in [1], [4] (as well as in papers on communication complexity, see, e.g. [6] and the references therein). For fixed ϵ, say $\epsilon = \frac{1}{20}$, an ϵ-ℓ_∞ approximation of the identity only requires rank $\Theta(\log n)$, while that of any Hadamard matrix (shifted to have 0/1 entries) requires rank $\Omega(n)$ (which is also the rank required for random 0/1 matrices). On the other hand, when approximating using the cut norm, the identity requires linear rank while Hadamard, random or pseudo-random matrices require rank 1 (the constant matrix provides a good approximation). Similarly, it is known that the binary n by n matrix $A = (a_{ij})$ defined by $a_{ij} = 1$ if $i \geq j$ and $a_{ij} = 0$ otherwise admits a $\frac{1}{20} - \ell_\infty$ approximation of rank $O(\log^3 n)$. This follows from the results in [8], see also [4]. On the other hand, our results here imply that if for this A there is a matrix B satisfying $||A-B||_C \leq \frac{n}{20}$, then $rank(B) \geq \Omega(n)$. This is because the identity matrix is $(A + A^t - J)/2$, where J is the all 1 matrix, and hence if B is a good approximation for it, then $(B + B^t - J)/2$ is a good approximation for the identity.

- An (n, d, λ)-graph is a d-regular graph on n vertices in which all eigenvalues but the first have absolute value at most λ. The adjacency matrix of any (n, d, λ)-graph with $\epsilon > \frac{\lambda}{d}$ can be approximated in the cut-norm in the sense considered here by a rank 1 constant matrix, by the Expander Mixing Lemma (see [5], Corollary 9.2.5). Similarly, for random d-regular graphs and $\epsilon > c\frac{\log d}{d}$, rank 1 suffices. On the other hand, if, say, $\epsilon < \frac{1}{100d^2}$ then for any d-regular graph we need rank $\Omega(n/d)$. The lower bound here comes from the fact that any such graph contains an induced matching with at least $\frac{n}{2d}$ edges, and hence the adjacency matrix contains a permutation matrix of size $\Omega(n/d)$ and the lower bound follows from Theorem 1.1.
- It may be interesting to determine the required dependence on ϵ for the minimum possible rank in the assertion of Theorem 1.2. As already mentioned it is not difficult to show that $\Omega(n^2/(\epsilon m))$ is required if $\epsilon m \geq n$, but it is possible that if $\epsilon^2 m \geq n$ one may actually need $\Omega(n^2/(\epsilon^2 m))$, which, if true, is tight.

Acknowledgment: I would like to thank Santosh Vempala for suggesting the problem considered here, and for helpful discussions, and thank Nicholas Wawrykow and Ronald Rietman for helpful comments.

References

[1] N. Alon, Perturbed identity matrices have high rank: proof and applications, Combinatorics, Probability and Computing 18 (2009), 3-15.

[2] N. Alon and P. Pudlak, Equilateral sets in l_p^n, Geometric and Functional Analysis 13 (2003), 467-482.

[3] N. Alon, W. F. de la Vega, R. Kannan and M. Karpinski, Random Sampling and Approximation of MAX-CSP Problems, JCSS 67 (2003), 212-243.

[4] N. Alon, T. Lee, A. Shraibman and S. Vempala, The approximate rank of a matrix and its algorithmic applications, Proc. STOC 2013, 675-684.

[5] N. Alon and J. H. Spencer, *The Probabilistic Method, Third Edition*, Wiley, 2008, xv+352 pp.

[6] H. Buhrman and R. de Wolf, Communication Complexity Lower Bounds by Polynomials, Proc. of the 16th IEEE Annual Conference on Computational Complexity, 2001, 120–130.

[7] A. M. Frieze and R. Kannan, Quick Approximation to matrices and applications, Combinatorica 19 (1999), 175-200.

[8] R. Mathias, The Hadamard operator norm of a circulant and applications, SIAM journal on matrix analysis and applications, 14(1993), 1152–1167.

[9] S. J. Szarek, On the best constants in the Khinchin Inequality, Studia Math. 58 (1976), 197-208.