Linear Algebra and its Applications 486 (2015) 409–418

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Approximating sparse binary matrices in the cut-norm

Noga Alon [a,b,*,1]

[a] *Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel*
[b] *School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, United States*

ARTICLE INFO

Article history:
Received 20 October 2014
Accepted 23 August 2015
Available online 2 September 2015
Submitted by R. Brualdi

MSC:
15A60

Keywords:
Cutnorm
Low rank approximation

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.

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.

* Correspondence to: Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail address: nogaa@tau.ac.il.

[1] 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 Oswald Veblen Fund.

http://dx.doi.org/10.1016/j.laa.2015.08.024
0024-3795/

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 Frobenius 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 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 $-\frac{4}{5n}$ gives $||A - B'||_C \leq n/5$. Therefore, an approximation up to cut norm $\frac{1}{5} \cdot n$ 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.