Rank minimization via the $\gamma_2$ norm

Troy Lee
Rutgers University

Adi Shraibman
Weizmann Institute
Rank Minimization Problem

- Consider the following problem

\[
\min_X \text{rank}(X) \\
\langle A_i, X \rangle \leq b_i \text{ for } i = 1, \ldots, k
\]

- Arises in many contexts: complexity theory, recommendation systems, control theory

- Known to be NP-hard

- Optimization problem over a nonconvex function
Example 1: From communication complexity

- Two parties Alice and Bob wish to evaluate a function $f : X \times Y \to \{-1, +1\}$ where Alice holds $x \in X$ and Bob $y \in Y$.

- How much communication is needed? Can consider deterministic $D(f)$, randomized $R_{\epsilon}(f)$, and even quantum versions $Q_{\epsilon}(f)$.

- Often convenient to work with $|X|$-by-$|Y|$ matrix $A$ known as communication matrix where $A(x, y) = f(x, y)$. Allows tools from linear algebra to be applied.
How a protocol partitions communication matrix

<table>
<thead>
<tr>
<th>Alice</th>
<th>0</th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
How a protocol partitions communication matrix

<table>
<thead>
<tr>
<th></th>
<th>00</th>
<th>01</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bob
Y

Alice
X
How a protocol partitions communication matrix

<table>
<thead>
<tr>
<th>Alice</th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>010</td>
</tr>
<tr>
<td>000</td>
<td>010</td>
</tr>
<tr>
<td>111</td>
<td>101</td>
</tr>
<tr>
<td>111</td>
<td>100</td>
</tr>
</tbody>
</table>

0, 1 represent communication.
Log rank bound

- As a successful protocol partitions the communication matrix into rank one matrices we find

\[ D(f) \geq \log \text{rank}(A_f) \]

- One of the greatest open problems in communication complexity is the log rank conjecture [LS88], which states that \( D(f) \leq (\log \text{rank}(A_f))^k \) for some constant \( k \).
Approximation rank

- As rank lower bounds deterministic communication complexity, the relevant quantity for randomized (and quantum) models is approximation rank.

- Given a target matrix $A$, find a low rank matrix entrywise close to $A$:

\[
\text{rank}_\epsilon(A) = \min_X \text{rank}(X) \quad \text{such that } |X(i,j) - A(i,j)| \leq \epsilon \text{ for all } i, j
\]

- Krause [Kra96] shows that

\[
R_\epsilon(A) \geq \log \text{rank}_\epsilon(A).
\]
Approximation rank

• We do not know if approximation rank remains NP-hard to compute, but can be difficult in practice.

• For the “disjointness” matrix $A$ with rows and columns labeled by $n$-bit strings where

$$A(x, y) = \begin{cases} 
-1 & \text{if } |x \cap y| = 0 \\
1 & \text{otherwise.}
\end{cases}$$

The $\epsilon = 1/3$ approximation rank is $2^{\Theta(\sqrt{n})}$ [Raz03, AA05].
Example 2: Matrix completion

- Popularly known as the “Netflix problem.” Think of a $M$-by-$N$ matrix where rows are labeled by users, columns are labeled by movies, and entries are “ratings.”

- From a partial filling of this matrix—ratings supplied by some users—would like to make predictions for other users, i.e. fill out the rest of the matrix.

- Motivating assumption: a user’s rating depends on only on a few factors, thus the completed matrix should have low rank.
Example 2: Matrix completion

- Given a set $\Omega \subseteq M \times N$ of constraints $X(i, j) = a_{i,j}$ for $(i, j) \in \Omega$, find the lowest rank completion of $X$

$$\min_X \text{rank}(X)$$

$$X(i, j) = a_{i,j} \text{ for all }(i, j) \in \Omega$$

- For applications, can think of a “hidden” low rank matrix $A$ in the background. The goal is to exactly recover this matrix.

- Basic observations: rank $r$ matrix has $O((M + N)r)$ degrees of freedom. In general, will need some assumptions for interesting results—think of matrix with only one nonzero entry.
Convex relaxations

- Part of the difficulty of the rank minimization problem is that it is an optimization problem over a nonconvex function.

- Much work has looked at substituting the rank function by a convex function.

- We will look at substituting rank function by different norms.
Matrix norms

• Define the \( i^{th} \) singular value as \( \sigma_i(X) = \sqrt{\lambda_i(XX^t)} \)

• Many useful matrix norms expressed in terms of vector of singular values \( \sigma(X) = (\sigma_1(X), \ldots, \sigma_n(X)) \).

\[
\|X\|_1 = \ell_1(\sigma(X)) \text{ “trace norm”}
\]
\[
\|X\|_\infty = \ell_\infty(\sigma(X)) \text{ “spectral norm”}
\]
\[
\|X\|_2 = \ell_2(\sigma(X)) = \sqrt{\text{Tr}(XX^t)} \text{ “Frobenius norm”}
\]
Trace norm heuristic

• Popular heuristic is to replace rank by the trace norm

\[
\min_X \|X\|_1 \quad \langle A_i, X \rangle \leq b_i \text{ for } i = 1, \ldots, k
\]

• Motivation: rank is equal to the number of nonzero singular values, thus

\[
\frac{\|X\|_1}{\|X\|_\infty} \leq \text{rank}(X).
\]

• Over matrices satisfying \(\|X\|_\infty \leq 1\), trace norm is the largest convex lower bound on rank [Faz02].
Trace norm heuristic for matrix completion

- There has recently been a lot of work on the trace norm heuristic for the matrix completion problem [Faz02, RFP07, CR08, CT09].

- These results are of the form: Say that $X$ is generated by taking $N$-by-$r$ random Gaussian matrices $Y$ and $Z$ and setting $X = YZ^t$.

- Let $|\Omega| \geq Nr \log^7 n$ consist of entries of $X$ sampled uniformly at random. Then with high probability the trace norm heuristic will exactly recover $X$. 
Trace norm heuristic for approximation rank

• In the context of approximation rank and communication complexity, often work with sign matrices.

• Here the trace norm heuristic is better motivated by another simple inequality:
  \[ \|X\|_1 = \sum_i \sigma_i(X) \leq \sqrt{\text{rank}(X)} \|X\|_2. \]

• For a \(M\text{-by-}N\) sign matrix \(A\) this simplifies nicely:
  \[ \text{rank}(A) \geq \frac{\|A\|_1^2}{MN}. \]
Trace norm bound on rank (example)

• Let $H_N$ be a $N$-by-$N$ Hadamard matrix (entries from $\{-1, +1\}$).

• Then $\|H_N\|_1 = N^{3/2}$.

• Trace norm method gives bound on rank of $N^3/N^2 = N$
• As a complexity measure, the trace norm bound suffers one drawback—it is not monotone.

\[
\begin{pmatrix}
H_N & 1_N \\
1_N & 1_N
\end{pmatrix}
\]

• Trace norm at most \( N^{3/2} + 3N \)

• Trace norm method gives

\[
\frac{(N^{3/2} + 3N)^2}{4N^2} = \frac{N}{4} + O(\sqrt{N})
\]

worse bound on whole than on \( H_N \) submatrix!
Trace norm method (a fix)

• We can fix this by considering

\[
\max_{u,v} \|A \circ uv^t\|_1 \\
\|u\|_2 = \|v\|_2 = 1
\]

• As rank(A \circ uv^t) \leq \text{rank}(A) we still have

\[
\text{rank}(A) \geq \left( \frac{\|A \circ uv^t\|_1}{\|A \circ uv^t\|_2} \right)^2
\]
The $\gamma_2$ norm

- This bound simplifies nicely for a sign matrix $A$

$$\text{rank}(A) \geq \max_{u,v: \|u\|_2 = \|v\|_2 = 1} \left( \frac{\|A \circ uv^t\|_1}{\|A \circ uv^t\|_2} \right)^2 = \max_{u,v: \|u\|_2 = \|v\|_2 = 1} \|A \circ uv^t\|_1^2$$

- We have arrived at the $\gamma_2$ norm introduced to communication complexity by [LMSS07, LS07]

$$\gamma_2(A) = \max_{u,v: \|u\|_2 = \|v\|_2 = 1} \|A \circ uv^t\|_1$$
$\gamma_2$ norm: Surprising usefulness

- $\gamma_2$ is a norm, though not a matrix norm. In matrix analysis known as “Schur/Hadamard product operator/trace norm.”

- Schur (1911) showed that $\gamma_2(A) = \max_i A_{ii}$ if $A$ positive semidefinite.

- $\gamma_2(A)$ can be written as a semidefinite program and so can be well approximated in time polynomial in the size of $A$.

- The dual norm $\gamma_2^*(A) = \max_B \langle A, B \rangle / \gamma_2(B)$ turns up in semidefinite programming relaxation of MAX-CUT of Goemans and Williamson, and is closely related to the discrepancy method in communication complexity.
Cross-section of $\gamma_2$ unit ball of symmetric matrices $[x,y,y,z]$
Approximation rank with $\gamma^2$

- Substitute rank by $\gamma^2$ in the approximation rank optimization problem:

$$\gamma^\epsilon_2(A) = \min_X \gamma_2(X)$$

$$|A(i, j) - X(i, j)| \leq \epsilon \text{ for all } (i, j).$$

- Main theorem: For any $M$-by-$N$ sign matrix $A$ and constant $0 < \epsilon < 1/2$

$$\frac{\gamma^\epsilon_2(A)^2}{(1 + \epsilon)^2} \leq \text{rank}_\epsilon(A) = O\left(\gamma^\epsilon_2(A)^2 \log(MN)\right)^3$$
Proof sketch

• We introduced $\gamma_2$ as a maximization problem.

• For the proof, we use an alternative characterization of $\gamma_2$ in terms of a minimization problem.

• Trace norm can be written as

$$\|X\|_1 = \min_{Y,Z: X = YZ^t} \|Y\|_2 \|Z\|_2$$

This follows from singular value decomposition: $X = U\Sigma V$ where $U, V$ unitary.
Min formulation of $\gamma_2$

$$\gamma_2(X) = \max_{u,v: \|u\|_2 = \|v\|_2 = 1} \|X \circ uv^t\|_1$$

$$= \max_{u,v} \min_{Y,Z} \|D(u)Y\|_2 \|Z^t D(v)\|_2$$

$$X = YZ^t$$

"= " \min_{Y,Z} \max_{u,v} \|D(u)Y\|_2 \|Z^t D(v)\|_2$$

$$X = YZ^t$$

$$= \min_{Y,Z} \|Y\|_r \|Z\|_r$$

$$X = YZ^t$$

where $\|Y\|_r$ is the largest $\ell_2$ norm of a row of $Y$. 
First step: dimension reduction

- Thus $\gamma_2$ looks for factorization $X = YZ^t$ where $Y, Z$ have short rows in terms of $\ell_2$ norm.

- Similarly rank looks for factorization $X = YZ^t$ where $Y, Z$ have short rows in terms of dimension.

- Use Johnson-Lindenstrauss lemma to project rows of $Y, Z$ to dimension about equal to $\gamma_2(X)^2$. Let $R$ be a random $K'$-by-$K$ matrix

\[
\Pr_{R} \left[ \langle Ru, Rv \rangle - \langle u, v \rangle \geq \frac{\delta}{2}(\|u\|^2 + \|v\|^2) \right] \leq 4e^{-\delta^2 K'/8}
\]
Second step: error reduction

• After the first step, we obtain a new matrix $X'$ of rank $O(\gamma^2(A)^2 \log N)$ but the approximation factor has worsened—$X'$ is only $2\epsilon$ close to $A$.

• Trick going back to Krivine: Apply a low degree polynomial entrywise to the matrix $X'$.

$$p(M) = a_0 J + a_1 M + \ldots + a_d M^\circ d.$$  

• See that $\text{rank}(p(M)) \leq (d+1) \text{rank}(M)^d$. Taking $p$ to be approximation to the sign function reduces error.
Polynomial for Error Reduction

\[ \frac{(7x - x^3)}{6} \]

Graph of the polynomial \( \frac{(7x - x^3)}{6} \) against the x and y axes.
Final result

• For any $M$-by-$N$ sign matrix $A$ and constant $0 < \epsilon < 1/2$

\[
\frac{\gamma_2^\epsilon(A)^2}{(1 + \epsilon)^2} \leq \text{rank}_\epsilon(A) = O\left(\gamma_2^\epsilon(A)^2 \log(MN)\right)^3
\]

• Logarithmic factor is necessary as (sign version of) identity matrix has approximation rank $\log(N)$ [Alo03] but constant $\gamma_2$.

• [BES02] used dimension reduction to upper bound sign rank by $\gamma_2^\infty(A)^2$. Interestingly, here the lower bound fails.
Extension to general rank minimization problem

• Consider again the general rank minimization problem

$$\alpha(A, b) = \min_X \text{rank}(X)$$

$$\langle A_i, X \rangle \leq b_i \text{ for } i = 1, \ldots, k$$

• Let $C = \{X : \langle A_i, X \rangle \leq b_i\}$ be the feasible set.

• Let $\ell_\infty(C) = \max_{X \in C} \ell_\infty(X)$. 
Extension to general rank minimization problem

• As argued before we have

\[ \alpha(A, b) \geq \min_{X \in C} \frac{\gamma_2(X)^2}{\ell_\infty(C)^2} \]

• Say we solve via semidefinite programming the program \( \min_{X \in C} \gamma_2(X) \)

• Then by doing dimension reduction on an optimal \( X^* \) we obtain a matrix \( Y \) of rank about \( \gamma_2(X^*)^2 \log(N) \) which is \( \epsilon \)-close to \( X^* \).

• This matrix will satisfy the \( i^{th} \) constraint up to a factor \( \epsilon \ell_1(A_i) \).
Application to the matrix completion problem

• In matrix completion, often is a natural bound on $\ell_\infty(C)$. For example, Netflix uses ratings \{1, 2, 3, 4, 5\} so matrix will be bounded.

• Also in the matrix completion problem each constraint matrix $A_i$ consists of a single entry so $\ell_1(A_i) = 1$.

• Thus if the lowest rank completion has rank $d$, via $\gamma_2$ we can find a rank $O(d \log(N))$ matrix which is $\epsilon$-close on the specified entries.
Open questions

• Approximation algorithm for the limiting case of sign rank?

• For matrix completion, can one show similar unconditional results for trace norm heuristic?

• Practical implementations of $\gamma_2$ for large matrices. Standard SDP solvers in Matlab top out around instances of size 100-by-100.