Complexity of Changing Matrix Rank

Jayalal Sarma M.N.

The Institute of Mathematical Sciences(IMSc),
Chennai, India

WCDPCA 2007

Moscow, Russia

August 29-31, 2007

Matrix Rank

Rank of a matrix $M \in \mathbb{F}^{n \times n}$ has the following equivalent definitions.

- The size of the largest submatrix with a non-zero determinant.
- The number of linearly independent rows/columns of a matrix.
- The smallest r such that $M = AB$ where $A \in \mathbb{F}^{n \times r}$, $B \in \mathbb{F}^{r \times n}$.

RANK BOUND: Given a matrix M and a value r, is $\text{rank}(M) < r$?

Matrix Rank

Rank of a matrix $M \in \mathbb{F}^{n \times n}$ has the following equivalent definitions.

- The size of the largest submatrix with a non-zero determinant.
- The number of linearly independent rows/columns of a matrix.
- The smallest r such that $M = AB$ where $A \in \mathbb{F}^{n \times r}, B \in \mathbb{F}^{r \times n}$.

RANK BOUND: Given a matrix M and a value r, is $\text{rank}(M) < r$?

Motivations from linear algebra, control theory, from algorithmics, complexity theory. In the context of seperating complexity classes, it might facilitate application of well developed algebraic techniques.

A Natural Optimisation Question

How "close" is M to a rank r matrix N?

- How does one define "closeness"? Various options are norm of $M - N$, hamming weight of $M - N$.
- Representation of the "close" Matrix.
- Bounds, Complexity of computing them, Approximations.

Several practical applications:

- Low dimensional representation of large volumes of data.
- Netflix problem : DVD rental table is of "low" rank.
- Arises in feedback control system.

Under various norms...

- Frobenius Norm:

$$||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{i=1}^{m} |a_{i,j}|^2}$$

  Studied under the name *low-rank approximations*. Sampling based approximations are known [Des07].

- General Matrix Norms:

$$||A||_{\alpha,\beta} = \max_{||x||_\alpha = 1} ||Ax||_\beta$$

  Nothing better is known.

- Related : p-norms (subspace approximation) : For a k-dimensional linear subspace H such that for column vectors $v_1 \ldots v_n$ minimize:

$$\left(\sum_i d(v_i, H)^p \right)^{\frac{1}{p}}$$

Hamming is different

Definition (Rigidity)

Given a matrix M and $r \leq n$, rigidity of the matrix M ($R_M(r)$) is the number of entries of the matrix that we need to change to bring the rank below r.

[Val77] Interesting in a circuit complexity theory setting. If for some $\epsilon > 0$ there exists a $\delta > 0$ such that an $n \times n$ matrix M_n has rigidity $R_{M_n}(\epsilon n) \geq n^{1+\delta}$ over a field $\mathbb{F}$, then the transformation $x \rightarrow Mx$ cannot be computed by linear size logarithmic depth linear circuits.

[Raz89] For an explicit infinite sequence of (0,1)-matrices $\{M_n\}$ over a finite field $\mathbb{F}$, if $R_M(r) \geq \frac{n^2}{2^{(\log r)^{o(1)}}}$ for some $r \geq 2^{(\log \log n)^{\omega(1)}}$, then there is an explicit language $L_M \notin \mathrm{PH}^{cc}$, where PH^{cc} is the analog of PH in the communication complexity setting.

Lower bounds attempts

Find an explicit family of matrices $\{M_n\}$ such that

$$R_{M_n}(\epsilon n) \geq n^{1+\delta}$$

For any r, $\text{rank}(M) - r \leq R_M(r) \leq (n-r)^2$

Matrices	$\Omega(.)$	References
Vandermonde	$\frac{n^2}{r}$	Razborov '89, Pudlak '94 Shparlinsky '97, Lokam '99
Hadamard	$\frac{n^2}{r}$	Kashin-Razborov '98
Parity Check	$\frac{n^2}{r}\log(\frac{n}{r})$	Friedman '93 Pudlak-Rodl '94
$\sqrt{p_{ij}}$	n^2	Lokam '06

Computing Rigidity - How "natural" is Valiant's proof?

- Razborov-Rudich defined the concept of natural proofs for lower-bound proofs. They defined the notion of a combinatorial property being Γ-natural against a class Δ.
- Valiant's reduction [Val77] identifies "high rigidity" as a a combinatorial property of the matrices (which defines the function computed) based on which he proves linear size lower bounds for log-depth circuits. Among the $n \times n$ matrices, the density of "rigid" matrices is high.
- Two requirements:
  - The notion of natural proofs in arithmetic circuits?
  - Tighten the default parameters : polynomial factors.

Computing Rigidity

RIGID(M, r, k): Given a matrix M, values r and k, is $R_M(r) \leq k$?

Field $\mathbb{F}$	restriction	bound
$\mathbb{F}$	-	in NP
$\mathbb{F}_2$	-	NP -complete [Des07]
$\mathbb{Z}$ or $\mathbb{Q}$	Boolean, constant k	$C_=L$-complete
$\mathbb{Z}$ or $\mathbb{Q}$	constant k	$C_=L$-hard
$\mathbb{F}_p$	constant k	$\text{Mod}_p L$-complete
$\mathbb{Q}$	$r = n$	$C_=L$-complete witness-search in L^{GapL}
$\mathbb{Z}$	$r = n$ and $k = 1$	in L^{GapL}

For constant k, for 0-1 matrices, $\text{RIGID}_k \leq_m \text{SINGULAR}$

$\text{RIGID}(M, r, k)$: we need to test if if there is a set of $0 \leq s \leq k$ entries of M, which, when flipped, yield a matrix of rank below r.

The number of such sets is bounded by $\Sigma_{s=0}^{k} \binom{n}{s} = t \in n^{O(1)}$.

Let the corresponding matrices be $M_1, M_2 \ldots M_t$; these can be generated from M in logspace. Now,

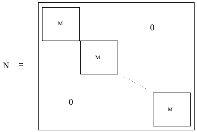
$$\begin{aligned}(M, r) \in \text{RIGID}(k) &\iff \exists i : (M_i, r) \in \text{RANK BOUND}(\mathbb{Z}) \\ &\iff \exists i : (N_i, r) \in \text{SINGULAR} \\ &\iff (N', r') \in \text{RANK BOUND}(\mathbb{Z})\end{aligned}$$

where N_is can be obtained in logspace from M_is and N' and r' can be generated using standard techniques.

For constant k, for 0-1 matrices, $\textsc{Singular} \leq_m \textsc{Rigid}_k$

For constant k, for 0-1 matrices, $\text{SINGULAR} \leq_m \text{RIGID}_k$

$\text{RIGID}(M, n, 0)$ tests if the matrix is singular.
To prove it for arbitrary k, tensor it with I_{k+1}, the rigidity gets amplified by a factor of k.



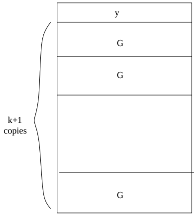
$$M \in \text{SINGULAR}(\mathbb{Z}) \implies (N, n(k+1)-k) \in \text{RIGID}(N, n(k+1)-k, 0) \subseteq \text{RIGID}(N, n(k+1)-k, k)$$

$$M \notin \text{SINGULAR}(\mathbb{Z}) \implies (N, n(k+1)-k) \notin \text{RIGID}(N, n(k+1)-k, k)$$

NP-hardness over $\mathbb{F}_2$

NEAREST CODEWORD PROBLEM(NCP): Fix a linear code $\mathcal{C} : \{0,1\}^m \to \{0,1\}^n$ over $\mathbb{F}_2$ with generator matrix $G_{m\times n}$, a received vector y, and distance k, check if there is a codeword x such that $\Delta(x,y) \leq k$.

NCP is NP-hard.
REDUCTION: Given $NCP(G,k,y)$ define M as



Claim : $R_M(m-1) \leq k \iff NCP(G,k,y)$.

Inapproximability results for RIGID

Theorem

Over $\mathbb{F}_2$, for any constant $\alpha > 1$, given a matrix $M \in \mathbb{F}_2^{m \times n}$ of rank r, deciding if $R_M(r-1) \leq k$ or $R_M(r-1) \geq \alpha k$ is NP*-hard.*

Theorem

Assuming NP *is not contained in* DTIME$(n^{\log n})$*, over $\mathbb{F}_2$, for any $\epsilon > 0$, for $\alpha \leq 2^{n \log^{0.5-\epsilon}}$, given a matrix $M \in \mathbb{F}_2^{m \times n}$, of rank r it is impossible to distinguish between the following two cases:*

1. $R_M(r-1) \leq k$.
2. $R_M(r-1) \geq \alpha k$.

The MinRank problem

Let $\mathbb{F}$ be a field. with $E, S \subseteq \mathbb{F}$.

MinRank: Given a matrix M with entries from $\mathbb{E} \cup \{x_1 \ldots x_k\}$,

$$(\mathsf{E}, \mathsf{S}) - \mathrm{minrank}_{\mathbb{F}}(M) = \min_{(\alpha_1, \ldots \alpha_k) \in S^k} rank_{\mathbb{F}}(M(\alpha_1, \ldots \alpha_k))$$

The MINRANK problem

Let $\mathbb{F}$ be a field. with $E, S \subseteq \mathbb{F}$.

MINRANK: Given a matrix M with entries from $\mathbb{E} \cup \{x_1 \ldots x_k\}$,

$$(\mathrm{E}, \mathrm{S}) - \mathrm{minrank}_{\mathbb{F}}(M) = \min_{(\alpha_1, \ldots \alpha_k) \in S^k} rank_{\mathbb{F}}(M(\alpha_1, \ldots \alpha_k))$$

rank computed over $\mathbb{F}$	Entries of M from E	Solution from S	Complexity
$\mathbb{Q}$	$\mathbb{Q}$	$\mathbb{Z}$	Undecidable
$\mathbb{F}_2$	$\mathbb{F}_2$	$\mathbb{F}_2$	NP-complete
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$	in NP
$\mathbb{R}$	$\mathbb{Q}$	$\mathbb{R}$	NP-hard, PSPACE

The MINRANK problem

Let $\mathbb{F}$ be a field. with $E, S \subseteq \mathbb{F}$.

MINRANK: Given a matrix M with entries from $\mathbb{E} \cup \{x_1 \dots x_k\}$,

$$(\mathsf{E},\mathsf{S}) - \mathsf{minrank}_{\mathbb{F}}(M) = \min_{(\alpha_1,\dots\alpha_k)\in S^k} rank_{\mathbb{F}}(M(\alpha_1,\dots\alpha_k))$$

$$\text{RIGID} \in \mathsf{NP}^{1-Minrank}$$

1-Minrank : Every variable occurs exactly once.

Guess the entries of the matrix to be changed and replace them with distinct variables.

The MinRank problem

Let $\mathbb{F}$ be a field. with $E, S \subseteq \mathbb{F}$.

MinRank: Given a matrix M with entries from $\mathbb{E} \cup \{x_1 \ldots x_k\}$,

$$(\mathsf{E}, \mathsf{S}) - \mathsf{minrank}_{\mathbb{F}}(M) = \min_{(\alpha_1, \ldots \alpha_k) \in S^k} rank_{\mathbb{F}}(M(\alpha_1, \ldots \alpha_k))$$

$$\text{RIGID} \in \mathsf{NP}^{1-Minrank}$$

1-Minrank : Every variable occurs exactly once.

Guess the entries of the matrix to be changed and replace them with distinct variables.
We don't know much for the 1-Minrank problem either.

Bounded Rigidity

Definition (Bounded Rigidity)

Given a matrix M and $r < n$, bounded rigidity of the matrix M ($R_M(b, r)$) is the number of entries of the matrix that we need to change to bring the rank below r, if the change allowed per entry is atmost b.

- B-RIGID(M, r, k, b): Given a matrix M, values b, r and k, is $R_M(b, r) \leq k$?
- Another formulation : Define an interval of matrices $[A]$ where

$$m_{ij} - b \leq a_{ij} \leq m_{ij} + b$$

Question : Is there a rank r matrix $B \in [A]$ such that $M - B$ has atmost k non-zero entries?

Why should there be?

Consider the matrix

$$\begin{bmatrix} 2^k & 0 & 0 & 0 & 0 \\ 0 & 2^k & 0 & 0 & 0 \\ 0 & 0 & 2^k & 0 & 0 \\ 0 & 0 & 0 & 2^k & 0 \\ 0 & 0 & 0 & 0 & 2^k \end{bmatrix}$$

- $R_M(b, n-1)$ is undefined unless $b \geq \frac{2^k}{n}$.
- Question : For a given matrix M, bound b, target rank r, can we efficiently test whether $R_M(b, r)$ is defined ?

Why should there be?

Consider the matrix

$$\begin{bmatrix} 2^k & 0 & 0 & 0 & 0 \\ 0 & 2^k & 0 & 0 & 0 \\ 0 & 0 & 2^k & 0 & 0 \\ 0 & 0 & 0 & 2^k & 0 \\ 0 & 0 & 0 & 0 & 2^k \end{bmatrix}$$

- $R_M(b, n-1)$ is undefined unless $b \geq \frac{2^k}{n}$.
- Question : For a given matrix M, bound b, target rank r, can we efficiently test whether $R_M(b, r)$ is defined ?

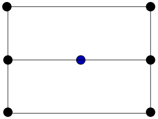
It is NP-hard.

NP-completeness for a restricted case

For a given matrix M, bound b, testing whether $R_M(b, n-1)$ is defined, is NP-complete.

MEMBERSHIP:

- The bound b defines an interval for each entry of the matrix.
- Determinant: a multlilinear polynomial in the entries of M.
- ZERO-ON-AN-EDGE LEMMA: For a multilinear polynomial $p(x_1, x_2 \dots x_t)$, consider the hypercube defined by the interval of each of the x_is. If there is a zero of the polynomial in the hypercube then there is a zero on an edge of the hypercube

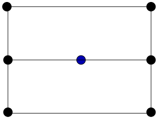


NP-completeness for a restricted case

For a given matrix M, bound b, testing whether $R_M(b, n-1)$ is defined, is NP-complete.

MEMBERSHIP:

- The bound b defines an interval for each entry of the matrix.
- Determinant: a multlilinear polynomial in the entries of M.
- ZERO-ON-AN-EDGE LEMMA: For a multilinear polynomial $p(x_1, x_2 \ldots x_t)$, consider the hypercube defined by the interval of each of the x_is. If there is a zero of the polynomial in the hypercube then there is a zero on an edge of the hypercube



- NP algorithm : Guess the edge of the hypercube where the zero occurs and verify if the signs of determinant at each end point are opposite.

NP-completeness for a restricted case

HARDNESS: The interval $[M - \theta J, M + \theta J]$ is singular if and only if $R_M(n, \theta)$ is defined.
By a reduction from MAXCUT problem, [PR93] showed that that checking interval singularity is NP-hard. Hence the hardness follows in our case too.

Increasing the Rank : MaxRank problem

MaxRank: Given a matrix M with entries from $\mathbb{F} \cup \{x_1 \ldots x_k\}$,

$$\text{MaxRank}(M) = \max_{(\alpha_1,\ldots\alpha_k)\in\mathbb{F}^k} rank(M(\alpha_1,\ldots\alpha_k))$$

Increasing the Rank : MaxRank problem

MaxRank: Given a matrix M with entries from $\mathbb{F} \cup \{x_1 \ldots x_k\}$,

$$\text{MaxRank}(M) = \max_{(\alpha_1, \ldots \alpha_k) \in \mathbb{F}^k} rank(M(\alpha_1, \ldots \alpha_k))$$

- Over $\mathbb{Z}$, there is a randomized polynomial time algorithm which can test if MaxRank is less than k.
- Over $\mathbb{F}_2$, the problem is NP-complete.

Increasing the Rank : MaxRank problem

MaxRank: Given a matrix M with entries from $\mathbb{F} \cup \{x_1 \ldots x_k\}$,

$$\text{MaxRank}(M) = \max_{(\alpha_1, \ldots \alpha_k) \in \mathbb{F}^k} rank(M(\alpha_1, \ldots \alpha_k))$$

- Over $\mathbb{Z}$, there is a randomized polynomial time algorithm which can test if MaxRank is less than k.
- Over $\mathbb{F}_2$, the problem is NP-complete.

$$\text{Max version of} \text{RIGID} \in \text{NP}^{1-Maxrank}$$

1-Maxrank : Every variable occurs exactly once.

Increasing the Rank : MaxRank problem

MaxRank: Given a matrix M with entries from $\mathbb{F} \cup \{x_1 \ldots x_k\}$,

$$\text{MaxRank}(M) = \max_{(\alpha_1, \ldots \alpha_k) \in \mathbb{F}^k} rank(M(\alpha_1, \ldots \alpha_k))$$

- Over $\mathbb{Z}$, there is a randomized polynomial time algorithm which can test if MaxRank is less than k.
- Over $\mathbb{F}_2$, the problem is NP-complete.

$$\text{Max version of} \textsc{Rigid} \in \mathsf{NP}^{1-Maxrank}$$

1-Maxrank : Every variable occurs exactly once.
MaxRank is in P (Geelen '93).

Open Problems

- A better upper bound for computing rigidity over $\mathbb{Q}$.
- Is there an efficient algorithm when r is a constant?
- An NP upper bound for bounded rigidity - a generalisation of the zero-on-an-edge lemma to arbitrary rank.

Thank You

Jonathan F. Buss, Gudmund Skovbjerg Frandsen, and Jeffrey Shallit.
The computational complexity of some problems of linear algebra (extended abstract).
In *Proc. STACS*, volume 1200 of *LNCS*, pages 451–462, 1997. JCSS 1999, 58 pp 572-596.

Amit Deshpande.
Sampling-based dimension reduction algorithms.
PhD thesis, MIT, May 2007.

S. Poljak and J. Rohn.
Checking robust nonsingularity is NP-hard.
Math. Control Signals Systems, 6:1–9, 1993.

A. A. Razborov.
On rigid matrices.
manuscript in russian, 1989.

L. G. Valiant.
Graph theoretic arguments in low-level complexity.

In *Proc. 6th MFCS*, volume 53 of *LNCS*, pages 162–176. Springer, Berlin, 1977.