A FAST AND EFFICIENT HEURISTIC NUCLEAR-NORM ALGORITHM FOR AFFINE RANK MINIMIZATION

Thong T. Do[†]*, Yi Chen*[†]*, Nam Nguyen*[†]*, Lu Gan*[‡] *and Trac D. Tran*[†] *

[†] Department of Electrical and Computer Engineering
The Johns Hopkins University
[‡]School of Engineering and Design
Brunel University, UK

ABSTRACT

The problem of affine rank minimization seeks to find the minimum rank matrix that satisfies a set of linear equality constraints. Generally, since affine rank minimization is NP-hard, a popular heuristic method is to minimize the nuclear norm that is a sum of singular values of the matrix variable [1]. A recent intriguing paper [2] shows that if the linear transform that defines the set of equality constraints is *nearly isometrically distributed* and the number of constraints is at least $\mathcal{O}(r(m+n)\log mn)$, where r and $m \times n$ are the rank and size of the minimum rank matrix, minimizing the nuclear norm yields exactly the minimum rank matrix solution. Unfortunately, it takes a large amount of computational complexity and memory buffering to solve the nuclear norm minimization problem with known nearly isometric transforms. This paper presents a fast and efficient algorithm for nuclear norm minimization that employs structurally random matrices [3] for its linear transform and a projected subgradient method that exploits the unique features of structurally random matrices to substantially speed up the optimization process. Theoretically, we show that nuclear norm minimization using structurally random linear constraints guarantees the minimum rank matrix solution if the number of linear constraints is at least $\mathcal{O}(r(m+n)\log^3 mn)$. Extensive simulations verify that structurally random transforms still retain optimal performance while their implementation complexity is just a fraction of that of completely random transforms, making them promising candidates for large scale applications.

Index Terms— Rank minimization, nuclear norm heuristic, compressed sensing, system identification, structurally random transforms, random matrices

1. INTRODUCTION

The affine rank minimization problem involves finding the minimum rank matrix $\mathbf{X}_r$ that satisfies a given system of linear equation constraints.

$$\mathbf{X}_r = \text{argmin rank}(\mathbf{X}) \qquad \text{s.t.} \qquad \mathcal{A}(\mathbf{X}) = \mathbf{b} \tag{1}$$

where $\mathbf{b}$ is an observation vector in $\mathbb{R}^M$, $\mathbf{X}$ is a matrix variable in $\mathbb{R}^{m\times n}$ and $\mathcal{A}$ is a linear mapping defining the linear equality constraints from $\mathbb{R}^{m\times n}$ to $\mathbb{R}^M$. When a set of feasible models is affine in the matrix variable, the above minimization is equivalent to finding the simplest model satisfying a given set of constraints. Many engineering problems can be formulated as affine rank minimization such as: minimal order system realization, reduced order controller design, low dimensional Euclidean embedding and interference with partial information, etc [4].

Recently, the authors of an inspiring paper [2] show the connection between compressed sensing [5] and affine rank minimization. In particular, they show that the minimum rank solution $\mathbf{X}_r$ can be exactly recovered by solving the minimization of the nuclear norm, a sum of singular values of the matrix, over the given affine space if the linear mapping is *nearly isometrically distributed* and the number of linear constraints p is at least on the order of $\mathcal{O}(r(m+n)\log mn)$, where r is the rank of $\mathbf{X}_r$:

$$\mathbf{X}_{\text{nuc}} = \text{argmin}\|X\|_* \quad s.t. \quad \mathcal{A}(\mathbf{X}) = \mathbf{b} \tag{2}$$

where $\|X\|_* = \sum_1^r \sigma_i(\mathbf{X})$ and $\sigma_i(\mathbf{X})$ are nonzero singular values of the matrix variable $\mathbf{X}$. A nearly isometric mapping $\mathcal{A}$ from $\mathbb{R}^{m\times n}$ to $\mathbb{R}^M$ is defined as the linear mapping that satisfies the following three conditions with all matrices $\mathbf{X} \in \mathbb{R}^{m\times n}$ and $\|\mathbf{X}\|_F^2 = 1$:

$$E\{\mathcal{A}(\|\mathbf{X}\|)^2\} = 1, \tag{3}$$

and for all $0 < \epsilon < 1$

$$P(|\|\mathcal{A}(\mathbf{X})\|^2 - 1| \geq \epsilon) \leq 2\exp(\frac{-M}{2}(\epsilon^2/2 - \epsilon^3/3)), \tag{4}$$

and for all $t > 0$ and some positive constant c

$$P(\|\mathcal{A}\| \geq 1 + \sqrt{\frac{mn}{M}} + t) \leq \exp(-cMt^2). \tag{5}$$

A typical example of a nearly isometric mapping is a random matrix of Gaussian or Bernoulli i.i.d entries. Because the nuclear norm is a convex function, its minimization can be achieved tractably via several popular algorithms, such as semidefinite programming, projected subgradient method, or low-rank parametrization [2].

Structurally Random Matrices (SRM) were first proposed [3] as a fast and highly efficient ensemble for compressed sensing. A structurally random matrix $\mathbf{A}$ (using the local randomizer) is a product of three matrices:

$$\mathbf{A} = \sqrt{\frac{d}{M}}\mathbf{DFR} \tag{6}$$

where

- $\mathbf{R}$, the local randomizer, is a $d \times d$ random diagonal matrix whose diagonal entries R_{ii} are i.i.d Bernoulli random variables $\text{P}(R_{ii} = \pm 1) = \frac{1}{2}$.

*This work has been supported in part by the National Science Foundation under Grant CCF-0728893.

- $\mathbf{F}$ is a $d \times d$ orthonormal matrix whose absolute magnitude of all entries are on the order of $\mathcal{O}(\frac{1}{\sqrt{d}})$. In practice, only $\mathbf{F}$ with fast computation and efficient implementation such as the (normalized) FFT, the DCT, the (normalized) WHT... are chosen. Finally,
- $\mathbf{D}$, a uniformly random downsampler, is a matrix composed of nonzero rows of a random diagonal matrix whose diagonal entries D_{ii} are i.i.d binary random variables with $\mathrm{P}(D_{ii} = 1) = \frac{M}{d}$. On average, $\mathbf{D}$ contains M nonzero rows and thus, $\mathbf{\Phi}$ is a $M \times d$ matrix.

Algorithmically, the projection can be acquired efficiently as the following: *(i)* pre-randomizing a target signal by randomly flipping sign of its entries*(ii)* applying a fast transform to the randomized signal and *(iii)* finally, randomly keeping M of those transformed coefficients.

In this paper, we propose a fast and efficient algorithm to solve the nuclear norm minimization problem, employing SRM as an alternative linear transform and a projected subgradient method that efficiently takes advantage of the highly structural property of the linear transform. The proposed algorithm has computational complexity and memory requirement only on the order of $\mathcal{O}(mn \log mn)$. The projected subgradient method exploits two unique features of SRM. First, both SRM and its adjoint can be easily implemented as serial operators without the need of explicit storing of the linear transform. Second, with a structurally random matrix $\mathbf{A}$, it can be shown that $\mathbf{AA^T} = \frac{\mathbf{d}}{\mathbf{M}}\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, simplifying substantially the orthogonal projection step of the solution approximation to the constraints subspace.

2. ALGORITHM DESCRIPTION

First, note that the notation $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ is equivalent to $\mathbf{A}\mathrm{vec}(\mathbf{X}) = \mathbf{b}$, where $\mathbf{A}$ is the matrix representation of the linear map $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ and $\mathrm{vec}(\mathbf{X})$ denotes a vectorized version of the matrix $\mathbf{X}$.

We use a projected subgradient method [2] to solve the linearly-constrained nuclear norm minimization problem (2). This method computes a sequence of feasible points $\{\mathbf{X}_k\}$ by iteratively updating $\mathbf{X}_k$ by

$$\mathbf{X}_{k+1} = \mathbf{X}_k - s_k \Pi_{\mathcal{N}(\mathcal{A})} \mathbf{Y}_k, \quad \mathbf{Y}_k \in \partial \left\|\mathbf{X}_k\right\|_*, \tag{7}$$

where $s_k > 0$ is the stepsize and $\Pi_{\mathcal{N}(\mathcal{A})}$ denotes the orthogonal projection onto the kernel of $\mathcal{A}$. The set $\partial \left\|\mathbf{X}_k\right\|_*$ is the sub-differential of the nuclear norm at $\mathbf{X}_k$ given by

$$\partial \left\|\mathbf{X}_k\right\|_* = \left\{\mathbf{U}\mathbf{V}^T + \mathbf{W} : \mathbf{X}\mathbf{W}^T = \mathbf{W}^T\mathbf{X} = \mathbf{0} \text{ and } \|\mathbf{W}\| \le 1\right\},$$

where $\mathbf{U} \in \mathbb{R}^{m\times r}$, $\mathbf{V} \in \mathbb{R}^{n\times r}$, and $\mathbf{X}_k = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ depicts the singular value decomposition of $\mathbf{X}_k$.

For general choices of the linear map $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^M$, this method is quite computationally intensive for large-scale problems as we have to solve a least-square problem in each iteration to find the orthogonal projection. Now, we will show how to simplify this algorithm by exploiting the structure of the SRM $\mathbf{A}$.

Proposition 2.1. *If $\mathbf{A}$ is a SRM (using the local randomizer or the global randomizer), $\mathbf{AA}^T = \frac{d}{M}\mathbf{I}$.*

Proof. It is easy to verify that $\mathbf{RR}^T = \mathbf{I}$, $\mathbf{FF}^T = \mathbf{I}$, and $\mathbf{DD}^T = \mathbf{I}$. Therefore, we have $\mathbf{AA}^T = \frac{d}{M}\mathbf{I}$. □

This property of $\mathbf{A}$ greatly simplifies the projected subgradient method as follows. First we show how to choose an initial point $\mathbf{X}_0$ that satisfies the linear constraint $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. We can initialize it to be $\mathbf{A}^+\mathbf{b}$, where $\mathbf{A}^+ = \mathbf{A}^T$ is the Moore-Penrose pseudoinverse of $\mathbf{A}$. Alternatively, we also can choose a random point $\mathbf{X}_0$ and then project it to the feasible region. Due to the structure of $\mathbf{A}$, the projection can be done as follows.

$$\Pi_{\{\mathbf{X}:\mathcal{A}(\mathbf{X})=\mathbf{b}\}}\mathbf{X}_0 = \mathbf{X}_0 - \frac{M}{d}\mathcal{A}_t\big(\mathcal{A}\left(\mathbf{X}_0\right) - \mathbf{b}\big)$$

where $\Pi_{\{\mathbf{X}:\mathcal{A}(\mathbf{X})=\mathbf{b}\}}$ denotes the orthogonal projection onto the affine subspace $\{\mathbf{X} : \mathcal{A}(\mathbf{X}) = \mathbf{b}\}$.

Similarly, the orthogonal projection in (7) can be simplified as follows

$$\Pi_{\mathcal{N}(\mathcal{A})}\mathbf{Y}_k = \mathbf{Y}_k - \frac{M}{d}\mathcal{A}_t\left(\mathcal{A}\left(\mathbf{Y}_k\right)\right).$$

Since both structurally random operator $\mathcal{A}$ and its adjoint $\mathcal{A}_t$ have fast and efficient implementation, the update step can be performed efficiently.

There are various choices for the stepsize s_k. If the nuclear norm of the optimal solution $\mathbf{X}_r$ is known or can be estimated, then we can select the optimal stepsize [6]

$$s_k = \frac{\left\|\mathbf{X}_k\right\|_* - \left\|\mathbf{X}_r\right\|_*}{\left\|\Pi_{\mathcal{N}(\mathcal{A})}\mathbf{Y}_k\right\|_F^2}.$$

Otherwise, the stepsize sequence $\{s_k\}$ can be chosen as a non-summable diminishing sequence.

The nuclear norms of the sequence $\{\mathbf{X}_k\}$ is usually not monotonically decreasing. We may terminate the algorithm when the relative change in the objective function is less than a prescribed threshold, or the singular values of the current matrix achieve satisfactory values.

The proposed method of projected subgradient using SRM is summarized below.

Step 1 Select a feasible initial point $\mathbf{X}_0$ as described above.

Step 2 At the kth iteration, update the current solution by

$$\mathbf{X}_{k+1} = \mathbf{X}_k - s_k\left(\mathbf{Y}_k - \frac{M}{d}\mathcal{A}_t\left(\mathcal{A}\left(\mathbf{Y}_k\right)\right)\right), \quad \mathbf{Y}_k \in \partial\left\|\mathbf{X}_k\right\|_*.$$

Step 3 If the stopping criterion is satisfied, then output the current matrix as the solution and stop. Otherwise, go to Step 2.

3. THEORETICAL ANALYSIS

We now claim the main theoretical result of this paper:

Theorem 3.1. *If $\mathcal{A}$ is a structurally random operator (using the local randomizer), solving the nuclear norm minimization (2) guarantees to yield the minimum rank matrix solution (1) , i.e. $\mathbf{X}_{nuc} = \mathbf{X}_r$ if the number of linear equality constraints M is at least $\mathcal{O}(r(m+n)\log^3 mn)$, where r and $m \times n$ are the rank and size of the minimum rank matrix $\mathbf{X}_r$.*

Compared with a nearly isometric mapping (e.g., a random matrix of Gaussian or Bernoulli i.i.d entries), structurally random operators require a slightly larger number of linear constraints that is scaled up by a factor of $\log^2 mn$. However, as one can clearly see in our followed-up numerical experiments, the performance of SRM is completely comparable to, if not slightly better than, that of a nearly isometric mapping.

Proof. Although we could not prove that SRM is nearly isometrically distributed, we will show that it satisfies the inequality (3) and (5) and a modified version of (4).

First, denote $\mathbf{y} = \mathbf{FRx}$, where $\mathbf{x} = \text{vec}(\mathbf{X})$ and without loss of generality, assume that $\|\mathbf{X}\|_F^2 = \|\mathbf{x}\|^2 = 1$. Due to the orthonormality property of $\mathbf{F}$ and $\mathbf{R}$, $\|\mathbf{y}\|^2 = 1$. Note that $\mathbf{Ax} = \mathbf{Dy}$ can be re-written as a sum of independent random variables as the following:

$$\|\mathbf{Ax}\|^2 = \frac{d}{M}\|\mathbf{Dy}\|^2 = \frac{d}{M}\sum_{i=1}^{d}\rho_i y_i^2$$

where ρ_i are i.i.d binary random variables $P(\rho_i = 1) = \frac{M}{d}$ and $E\{\rho_i\} = \frac{M}{d}$. Thus,

$$E\{\|\mathbf{Ax}\|^2\} = \sum_{1\le i\le d} y_i^2 = \|\mathbf{y}\|^2 = 1$$

which implies (3). In addition, it is easy to verify (5) as the spectral norm of $\mathbf{A} = \sqrt{\frac{d}{M}}$.

The next proposition states that $\mathbf{A}$ has a property roughly similar to (4):

Proposition 3.1. *Let $\mathcal{A}$ be a structurally random operator from $\mathbb{R}^{m\times n}$ to $\mathbb{R}^M$. Assume that $\|\mathbf{X}\|_F^2 = \|vec(\mathbf{X})\|^2 = 1$. Denote $\mathbf{y} = \mathbf{FR}vec(\mathbf{X})$ and $K = \max_{i\le i\le d} d|\mathbf{y}_i|^2$, where $d = mn$. Then, for all $0 < \epsilon < 1$ and for some positive constant c,*

$$P(|\|\mathcal{A}(\mathbf{X})\|^2 - 1| \ge \epsilon) \le 2\exp(\frac{-Mc\epsilon^2}{K^2}) \tag{8}$$

Moreover, the value of K can be bounded as shown in the following proposition:

Proposition 3.2. *With a vector $\mathbf{x} \in \mathbb{R}^d$ and $\|\mathbf{x}\| = 1$, denote $\mathbf{y} = \mathbf{FRw}$. Let c be a positive constant such that $\max_{1\le i,j\le d}|F_{ij}| = \sqrt{\frac{c}{d}}$. Then,*

$$P\{\max_{1\le i\le d}|\mathbf{y}_i| \ge \sqrt{\frac{2c\log(2d/\alpha)}{d}}\} \le \alpha. \tag{9}$$

Theorem 3.1 can be easily derived from the following lemma and Theorem 3.3 in [2] which asserts that $\mathbf{X}_{\text{nuc}} = \mathbf{X}_r$ if the restricted isometry constant $\delta_{5r} \le 0.1$, where the restricted isometry constant $\delta_r(\mathcal{A})$ is defined as the smallest number such that

$$(1-\delta_r) \le \|\mathcal{A}(\mathbf{X})\| \le (1+\delta_r) \tag{10}$$

holds for all matrices $\mathbf{X}$ of at most rank r and $\|\mathbf{X}\|_F^2 = 1$.

Lemma 3.1. *Assume that $\mathcal{A}$ is a SRM. For a fixed number δ, $0 \le \delta \le 1$, with probability at least $1 - \frac{1}{mn}$, the restricted isometry constant $\delta_r(\mathcal{A}) \le \delta$ if $M \ge c_0 r(m+n)\log^3 mn$, where c_0 depends only on δ.*

Proof. Define probabilistic events $\mathcal{Q}_{\mathbf{X}} = \{|\|\mathcal{A}(\mathbf{X})\|^2 - 1| \ge \frac{\delta}{2}\}$ and $\mathcal{K} = \{K \le 2c\log 2d/\alpha\}$, where c and α are constants in (9). Following the same arguments of the proof of Theorem 4.2 in [2], we have:

$$P(\delta_r(\mathcal{A}) \ge \delta) \le P(\bigcup_{\mathbf{X}}\mathcal{Q}_{\mathbf{X}}) + P(\|\mathcal{A}\| \ge \frac{\delta}{2\epsilon} - 1) \tag{11}$$

Conditioning on the event $\mathcal{K}$,

$$P(\delta_r(\mathcal{A}) \ge \delta|\mathcal{K}) \le P(\bigcup_{\mathbf{X}}\mathcal{Q}_{\mathbf{X}}|\mathcal{K}) + P(\|\mathcal{A}\| \ge \frac{\delta}{2\epsilon} - 1|\mathcal{K}) \tag{12}$$

and thus,

$$P(\delta_r(\mathcal{A}) \ge \delta) \le P(\bigcup_{\mathbf{X}}\mathcal{Q}_{\mathbf{X}}|\mathcal{K}) + P(\overline{\mathcal{K}}) + P(\|\mathcal{A}\| \ge \frac{\delta}{2\epsilon} - 1|\mathcal{K}) \tag{13}$$

On the right-hand side of (13), the last term can be made to be zero if we choose $\epsilon < \frac{\delta}{4}(\sqrt{mn/M}+1)^{-1}$ because $\|\mathcal{A}\| = \sqrt{mn/M}$ regardless of the event $\mathcal{K}$. The second term can be shown to be less than $1/2mn$ if we choose $\alpha = 1/2mn$ by Proposition (3.2). We fix these values of α and ϵ. Due to the lemma 4.3 and the lemma 4.5 of computing the covering number in [2] and the proposition 3.1 above:

$$P(\delta_r(\mathcal{A}) \ge \delta) \le 2(\frac{2c_0}{\epsilon})^{r(m+n-2r)}(\frac{24}{\delta})\exp(\frac{-Mc\delta^2}{K}) + \frac{1}{2mn}$$

where c_0 is some contant. With the choice of ϵ and α above and $K \le 2c\log 2d/\alpha = 2c\log 4(mn)^2$,

$$P(\delta_r(\mathcal{A}) \ge \delta) \le 2(\frac{2c_0}{\epsilon})^{r(m+n-2r)}(\frac{24}{\delta})^{r^2}\exp(\frac{-Mc\delta^2}{4c^2\log^2 4(mn)^2}) + \frac{1}{2mn}$$

Finally, we can make the first term in the right-hand side the above inequality less than $1/2mn$ if $M \ge \mathcal{O}(r(m+n)\log^3 mn)$ and thus, derive the lemma 3.1. □

Detailed proofs of Lemma 3.1 and Propositions 3.1, 3.2 will be provided in the journal version of this paper due to space limitation. □

4. NUMERICAL RESULTS

We compare the performance and computational time of the projected subgradient algorithm described in Section 2 between a completely random matrix and a SRM for the linear transform. For simplicity, we assume the nuclear norm of the original signal $\mathbf{X}_r$ is known and use that value for the optimal step size in the projected subgradient algorithm.

Experiment 1: We adopt the MIT logo image [2] that has size of 46×81 ($d = 3726$) and rank $r = 5$ as the input signal $\mathbf{X}_r$. We sample it using a Gaussian i.i.d measurement matrix and a SRM with various number of measurements $M = \{700, 750, 800, \ldots, 1500\}$. The projected subgradient algorithm is used to find a solution of the nuclear norm minimization $\mathbf{X}_{\text{nuc}}$. Fig. 1 depicts the performance curves of these two measurement matrices. The numerical values on the x-axis represent the number of linear constraints (or measurements) M while those on the y-axis represent the Signal to Noise Ratio (SNR in dB) between $\mathbf{X}_r$ and $\mathbf{X}_{\text{nuc}}$. Visually reconstructed MIT logo images from 1100 measurements of the i.i.d Gaussian measurement matrix and the SRM are also shown in Figs. 3(a) and (b), respectively. As the reader can observe, both performance curve and visually reconstructed image using the SRM method are slightly better than those from the i.i.d Gaussian method.

Experiment 2: $n \times n$ matrices of rank r are generated by choosing Gaussian random matrices $\mathbf{U}$, $\mathbf{V}$ of size $n \times r$ and setting $\mathbf{X_r} = \mathbf{UV}^T$, where the rank $r = 5$ is fixed and the dimension $n \in \{40, 50, 60, 70\}$. For each matrix $\mathbf{X}_r$, we take $M = n^2/2$ measurements using i.i.d Gaussian measurement matrix and SRM. The projected subgradient algorithm is used to exactly recover $\mathbf{X}_r$ by solving the nuclear norm minimization problem. The algorithm iterates until the SNR between the approximation and the original signal is greater than or equal to 40 dB that is regarded as exact recovery. For i.i.d Gaussian matrix, the projection matrices are

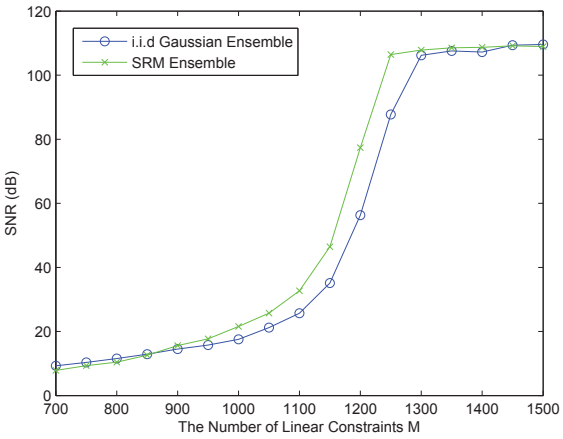


Fig. 1. Performance curves of i.i.d Gaussian and SRM measurement matrices: SNR vs. the number of measurements.

pre-computed and stored. The time for these computations (which is significant) is excluded from the running time in our comparison. Fig. 2 illustrates the amount of time required for exact recovery vs. the dimension n of low-rank matrices. Each point is obtained by averaging the running time over 10 different $\mathbf{X}_r$ of the same size. In this experiment, the amount of time required for exact recovery using SRM slightly decreases when the dimension n increases. This is because when the number of measurements $M = n^2/2$ increases and the rank $r = 5$ is kept fixed, the algorithm converges faster although there is a small increase of computational complexity at each iteration. One can clearly see a substantial improvement of speed when using SRM over using a completely random matrix.

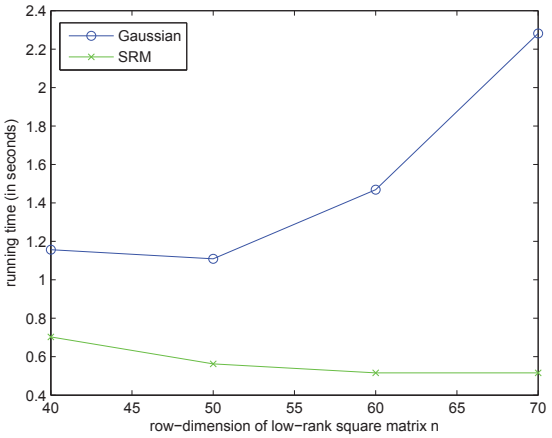
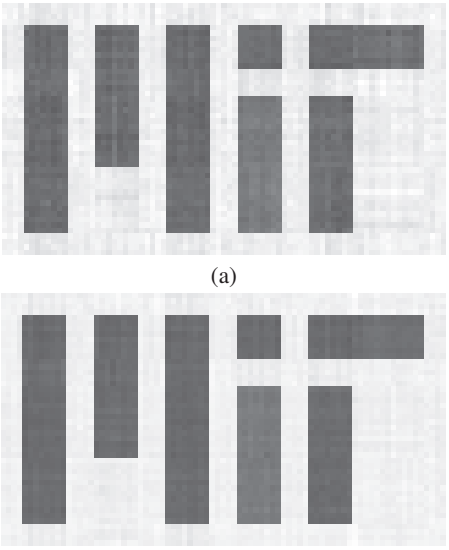


Fig. 2. Complexity comparison of the reconstruction algorithm using i.i.d Gaussian and SRM measurement matrices: running time vs. a row-dimension of a low-rank square matrix.

5. CONCLUSIONS

This paper presents a fast and efficient algorithm of nuclear norm minimization for finding the minimum rank matrix with complexity on the order of $\mathcal{O}(mn \log mn)$, where $m \times n$ is the size of the minimum rank matrix. It is based on the SRM ensemble and a projected subgradient method that specifically exploits the efficient features of SRM. The algorithm theoretically guarantees to provide the minimum rank solution if the number of linear constraints or measurements is at least $\mathcal{O}(r(m+n)\log^3 mn)$, where r is the rank of the minimum rank matrix. Simulation results verify that its performance is comparably optimal while its computational complexity is substantially lower comparing to that of a completely random measurement matrix, making it a very promising candidate for large scale applications.



(a)

(b)

Fig. 3. Reconstructed MIT logo from 1100 measurements using (a) the i.i.d Gaussian matrix and (b) the SRM.

6. REFERENCES

[1] M. Fazel, H. Hindi, and S.P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," *In Proceedings of American Control Conference*, vol. 6, pp. 4734–4739, 2001.

[2] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization," *Submitted to SIAM*, 2007.

[3] T. T. Do, T. D. Tran, and L. Gan, "Fast compressive sampling with structurally random matrices," *Proceedings of Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pp. 3369–3372, May 2008.

[4] M. Fazel, "Matrix rank minimization with applications," *Phd thesis, Stanford University*, 2002.

[5] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, pp. 489 – 509, Feb. 2006.

[6] B. Polyak, *Introduction to Optimization*, Optimization Software, Inc., 1987.