Fast SVD Free Low-rank Matrix Recovery: Application to Dynamic MRI Reconstruction

[1]*Angshul Majumdar and* [2]*Rabab Ward*

[1]Indraprastha Institute of Information Technology – Delhi
[2]Department of Electrical and Computer Engineering, University of British Columbia

ABSTRACT

This study proposes a new algorithm for recovering low-rank matrices from their under-sampled projections. Such algorithms are traditionally based on the Schatten-p *(*0<p≤1*)* norm minimization. The minimization problem is solved directly, requiring the computing of a singular value decomposition (SVD) at each iteration. This is time consuming and greatly limits the speed of the algorithms and its applicability to real life problems. To overcome this problem, we replace the Schatten-p norm by its equivalent Ky-Fan norm. For minimizing the said norm, we derive an algorithm that does not require computing SVD's. Instead, it computes a Cholesky decomposition – which requires many less computations than SVD. Our method yields an order of magnitude improvement in speed over existing techniques. We apply our proposed algorithm on the dynamic MRI reconstruction problem and obtain significant improvement in computational speed over the existing technique.

Index Terms— Matrix completion, dynamic MRI.

1. INTRODUCTION

The problem of recovering a low-rank matrix from its under-sampled projections arises in many areas in signal processing and machine learning. It is a typical inverse problem:

$$y_{m\times 1} = A(X_{n\times n}) + \eta_{m\times 1}, A: \square^{n\times n} \rightarrow \square^{m} \quad (1)$$

where y is the measurement, X is the signal to be reconstructed – assumed to be of low-rank and A is the mapping from a higher dimensional space of $\square^{n\times n}$ to a lower dimensional space of $\square^{m}$.

The conventional approach to solve (1) uses factorization; the matrix *X* is representedas *X=UV* where *U* and *V* are low-rank matrices. The problem is solved via:

$$\min_{U,V} \left\| y - A(UV) \right\|_2^2 + \lambda \left(\left\| U \right\|_2 + \left\| V \right\|_2 \right) \quad (2)$$

Unfortunately, this is a bilinear problem and naturally non-convex. This is usually solved via the method of alternating least squares. At every iteration, *U* is solved assuming *V* is fixed and vice versa. This technique is simple to implement but is not optimal. Also one needs to know the rank of *X*, which is unknown for most practical problems. But the main issue with this approach is that the problem is non-convex, thus there are no convergence guarantees. It is often observed that the objective function decreases initially but keeps on increasing after a while.

Ideally, searching for the minimum rank solution requires solving the following optimization problem:

$$\min_X rank(X) \text{ subject to } \left\| y - A(X) \right\|_2^2 \leq \varepsilon \quad (3)$$

where the noise parameter ε is dependent on the variance of η and is assumed to be known.

Solving this rank minimization problem is NP hard and henceforth not practical. Theoretical studies in applied mathematics [1, 2] have shown that it is possible to guarantee the recovery of the exact solution by relaxing the NP hard rank minimization by its closest convex surrogate the nuclear norm minimization:

$$\min_X \left\| X \right\|_* \text{ subject to } \left\| y - A(X) \right\|_2^2 \leq \varepsilon \quad (4)$$

where the nuclear norm $\left\| X \right\|_*$ is defined as the sum of the singular values of X. This is a convex problem that can be solved by semi-definite programming. However more efficient algorithms exist.

Taking cues from non-convex compressed sensing, it has been further observed that instead of employing the convex nuclear norm, better recovery can be achieved using the non-convex Schatten-p norm (*0<p<1*) [3-5]:

$$\min_X \left\| X \right\|_{Sp} \text{ subject to } \left\| y - A(X) \right\|_2^2 \leq \varepsilon \quad (5)$$

The Schatten-p norm $\left\| X \right\|_{Sp}$ is defined as the l_p-norm of the singular values of X.

The price to be paid for applying the theoretically sound solution (4)-(5) instead of the factorization based approach (2) is the speed. Schatten-p norm minimization problems however require computing the singular value decomposition (SVD) of the matrix at every iteration; this is computationally complex and is the main source of delay. Our aim is to solve the same Schatten-p norm minimization problem, but without computing the SVD at each iteration. We propose to replace the Schatten-p norm by its equivalent Ky-Fan norm and we derive an algorithm for its minimization**.** The Ky-Fan norm minimization does not require computing SVDs. It relies on the Cholesky Decomposition, which is much cheaper to compute. This naturally speeds up the algorithm.

Low rank matrix recovery has several applications. An important problem is dynamic MRI reconstruction. Casorati matrix formed by stacking the dynamic MRI frames as columns is low-rank owing to the temporal correlation between frames. The problem is to recover this low-rank matrix given the under-sampled Fourier frequency data of each frame. We apply the proposed matrix recovery algorithm for the said problem.

In the paper, the matrix recovery algorithm is proposed in section 2. Its application to the dynamic MRI application is described in section 3. The results are examined in section 4 and the conclusions are presented in section 5.

2. MATRIX RECOVERY ALGORITHM

In a low-rank matrix recovery problem, we are required to solve (5). However solving the constrained problem directly is difficult. Therefore we propose to solve the following problem instead:

$$\min_X \frac{1}{2}\|Y - A(X)\|_2^2 + \lambda \|X\|_{Sp}^p \tag{6}$$

The problems (5) and (6) are the same for the proper choice of λ and ε. In this work, we assume λ is known.

Since *A(.)* is a linear function, it is possible to write (6) in the following matrix-vector notation:

$$\min_X \frac{1}{2}\|y - Ax\|_2^2 + \lambda \|X\|_{Sp}^p \tag{7}$$

Here *x=vec(X)* and *y=vec(Y)*.

Following the Majorization-Minimization approach [6, 7], the above form (7) can be expressed at each iteration (k) in the following fashion:

$$\min_X \frac{1}{2}\|b - x\|_2^2 + \frac{\lambda}{\alpha}\|X\|_{Sp}^p \quad \text{where } b = x^{(k)} + \frac{1}{\alpha}A^T(y - Ax^{(k)}) \tag{8}$$

Here α is slightly higher than the maximum eigenvalue of A^TA in order to ensure convergence.

Reshaping the vectors *b* and *x* to their matrix forms *B* and *X* respectively, (8) is expressed as:

$$\min_X \frac{1}{2}\|B - X\|_2^2 + \frac{\lambda}{\alpha}\|X\|_{Sp}^p \tag{9}$$

There exists a direct solution to this problem [4]; however this requires computing the singular value decomposition (SVD) of *B* at each iteration. Computing SVD's is time consuming and is the main computational bottleneck for all low-rank matrix recovery algorithms. The computational complexity for SVD is $2mn^2+11n^3$ [8], where m and n are the dimensions of the matrix. Almost all efficient algorithms for SVD based low-rank matrix recovery problems exploit the fact that the matrix to be recovered is low rank and hence computing the full SVD is not necessary, computing the partial SVD using LAPACK routines slightly improves the computational cost. However, in order to compute the partial SVD, one needs to specify the number of singular values to be computed, i.e. one needs to have an estimate of the rank of the matrix to be recovered. This rank estimation is a heuristic step that either yields poor reconstruction results like Singular Value Thresholding (SVT) [9] or is markedly slow like Fixed Point Continuation [10].

In this work we aim to speed up low-rank matrix recovery techniques by about an order of magnitude compared to previous SVD based methods. To reach that goal, we replace the Schatten-p norm in (9) by its equivalent Ky-Fan norm [3],

$$\min_X g(X) : g(X) = \frac{1}{2}\|B - X\|_2^2 + \frac{\lambda}{\alpha} Tr(X^TX)^{\frac{p}{2}-1} \tag{10}$$

Taking the derivative of *g(X)* and equating it to zero (the intermediate steps are easy to compute; it just requires applying the chain rule for multivariate calculus):

$$X - B + \frac{\lambda}{\alpha} p(X^TX)^{\frac{p}{2}-1} X = 0 \tag{11}$$

Re-arranging (11) as:

$$\left(I + \frac{\lambda}{\alpha} p(X^TX)^{\frac{p}{2}-1}\right) X = B \tag{12}$$

the term inside the brackets $I + \frac{\lambda}{\alpha} p(X^TX)^{\frac{p}{2}-1}$ is a positive definite matrix which is easy to invert using the Cholesky decomposition: $I + \frac{\lambda}{\alpha} p(X^TX)^{\frac{p}{2}-1} = R^TR$. Since the algorithm is iterative, the term X^TX is computed from the previous iteration. Using the Cholesky decomposition, *X* becomes easy to solve from (7) in two steps:

$$R^T \underbrace{RX}_{Z} = B$$
$$\text{step 1. } Solve: R^TZ = B \tag{13}$$
$$\text{step 2. } Solve: RX = Z$$

Since *R* is a triangular matrix, *X* can be solved very fast.

The Cholesky decomposition is a computationally demanding operation in our algorithm. However the cost of computing the Cholesky decomposition is only $1/3n^3$ [8]; this is significantly less compared to the cost of computing an SVD. Since $I + \frac{\lambda}{\alpha} p(X^TX)^{\frac{p}{2}-1}$ is always full rank, there is no need for rank estimation heuristics; even a full Cholesky decomposition is much faster than a partial or complete SVD. To give an example of the increase in speed, we provide some numerical results computed in Matlab.

For a given matrix size, a symmetric positive definite matrix is generated. Its SVD is computed using the default in-built Matlab routine. The partial SVD's are computed using Lansvd [11]. For the partial SVD case, 'k%' means that SVD corresponds to the top k% of the top singular values. The Cholesky decomposition is computed using the default Matlab in-built routine. For

each matrix size, 100 such matrices were generated and the different decompositions were carried out on each of them. The average reconstruction times are reported in Table 1.

Table 1. Decomposition Time in Seconds

MATRIX Size	SVD (Full)	Lansvd (1%)	Lansvd (10%)	Cholesky
1000X1000	0.53	0.25	1.31	0.02
2500X2500	9.83	2.00	17.82	0.25
5000X5000	77.92	18.01	138.47	1.59

It is easy to note that Cholesky decomposition is at least more an order of magnitude faster than SVD. Note that, these decompositions need to be computed AT EVERY ITERATION in their corresponding algorithms.

In a succinct fashion, the pseudo-code for our proposed algorithm is expressed as follows:

Initialize: $X = \min_X \|Y - A(X)\|_{Fro}^2$

In each iteration (k):

Cholesky Decomposition - $I + \frac{\lambda}{\alpha} p(X^T X)^{\frac{p}{2}-1} = R^T R$

Solve X - *step* 1. $R^T Z = B$

step 2. $RX = Z$

The complexity for solving X is only $2n^2$ (nxn being the size of X) by backward and forward substitutions. The complexity for computing $I + \frac{\lambda}{\alpha} p(X^T X)^{\frac{p}{2}-1}$ is governed by the matrix-matrix product which requires n^3 flops. We already mentioned that the complexity of the Cholesky decomposition is $1/3n^3$. Thus the total complexity of each iteration of our algorithm is $2n^2+4/3n^3$, which is considerably less than the previous SVD based recovery techniques. A low-rank matrix recovery algorithm such as [3] requires an SVD, a shrinkage and a matrix-matrix product in every iteration. Therefore the total cost is $13n^3$ (for SVD assuming a square matrix) + n^3 (matrix-matrix product); we can ignore the shrinkage operation as it is of linear complexity. Other algorithms [9, 10] have similar costs per iterations. Thus it is easy to note that the computational cost per iteration for our algorithm is much less when compared to competing algorithms.

3. DYNAMIC MRI RECONSTRUCTION

In Magnetic Resonance Imaging (MRI), the data acquisition model is expressed as:

$$b = Fx + \eta \quad (14)$$

Here x is the underlying image, b is the acquired K-space (Fourier frequency) data and F is the Fourier mapping from the spatial domain to the K-space. The noise η is assumed Gaussian.

This is a classic inverse problem. When the K-space is fully sampled on a uniform Cartesian grid, the reconstruction / inversion is trivial (requires applying an inverse FFT on *b)*. However, such uniform Cartesian sampling is time consuming. In order to accelerate the scans, the K-space is partially sampled; but this makes the inverse problem (9) under-determined and the reconstruction challenging.

In dynamic MRI, the data is acquired for each time frame. Thus the data acquisition is expressed as:

$$b_t = F_t x_t + \eta, \text{ t=1...T} \quad (15)$$

Considering all the frames, this can be expressed succinctly in the following form:

$$vec(B) = F\ vec(X) + \eta \quad (16)$$

Where B is a Casorati matrix formed by stacking the b_t's as columns, X is a Casorati matrix formed by stacking the x_t's as columns and F is a block diagonal matrix with corresponding F_t's.

In [12, 13] it is argued that the matrix X is rank-deficient. This is because, the MRI frames are temporally correlated and hence the columns of X are not independent. This rank deficiency can be exploited in order to recover X from (16). In [12, 13] a matrix factorization based technique [14] was used for recovering the low-rank solution. However in principle other low-rank matrix recovery techniques like Schatten-p norm minimization can be employed as well.

4. EXPERIMENTAL EVALUATION

4.1. Experiments on Synthetic Data

To test the performance of our algorithm, we compared it with the singular value shrinkage (SVS) technique [4]. We chose SVS because it has been shown to outperform other methods like SVT and FPC both in terms of speed and accuracy [4]. Besides, both the proposed algorithm and SVS solve the same Schatten-p norm minimization problem while SVT and FPC can only solve the nuclear norm minimization problem.

For this work, we have taken $p = 1$ and $\lambda = 0.1$. We are interested in the speed of the algorithms; to test this we took three examples of Matrix Completion –

Example 1. Matrix Size – 1000 X 1000, rank – 10

Example 2. Matrix Size – 1000 X 1000, rank – 100

Example 3. Matrix Size – 2500 X 2500, rank – 25

For all three examples, the sub-sampling ratio was fixed at 25%; i.e. only 25% of the randomly selected entries in the matrix are observed.

For both algorithms, the exit criterion is the same, i.e. the algorithm stops when it reaches a prescribed number of iterations (50) or it stops when the objective function does not change significantly (tolerance is 10^{-6}).

The experiments were carried using Matlab 2012a running on an Intel core i5 CPU with 4GB RAM. For each of the above examples, the matrix was generated 100 times, and the average reconstruction time was reported The recovery is considered successful, if the average

normalized mean squared error (NMSE) is less than 10^{-3}. The consolidated results are shown in Table 2.

Table 2. Recovery Results in seconds

Example	SVS	Proposed	Successful?
#1	31.25	2.82	Yes
#2	550.45	54.78	No
#3	221.87	18.91	Yes

We find that our proposed method delivers the promised speed-up. For Example1 and Example3, when the algorithms actually converge to the desired solution, we notice that the proposed algorithm converges faster (in less than 50 iterations, but SVS takes all 50 iterations), that is why the improvements in speed is more significant compared to SVS. For Example2, when the solution does not converge, our algorithm runs for all the 50 iterations and hence speed up is only about an order of magnitude.

4.2. Experiments on Dynamic MRI Reconstruction

We apply the proposed matrix recovery technique for dynamic MRI reconstruction and compare it with the reconstruction algorithm [12] which uses the IRPF (Incremental Rank Power Factorization). For our experiments, the value of p (of Schatten-p norm) is kept fixed at 1 and λ is kept at 0.2.

DCE-MRI experiments were performed on female tumour bearing non-obese diabetic/severe combined immune-deficient mice. All animal experimental procedures were carried out in compliance with the guidelines of the Canadian Council for Animal Care and were approved by the institutional Animal Care Committee. Tumour xenografts were implanted subcutaneously on the lower back region.

All images were acquired on a 7T/30 cm bore MRI scanner (Bruker, Germany). Mice were anaesthetized with isofluorane, temperature and respiration rate were monitored throughout the experiment. FLASH was used to acquire fully sampled 2D DCE-MRI data from the implanted tumour with 42.624 × 19.000 mm field of view, 128 × 64 matrix size TR/TE = 35/2.75 ms, 40° flip angle. 1200 repetitions were performed at 2.24 s per repetition. The 2D DCE1 dataset was acquired from a mouse bearing HCT-116 tumour (human colorectral carcinoma). The animal was administered 5 μL/g Gadovist® (Leverkusen, Germany) at 60 mM. The 2D DCE2 dataset was acquired from a mouse bearing MDA435/LCC6 tumour (human breast cancer). The animal was administered 6 μL/g hyperbranched polyglycerol-Gd (synthesized in the Faclty of Pharmaceutical Sciences at the University of British Columbia) at 0.2 mM. 3D DCE-MRI data was also acquired using a FLASH sequence, with 38.4 × 21.6 × 24.0 mm field of view, 128 × 72 × 24 matrix size, TR/TE = 9/2.66 ms, 25° flip angle. 170 repetitions were performed at 15.55 s per repetition.

Under-sampling of the K-space was simulated using Variable Density random sampling. Acceleration factors of 2 (50% sampling) and 4 (25% sampling) were used. The results are tabulated in the following Table 3.

Table 3. Reconstruction Error (NMSE)

Method	IRPF		Proposed	
	25%	50%	25%	50%
2D DCE 1	0.19	0.27	0.15	0.21
2D DCE 1	0.18	0.27	0.15	0.20
3D DCE	0.17	0.25	0.13	0.19

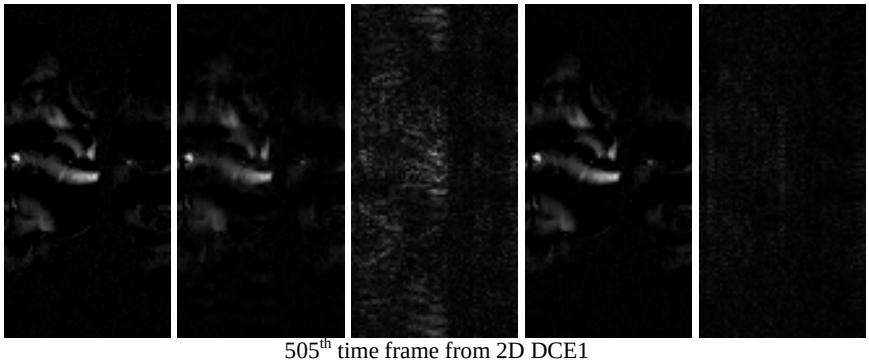
The numerical and qualitative results show that our proposed method yields considerable construction improvement over the existing technique.

For qualitative evaluation we show the ground-truth images, reconstructed images and difference (between ground-truth and reconstructed) in Fig. 1. Visual evaluation shows that our method is considerably superior than IRPF reconstruction. For example in Fig. 1, the lower portion of the frames for 2D DCE1 and 2D DCE2 sequence show discernible reconstruction artifacts for IRPF; these artifacts are almost absent in our proposed technique. A better understanding of reconstruction accuracy can be seen in the difference images. Difference images from IRPF are considerably brighter than those of our proposed method. This means that the disparity between the original and the reconstructed images is higher for IRPF as compared to our offline method.

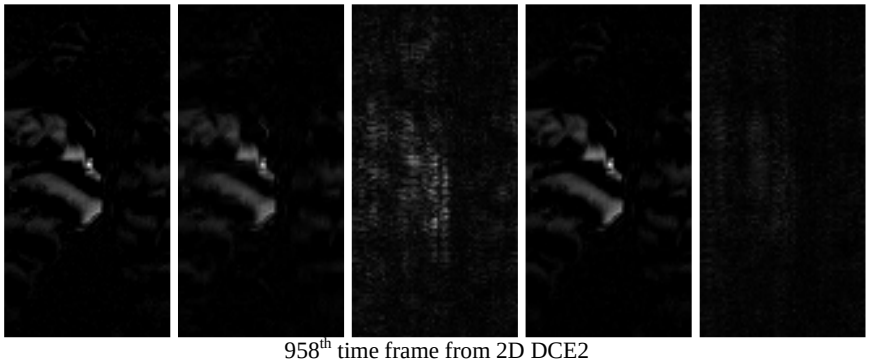
5. CONCLUSION

This work presents a solution to the Schatten-p ($0<p\leq 1$) norm minimization problem. Direct minimization of this norm leads to an algorithm that requires computing a singular value decomposition (SVD) at each iteration [4]. Computing the SVD is computationally complex; and is the main computational bottleneck for all low-rank matrix recovery algorithms [4, 8, 9]. In this work we propose an alternate technique to address the limitations of speed. We replace the Schatten-p norm by its equivalent Ky-Fan norm. The algorithm is derived in a way that precludes computing the SVD, rather the Cholesky decomposition, which is computationally much cheaper than SVD, is used. The other steps of our algorithm require a matrix-matrix product and solving of two triangular systems – both of which can be computed efficiently. We show that our algorithm is about an order of magnitude faster than [4] and does not compromise the accuracy.

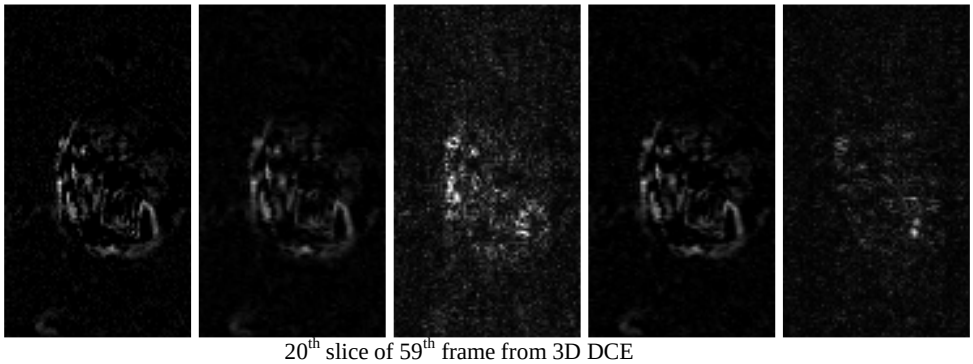
We apply our new algorithm on the problem of dynamic MRI reconstruction [12]. It has been shown that when dynamic MRI frames are stacked as columns of a Casorati matrix, the resulting matrix is low-rank since the frames are temporally correlated. This allows for recovery of the frames using low-rank recovery techniques. We show that our proposed algorithm yields significantly better results than the previous approach.



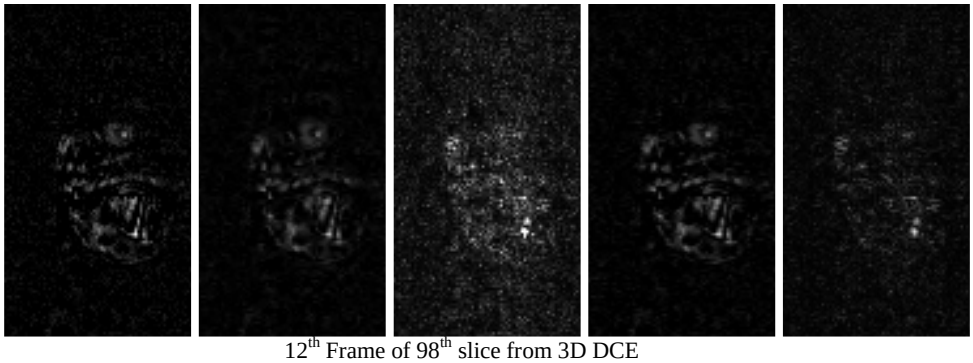
505th time frame from 2D DCE1



958th time frame from 2D DCE2



20th slice of 59th frame from 3D DCE



12th Frame of 98th slice from 3D DCE

Fig. 1. Reconstructed and Difference Images Images. From Left to Right: Ground-truth, IRPF Reconstruction, IRPF Difference Image, Proposed Reconstruction and Proposed Difference Image.

REFERENCES

[1] B. Recht, W. Xu and B. Hassibi, “Null Space Conditions and Thresholds for Rank Minimization”, Mathematical Programming. Ser B. Vol. 127, Pages 175-211. 2011.

[2] E. J. Candes and B. Recht, “Exact Matrix Completion Via Convex Optimization”, Foundations of Computational Mathematics. Vol 9. pp. 717-772. 2009.

[3] K. Mohan, M. Fazel, Iterative Reweighted Algorithms for Matrix Rank Minimization. Journal of Machine Learning Research (JMLR), 2012.

[4] A. Majumdar and R. K. Ward, “Some Empirical Advances in Matrix Completion”, Signal Processing, Vol. 91 (5), pp. 1334-1338, 2011.

[5] A. Majumdar, “FOCUSS Based Schatten-p Norm Minimization for Real-Time Reconstruction of Dynamic Contrast Enhanced MRI”, IEEE Signal Processing Letters, Vol. 9(5), pp. 315-318, 2012.

[6] http://cnx.org/content/m32168/latest/

[7] http://cnx.org/content/m44991/latest/

[8] Lloyd N. Trefethen and David Bau, III, “Numerical Linear Algebera”, SIAM, 1997.

[9] J.F. Cai, E.J. Candes and Z.W. Shen, “A singular value thresholding algorithm for matrix completion,” Technical report, Septermber 2008.

[10] S. Ma, D. Goldfarb and L. Chen, “Fixed point and Bregman iterative methods for matrix rank minimization", Mathematical Programming, Vol 128(1-2), pp.321-353, June 2009.

[11] http://soi.stanford.edu/~rmunk/PROPACK/

[12] B. Zhao, J. P. Haldar, C. Brinegar, Z. P. Liang, “Low rank matrix recovery for real-time cardiac MRI”, International Symposium on Biomedical Imaging, pp.996~999, 2010.

[13] J. P. Haldar, Z.-P. Liang, “Low-Rank Approximations for Dynamic Imaging”, IEEE International Symposium on Biomedical Imaging, pp. 1052-1055, 2011.

[14] J. P. Haldar, D. Hernando, “Rank-Constrained Solutions to Linear Matrix Equations using Power Factorization”, IEEE Signal Processing Letters, Vol. 16, pp. 584-587, 2009.