Real-time cardiac MRI using low-rank and sparsity penalties

Sajan Goud, Yue Hu, Mathews Jacob

***Abstract*— We introduce a novel algorithm to reconstruct real-time cardiac MRI data from undersampled radial acquisitions. We exploit the fact that the spatio-temporal data can be represented as the linear combination of a few temporal basis functions. The current approaches that capitalize this property estimate the basis functions from central phase encodes, acquired with a fine temporal sampling rate. In contrast, we estimate the basis functions from the entire under-sampled data. By eliminating the need for training data, the proposed method can achieve potentially high acceleration factors. More importantly, the estimation of the temporal functions from the entire data significantly improves the quality of the basis functions, which inturn improves the quality of the reconstructions. Experiments on numerical phantoms show a significant reduction in artifacts at high acceleration factors, in comparison to current schemes.**

I. Introduction

Classical cardiac MRI (CMRI) techniques aim to image the heart by freezing the motion through ECG gating and breath-holding. Recently, several researchers have exploited the structure in $x-f$ space to reconstruct the cardiac data from under-sampled acquisitions, thus significantly accelerating Cine CMRI [1], [2], [3], [4]. Since respiration and motion can alter the x-f space structure, these schemes are usually applied to breath-hold acquisitions.

Recently, algorithms that model the spatial-spectral signal as the linear combination of arbitrary temporal basis functions, which are estimated from the data itself, were introduced [5], [6]. These methods are based on the theoretical frame work developed in the Partially Separable Functions (PSF) model [7]. The flexibility of these schemes enable them to account for respiration and heart beat variations; it has the potential to enable real-time CMRI. In practical implementations, these methods involve the estimation of the temporal basis functions from the central phase encodes, acquired with a high temporal sampling rate. The ability of the temporal basis functions to represent the spatio-temporal signal is dependent on the number of phase encodes used to estimate them. For example, if only the central lines are used as in [6], the estimated temporal functions will fail to capture intermediate vertical shifts in the heart during the acquisition (eg. due to respiration or motion). This can be mitigated by acquiring more training data as in [5]. However, the acquisition of a large number of phase encodes at a fine sampling rate will significantly limit the achievable acceleration factor.

To overcome these problems, we propose to simultaneously estimate the signal and the basis functions from the entire under-sampled k-t space data. Since the temporal functions are estimated from significantly more k-t space samples than central phase encodes, they will be more representative of the data, resulting in higher quality reconstructions. In addition, by freeing up the time for collecting the training data at a fine temporal resolution, this approach can significantly improve the achievable acceleration.

The simultaneous estimation of the signal and basis functions is enabled by re-interpreting the modeling of the spatio-temporal signal as the linear combination of very few basis functions. This representation is equivalent to constraining the signal to be a low-rank matrix, whose rows correspond to the voxels and the columns correspond to the temporal samples. In contrast to the two-step approach of first determining the temporal basis functions, followed by the estimation of the spatial weights, we propose to simultaneously estimate them by formulating it as a nuclear norm minimization algorithm [8]. Since the achievable acceleration with nuclear norm minimization alone is not sufficient for real-time cardiac applications, we propose to additionaly penalize the sparsity of the signal in $x-f$ space. The class of signals that are simultaneously sparse and low-rank are much smaller than the class of sparse signals or the class of low-rank signals. Hence, this approach enables us to reconstruct the signal reliably even at high acceleration factors.

The classical semi definite programming (SDP) based algorithms to minimize the nuclear norm are computationally inefficient for large scale problems [8]. Moreover, algorithms designed for nuclear norm minimization are not directly applicable to our case, since our criterion has both nuclear norm and sparsity penalties. We introduce a new majorize-minimize algorithm to minimize the proposed penalty. The proposed algorithm derives the solution by iterating two simple steps. The first step involves shrinking the singular values and the $x-f$ space signal to obtain two auxillary variables. The second step uses the auxillary variables to derive the signal estimate using a simple Fourier domain replacement algorithm. The run time of the algorithm is a few minutes, even for large datasets.

We study the performance of the proposed scheme, in comparison to classic low rank matrix recovery algorithms based on temporal basis functions [5], [6]. Using numerical experiments, we show that the proposed scheme can significantly improve the acceleration factors in cardiac MRI applications.

Sajan Goud and Mathews Jacob are with the Department of Biomedical Engineering, University of Rochester, NY, USA. e-mail: (see http://www.cbig.rochester.edu)

Yue Hu is with the Dept of Electrical and Computer Engineering, University of Rochester, NY, USA

This work is supported by NSF award CCF-0844812.

II. CMRI RECONSTRUCTION AS AN OPTIMIZATION

We will focus on the reconstruction of a single cardiac slice in this paper; the extension of this scheme to 3-D imaging is straightforward. We denote the spatio-temporal signal as $\gamma(\mathbf{x},t)$, where $\mathbf{x}=(x,y)$ is the spatial location. The temporal Fourier transform of γ is indicated as $\hat{\gamma}(\mathbf{x},f)$. The goal of the paper is to recover the signal from its sparse $k-t$ space samples, which are corrupted by noise

$$\mathbf{y}_i = \int_{\mathbf{x}} \gamma(\mathbf{x},t_i)\exp\left(j\mathbf{k}_i^T\mathbf{x}\right)+\mathbf{n}_i. \tag{1}$$

Here, $(\mathbf{k}_i,t_i)$ indicates the i^{th} sampling location in k-t space. We denote the set of sampling locations as $\Xi=\{(\mathbf{k}_i,t_i),i=0,..,L-1\}$. The above expression can be rewritten in the vector form as $\mathbf{y}=\mathcal{A}(\gamma)+\mathbf{n}$. Here, $\mathcal{A}$ is the Fourier sampling operator.

The spatio-temporal signal $\gamma(\mathbf{x},t)$ can be re-arranged in a matrix form, where the rows correspond to the voxels and the columns correspond to the time samples as in the PSF model [7]

$$\Gamma=\begin{bmatrix}\gamma(\mathbf{x}_0,t_0) & \dots & \gamma(\mathbf{x}_0,t_n)\\ \vdots & & \\ \gamma(\mathbf{x}_m,t_0) & \dots & \gamma(\mathbf{x}_m,t_n)\end{bmatrix} \tag{2}$$

The signal is modeled as a linear combination of a few temporal basis functions.This representation is equivalent to assuming that Γ is a low rank matrix. If the $m\times n$ matrix Γ (corresponding to m voxels and n time points) is a matrix of rank r, its singular value decomposition is given by

$$\Gamma=\underbrace{\mathbf{U}}_{m\times r}\underbrace{\mathbf{S}}_{r\times r}\underbrace{\mathbf{V}^*}_{r\times n} \tag{3}$$

The above decomposition implies that γ can be expressed as the linear combinations of r temporal basis functions:

$$\gamma(\mathbf{x},t)=\sum_{i=0}^{r-1}\rho_i(\mathbf{x})\,v_i(t). \tag{4}$$

The temporal basis functions $v_i(t)$ are the columns of the matrix $\mathbf{V}$ in (3), while the spatial weighting functions $\rho_i(\mathbf{x})$ correspond to the rows of the $m\times r$ matrix $\mathbf{US}$.

Classical schemes based on low rank matrix recovery estimate $v_i(t)$ using the PCA of a low-spatial resolution training data set, acquired with a high temporal sampling rate [5], [6]. If the number of phase encodes in the training dataset are small, this approach may fail to capture the temporal variations in the high-resolution signal. To avoid this problem, we propose to recover $\rho_i(\mathbf{x})$ and $v_i(t)$ simultaneously from the undersampled k-t space data. We propose to use the penalized optimization algorithm:

$$\gamma=\arg\min_{\gamma}\ \underbrace{\|\mathcal{A}(\gamma)-\mathbf{y}\|^2}_{\text{data consistency}}+\underbrace{\mu_1\,\|\Gamma\|_*}_{\text{nuclear norm}}+\underbrace{\mu_2\,\|\hat{\gamma}\|_{\ell_1}}_{\ell_1\text{ norm}}, \tag{5}$$

where $\|\Gamma\|_*=\sum_{i=1}^{\min(m,n)}|\sigma_i|$ is the nuclear norm of Γ and $\|\hat{\gamma}\|_{\ell_1}$ is the ℓ_1 norm of the temporal Fourier transform of ρ. The nuclear norm is a surrogate for the rank of the matrix, while the ℓ_1 norm is a surrogate for the cardinality of ρ in the $x-f$ space. The use of the surrogates results in a convex optimization problem that has a unique global minimum.

The recovery of the temporal basis functions from the entire data ensures that all the temporal variations will be captured. At high acceleration factors, the use of the nuclear norm alone might not be sufficient. Hence, we exploit the smaller size of the class of signals that are simultaneously sparse and low-rank in comparison to signals that are either sparse or low-rank; the use of the joint penalty can reliably recover the signal even when the acceleration factor is high. Additionally, we will use radial trajectories with pseudo-random angles. Pseudo-random radial trajectories result in incoherent spatio-temporal aliasing patterns. Since the aliasing pattern varies from voxel to voxel, it is unlikely that the singular vectors corresponding to high singular values are influenced by them.

III. NUMERICAL OPTIMIZATION ALGORITHM

Classical nuclear norm minimization algorithms are not readily applicable to our case since our criterion has two penalty terms: the nuclear norm and the ℓ_1 norm. Hence, we now introduce a computationally efficient majorize-maximize algorithm to derive the minimum of (5)

A. Approximation induced by the Huber function

We introduce a novel majorize maximize algorithm to minimize (5). The central idea is to approximate the nuclear norm by a Huber induced matrix penalty, while the ℓ_1 norm

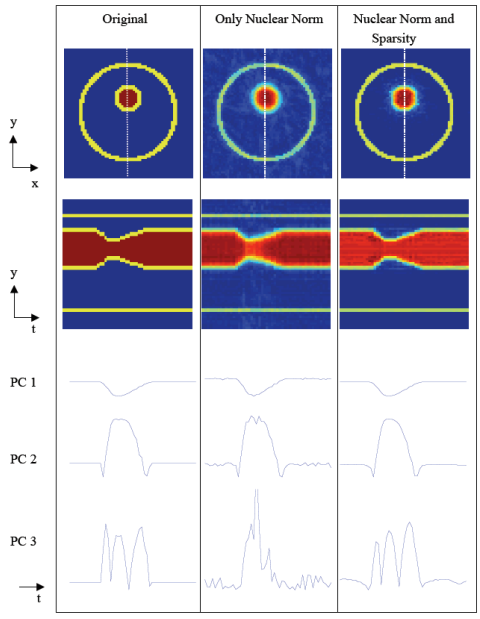

Fig. 1. Dependence of the reconstruction on the regularization parameters at high acceleration factors: *(First row: A frame at mid systole, Second row: A time profile for the y dimension at x = 32. Third, Fourth and Fifth rows show respectively the first three principal components i.e, $v_i(t)$, for $i=1$ to 3.* If we rely only on the nuclear norm penalty (μ_1) when $A=12.8$, the temporal basis functions are distorted as shown in the second column. The basis functions capture the alias components, thus resulting in poor reconstructions. In contrast, the addition of the sparsity penalty (μ_2) significantly improve the quality of the temporal basis functions and hence the reconstructions, as seen from the last column.

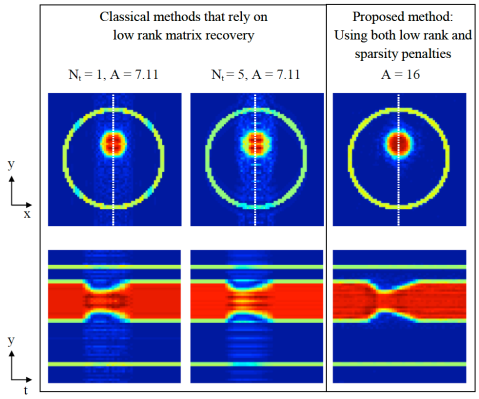

Fig. 2. Comparison of the proposed reconstructions with classical methods that rely on low rank matrix recovery with different training data sizes on Cine data: The ability of the proposed scheme to estimate the temporal basis functions from the entire k-space data eliminated the need for training data, acquired at high temporal sampling rates. In addition, the quality of the estimated temporal functions using both the low rank and sparsity penalties are higher. Specifically, the error at the dynamic region(heart) of the image is significantly lower for the proposed method in comparison to classical schemes.

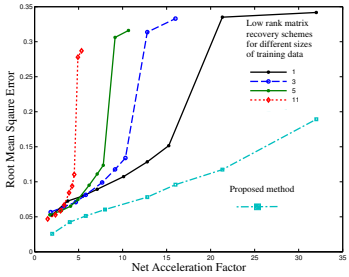

Fig. 3. Root Mean Square Error (RMSE) as a function of net acceleration factor for Cine data. The proposed scheme which simultaneously estimates the temporal basis and the spatial weights using joint low rank and sparsity penalties produce lower errors consistently and extends over a wide range of net acceleration factors.

is approximated by the Huber norm to obtain

$$\gamma = \arg\min_{\gamma} \underbrace{\|\mathcal{A}(\gamma) - \mathbf{y}\|^2}_{\text{data consistency}} + \mu_1 \underbrace{\varphi_{\beta_1}(\Gamma)}_{\approx\|\Gamma\|_*} + \mu_2 \underbrace{\psi_{\beta_2}(\hat{\gamma})}_{\approx\|\hat{\gamma}\|_{\ell_1}} \quad (6)$$

Here, $\varphi_\beta(\Gamma) = \sum_{i=1}^{\min(m,n)} \psi_\beta(\sigma_i)$ is the Huber-induced matrix penalty and $\psi_\beta(\hat{\gamma}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \psi_\beta(\hat{\gamma}(\mathbf{x}_i, f_j))$ is the Huber norm of $\hat{\gamma}$. The Huber function is specified as

$$\psi_\beta(x) = \begin{cases} |x| - 1/2\beta & \text{if} \quad x \geq \frac{1}{\beta} \\ \beta |x|^2 / 2 & \text{else} \ . \end{cases} \quad (7)$$

The parameters β_1 and β_2 specifies the quality of the approximation of (6) to (5). (6) tends to (5), as $\beta_1 = \beta_2 \to \infty$. Similarly, when $\beta_1 = \beta_2 = 0$, (5) simplifies to the standard Tikhonov regularized least squares problem, which can be solved analytically. We will use this approximation property to derive a fast continuation scheme, which is discussed in Section III-C.

B. Majorize maximize algorithm for Huber induced penalty

We majorize the two penalty terms in (6) as

$$\varphi_\beta(\Gamma) = \min_{\Lambda} \frac{\beta}{2}\|\mathbf{\Gamma} - \Lambda\|_F^2 + \|\Lambda\|_* \quad (8)$$

$$\psi_\beta(\hat{\gamma}) = \min_{\hat{\theta}} \frac{\beta}{2}\|\hat{\gamma} - \hat{\theta}\|_2^2 + \|\hat{\theta}\|_{\ell_1} \quad (9)$$

Here Λ and $\hat{\theta}$ are auxillary varibles. Using these expressions for φ_{β_1} and ψ_{β_2}, we rewrite (6) as

$$\gamma = \arg\min_{\gamma,\Lambda,\hat{\theta}} \mathcal{C}\left(\gamma, \Lambda, \hat{\theta}\right), \quad (10)$$

where

$$\mathcal{C} = \|\mathcal{A}(\gamma) - \mathbf{y}\|^2 + \frac{\mu_1\beta_1}{2}\|\Gamma - \Lambda\|_{\mathrm{F}}^2 + \mu_1\|\Lambda\|_* + \frac{\mu_2\beta_2}{2}\|\hat{\gamma} - \hat{\theta}\|_2^2 + \mu_2\|\hat{\theta}\|_{\ell_1} \quad (11)$$

(10) requires the minimization with respect to three variables. This might sound more complicated than the original problem (6). However, we show that its minimization with respect to each of the variables can be performed analytically, assuming the other variables to be fixed. These steps are described below.

1) solve for Λ and $\hat{\theta}$, assuming γ to be fixed: the minimization of (11), with respect to Λ and $\hat{\theta}$, yields

$$\Lambda = \mathbf{U}\left(\mathbf{S} - \frac{\mathbf{I}}{\beta_1}\right)_+ \mathbf{V}^*$$
$$\hat{\theta} = \frac{\hat{\gamma}}{|\hat{\gamma}|}\left(|\hat{\gamma}| - \frac{1}{\beta_2}\right)_+ \quad (12)$$

Here,

$$X_+ = \begin{cases} X & \text{if} \quad X \geq 0 \\ 0 & \text{else} \ . \end{cases} \quad (13)$$

and $\Gamma = \mathbf{USV}^*$ is the singular value decomposition of the matrix Γ. The above steps corresponds to the singular value soft thresholding of Γ and the soft-thresholding of $\hat{\gamma}$ respectively.

2) solve for γ, assuming Λ and θ to be fixed. Minimizing (11) with respect to γ gives

$$\mathcal{A}^*\mathcal{A}\gamma + (\mu_1\beta_1 + \mu_2\beta_2)\,\gamma = \mathcal{A}^*(\mathbf{y}) + \mu_1\beta_1\lambda + \mu_2\beta_2\theta. \quad (14)$$

Since $\mathcal{A}$ is a Fourier sampling operator, we can solve for γ analytically in k-t space

$$\gamma(\mathbf{k}, t)^* = \begin{cases} \frac{y_{\mathbf{k},t} + \mu_1\beta_1\,\widetilde{\lambda}(\mathbf{k},t) + \mu_2\beta_2\,\widetilde{\theta}(\mathbf{k},t)}{1 + \mu_1\beta_1 + \mu_2\beta_2} & \text{if} \quad \mathbf{k}, t \in \Xi \\ \frac{\mu_1\beta_1\,\widetilde{\lambda}(\mathbf{k},t) + \mu_2\beta_2\,\widetilde{\theta}(\mathbf{k},t)}{\mu_1\beta_1 + \mu_2\beta_2} & \text{else} \end{cases} \quad (15)$$

Here, $\widetilde{\lambda}$ and $\widetilde{\theta}$ are the 2-D Fourier transforms of $\lambda(\mathbf{x}, t)$ and $\theta(\mathbf{x}, t)$ respectively.

Thus, the minimization of (6) can be performed by iterating the shrinkage steps and the Fourier domain replacement.

C. Solving (5) using Huber continuation

The majorize maximize algorithm can provide fast convergence to the solution of the Huber induced criterion specified by (6), when β_1 and β_2 are small. However, the approximation of the original criterion specified by (5) is poor in this case. While, the Huber induced penalty is a good approximation of (5) as $\beta_1, \beta_2 \to \infty$, the convergence of the resulting majorize-maximize algorithm is poor.

To overcome this tradeoff between recovery rates and convergence, we propose to use a continuation scheme. We progressively increase β_1 and β_2, starting with small values. The solution at each step is used as the initialization for the next step.

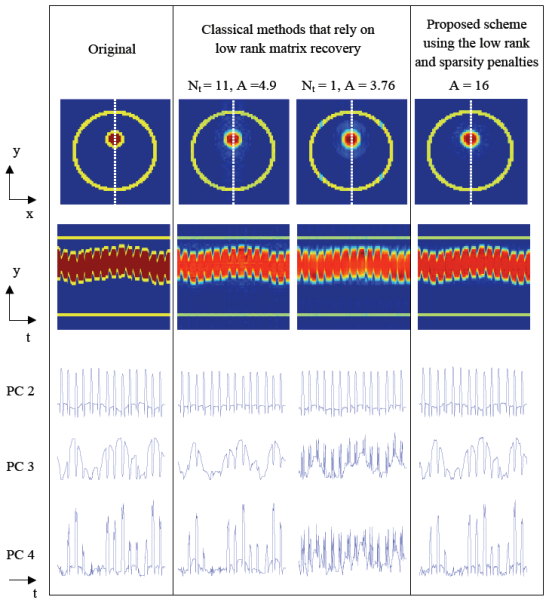

Fig. 4. Comparison of the reconstructions with classical methods that rely on low rank matrix recovery with different sizes of training data on free-breathing cardiac MRI simulations: *(First row: A mid systolic frame of the first heart beat, Second row: Time profile for the y dimension at x = 32. Third, Fourth and Fifth rows show respectively the second, third and fourth principal components,i.e,* $v_i(t)$*, for* $i = 2$ *to* 4.The ability of classical schemes to capture the respiratory motion which is in the up-down direction decreases with the size of training data. Specifically, when $N_t = 1$, even the third principal component is significantly distorted in contrast to $N_t = 11$. The use of these principal components to fit the under-sampled data results in higher errors as shown in Fig. 5. In contrast, the proposed scheme is able to reliably estimate the principal components even when $A = 16$.

IV. Results

Our studies were performed based on a slice of a numerical phantom which represents the heart. The cross section of the phantom has an inner double layered dynamic circular structure representing the beating heart and an outer stationary chest wall. We simulated two scenarios: Cine and real time imaging. The Cine data had a total of 50 frames with each image frame of size 64 x 64 while the real time data had 700 frames with no constrains on the respiration and heart rates, i.e free breathing cardiac data with heart rate variability. A pseudo random radial trajectory was employed. The accelerated data was simulated by subsampling the k-space by using fewer radial lines. The experiments were performed for different numbers of radial lines ranging from 2 to 32.

We initially demonstrate in Figure 1 that incorporating both the low rank and sparsity penalties produce superior and reliable reconstructions at higher accelerations as opposed to using the low rank penalty alone. We chose the parameters μ_1 and μ_2 to be 0.5 based on a rigorous trial and error procedure where in the Root Mean Square Error(RMSE) between the reconstructions and the simulated data were studied.

We later compare our reconstructions with the classical methods that rely on low rank matrix recovery for different training sizes, N_t [5], [6]. The classical methods were implemented using 10 and 20 principal components respectively for the Cine and real time data sets. The RMSE between the reconstructions and the simulated data was used as a metric

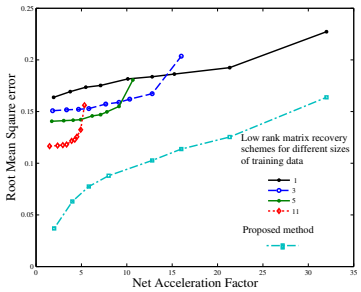

Fig. 5. Root Mean Square Error(RMSE) as a function of net acceleration factor for real time data: We observe similar behavioral patterns as with Cine data where the proposed method with its simultaneous estimation and joint penalty scheme performs significantly better than classic methods at wider range of net acceleration factors

for the comparisons. The comparisons between the proposed method and classical schemes were done over a wide range of net acceleration factors, (A). A is defined as the effective acceleration level one could achieve after taking both the training and acquired data into account.The results of the comparisons are demonstrated in figures 2 to 5.

V. Conclusion

We introduced a novel algorithm to real-time cardiac MRI data from undersampled radial acquisitions. We significantly accelerated the acquisitions by modeling the spatio-temporal data to be the linear combination of a few temporal basis functions. In contrast to classical schemes that estimate the basis functions from central phase encodes acquired with a high temporal sampling rate, we estimated them from the entire undersampled data. The use of the entire data to estimate the temporal functions resulted in significant improvement in accuracy. By removing the need for collecting training data at high temporal sampling rate, the proposed scheme achieved higher acceleration factors without significant artifacts.

References

[1] J. Tsao, S. Kozerke, P. Boesiger, and K. P. Pruessmann, "Optimizing spatiotemporal sampling for k-t BLAST and k-t SENSE: application to high-resolution real-time cardiac steady-state free precession," *Magn Reson Med*, vol. 53, no. 6, pp. 1372–1382, Jun 2005.

[2] H. Jung, J. Park, J. Yoo, and J. C. Ye, "Radial k-t FOCUSS for high-resolution cardiac cine MRI," *Magn Reson Med*, Oct 2009.

[3] B. Madore, "Using UNFOLD to remove artifacts in parallel imaging and in partial-Fourier imaging," *Magn Reson Med*, vol. 48, no. 3, pp. 493–501, Sep 2002.

[4] B. Sharif and Y. Bresler, "Adaptive Real-time cardiac MRI using PARADISE: Validation by the Physiologically Improved NCAT Phantom," in *ISBI*, 2007.

[5] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t PCA: temporally constrained k-t BLAST reconstruction using principal component analysis," *Magn Reson Med*, vol. 62, no. 3, pp. 706–716, Sep 2009.

[6] C. Brinegar, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, C. Ho, and Z.-P. Liang, "Real-Time Cardiac MRI Without Triggering, Gating, or Breath Holding," in *EMBC*, 2008.

[7] Z.-P. Liang, "Spatiotemporal Imaging with Partially Separable Functions," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, April 2007, pp. 988–991.

[8] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization," 06 2007.