SPATIOTEMPORAL IMAGING WITH PARTIALLY SEPARABLE FUNCTIONS: A MATRIX RECOVERY APPROACH

Justin P. Haldar and Zhi-Pei Liang

Department of Electrical and Computer Engineering, Beckman Institute, University of Illinois at Urbana-Champaign

ABSTRACT

There has been significant recent interest in fast imaging with sparse sampling. Conventional imaging methods are based on Shannon-Nyquist sampling theory. As such, the number of required samples often increases exponentially with the dimensionality of the image, which limits achievable resolution in high-dimensional scenarios. The partially-separable function (PSF) model has previously been proposed to enable sparse data sampling in this context. Existing methods to leverage PSF structure utilize tailored data sampling strategies, which enable a specialized two-step reconstruction procedure. This work formulates the PSF reconstruction problem using the matrix-recovery framework. The explicit matrix formulation provides new opportunities for data acquisition and image reconstruction with rank constraints. Theoretical results from the emerging field of low-rank matrix recovery (which generalizes theory from sparse-vector recovery) and our empirical results illustrate the potential of this new approach.

Index Terms— Magnetic Resonance Imaging, Low-Rank Matrix Recovery, Partially-Separable Functions

1. INTRODUCTION

Spatiotemporal imaging encompasses a wide variety of experiments, including dynamic contrast enhanced (DCE) imaging, spectroscopic imaging, and fMRI. In the context of Fourier imaging, spatiotemporal data acquisition is modeled as

$$s\left(\mathbf{k},t\right)=\int\rho\left(\mathbf{x},t\right)\exp\left(-\imath2\pi\mathbf{k}\cdot\mathbf{x}\right)d\mathbf{x}+\eta\left(\mathbf{k},t\right),\qquad(1)$$

where the measured data corresponds to samples of $s\left(\mathbf{k},t\right)$, $\rho\left(\mathbf{x},t\right)$ is the desired spatiotemporal function, and $\eta\left(\mathbf{k},t\right)$ is a noise process.

A common objective in spatiotemporal imaging is to reconstruct $\rho\left(\mathbf{x},t\right)$ with both high spatial and high temporal resolution. However, if $\rho\left(\mathbf{x},t\right)$ is modeled as a support-limited function, then the need to sample $s\left(\mathbf{k},t\right)$ at the Nyquist rate places practical limits on the types of spatiotemporal experiments that are possible. There are many different approaches that have been proposed to overcome Nyquist limits, including generalized support [1] and sparsity [2] models; in this work, we will focus on the *partially-separable function* (PSF) model [3], which can also be leveraged to considerably reduce data acquisition requirements. In particular, we will consider the case where

$$s\left(\mathbf{k},t\right)=\sum_{\ell=1}^{L}\psi_{\ell}\left(\mathbf{k}\right)\phi_{\ell}\left(t\right)\qquad(2)$$

This work was supported in part by the department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign and by research grants NIH-P41-RR023953-01, NIH-P41-EB001977-21 and NSF-CBET-07-30623. The authors would like to thank D. Hernando, B. Zhao, C. Brinegar, and K. Lee for useful discussions.

for some (possibly signal-dependent) functions $\{\psi_{\ell}\left(\cdot\right)\}_{\ell=1}^{L}$ and $\{\phi_{\ell}\left(\cdot\right)\}_{\ell=1}^{L}$. In this context, $s\left(\mathbf{k},t\right)$ is said to be Lth-order partially-separable. In addition, due to the linearity of the Fourier transform, the function $\rho\left(\mathbf{x},t\right)$ will also be Lth-order partially-separable in the absence of noise.

While strict partial separability (i.e., $L=1$) applies only to a "small" set of signals, higher-order partial separability ($L>1$) significantly enhances the representational power of the model. For example, low-order partially-separable representations have been used in the context of dynamic cardiac MRI (see [4] and references), dynamic MRI, PET, and SPECT imaging of contrast kinetics [3, 5], relaxation experiments [6, 7], diffusion experiments [7], fMRI [8], and spectroscopic imaging [9].

Earlier work leveraging the PSF model to enable sparse sampling not only assumed low-rank structure for $\mathbf{C}$, but also required significant additional constraints on both the data acquisition procedure and on the solution [3, 4]. This paper removes these constraints by formulating the reconstruction as a low-rank matrix recovery problem.

2. MATRIX RECOVERY APPROACH

2.1. Matrix Formulation

A consequence of Lth-order partial separability is that for any sets of k-space locations $\{\mathbf{k}_n\}_{n=1}^{N}$ and time points $\{t_m\}_{m=1}^{M}$, the $N\times M$ *Casorati matrix* $\mathbf{C}$ formed as

$$\mathbf{C}=\begin{bmatrix}s\left(\mathbf{k}_1,t_1\right) & \cdots & s\left(\mathbf{k}_1,t_M\right)\\ \vdots & \ddots & \vdots\\ s\left(\mathbf{k}_N,t_1\right) & \cdots & s\left(\mathbf{k}_N,t_M\right)\end{bmatrix}\qquad(3)$$

has at most rank L [3]. Reconstructing $s\left(\mathbf{k},t\right)$ on $\{\mathbf{k}_n\}_{n=1}^{N}\times\{t_m\}_{m=1}^{M}$ from sparse data is equivalent to reconstructing $\mathbf{C}$ from sparse data. However, posing the problem in terms of recovery of $\mathbf{C}$ allows us to take advantage of its low-rank structure. Note that an $N\times M$ complex matrix of rank L has NM entries (i.e., $2NM$ real values), but only $2\left(N+M-L\right)L$ degrees of freedom. As a result, there is considerable potential to use PSF structure to reduce data sampling requirements when L is small relative to M and N.

Independently of PSF work, low-rank matrix recovery has recently received significant attention, motivated by emerging theoretical results for the guaranteed success of different reconstruction methods (e.g., see [10, 11, 12, 13, 14]), which are related to earlier results from sparse-vector recovery [2]. In cases when L is known *a priori*, the maximum-likelihood (under white Gaussian noise) matrix recovery problem can be posed as

$$\hat{\mathbf{C}}=\arg\min_{\substack{\mathbf{C}\in\mathbb{C}^{N\times M}\\ \mathrm{rank}(\mathbf{C})\leq L}}\left\|\mathcal{A}\left(\mathbf{C}\right)-\mathbf{b}\right\|_{\ell_2},\qquad(4)$$

where $\mathcal{A} : \mathbb{C}^{N\times M} \rightarrow \mathbb{C}^P$ is a linear measurement operator (with $P \ll NM$ for sparse sampling), and $\mathbf{b} \in \mathbb{C}^P$ is a data vector. However, determination of the exact rank is not trivial in general, particularly in the presence of noise. As a result, another approach is to enforce the low-rank structure in a softer way. Examples of this include the rank-minimization problem:

$$\hat{\mathbf{C}} = \arg \min_{\substack{\mathbf{C}\in\mathbb{C}^{N\times M} \\ \|\mathcal{A}(\mathbf{C})-\mathbf{b}\|_{\ell_2}\leq\varepsilon}} \operatorname{rank}(\mathbf{C}), \tag{5}$$

where ε represents a noise tolerance, or more general regularization-based formulations:

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C}\in\mathbb{C}^{N\times M}} \|\mathcal{A}(\mathbf{C}) - \mathbf{b}\|_{\ell_2}^2 + \lambda R(\mathbf{C}), \tag{6}$$

where $R(\cdot)$ is a regularization functional that favors matrices with low rank, and λ is a regularization parameter. Choices of $R(\cdot)$ that have been used previously include the nuclear norm (NN) [10] ($R_{\text{NN}}(\mathbf{C}) = \sum_i \sigma_i$, where σ_i are the singular values of $\mathbf{C}$) and the log-determinant functional [15] ($R_{\text{logdet}}(\mathbf{C}) = \sum_i \log(\sigma_i + \delta)$, where δ is a small constant), though other reasonable choices could include model-order selection criteria such as a transformation of the Akaike Information Criterion (AIC) [16].

2.2. Sampling Considerations

While the formulations presented in the previous subsection are interesting, an associated practical question is whether such formulations can lead to significant improvements in data acquisition and image reconstruction. A case that is easy to analyze is that of the basic PSF procedure [3, 4]. The basic PSF procedure can be viewed as a special case of Eq. (4), but with the additional constraint that the L-dimensional row space of $\mathbf{C}$ is predetermined from training/navigator data. In previous work [3, 4], this training data has been assembled by direct measurement of at least L full rows of $\mathbf{C}$, and forming a basis for the row space via principal component analysis of these rows. In principle, if the true matrix were exactly rank-L and if the L-dimensional row space could be precisely identified, then the sampling requirements for reconstructing $\mathbf{C}$ exactly (in the case of noiseless data) are easy to determine: in most cases of interest, $\mathbf{C}$ could be recovered using simple linear algebra if each row is directly sampled at least L times [3].

In contrast to the basic PSF procedure [3], which only uses a subset of the acquired data to estimate the structure of the full matrix, reconstruction using any of Eqs. (4)-(6) would use all of the measured data to estimate the matrix structure. In addition, while the original data acquisition strategy is fully compatible with image reconstruction under the matrix recovery framework, Eqs. (4)-(6) give the potential to use significantly more "arbitrary" sampling schemes. The price that is paid for this increased flexibility is that the optimization problem becomes significantly more complicated, and that sampling requirements are no longer as easy to analyze. Despite this, theoretical results have been derived [10, 11, 12, 13, 14] that provide sufficient conditions on $\mathcal{A}$ and $\mathbf{C}$ such that the various algorithms for solving Eqs. (4)-(6) and their variations are guaranteed to be successful. These sufficient conditions often require that $\mathcal{A}$ obeys a *rank-restricted isometry property* (RIP), i.e., that $\|\mathcal{A}(\mathbf{C})\|_{\ell_2} \approx \|\mathbf{C}\|_F$ for any sufficiently low-rank $\mathbf{C}$. While computing RIPs is computationally intractable, it is known that certain classes of randomized sampling operators have good RIP properties, as long as P is large enough [10, 12].

For example, consider the case where we directly sample P entries of $\mathbf{C}$, and represent this sampling set as

$$\Omega = \{(n, m) : \mathbf{C}_{nm} \text{ is sampled}\}. \tag{7}$$

A representative theoretical result for this context is that if the sampling set Ω is chosen uniformly at random and if the number of measurements is $O\left(L^2 K \log K\right)$ with $K = \max(M, N)$, then the corresponding sampling operator $\mathcal{A}$ has a high-probability of satisfying an RIP for a class of suitably-regular low-rank matrices [12].[1] This result suggests that recovery of PSFs from undersampled data can be attained without utilizing specialized sampling patterns, as long as appropriate algorithms are used for reconstruction.

2.3. Algorithm Considerations

The problems in Eqs. (4)-(6) are often nonconvex, and the problem in Eq. (5) can be NP-hard [10]. Despite these difficulties, many different methods exist to solve such problems.

Solving Eq. (4) has been performed using alternating least-squares algorithms [17, 18], gradient descent and expectation-maximization algorithms [19], optimization over Grassmann manifolds [13, 20], and projected gradient algorithms [12]. While many of these algorithms are heuristic and might not achieve global optimization for general $\mathcal{A}$, some choices of $\mathcal{A}$ can guarantee the global optimality of certain algorithms [12, 13, 14, 18].

Due to the complexity of addressing Eq. (5) directly, common approaches to this problem involve the use of surrogate optimization problems. For example, Eq. (5) can be approached using the regularization framework of Eq. (6). One convenient choice of $R(\cdot)$ for this case is the NN [10], since the NN is the tightest convex relaxation of matrix rank, in just the same way that the ℓ_1-norm is the tightest convex relaxation of the ℓ_0-norm in the context of sparse-vector recovery [2]. Due to convexity, problems involving the NN can be solved globally [10, 21, 22]. And, under appropriate RIP constraints, the solution using NN minimization can be proven to be equivalent to the original problem. However, when RIPs are not satisfied, the resulting solutions can have higher rank than those that would have been obtained when solving Eq. (5) directly. Algorithms for solving Eq. (4) can also be used to solve Eq. (5), by noting that $\|\mathcal{A}(\hat{\mathbf{C}}) - \mathbf{b}\|_{\ell_2}$ is monotonically decreasing in L. As a result, the solution to Eq. (5) can be achieved by solving Eq. (4) for increasing values of L until the data-consistency constraints are satisfied.

In the example shown later in this work, we use the incremented-rank PowerFactorization (IRPF) algorithm in the context of Eq. (4). While guarantees for global optimality with this algorithm have only been established for the fully-sampled data scenario [18], it has been observed empirically that IRPF is both fast and can give better solutions than alternatives like NN minimization [17] for solving both Eqs. (4) and (5). PowerFactorization is an alternating least-squares approach that makes use of the factorization $\mathbf{C} = \mathbf{UV}$, with $\mathbf{U} \in \mathbb{C}^{N\times L}$ and $\mathbf{V} \in \mathbb{C}^{L\times M}$, to enforce rank-$L$ structure implicitly. Subsequently, we seek a local minimum of

$$\left\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\right\} = \arg \min_{\substack{\mathbf{U}\in\mathbb{C}^{N\times L} \\ \mathbf{V}\in\mathbb{C}^{L\times M}}} \|\mathcal{A}(\mathbf{UV}) - \mathbf{b}\|_{\ell_2}. \tag{8}$$

Minimization is performed by optimizing over $\mathbf{U}$ and $\mathbf{V}$ in alternation. The optimizations involved can be represented as standard

[1] Regularity in this context means that the magnitude of the largest element of the matrix is not significantly larger than the root-mean-squared magnitude of all the elements; see [12] for a more formal description.

linear-least squares problems, and can be solved using standard solvers. In the case where $\mathcal{A}$ directly subsamples $\mathbf{C}$ on the set Ω, it can be shown that the corresponding normal equations are block diagonal, with $L \times L$ blocks. This leads to efficient computation, either by direct inversion of each $L \times L$ block separately, or by using the iterative conjugate gradient algorithm, which would find optimal solutions (assuming infinite precision) after L iterations.

As described in [17], we have observed that using an incremented rank strategy improves the performance of PowerFactorization and helps to avoid many of the local minimizers of Eq. (8). IRPF uses a continuation scheme, and starts by solving Eq. (8) with $L = 1$. Subsequently, L is incremented, with the results of rank-L optimization used to initialize the rank-$(L + 1)$ problem. In this fashion, the procedure is iterated until the final desired L is achieved [17].

3. APPLICATION EXAMPLE

In this section, we present a practical example of the PSF model and the IRPF algorithm applied to DCE breast imaging. In this context, an exogenous contrast agent is injected into the bloodstream, and the contrast kinetics help to localize a tumor and provide information regarding its physiological and morphological properties. Fourier data was simulated from a series of magnitude images corresponding to 18 different time-points from a real DCE experiment. These 18 frames were subsequently interpolated onto a set of $M = 52$ equally-spaced time points. The simulated data acquisition assumed a 256×256 Cartesian *k*-space sampling grid (i.e., $N = 256^2$).

Elements of $\mathbf{C}$ were sampled directly, with the sampling set Ω constructed by selecting 25% of the entries in $\mathbf{C}$ uniformly at random. This was augmented by an additional set of samples chosen to ensure that each row and each column of $\mathbf{C}$ was sampled at least 10 times,[2] resulting in a total of $\sim 27\%$ of the entries being sampled.

Two temporal-frames from an IRPF reconstruction with $L = 5$ are shown in Fig. 1. These results illustrate that high-quality PSF reconstruction is possible from highly-undersampled data without the use of specialized sampling.

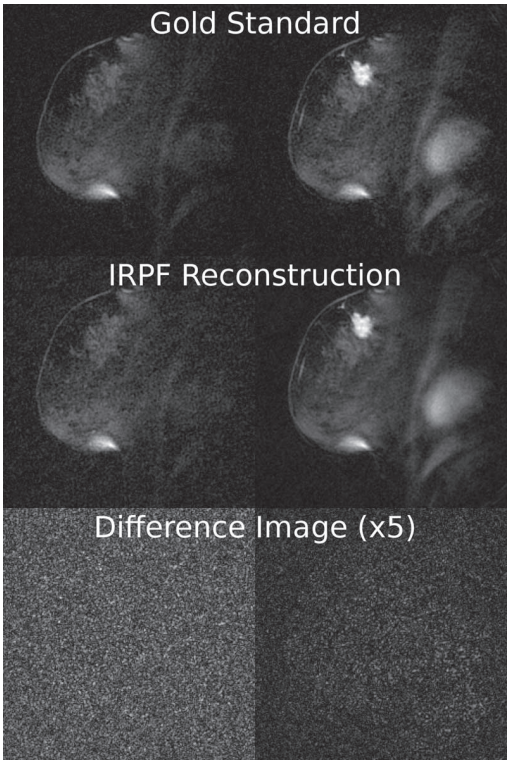

Fig. 1. Simulated PSF spatiotemporal reconstruction (L=5) using IRPF with $\sim 27\%$ of the full data. The left and right sides of the figure show different temporal frames.

4. DISCUSSION

The preceding example illustrated the potential of using IRPF for "arbitrarily"-sampled PSF spatiotemporal images. While purely random sampling was used in this example, this type of scheme is not necessarily well-suited to real experiments. First, purely random Fourier sampling is generally not practical for traditional MRI readouts, which typically sample along a set of smooth curves in *k*-space. Second, it is well-known that *k*-space energy is often highly-concentrated at low spatial frequencies. As a result, purely random sampling can incur an SNR penalty relative to sampling schemes that acquire more low-frequency data. In addition, the concentration of energy on certain rows of $\mathbf{C}$ could indicate that $\mathbf{C}$ might not be regular enough for a completely random sampling set Ω to work well with high probability. Thus, optimized sampling for different spatiotemporal MRI applications is an interesting open problem.

One important consideration for image reconstruction using the PSF model is that the true rank L needs to be small enough relative to both M and N that a constraint on $\mathrm{rank}\,(\mathbf{C})$ significantly reduces the degrees of freedom. In practice, this can mean that more significant accelerations are possible for large-scale reconstruction problems where the number of reconstructed time points M is large. For example, a complex rank-5 matrix of size $256^2 \times 18$ has 6.6×10^5 real degrees of freedom and 1.2×10^6 complex entries, meaning that it would be necessary to sample at least 25% of the matrix to have any hope of successful reconstruction. In contrast, a complex rank-5 matrix of size $256^2 \times 52$ still has roughly 6.6×10^5 degrees of freedom, but has significantly more (i.e., 3.4×10^6) entries, so that accurate reconstruction is conceivable with only 10% of the data.

Another practical issue for Eq. (4) is the choice of L, which can represent a trade-off between the expressive power of the signal model and the conditioning of the inverse problem, as illustrated in Fig. 2. While the total error is low for $L = 3$,[3] this choice fails to capture all of the signal dynamics. We prefer to use $L = 5$, since this choice reconstructs local contrast kinetics more faithfully, despite higher total error. In this light, choice of an appropriate L is nontrivial, and should be made based on the imaging context. Similar comments can be made with regard to the choice of ε in Eq. (5).

Finally, a number of extensions to the proposed PSF reconstruction scheme are possible. For example, by invoking general linear sampling operators $\mathcal{A}$, it becomes possible to incorporate any prior information that might be available regarding the known spatial-spectral support of the spatiotemporal image [1, 4, 5]. A preliminary investigation of this is presented in [23], in the context of cardiac MRI. In addition, a more general choice of $\mathcal{A}$ makes it possible to model non-Fourier acquisition physics, which could be useful for a variety of imaging contexts. It is also relatively straightforward to include additional regularization in the formulation of the problem, if

[2] A necessary condition for PSF reconstruction with Eq. (4) to be well-posed in this context is that each row and each column is sampled at least L times. Sampling more than this will improve the conditioning of the problem.

[3] This is the rank selected by the small-sample corrected AIC [16].

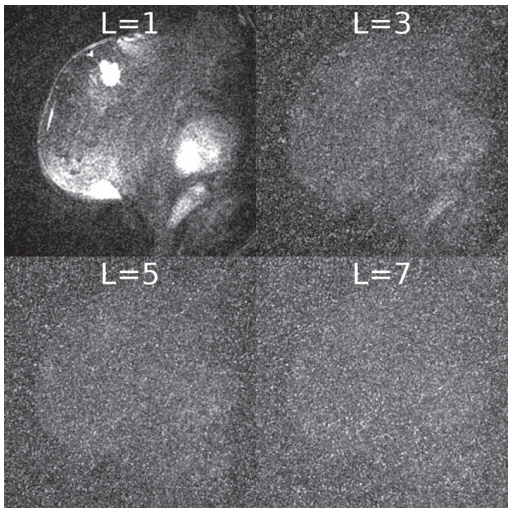

Fig. 2. Mean-squared reconstruction error for different model orders. As L increases, the signal model has better capability to represent the true signal, leading to fewer signal features being apparent in the error map. However, this additional flexibility comes at the expense of conditioning, which is reflected by the increasing "noise" component of the error maps with increasing L.

additional prior information is available. Another interesting extension is the generalization to low-rank tensor recovery [24], as this can enable accelerated reconstruction of PSFs with higher-dimensional structure (e.g., see Eq. (4) of [3]).

5. CONCLUSIONS

This paper has presented a matrix recovery approach to estimating spatiotemporal images from sparsely sampled data, based on the assumption that the underlying function is partially separable. Partial separability leads to the formulation of a low-rank matrix recovery problem, which can be solved to yield high-quality reconstructions from "arbitrarily"-sampled data. Thus, the combined PSF/matrix-recovery approach provides a flexible way to reconstruct partially-separable spatiotemporal images from limited data.

6. REFERENCES

[1] Y. Bresler, N. Aggarwal, and B. Sharif, "Patient-adaptive spatio-temporal MRI: From PARADIGM to PARADISE and beyond," in *Proc. IEEE ISBI*, 2007, pp. 980–983.

[2] E. J. Candès and T. Tao, "Reflections on compressed sensing," *IEEE Info. Theory Soc. News.*, vol. 58, pp. 14–17, 2008.

[3] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. IEEE ISBI*, 2007, pp. 988–991.

[4] C. Brinegar, H. Zhang, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, D. Pocci, F. Lam, C. Ho, and Z.-P. Liang, "Real-time cardiac MRI using prior spatial-spectral information," in *Proc. IEEE EMBC*, 2009, pp. 4383–4386.

[5] J. G. Brankov, M. N. Wernick, M. A. King, Y. Yang, and M. V. Narayanan, "Spatially adaptive temporal smoothing for reconstruction of dynamic image sequences," *IEEE Trans. Nuc. Sci.*, vol. 53, pp. 2769–2777, 2006.

[6] M. Bydder and J. Du, "Noise reduction in multiple-echo data sets using singular value decomposition," *Magn. Reson. Imaging*, vol. 24, pp. 849–856, 2006.

[7] M. Doneva, J. Sènègas, P. Börnert, H. Eggers, and A. Mertins, "Accelerated MR parameter mapping using compressed sensing with model-based sparsifying transform," in *Proc. Int. Symp. Magn. Reson. Med.*, 2009, p. 2812.

[8] H. Jung, J. C. Ye, and E. Y. Kim, "Improved k-t BLAST and k-t SENSE using FOCUSS," *Phys. Med. Biol.*, vol. 52, pp. 3201–3226, 2007.

[9] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, and L. C. Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Trans. Med. Imaging*, vol. 23, pp. 1453–1465, 2004.

[10] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *Preprint*, 2007, http://arxiv.org/abs/0706.4138.

[11] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *Preprint*, 2009, http://arxiv.org/abs/0903.1476.

[12] R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," *Preprint*, 2009, http://arxiv.org/abs/0909.5457.

[13] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *Preprint*, 2009, http://arxiv.org/abs/0901.3150.

[14] K. Lee and Y. Bresler, "ADMiRA: Atomic decomposition for minimum rank approximation," *Preprint*, 2009, http://arxiv.org/abs/0901.1898.

[15] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," 2003, pp. 2156–2162.

[16] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer-Verlag, New York, 2nd edition, 2002.

[17] J. P. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using PowerFactorization," *IEEE Signal Process. Lett.*, vol. 16, pp. 584–587, 2009.

[18] R. Hartley and F. Schaffalitzky, "PowerFactorization: 3D reconstruction with missing or uncertain data," in *Australia-Japan Advanced Workshop on Computer Vision*, 2003.

[19] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 720–727.

[20] W. Dai and O. Milenkovic, "SET: An algorithm for consistent matrix completion," in *Proc. IEEE ICASSP*, 2010.

[21] J.-F. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *Preprint*, 2008, http://arxiv.org/abs/0810.3286.

[22] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program., Ser. A*, 2010.

[23] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang, "Low rank matrix recovery for real-time cardiac MRI," in *Proc. IEEE ISBI*, 2010.

[24] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, pp. 455–500, 2009.