A Fast Majorize–Minimize Algorithm for the Recovery of Sparse and Low-Rank Matrices

Yue Hu, *Student Member, IEEE*, Sajan Goud Lingala, *Student Member, IEEE*, and Mathews Jacob, *Member, IEEE*

***Abstract*—We introduce a novel algorithm to recover sparse and low-rank matrices from noisy and undersampled measurements. We pose the reconstruction as an optimization problem, where we minimize a linear combination of data consistency error, nonconvex spectral penalty, and nonconvex sparsity penalty. We majorize the nondifferentiable spectral and sparsity penalties in the criterion by quadratic expressions to realize an iterative three-step alternating minimization scheme. Since each of these steps can be evaluated either analytically or using fast schemes, we obtain a computationally efficient algorithm. We demonstrate the utility of the algorithm in the context of dynamic magnetic resonance imaging (MRI) reconstruction from sub-Nyquist sampled measurements. The results show a significant improvement in signal-to-noise ratio and image quality compared with classical dynamic imaging algorithms. We expect the proposed scheme to be useful in a range of applications including video restoration and multidimensional MRI.**

***Index Terms*—Dynamic magnetic resonance imaging (MRI), low rank, majorize minimize, matrix recovery, sparse.**

I. Introduction

THE RECOVERY of a low-rank or approximately low-rank matrix from very few measurements of its entries has received a lot of attention in recent years, mainly due to its application in machine learning, computer vision, and recommendation systems [1], [2]. Our own motivation in this area is to use this framework to recover dynamic imaging/video data from sparse and noisy measurements. Several researchers have posed the reconstruction of dynamic image data as the recovery of a low-rank Casorati matrix, whose columns correspond to spatial pixels and rows correspond to temporal intensity variations of the pixels [3]–[7]. In addition, the low-rank matrix may be known to be additionally sparse in a specified domain. For example, each of the frames of a dynamic imaging data set is a natural image and can hence have sparse wavelet coefficients or gradients [8]. It is known that sparsity and low-rank properties are somewhat complementary [1]; most randomly selected low-rank matrices are not sparse, and most sparse random matrices are not of low rank [9]. However, there are matrices that are simultaneously sparse and of low rank. For example, a matrix with only one nonzero entry will have a sparsity of one and will be of unit rank. We exploit the fewer degrees of the freedom of the set of matrices that are simultaneously sparse and of low rank, as compared with only sparse or only low-rank matrices, to significantly reduce the number of measurements. We, as well as other researchers, have demonstrated the utility of modeling dynamic imaging data set as sparse and low-rank Casorati matrices [5], [6], [10], [11].

In this paper, we focus on nonconvex spectral and sparsity penalties to further decrease the number of measurements required for recovery. Specifically, we consider Schatten-p functionals, which are the extensions of the classical nuclear-norm spectral penalty. This choice is inspired by the recent success of nonconvex compressed sensing schemes that rely on $\ell_p; p < 1$ penalties [12], [13]. Similar to the ℓ_p vector penalty, the Schatten-p functions cease to be norms and are nonconvex for $p < 1$. We also use nonconvex sparsity penalties as in [12]–[14]. Thus, we pose the recovery of the sparse and low-rank matrices as a nonconvex optimization problem, where the cost function is a linear combination of the data consistency term, the nonconvex spectral penalty, and the nonconvex sparsity penalty. Most of the current matrix recovery algorithms are variants of iterative singular-value thresholding (IST) [15]–[17]. Unfortunately, it is not easy to adapt this scheme to our case since our penalty is a linear combination of nonconvex spectral and sparsity functions; it is difficult to efficiently evaluate the proximal mapping of the linear combination of penalties. It may be possible to extend the algorithm in [18], which relies on multiple proximal projections, to solve the matrix recovery problem with convex priors. However, the proximal projections will not have closed-form expressions when nonconvex penalties are used, making the efficient implementation of these schemes challenging.

We introduce a novel majorize–minimize (MM) algorithm to recover sparse and low-rank matrices from undersampled measurements. In contrast with current matrix recovery schemes that majorize the data consistency term, we majorize each of the penalty terms with quadratic functions. We use the property of unitarily invariant matrix penalties to majorize the spectral penalty. This majorization of the penalty terms enables us to solve for the matrix using a three-step alternating minimization scheme with closed-form shrinkage rules. The iterative algorithm alternates between the three simple steps: 1) the solution

Manuscript received March 14, 2011; revised June 10, 2011 and July 28, 2011; accepted July 29, 2011. Date of publication August 22, 2011; date of current version January 18, 2012. This work is supported by the National Science Foundation Award CCF-0844812 and Award CCF-1116067. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ali Bilgin.

Y. Hu is with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 014627 USA (e-mail: yue.hu@rochester.edu).

S. G. Lingala and M. Jacob are with the Department of Biomedical Engineering and the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: sajangoud-lingala@uiowa.edu; mathews-jacob@uiowa.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2165552

of a linear system of equations; 2) a singular value shrinkage; and 3) a gradient shrinkage. The linear system of equations can be solved using either analytical expressions or a few conjugate gradient (CG) steps. Both the shrinkage steps are obtained as the proximal mappings [19] of new matrix functions, which are related to the original spectral and sparsity penalties. Due to the property of convex conjugate matrix functions, these shrinkage rules have analytical expressions, even when nonconvex penalties are used. Since each step of the algorithm can be evaluated analytically or using fast schemes, the algorithm is computationally very efficient. The proposed MM scheme is equivalent to the variable splitting (VS) scheme, which was introduced in our earlier work [6], when nuclear-norm and total-variation (TV) penalties are used. However, the proximal projections in the VS scheme do not have closed-form expressions when nonconvex penalties are used. While the proximal projections can be approximated, as shown in [20], we observe that the resulting approximate VS algorithm is significantly slower than the corresponding MM scheme. We also introduce a continuation scheme to accelerate the convergence of the algorithm. In addition to providing fast algorithms, this approach makes the algorithm robust to local minima. Specifically, we use a sequence of criteria with gradually increasing complexity while using the solution from the previous iteration to initialize the new criterion. Similar homotopy continuation schemes are used in nonconvex compressed sensing to minimize local-minimum effects [21].

The rest of this paper is organized as follows: We briefly review the background literature in Section II to make this paper self-contained. We introduce the MM algorithm in Section III, whereas its numerical implementation is described in Section IV. We study the convergence of the algorithm and its utility in practical applications in the results in Section V.

II. Background

A. Matrix Recovery Using Nuclear-Norm Minimization

Current theoretical results indicate that matrix $\mathbf{\Gamma} \in \mathbb{R}^{m\times n}$ of rank $r; r \leq \min(m,n)$ can be perfectly recovered from its linear measurements $\mathbf{b} = \mathcal{A}(\mathbf{\Gamma})$ [1], [2]. This recovery can be formulated as the following constrained optimization problem:

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 \text{ such that rank } (\mathbf{\Gamma}) \leq r. \quad (1)$$

To realize computationally efficient algorithms, the aforementioned problem is often reformulated as an unconstrained convex optimization scheme, i.e.,

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma}} \underbrace{\|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 + \lambda \|\mathbf{\Gamma}\|_*}_{\mathcal{C}(\mathbf{\Gamma})} \quad (2)$$

where $\|\mathbf{\Gamma}\|_*$ is the nuclear norm of matrix $\mathbf{\Gamma} = \sum_{i=1}^{\min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^H$. This penalty is the convex relaxation of the rank and is defined as the sum of the singular values of $\mathbf{\Gamma}$: $\|\mathbf{\Gamma}\|_* = \sum_{i=1}^{\min(m,n)} \boldsymbol{\sigma}_i$.

B. MM Algorithms

MM algorithms rely on a surrogate function $\mathcal{S}(\mathbf{\Gamma}, \mathbf{\Gamma}_n)$ that majorizes the objective function $\mathcal{C}(\mathbf{\Gamma})$ using the current iterate $\mathbf{\Gamma}_n$. The surrogate function is equal to the objective function at tangent $\mathbf{\Gamma}_n$ [i.e., $\mathcal{S}(\mathbf{\Gamma}_n, \mathbf{\Gamma}_n) = \mathcal{C}(\mathbf{\Gamma}_n)$] and is larger than the objective function elsewhere, i.e.,

$$\mathcal{S}(\mathbf{\Gamma}, \mathbf{\Gamma}_n) \geq \mathcal{C}(\mathbf{\Gamma}) \qquad \forall \mathbf{\Gamma}. \quad (3)$$

The successive minimization of the majorant function $\mathcal{S}(\mathbf{\Gamma}, \mathbf{\Gamma}_n)$ ensures that the cost function $\mathcal{C}(\mathbf{\Gamma})$ monotonically decreases. This property guarantees global convergence for convex cost functions. We rely on homotopy continuation schemes to minimize local-minimum problems, when nonconvex cost functions are used.

C. Matrix Recovery Using IST

The common approaches to solve for (2) involve different flavors of IST [15]–[17]. These schemes majorize the data-consistency term in (2) with the following quadratic expression:

$$\|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 \leq \tau \|\mathbf{\Gamma} - \mathbf{Z}_n\|^2 + c_n. \quad (4)$$

Here, τ is a constant such that $\tau \mathcal{I} \geq \mathcal{A}^t\mathcal{A}$, c_n is a constant that is independent of $\mathbf{\Gamma}$, and $\mathbf{Z}_n = \mathbf{\Gamma}_n - \mathcal{A}^t(\mathcal{A}(\mathbf{\Gamma}_n) - \mathbf{b})/\tau$. Here, $\mathcal{I}$ is the identity operator. Thus, we have

$$\mathcal{C}(\mathbf{\Gamma}) \leq \underbrace{\tau\|\mathbf{\Gamma} - \mathbf{Z}_n\|^2 + \lambda\|\mathbf{\Gamma}\|_*}_{\mathcal{C}_{\text{maj}}(\mathbf{\Gamma})} + c_n. \quad (5)$$

The minimization of the aforementioned expression is termed as the proximal mapping of $\mathbf{Z}_n$, associated with the nuclear-norm penalty [19]. This proximal mapping has an analytical solution [15], i.e.,

$$\mathbf{\Gamma}_{n+1} = \sum_{i=1}^{\min(m,n)} (\sigma_i - \lambda/2\tau)_+ \mathbf{u}_i \mathbf{v}_i^H \quad (6)$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the singular vectors and σ_i values are the singular values of $\mathbf{Z}_n$. The thresholding function in (6) is defined as

$$(\sigma)_+ = \begin{cases} \sigma, & \text{if } \sigma \geq 0 \\ 0, & \text{else.} \end{cases} \quad (7)$$

Unfortunately, it is not straightforward to adapt this algorithm to optimization schemes with multiple nondifferentiable penalty terms (e.g., spectral and sparsity penalties), as previously discussed.

D. Unitarily Invariant Matrix Functions

We focus on the general class of unitarily invariant spectral penalties, which satisfy the following property:

$$\phi(\mathbf{\Gamma}) = \phi(\mathbf{V}\mathbf{\Gamma}\mathbf{U}) \quad \forall \mathbf{\Gamma} \in \mathbb{R}^{m\times n} \quad \forall \mathbf{U} \in \mathcal{U}_n \quad \forall \mathbf{V} \in \mathcal{U}_m. \quad (8)$$

Here, $\mathcal{U}_m$ is the set of $m \times m$ unitary matrices. The aforementioned definition implies that $\phi(\mathbf{\Gamma})$ is invariant to pre- and postmultiplication of $\mathbf{\Gamma}$ by unitary matrices. This class has several attractive properties, which are valuable in realizing fast algorithms. Specifically, Lewis has shown that unitarily invariant convex matrix functions are fully characterized by composite functions of the singular values of the matrix [22]. If $\boldsymbol{\sigma}(\mathbf{\Gamma})$ denotes the $\min(m,n)$ dimensional vector of singular values of $\mathbf{\Gamma}$, any unitarily invariant penalty can be specified as

$$\phi(\mathbf{\Gamma}) = \rho\left(\boldsymbol{\sigma}(\mathbf{\Gamma})\right). \tag{9}$$

Here, $\rho : \mathbb{R}^{min(m,n)} \rightarrow \mathbb{R}$ is a function that is invariant under sign changes and permutations of the elements of $\boldsymbol{\sigma}(\mathbf{\Gamma})$. An interesting case is when $\rho(\boldsymbol{\sigma}) = \sum_{i=1}^{\min(m,n)} \mu(\sigma_i)$, where $\mu : \mathbb{R} \rightarrow \mathbb{R}$ is a function that is invariant to sign of σ_i. This class includes most of the spectral penalties of practical interest. For example, the Schatten-p spectral norms correspond to $\mu(\sigma) = |\sigma|^p$. These penalties are convex and are norms when $p \geq 1$. Clearly, the nuclear norm is a special case of Schatten-p norms, when $p = 1$. Due to the results in [22], many of the properties of vector functions can be extended to matrices. For example, the subgradient of $\phi(\mathbf{\Gamma})$ is given by [22] $\partial\phi(\mathbf{\Gamma}) = \{\mathbf{U}\,\mathrm{diag}(\delta)\mathbf{V}^* | \delta \in \partial\rho(\boldsymbol{\sigma}(\mathbf{\Gamma}))\}$. Here, $\mathbf{\Gamma} = \mathbf{U}\,\mathrm{diag}(\boldsymbol{\sigma}(\mathbf{\Gamma}))\mathbf{V}^*$ is the singular-value decomposition of $\mathbf{\Gamma}$. When $\rho(\boldsymbol{\sigma}) = \sum_i \mu(\sigma_i)$, we have

$$\partial\phi(\mathbf{\Gamma}) = \sum_{i=1}^{\min(m,n)} \partial\mu(\sigma_i)\mathbf{u}_i\,\mathbf{v}_i^H \tag{10}$$

where $\partial\mu(\sigma_i)$ is the subgradient of $\mu(\sigma_i)$. Similarly, the convex conjugate of unitarily invariant penalties can be easily derived in terms of the convex conjugates of the corresponding μ functions. We use these results to extend the MM algorithms, which are originally developed for vector recovery, to matrix recovery problems.

E. Dynamic Imaging Using Matrix Recovery Schemes

Our motivation in developing this algorithm is to use it in dynamic imaging and video restoration. We denote the spatiotemporal signal as $\gamma(\mathbf{x}, t)$, where $\mathbf{x}$ is the spatial location and t denotes time. We denote the sparse and noisy measurements to be related to $\boldsymbol{\gamma}$ as $\mathbf{b} = \mathcal{A}(\gamma) + \mathbf{n}$, where $\mathcal{A}$ is the measurement operator and $\mathbf{n}$ is the noise process. The vectors corresponding to the temporal profiles of the voxels are often highly correlated or linearly dependent. The spatiotemporal signal $\gamma(\mathbf{x}, t)$ can be rearranged as a Casorati matrix to exploit the following correlations [3], [4], [6], [7]:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma(\mathbf{x}_0, t_0) & \dots & \gamma(\mathbf{x}_0, t_{n-1}) \\ \vdots & & \\ \gamma(\mathbf{x}_{m-1}, t_0) & \dots & \gamma(\mathbf{x}_{m-1}, t_{n-1}) \end{bmatrix}. \tag{11}$$

The ith row of $\mathbf{\Gamma}$ corresponds to the temporal intensity variations of voxel $\mathbf{x}_i$. Similarly, the jth column of $\mathbf{\Gamma}$ represents the image at the time point t_j. Since the rows of this $m \times n$ matrix are linearly dependent, the rank of $\mathbf{\Gamma}$ is given by $r < \min(m,n)$. We will refer to the dynamic imaging data set either as $\gamma(\mathbf{x}, t)$ or as $\mathbf{\Gamma}$ in the remaining sections. The low-rank structure of dynamic imaging data sets was used to recover them from undersampled Fourier measurements by several authors [3], [4], [6], [7], [23]. These schemes rely on either simpler two-step algorithms, which are relatively inefficient at high acceleration factors, or greedy low-rank decomposition schemes. In contrast with these methods, the proposed scheme is computationally efficient, accurate, highly flexible, and capable of using multiple nonconvex spectral and sparsity priors.

III. MM Algorithm for Matrix Recovery

We introduce the problem formulation and the algorithm here. The details of the numerical implementation are covered in Section IV.

A. Matrix Recovery Using Sparsity and Spectral Penalties

To exploit the low rank and the sparsity of the matrix in the transform domain (specified by $\mathbf{R}$ and $\mathbf{C}$), we formulate the matrix recovery as the following constrained optimization scheme:

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 \quad \text{such that } \left\{\mathrm{rank}(\mathbf{\Gamma}) \leq r, \|\mathbf{R}^H\mathbf{\Gamma}\mathbf{C}\|_{\ell_0} < K\right\}. \tag{12}$$

We rewrite the aforementioned constrained optimization problem using Lagrange's multipliers and relax the penalties to obtain

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 + \lambda_1\,\phi(\mathbf{\Gamma}) + \lambda_2\,\psi(\mathbf{\Gamma}). \tag{13}$$

Here, the spectral penalty ϕ is the relaxation of the rank constraint. We choose it as the class of Schatten-p matrix penalties ($\mu(\sigma) = (\sigma)^p$), specified by

$$\phi(\mathbf{\Gamma}) = \sum_{i=1}^{\min(m,n)} \sigma_i(\mathbf{\Gamma})^{p_1}. \tag{14}$$

Similarly, we specify the sparsity penalty as $\psi(\mathbf{\Gamma}) = \|\mathbf{R}^H\mathbf{\Gamma}\mathbf{C}\|_{\ell_{p2}}^{p_2}$, which is the ℓ_{p_2} norm of the matrix entries, specified by

$$\|\mathbf{\Gamma}\|_{\ell_p}^p = \sum_{i,j} |\mathbf{\Gamma}_{i,j}|^p. \tag{15}$$

When p_1 and $p_2 \geq 1$, the cost function (13) is convex and hence has a unique minimum. We now generalize the sparsity penalty to account for nonseparable convex and nonconvex TV-like penalties, i.e.,

$$\psi(\gamma) = \int_{\mathbb{R}^3} \|\nabla\gamma(\mathbf{x}, t)\|_2^{p_2}\; d\mathbf{x}dt \tag{16}$$

which are widely used in imaging applications [24], [25]. The aforementioned penalty is often implemented using finite-difference operators. Rewriting the aforementioned expression in terms of matrix $\mathbf{\Gamma}$, we get

$$\psi(\mathbf{\Gamma}) = \varphi(\mathbf{P}) = \|\mathbf{P}\|_{\ell_{p2}}^{p_2} \tag{17}$$

where $\mathbf{P} = \sqrt{\sum_{i=1}^{q} |\mathbf{R}_i^H \mathbf{\Gamma} \mathbf{C}_i|^2}$. Here, $\mathbf{R}_i$ and $\mathbf{C}_i$, $i = 1, \ldots, q$, are matrices that operate on the rows and the columns of $\mathbf{\Gamma}$, respectively. The nonseparable gradient penalty in (16) is thus obtained when $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ correspond to the finite differences of $\gamma(\mathbf{x}, t)$ along x, y, and t, respectively; $\mathbf{R}_i$ and $\mathbf{C}_i$ are the corresponding finite-difference matrices.

Gao *et al.*, have recently used a linear combination of sparse and low-rank matrices [26] to model dynamic imaging data set and recover it from undersampled measurements. They chose the regularization parameters such that the low-rank component is the static background signal. The dynamic components are assumed to be sparse in a preselected basis/frame, which is enforced by a convex sparsity prior. The use of a sparse model to capture the dynamic components is conceptually similar to classical compressed sensing dynamic imaging schemes [27]–[29]. We have shown that the basis functions estimated from the data itself (using low-rank recovery) are more effective in representing the data compared with preselected basis functions, particularly when the significant respiratory motion is present [6]. We plan to compare the proposed scheme with the model in [26] and other state-of-the-art dynamic imaging schemes in the future.

B. Algorithm Formulation

We now derive a fast MM algorithm to solve (13). Specifically, we majorize the penalty terms by quadratic functions of $\bar{\mathbf{\Gamma}}$, i.e.,

$$\phi(\mathbf{\Gamma}) = \min_{\mathbf{W}} \frac{\beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}\|_F^2 + \eta(\mathbf{W}) \tag{18}$$

$$\psi(\mathbf{\Gamma}) = \min_{\{\mathbf{Q}_i, i=1,\ldots,q\}} \frac{\beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma} \mathbf{C}_i - \mathbf{Q}_i\|_F^2 + \theta\left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}\right). \tag{19}$$

Here, $\mathbf{W}$ and $\mathbf{Q}_i, i = 1, \ldots, q$, are auxiliary matrix variables, and $\|\mathbf{\Gamma}\|_F$ is the Frobenius norm of $\mathbf{\Gamma}$. By definition, $\eta(\mathbf{W})$ and $\theta(\sqrt{\sum_i |\mathbf{Q}_i|^2})$ are matrix functions that are dependent on $\phi(\mathbf{W})$ and $\varphi(\mathbf{P})$, respectively.

Analytical expressions for η and θ can be derived in many cases, as shown below. However, we find in Section IV that analytical expressions for η and θ are not required for efficient implementation. Using the aforementioned majorizations, we simplify the original cost function in (13) as

$$(\mathbf{\Gamma}, \mathbf{W}, \{\mathbf{Q}_i\})_{\text{opt}} = \arg \min_{\mathbf{\Gamma}, \mathbf{W}, \{\mathbf{Q}_i\}} \mathcal{C}(\mathbf{\Gamma}, \mathbf{W}, \mathbf{Q}_i) \tag{20}$$

where

$$\mathcal{C} = \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{y}\|^2 + \frac{\lambda_1 \beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}\|_F^2 + \frac{\lambda_2 \beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma} \mathbf{C}_i - \mathbf{Q}_i\|_F^2 + \lambda_1\, \eta(\mathbf{W}) + \lambda_2\, \theta\left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}\right). \tag{21}$$

We propose to use an iterative alternating minimization scheme to minimize the aforementioned criterion. Specifically, we alternatively minimize (21) with respect to each of the variables, assuming others to be fixed. We denote the nth iterate of these variables as $\mathbf{\Gamma}_n$, $\mathbf{W}_n$, and $\mathbf{Q}_{i,n}; i = 1, \ldots, q$, respectively. One iteration of this scheme is described below.

1) Derive $\mathbf{\Gamma}_{n+1}$, assuming $\mathbf{W} = \mathbf{W}_n$ and $\mathbf{Q}_i = \mathbf{Q}_{i,n}$, i.e.,

$$\mathbf{\Gamma}_{n+1} = \arg \min_{\mathbf{\Gamma}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{y}\|^2 + \frac{\lambda_1 \beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}_n\|_F^2 + \frac{\lambda_2 \beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma} \mathbf{C}_i - \mathbf{Q}_{i,n}\|_F^2. \tag{22}$$

Since this expression is quadratic in $\mathbf{\Gamma}$, we derive the analytical solutions for many measurement operators in Section IV.

2) Derive $\mathbf{W}_{n+1}$, assuming $\mathbf{\Gamma} = \mathbf{\Gamma}_{n+1}$, i.e.,

$$\mathbf{W}_{n+1} = \arg \min_{\mathbf{W}} \frac{\beta_1}{2} \|\mathbf{\Gamma}_{n+1} - \mathbf{W}\|_F^2 + \eta(\mathbf{W}). \tag{23}$$

The optimal $\mathbf{W}$ is thus obtained as the proximal mapping of $\mathbf{\Gamma}_{n+1}$, corresponding to the spectral penalty η. We derive analytical expressions for this step for the widely used nuclear-norm and Schatten-p functionals in Section IV.

3) Derive $\mathbf{Q}_{i,n+1}$, assuming $\bar{\mathbf{\Gamma}} = \bar{\mathbf{\Gamma}}_{n+1}$, i.e.,

$$\mathbf{Q}_{i,n+1} = \arg \min_{\mathbf{Q}_i} \frac{\beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma}_{n+1} \mathbf{C}_i - \mathbf{Q}_i\|_F^2 + \theta\left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}\right). \tag{24}$$

The optimal $\{\mathbf{Q}, i = 1, \ldots, q\}$ is thus the proximal mapping of $\{\mathbf{R}_i \mathbf{\Gamma}_{n+1} \mathbf{C}_i; i = 1, \ldots, q\}$, associated with the matrix penalty θ. Since θ is nonseparable, the corresponding shrinkage involves the simultaneous processing of the component matrices $\mathbf{R}_i \mathbf{\Gamma}_{n+1} \mathbf{C}_i; i = 1, \ldots, q$. This step also has analytical expressions, as shown in Section IV.

C. Expression of $\eta(\mathbf{W})$

We now focus on determining function η, such that the majorization of the spectral penalty term in (18) holds. Since analytical expressions for η and θ are not essential to realize an efficient algorithm, readers may skip this section and go directly to Section IV.

We reorder the terms in (18) to obtain

$$\underbrace{\frac{\|\mathbf{\Gamma}\|_F^2}{2} - \frac{\phi(\mathbf{\Gamma})}{\beta_1}}_{g(\mathbf{\Gamma})} = \max_{\mathbf{W}} \left[\langle \mathbf{\Gamma}, \mathbf{W} \rangle - \underbrace{\left(\|\mathbf{W}\|_F^2/2 + \eta(\mathbf{W})/\beta_1 \right)}_{f(\mathbf{W})} \right]. \tag{25}$$

Here, $\langle \mathbf{W}, \mathbf{\Gamma} \rangle = \text{trace}(\mathbf{W}^T \mathbf{\Gamma})$ is the inner product of two matrices. From the theory in [22], the aforementioned relation is satisfied if $g(\mathbf{\Gamma})$ is a convex function and $f = g^*$ is the convex dual of g, i.e.,

$$g^*(\mathbf{W}) = \max_{\mathbf{\Gamma}} \left(\langle \mathbf{W}, \mathbf{\Gamma} \rangle - g(\mathbf{\Gamma}) \right). \tag{26}$$

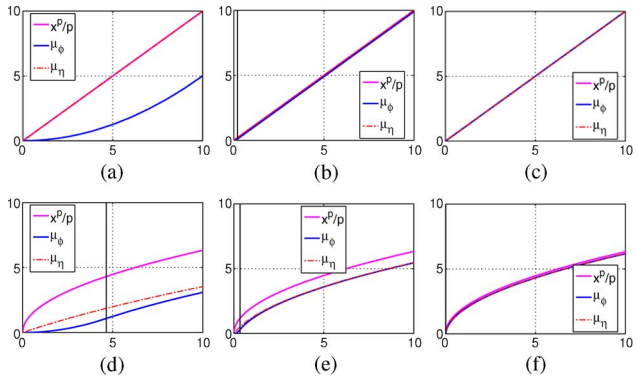


Fig. 1. Huber approximation of the spectral penalty and the corresponding η function. Note that the approximation of the original spectral penalty by the Huber function improves with increasing values of β. Clearly, large values of β are required to approximate Schatten-p norms; $p < 1$. It is observed that $\eta(x) = |x|$, $\forall \beta$, when $p = 1$. Hence, the VS interpretation (see Section IV-E) is equivalent to the MM scheme. However, this equivalence breaks down when $p < 1$. Specifically, $\eta(\mathbf{\Gamma}) \to \|\mathbf{\Gamma}\|_p^p$ only when $\beta \to \infty$. (a) $p = 1; \beta_1 = 0.1$. (b) $p = 1; \beta_1 = 5$. (c) $p = 1; \beta_1 = 1000$. (d) $p = 0.5; \beta_1 = 0.1$. (e) $p = 0.5; \beta_1 = 5$. (f) $p = 0.5; \beta_1 = 1000$.

Note that ϕ need not be convex for the aforementioned relation to hold. This majorization is valid if $g(\mathbf{\Gamma})$ is convex, which is possible even when ϕ is concave. Due to the property of unitarily invariant functions, the dual of a specified matrix function $g(\mathbf{\Gamma}) = \sum \mu_g(\sigma_i(\mathbf{\Gamma}))$ is obtained as

$$f(\mathbf{W}) = \sum \mu_g^*(\sigma_i(\mathbf{W})). \tag{27}$$

Thus, $\mu_f(\cdot) = \mu_g^*(\cdot)$ is the convex conjugate of $\mu_g(\cdot)$. From the aforementioned relations, we have $\eta(\mathbf{W}) = \sum \mu_\eta(\sigma_i(\mathbf{W}))$, where $\mu_\eta(x) = \beta_1(\mu_f(x) - x^2/2)$.

We now approximate the nondifferentiable ϕ penalties by continuously differentiable Huber functionals. These approximations are required to ensure that $g(\mathbf{\Gamma})$ is convex. In addition, the differentiability of ϕ also provides additional simplifications.

1) Nuclear norm: We approximate the nuclear-norm penalty $\|\mathbf{\Gamma}\|_* = \sum_i \sigma_i(\mathbf{\Gamma})$ as $\phi_{\beta_1}(\mathbf{\Gamma}) = \sum_i \mu_{\phi_{\beta_1}}(\sigma_i(\mathbf{\Gamma}))$. Here, $\mu_{\phi_{\beta_1}}(x)$ is the standard scalar Huber function, i.e.,

$$\mu_{\phi_{\beta_1}}(x) = \begin{cases} |x| - 1/2\beta_1, & \text{if } x \geq \frac{1}{\beta_1} \\ \beta_1 |x|^2/2, & \text{else.} \end{cases} \tag{28}$$

Note that $\phi_{\beta_1}(\mathbf{\Gamma}) \to \|\mathbf{\Gamma}\|_*$ as $\beta_1 \to \infty$. With this choice, the corresponding $g(\mathbf{\Gamma}) = \sum \mu_g(\sigma_i(\mathbf{\Gamma}))$ is given by

$$\mu_g(x) = \begin{cases} \frac{1}{2}\left(x - \frac{1}{\beta_1}\right)^2, & \text{if } x \geq \frac{1}{\beta} \\ 0 & \text{else.} \end{cases} \tag{29}$$

Note that g is convex for any β_1. Using the property of convex conjugate functions described earlier, we find in the Appendix that $\mu_\eta = w$. Thus, we have $\eta(\mathbf{W}) = \|\mathbf{W}\|_*$, $\forall \beta_1$.

2) Schatten-p norm: We approximate the Schatten-p matrix norm by the corresponding Huber matrix function, i.e.,

$$\mu_{\phi_{\beta_1}}(x) = \begin{cases} \frac{x^p}{p} - 1/\left(2\alpha\beta_1^\alpha\right), & \text{if } x \geq \beta_1^{1/(p-2)} \\ \beta_1 x^2/2, & \text{else.} \end{cases} \tag{30}$$

Here, $\alpha = p/(2-p)$. The threshold specified by $\beta_1^{1/(p-2)}$ and constant $1/(2\alpha\beta_1^\alpha)$ is chosen such that $\mu_{\phi_{\beta_1}}$ is continuously differentiable and μ_g is convex. The aforementioned formula is essentially an extension of the generalization proposed by [24] to matrix functionals. It is difficult to derive analytical expressions for $\mu_\eta(w)$ for arbitrary $p < 1$. However, we can numerically solve for $x = \partial_x \mu_g$ and evaluate $\mu_\eta(w)$ for specific values of w, as shown in Fig. 1. We show in the next section that analytical expressions for the proximal mapping, specified by (23), can be derived even if analytical expressions for η are not available.

We plot μ_ϕ, μ_η, and x^p/p for $p = 1$ and $p = 0.5$ for different values of β_1 in Fig. 1. Note that $\mu_\eta(x) = |x|$, $\forall \beta_1$, when $p = 1$. However, $\mu_\eta(x)$ is different from $|x|^p/p$ when $p < 1$. This implies that the VS interpretation of the MM algorithm breaks down when $p < 1$, as explained in Section IV.

D. Expression for θ

The Huber approximation of the TV norm $(p_2 = 1)$ was considered in [30], where they showed that

$$\theta\left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}\right) = \sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2} \qquad \forall \beta_2. \tag{31}$$

Analytical expressions of θ cannot be obtained when $p_2 < 1$. However, we derive the analytical expression for the shrinkage in Section IV, which will enable the efficient implementation.

IV. Numerical Algorithm

We now focus on the numerical implementation of the three main subproblems. Specifically, we show that all of the three steps can be solved either analytically or using efficient algorithms for most penalties and measurement operators of practical interest. This enables us to realize a computationally efficient algorithm. We also introduce a continuation scheme to accelerate the convergence of the algorithm.

A. Quadratic Subproblem Specified by (22)

Since subproblem (22) is entirely quadratic, we rewrite it as a Tikhonov regularized image recovery problem, i.e.,

$$\boldsymbol{\gamma}_{n+1} = \arg\min_{\boldsymbol{\gamma}} \|\mathcal{A}(\boldsymbol{\gamma}) - \mathbf{y}\|^2 + \frac{\lambda_1\beta_1}{2}\|\boldsymbol{\gamma} - \mathbf{w}_n\|^2 + \frac{\lambda_2\beta_2}{2}\sum_{i=1}^{q}\|\mathcal{G}_i(\boldsymbol{\gamma}) - \mathbf{q}_{n,i}\|^2. \tag{32}$$

Here, $\boldsymbol{\gamma} \leftrightarrow \mathbf{\Gamma}$ and $\boldsymbol{q}_i \leftrightarrow \mathbf{Q}_i$ are the 3-D data sets corresponding to the corresponding Casorati matrices. Similarly, $\mathcal{G}_i$ is the linear operator such that $\mathcal{G}_i(\boldsymbol{\gamma}) \leftrightarrow \mathbf{R}_i^*\mathbf{\Gamma}\mathbf{C}_i$. We obtain the Euler–Lagrange equation of this variational problem as

$$\left(\mathcal{A}^*\mathcal{A} + \lambda_1\beta_1\mathcal{I} + \lambda_2\beta_2\sum_{i=1}^{q}\mathcal{G}_i^*\mathcal{G}_i\right)\boldsymbol{\gamma}_{n+1} = \mathcal{A}^*\mathbf{y} + \lambda_1\beta_1\,\mathbf{w}_n + \lambda_2\beta_2\sum_{i=1}^{q}\mathcal{G}_i^*\mathbf{q}_{n,i}. \tag{33}$$

Here, $\mathcal{I}$ is the identity operator. Note that the variables in the right-hand side of (33) are fixed. Thus, this step involves the solution to a linear system of equations. In the general setting, this system of equations can be efficiently solved using iterative algorithms such as the CG. A few CG steps are often sufficient for good convergence since the algorithm is initialized by the previous iterate $\boldsymbol{\gamma}_n$. We now show that analytical solutions of (32) do exist for many measurement operators. When the TV penalty is used, the aforementioned equation can be rewritten as

$$(\mathcal{A}^*\mathcal{A}+\lambda_1\beta_1\,\mathcal{I}+\lambda_2\beta_2\,\triangle)\boldsymbol{\gamma}_{n+1}=\mathcal{A}^*\mathbf{y}+\lambda_1\beta_1\,\mathbf{w}_n+\lambda_2\beta_2\,p_n. \tag{34}$$

Here, $\triangle\boldsymbol{\gamma}_{n+1}$ is the 3-D Laplacian of $\boldsymbol{\gamma}_{n+1}$, and $p_n=\nabla\cdot\mathbf{q}_n$ is the divergence of the vector field $\mathbf{q}_n$.

- Fourier sampling: An analytical expression can be derived for (32), when the measurements are Fourier samples of $\boldsymbol{\gamma}$ on a Cartesian grid. Specifically, we assume that the index set corresponding to the measured samples to be indicated by Λ and the corresponding measurements to be b_i; i.e., $(b_i=\hat{\gamma}(\boldsymbol{\omega}_i)$, where $\hat{\gamma}$ is the discrete Fourier transform of $\gamma)$. We split the frequency samples, specified by ω, into two sets Λ and Λ^c and evaluate the discrete Fourier transform of both sides of (34) to obtain

$$\hat{\gamma}_{n+1}(\boldsymbol{\omega}_i)=\begin{cases}\frac{b_i+\lambda_1\beta_1\,\hat{w}_n(\boldsymbol{\omega}_i)+\lambda_2\beta_2\,\hat{p}_n(\boldsymbol{\omega}_i)}{1+\lambda_1\beta_1+\lambda_2\beta_2\|\boldsymbol{\omega}_i\|^2}, & \text{if } \boldsymbol{\omega}_i\in\Lambda\\ \frac{\lambda_1\beta_1\,\hat{w}_n(\boldsymbol{\omega}_i)+\lambda_2\beta_2\,\hat{p}_n(\boldsymbol{\omega}_i)}{\lambda_1\beta_1+\lambda_2\beta_2\|\boldsymbol{\omega}_i\|^2}, & \text{else.}\end{cases} \tag{35}$$

  Here, $p_n=\nabla\cdot\mathbf{q}_n$.
- Deconvolution: Convolution can be posed as a multiplication in the Fourier domain. Considering the Fourier transform of the matrix, (34) can be solved in the Fourier domain as

$$\hat{\gamma}_{n+1}(\boldsymbol{\omega})=\frac{\hat{H}(\boldsymbol{\omega})^*\hat{b}(\boldsymbol{\omega})+\lambda_1\beta_1\,\hat{w}_n(\boldsymbol{\omega})+\lambda_2\beta_2\,\hat{p}_n(\boldsymbol{\omega})}{\left|\hat{H}(\boldsymbol{\omega})\right|^2+\lambda_1\beta_1+\ \lambda_2\beta_2\|\boldsymbol{\omega}\|^2}. \tag{36}$$

  Here, $H(\boldsymbol{\omega})$ is the transfer function of the blurring filter, $\boldsymbol{\omega}$ is the frequency, $\hat{b}(\boldsymbol{\omega})$ is the Fourier transform of the measured blurred image data set, and $\hat{p}_n(\boldsymbol{\omega})$ is the Fourier transform of $\nabla\cdot\mathbf{q}_n$.

B. Subproblem 2 Specified by (23)

We will now focus on (23) and derive the analytical expression for $\mathbf{W}_{n+1}$, i.e.,

$$\begin{aligned}\mathbf{W}_{n+1}&=\arg\min_{\mathbf{W}}\overbrace{\beta\|\boldsymbol{\Gamma}_{n+1}\|^2/2+\beta\|\mathbf{W}\|_F^2/2-\beta\langle\boldsymbol{\Gamma}_{n+1},\mathbf{W}\rangle}^{\frac{\beta}{2}\|\boldsymbol{\Gamma}_n-\mathbf{W}\|_F^2}\\&\quad+\eta(\mathbf{W})\\&=\arg\max_{\mathbf{W}}\beta\langle\boldsymbol{\Gamma}_{n+1},\mathbf{W}\rangle-\beta\underbrace{\left(\|\mathbf{W}\|_F^2/2+\eta(\mathbf{W})/\beta\right)}_{f(\mathbf{W})}.\end{aligned} \tag{37}$$

The minimizer of this expression satisfies

$$\boldsymbol{\Gamma}_{n+1}=\beta\,\nabla f(\mathbf{W}_{n+1}). \tag{38}$$

We used the differentiability of ϕ, and hence f to obtain the aforementioned expression. This is valid for the Huber approximations of the spectral penalties. Since f and g are convex conjugates, ∇f and ∇g are inverse functions [19], [31]. Thus, we obtain the optimal $\mathbf{W}$ that solves (37) as

$$\begin{aligned}\mathbf{W}_{n+1}=\nabla f^{-1}(\boldsymbol{\Gamma}_{n+1})&=\nabla g(\boldsymbol{\Gamma}_{n+1})\\&=\boldsymbol{\Gamma}_{n+1}-\partial\phi(\boldsymbol{\Gamma}_{n+1})/\beta.\end{aligned} \tag{39}$$

We used relation $g(\boldsymbol{\Gamma})=\|\boldsymbol{\Gamma}\|_F^2/2-\phi(\boldsymbol{\Gamma})/\beta$ in the second step. Thus, analytical expressions for η are not required to derive the shrinkage step, due to the property of convex conjugate functions. We now derive the shrinkage steps for specific spectral penalties.

- Special case: Nuclear norm, i.e.,

$$\partial\mu_\phi(x)=\begin{cases}\text{sign}(x), & \text{if } |x|\geq\frac{1}{\beta_1}\\ \beta_1 x, & \text{else.}\end{cases} \tag{40}$$

  We assume $\boldsymbol{\Gamma}=\sum_{i=1}^{\min(m,n)}\sigma_i\mathbf{u}_i\,\mathbf{v}_i^H$ to be the singular-value decomposition of $\boldsymbol{\Gamma}$. Substituting in (39), we get

$$\mathbf{W}^*=\sum_{i=1}^{\min(m,n)}(\sigma_i-1/\beta_1)_+\mathbf{u}_i\,\mathbf{v}_i^H \tag{41}$$

  where $\mathbf{u}_i$, $\mathbf{v}_i$, and σ_i are the singular vectors and values of $\boldsymbol{\Gamma}$.
- Schatten-p norms: Following the same steps, we obtain the shrinkage step for Schatten-p norms as

$$\mathbf{W}^*=\partial g(\boldsymbol{\Gamma})=\sum_{i=1}^{\min(m,n)}\left(\sigma_i-\sigma_i^{(p_1-1)}/\beta_1\right)_+\mathbf{u}_i\,\mathbf{v}_i^H. \tag{42}$$

C. Solving Subproblem 3 Specified by (24)

Problems similar to (24) have been addressed in the context of iterative algorithms for TV minimization [30] and its nonconvex variants [24]. The generalized shrinkage rule to derive $\mathbf{Q}_i; i=1,\ldots,q$ is specified by

$$\mathbf{Q}_{i,n+1}=\frac{(\mathbf{P}-\mathbf{P}^{(p_2-1)}/\beta_2)}{\mathbf{P}}\,\mathbf{R}_i\boldsymbol{\Gamma}\mathbf{C}_i \tag{43}$$

where $\mathbf{P}=\sqrt{\sum_{i=1}^q|\mathbf{R}_i\boldsymbol{\Gamma}_{n+1}\mathbf{C}_i|^2}$. $\mathbf{P}^p$ is the matrix whose elements are the pth power of the entries of $\mathbf{P}$.

D. Continuation to Improve the Convergence

The three-step alternating minimization algorithm involves a tradeoff between convergence and accuracy. Specifically, when $\beta_1=\beta_2=0$, (21) simplifies to three decoupled problems in $\boldsymbol{\Gamma}$, $\mathbf{W}$, and $\{\mathbf{Q}_i, i=1,\ldots,q\}$. Since all of these problems have analytical solutions, the entire algorithm converges in a single step to the minimum norm solution, which is a poor approximation of (13); this is expected since the corresponding Huber

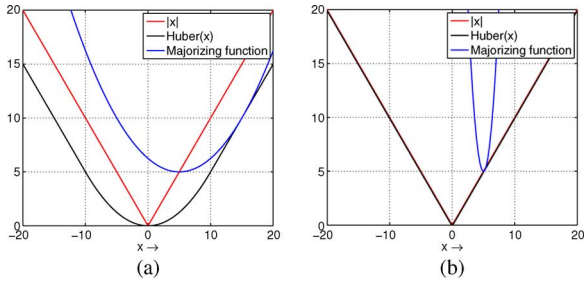


Fig. 2. Effect of β_1 on convergence and accuracy. We demonstrate the approximation of $|x|$ by the corresponding Huber penalty. Note that, for $\beta = 0.1$, the majorizing function approximates the Huber function well, resulting in fast convergence to the minimum of the corresponding penalty. However, in this case, the approximation of $|x|$ by the Huber function is poor, resulting in poor accuracy. In contrast, the approximation of $|x|$ by the Huber function is good when $\beta = 5$. In this case, the majorizing function of the Huber function is a poor approximation. Specifically, it is too narrow, resulting in slow convergence. (a) $\beta = 0.1$. (b) $\beta = 5$.

function is a poor approximation to the original cost function. In contrast, the approximation is exact when $\beta_1 = \beta_2 = \infty$. However, it is easy to see that the algorithm fails to converge in this case.

The aforementioned tradeoff between convergence and accuracy can be understood in terms of the ability of the Huber function to approximate the original penalty and the proximity of the majorizing function to the Huber approximation. We illustrate this issue in Fig. 2 in the context of the nuclear-norm penalty. Note that, for small values of β, the Huber approximation $\mu_\phi(\sigma)$ of $|\sigma|$ is poor. However, the corresponding quadratic majorizing function $\beta(x-w)^2 + \mu_\eta(\sigma)$ closely approximates μ_ϕ. Hence, the MM scheme converges fast to the minimum of the approximate penalty. In contrast, when $\beta \to \infty$, the Huber function approximates the spectral penalty well, resulting in the good accuracy of the final solution. However, the convergence is poor in this case since the approximation of the Huber function by the majorizing quadratic function is poor.

To overcome the aforementioned tradeoff, we introduce a continuation scheme. Specifically, we initialize β_1 and β_2 with small values and progressively increase them until convergence. The algorithm converges very fast for small values of β_1 and β_2, as discussed before. We use the solution at each step to initialize the next step. For each choice of continuation parameters, we iterate the algorithm to convergence (i.e., until the relative change in the cost function in (13) is less than a prespecified threshold). In all the experiments considered in this paper, we initialize the continuation parameters as $\beta_1 = \beta_2 = 5$ and increase them by a factor of five for each iteration of the outer loop.

E. Interpretation as a VS Scheme

The MM scheme to solve for the spectrally regularized matrix recovery may be interpreted as a VS strategy [6], similar to such schemes in compressed sensing [32], [33], i.e.,

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma},\mathbf{W}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 + \lambda_1\, \phi(\mathbf{W}) \quad \text{s.t. } \mathbf{\Gamma} = \mathbf{W}. \tag{44}$$

Here, $\mathbf{W}$ is an auxiliary variable, and the aforementioned constrained optimization problem is equivalent to (13), when $\lambda_2 = 0$. VS algorithms convert the aforementioned constrained optimization problem to an unconstrained problem by introducing an additional quadratic penalty, i.e.,

$$(\mathbf{\Gamma},\mathbf{W})^* = \arg\min_{\mathbf{\Gamma},\mathbf{W}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 + \lambda_1\, \phi(\mathbf{W}) + \lambda_1 \frac{\beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}\|_F^2. \tag{45}$$

This unconstrained problem is equivalent to (44), when $\beta_1 \to \infty$. Note that (45) is very similar to (21) with $\lambda_2 = 0$, except that ϕ is used instead of η. The VS and MM schemes are exactly the same for $p = 1$, since $\eta(\mathbf{W}) = \|\mathbf{W}\|_*$, $\forall \beta_1$. However, $\eta(\mathbf{W}) \neq \|\mathbf{W}\|_p^p$ when $p < 1$ (see Fig. 1). Hence, the two schemes are not equivalent for general Schatten-p norms.

The standard practice in VS compressed sensing schemes is to alternatively minimize the criterion with respect to each of the unknowns, assuming the other variable to be fixed [32], [33]. Thus, we obtain $\mathbf{W}_{n+1}$ as

$$\mathbf{W}_{n+1} = \arg\min_{\mathbf{W}} \frac{\beta_1}{2} \|\mathbf{\Gamma}_{n+1} - \mathbf{W}\|_F^2 + \phi(\mathbf{W}) \tag{46}$$

When $\phi(\mathbf{\Gamma}) = \|\mathbf{\Gamma}\|_*$ [1], this proximal mapping can be efficiently implemented using singular-value soft thresholding. However, analytical closed-form expressions for the aforementioned proximal mapping do not exist when nonconvex spectral penalties (e.g., Schatten-p norms; $\phi(\mathbf{\Gamma}) = \|\mathbf{\Gamma}\|_p^p$) are used. Iterative algorithms may be used to solve this proximal mapping. However, it will be computationally more expensive since it involves an additional inner loop. Ehler recently demonstrated that approximating the proximal mapping of nonconvex ℓ_p penalties with hard-thresholding shrinkage rules will provide computationally efficient solutions [20]. Adapting these approximate rules to nonconvex matrix penalties, we get

$$\mathbf{W}_{n+1} = \sum_{i=1}^{\min(m,n)} \chi(\sigma_i) \mathbf{u}_i\, \mathbf{v}_i^H. \tag{47}$$

The thresholding function in (47) is defined as

$$\chi(\sigma_i) = \begin{cases} \sigma_i - \frac{2p\, c_p}{\beta_1}, & \text{if } \sigma_i > \frac{2c_p}{\beta_1} \\ 0, & \text{else.} \end{cases} \tag{48}$$

where $c_p = 2^{p-2}((2-p)^{2-p}/(1-p)^{1-p})$ and $\mathbf{u}_i$, $\mathbf{v}_i$, and $\boldsymbol{\sigma}_i$ are the singular vectors and values of $\mathbf{\Gamma}_{n+1}$, respectively. A similar approximate rule can be used for the nonconvex TV penalty. In contrast with these approximate rules, analytical shrinkage formulas can be derived for most spectral penalties $\eta(\mathbf{\Gamma})$ in the MM framework [see (39)]. We compare the MM algorithm and the VS scheme with the approximate shrinkage rule in the results section; we observe that the proposed MM scheme provides faster convergence, due to the exact shrinkage rule.

The MM scheme in the nonconvex $(p_1 < 1; p_2 < 1)$ case may be alternatively interpreted as an approximate VS algorithm, where (42) and (43) are used to approximate the corresponding proximal mappings. Note that these approximations are not reported before and is only inspired by the MM framework. The aforementioned approximate VS scheme (with approximate MM-inspired proximal mappings) can be further accelerated using augmented-Lagrangian (AL) or split-Bregman

(SB) methods [34], [35], as shown in [36]. Unlike the conventional AL or SB method (originally developed for convex penalties), the proposed nonconvex schemes still require continuation since the approximation is only exact as $\beta \to \infty$ (see Fig. 1.)

V. Results

We will demonstrate the utility of the combined nonconvex penalty in reliably recovering sparse and low-rank matrices in Section V-A. The convergence of the algorithm and the utility of continuation will be studied in Section V-B. In Section V-C, we demonstrate the utility of the combined nonconvex penalty in recovering dynamic contrast-enhanced (DCE) magnetic resonance (MR) images from their undersampled Fourier measurements. The dynamic MR imaging (MRI) data set is only approximately of low rank and sparse.

A. Recovery of Low-Rank and Sparse Matrices

We first demonstrate the benefits in using the combination of two nonconvex penalties, as compared with widely used nuclear-norm scheme. We consider the recovery of the MIT logo from its sparse measurements to illustrate the algorithm. This image matrix (size of 46×81 with 3726 pixels) is ideal for our study since it is of low rank $(\text{rank} = 5)$ and also has sparse gradients. Note that the signal is naturally a matrix, and hence, we do not use Casorati matrix formulation as in (11). We use random measurement matrices and vary the number of measurements M from 100 to 1500 as in [1]. The matrix was then recovered from these measurements using (13) with six different parameter settings:

1. nuclear-norm penalty alone $(p_1 = 1; \lambda_2 = 0)$;
2. nonconvex spectral penalty alone $(p_1 = 0.5; \lambda_2 = 0)$;
3. standard TV penalty alone $(\lambda_1 = 0; p_2 = 1)$;
4. nonconvex gradient penalty alone $(\lambda_1 = 0; p_2 = 0.5)$;
5. combination of both convex penalties $(p_1 = p_2 = 1)$;
6. combination of both nonconvex penalties $(p_1 = p_2 = 0.5)$.

We repeat each experiment for ten different random measurement ensembles and evaluate the average signal-to-noise ratio (SNR), specified as

$$\text{SNR} = 20 \log \left(\frac{||\mathbf{\Gamma}_{\text{orig}}||_F}{||\mathbf{\Gamma}_{\text{rec}} - \mathbf{\Gamma}_{\text{orig}}||_F} \right). \tag{49}$$

For each setting, we optimize the regularization parameters (λ_1 and λ_2) with respect to SNR. Fig. 3 shows the SNR of the recovered image as a function of the number of measurements. It is shown that the SNR abruptly rises when the number of measurements exceeds a specified threshold. An SNR of 80 dB corresponds to almost perfect reconstruction.

We observe that the algorithm with the conventional nuclear-norm scheme can perfectly recover the image if the number of measurements is greater than 1300; these findings are consistent with the results in [1]. In contrast, the nonconvex spectral penalty alone requires only 900 measurements. Similarly, the algorithm with the nonconvex TV penalty alone requires only 400 measurements to perfectly recover the image, as compared with 800 with standard TV. These results demonstrate the benefit in using nonconvex penalties over convex schemes. We did not encounter any local-minimum issues. We believe that the continuation strategy, where the cost function is initialized as a quadratic criterion and gradually made nonconvex, minimizes the local-minimum problems.

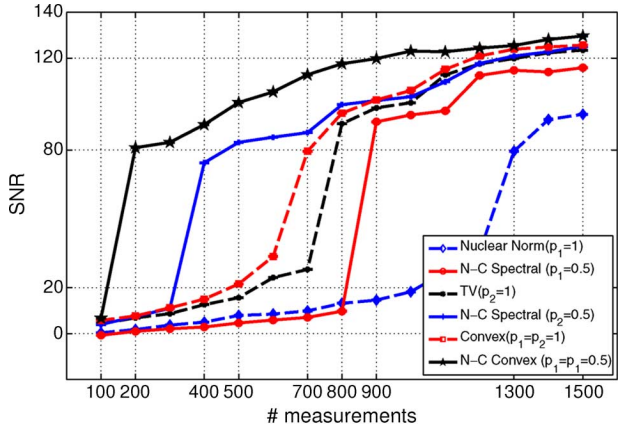


Fig. 3. Utility of the combination of nonconvex penalties. We plot the SNR as a function of the number of measurements on the MIT logo, recovered using the six algorithms. Note that the SNR abruptly increases when the number of measurements exceeds a specified threshold. It is seen that the algorithms using the (red solid curve) nonconvex spectral and (blue solid curve) nonconvex TV penalties alone reduce the number of measurements required to recover the image considerably over (blue dotted and black dotted curves, respectively) their convex counterparts. We also observe that the combination of the (red dotted) convex and (black solid) nonconvex penalties work much better than the individual penalties.

It is seen that the combined convex penalty (TV and nuclear norm) requires approximately 700 measurements, as compared with 1300 with the nuclear norm alone and 800 with TV alone. Similarly, the combination of the nonconvex penalties requires only 200 measurements, as compared with 900 with nonconvex spectral penalty alone and 400 with nonconvex gradient penalty alone. These experiments demonstrate a significant reduction in the number of measurements required to recover a matrix when sparsity and spectral penalties are combined. As described earlier, sparsity and low-rank properties are complementary; since the degrees of freedom of matrices that are simultaneously sparse and of low rank are small, the joint penalty is capable of significantly reducing the number of measurements. Sample images of recovered matrices for four different number of measurements are shown in Fig. 4.

B. Convergence of the Algorithm

We now study the effect of parameters β_1 and β_2 on the convergence of the algorithm and the accuracy of the solution. We consider the recovery of the MIT logo from $M = 1000$ measurements using the combined nonconvex penalties $(p_1 = p_2 = 0.5)$. We plot the evolution of the original cost function in (13) and the SNR with respect to the number of iterations in Fig. 5(a) and (b), respectively. It is observed that lower values of β_1 and β_2 result in fast convergence but yield solutions with higher cost and lower SNR. This is expected since (21) is a poor approximation to (13). In contrast, higher values of β_1 and β_2 approximate the original cost function well but result in slow convergence. We observe that the proposed continuation scheme, where β_1 and β_2 are initialized with small values and are gradually increased, offers the best compromise. In this

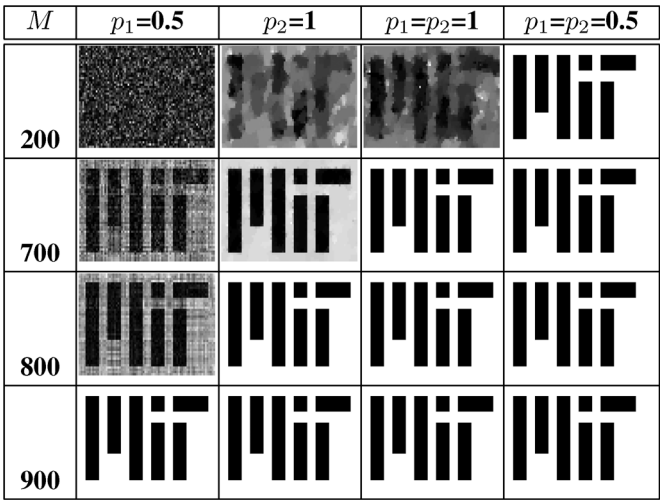


Fig. 4. Sample recovered images for four different number of measurements using nonconvex spectral penalty ($p_1 = 0.5$), standard TV ($p_2 = 1$), combination of convex penalties ($p_1 = p_2 = 1$), and combination of nonconvex penalties ($p_1 = p_2 = 0.5$). The rows correspond to different number of measurements. The results show that almost perfect reconstruction can be obtained when the number of measurements is larger than 200 by using the combination of both nonconvex penalties ($p_1 = p_2 = 0.5$). Note that this is six times lower than using only the nuclear-norm penalty (1300 measurements) and four times lower than standard TV (800 measurements). Similarly, we obtain a fourfold improvement over nonconvex spectral penalty alone.

specific example, the continuation scheme converged in 599 iterations. In contrast, the schemes with fixed values of β_1 and β_2 require far more number of iterations. The images recovered after 500 iterations using different parameter choices are shown in Fig. 6. We study the effect of different continuation schemes with different convergence rates in Fig. 7. We use fixed incrementing strategies in the first three experiments, where β_1 and β_2 are incremented by a fixed β_{inc}. We also tried several dynamic scheduling schemes. The fourth experiment shows the one with comparative best results, where β_{inc} is increased at each outer iteration, e.g., first iteration, $\beta_{\text{inc}} = 10$; second iteration, $\beta_{\text{inc}} = 20 \ldots$. The plots show that the different schemes converge to the same solution, but the convergence rates are dependent on the specific continuation strategy. Note that the convergence rate of the algorithms do not significantly vary with the specific continuation strategy, indicating that the algorithm is not too sensitive to β_{inc}.

We compare the proposed MM algorithm with the approximate VS scheme in the context of the spectrally regularized nonconvex criterion in Fig. 8; we set $\lambda_2 = 0$. The VS scheme relies on the approximate shrinkage rule, specified by (47). We compare the algorithms for $p_1 = 0.3$ and $p_1 = 0.5$, respectively. We consider $M = 1500$ measurements in Fig. 8(a), whereas $M = 1000$ measurements are studied in Fig. 8(b). It is observed that the MM scheme (blue and black curves) significantly outperforms the approximate VS scheme (red and magenta curves), particularly when p_1 and M are small. This experiments clearly demonstrate the utility of the MM scheme for nonconvex penalties. As discussed earlier, both the algorithms are essentially the same when $p = 1$.

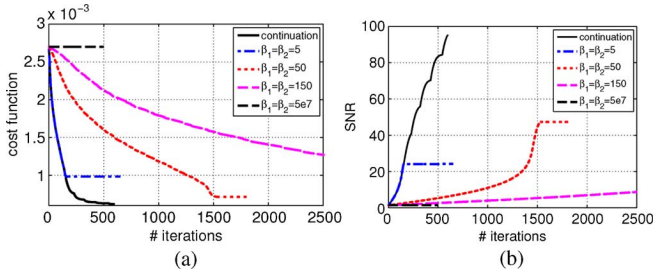


Fig. 5. Utility of the continuation scheme. We plot the cost and the SNR as a function of the number of iterations. We observe that lower values of the parameters ($\beta_1 = \beta_2 = 5$) result in a very fast convergence but yield a solution with higher cost and lower SNR. Higher values of the parameters improve the accuracy at the expense of the number of iterations. Note that the continuation strategy, where the parameters are initialized with $\beta_1 = \beta_2 = 5$ and increased by a factor of 5 within the outer loop, results in fast convergence and solutions with good SNR. We terminate the algorithm when the cost does not change, where the convergence is achieved with $\beta_1 = \beta_2 = 5e7$. We observe that the algorithm fails to converge if it is initialized with these parameters. (a) Cost function to iterations. (b) SNR to iterations.

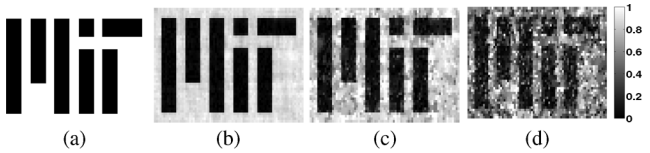


Fig. 6. Utility of continuation schemes in matrix recovery. We reconstruct the MIT logo with continuation and different fixed values of β. We show the reconstructions using (a) continuation scheme (initialized with $\beta_1 = \beta_2 = 5$ and gradually increased by a factor of 5) SNR $= 100.64$; (b) $\beta_1 = \beta_2 = 5$, SNR $= 22.41$; (c) $\beta_1 = \beta_2 = 50$, SNR $= 12.14$; and (d) $\beta_1 = \beta_2 = 150$, SNR $= 5.98$ after 500 iterations. Note that the continuation scheme provides almost perfect recovery, whereas the other methods result in artifacts.

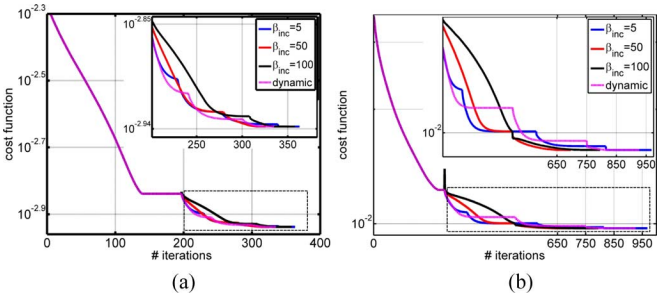


Fig. 7. Comparison of continuation strategies. We consider four schemes, which start with the same initial value of β_1 and β_2, but increase them by different factors (β_{inc}). In the first three schemes, β_{inc} values are fixed. In the fourth scheme (dynamic continuation), we increase β_{inc} at each outer iteration. We observe that the final result the algorithms converge to are the same. In addition, the convergence rate of the different continuation strategies does not significantly vary; this indicates that the algorithm does not require significant tuning of the continuation strategy. (a) $M = 1000$. (b) $M = 500$.

C. Accelerating Dynamic Contrast-Enhanced MRI

Here, we illustrate the utility of the proposed algorithm in accelerating DCE MRI. DCE MRI tracks the dynamic variations in the image intensity, resulting from the passage of a tracer bolus. Specifically, the paramagnetic tracer within the vasculature results in spin dephasing, hence resulting in decreased signal intensity. DCE MRI has shown great potential in diagnosing malignant lesions in the brain, breast, and other

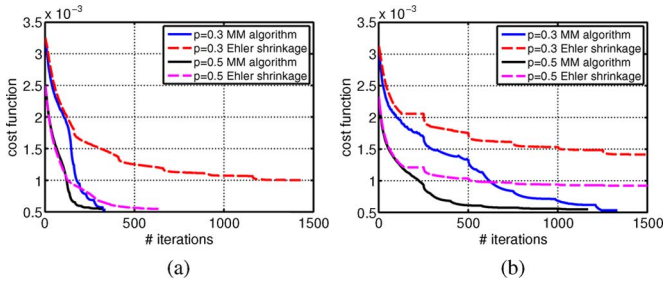


Fig. 8. Comparison of approximate VS and MM algorithms for the recovery of the MIT logo from different number of measurements, indicated by M. As discussed in the text, both algorithms are the same when $p = 1$. Note that the convergence rate of (black and blue curves) the MM scheme is considerably faster than (magenta and red curves) the approximate VS method that relies on the Ehler approximation, particularly for lower values of p. The differences between the two methods are larger when the number of measurements are smaller. (a) $M = 1500$. (b) $M = 1000$.

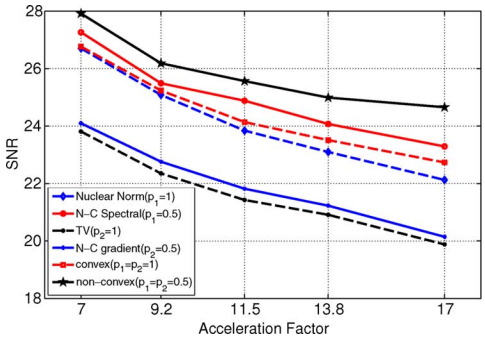


Fig. 9. SNR of the dynamic MRI data set as a function to the acceleration factor. We reconstruct the data set from its Fourier samples using six methods, i.e., nuclear norm ($p_1 = 1$), nonconvex spectral penalty ($p_1 = 0.5$), TV ($p_2 = 1$), nonconvex gradient penalty ($p_2 = 0.5$), combined convex penalties ($p_1 = p_2 = 1$), and combined nonconvex penalties ($p_1 = p_2 = 0.5$). Note that the combined nonconvex penalty provides significant gains in the SNR at almost all acceleration factors.

organs. High temporal resolution is required to accurately estimate the kinetic parameters, whereas high spatial resolution is required to visualize the lesion morphology. In addition, accelerated imaging can enable the simultaneous acquisition of two echoes (T1 weighted and T2* weighted), thus enabling the accurate quantification of microvascular density and vascular permeability; these parameters are highly correlated with malignancy and have been suggested as surrogate markers for angiogenesis. Several acceleration schemes have been proposed to accelerate DCE MRI [37], [38]. The accelerations offered by these schemes are modest (two- to threefold), leaving room for further improvement.

We demonstrate the utility of the proposed nonconvex scheme in significantly accelerating DCE MRI. The dynamic MRI measurements correspond to the samples of the signal in the Fourier $(k-t)$ space, corrupted by noise, i.e.,

$$\mathbf{b}_i = \int_{\mathbf{x}} \gamma(\mathbf{x}, t_i) \exp\left(-j\mathbf{k}_i^T \mathbf{x}\right) d\mathbf{x} + \mathbf{n}_i; \quad i = 1, \ldots, s.$$

Here, $(\mathbf{k}_i, t_i)$ indicates the ith sampling location. We denote the set of sampling locations as $\Xi = \{(\mathbf{k}_i, t_i), i = 1, \ldots, s\}$. The fully sampled 3-D data set of a single slice is shown in Fig. 10(a); the data corresponds to 60 time points, separated by TR $= 2$ sec; the matrix size is $128 \times 128 \times 60$. We retrospectively resample each slice of the data in the Fourier domain using a uniform radial trajectory. The trajectory is rotated by a random angle for each frame to obtain an incoherent pattern [see Fig. 10(b)]. The number of lines per slice is chosen depending on the specified acceleration. For example, 20 k-space lines approximately correspond to the acceleration factor of $A = 7$. We recover the dynamic imaging data set from its undersampled Fourier measurements using the proposed scheme. We use a few steps of the CG algorithm to solve for (22) at each iteration since the samples are not on the 3-D Cartesian grid. We use the previous iterate as an initial guess; thus, the CG algorithm converges to the solution of (22) in a few steps. The recovery of the DCE MRI data set using the MATLAB implementation of the proposed algorithm takes approximately 8 min on an Intel quad core processor with an NVDIA Tesla graphical processing unit. The computationally expensive components of the algorithm are implemented using Jacket [39].

The proposed algorithm has two regularization parameters λ_1 and λ_2 that control the quality of the reconstructions. The specific choice of the continuation parameters only influences the speed of convergence and will not determine the quality of the final solution, as shown in Fig. 7. Since the speed of convergence is not too sensitive to the specific continuation scheme, we do not expect to significantly improve the algorithm by dynamically updating the continuation parameters. Since the fully sampled data set is available in this experiment, we choose the regularization parameters that minimize the reconstruction error. Extensive literature is available on efficient schemes to choose the regularization parameters, when the ground truth is not available [40]–[42]; we plan to adapt these methods to determine the optimal regularization parameters in the future.

The SNR of the recovered 3-D data set as a function of the acceleration is plotted in Fig. 9. We observe that the best SNR is obtained when both the nonconvex penalties are used, which is around 1.5 dB better than nuclear norm alone and 6 dB superior than TV alone. It is also shown that the combined nonconvex penalty (solid black curve) gives reconstructions that are approximately 1–2 dB better than its convex counterpart (dotted red curve), particularly at higher accelerations. These experiments demonstrate the utility of the combination of nonconvex penalties in challenging practical applications. We show the slice corresponding to the peak of the perfusion contrast, recovered using TV, nuclear-norm, and combined nonconvex penalties, in Fig. 10(c)–(e). The corresponding error images are shown in Fig. 10(f)–(h), respectively. Here, we consider $A = 7$, which corresponds to 20 k-space lines per frame. We plot the average intensity variations of the recovered images from 5 pixels in the tumor region (green dot) and 5 pixels of the healthy tissue (red dot) in Fig. 10(i)–(k), respectively. Note that the curve from the tumor region has a larger dip and a larger width compared with that of the healthy tissue. This is due to the higher microvessel density and the increased tortuosity of the vessels in the tumor regions. We observe that the combination of nonconvex penalties gives good fit to the measured data. The sevenfold acceleration without significant degradation

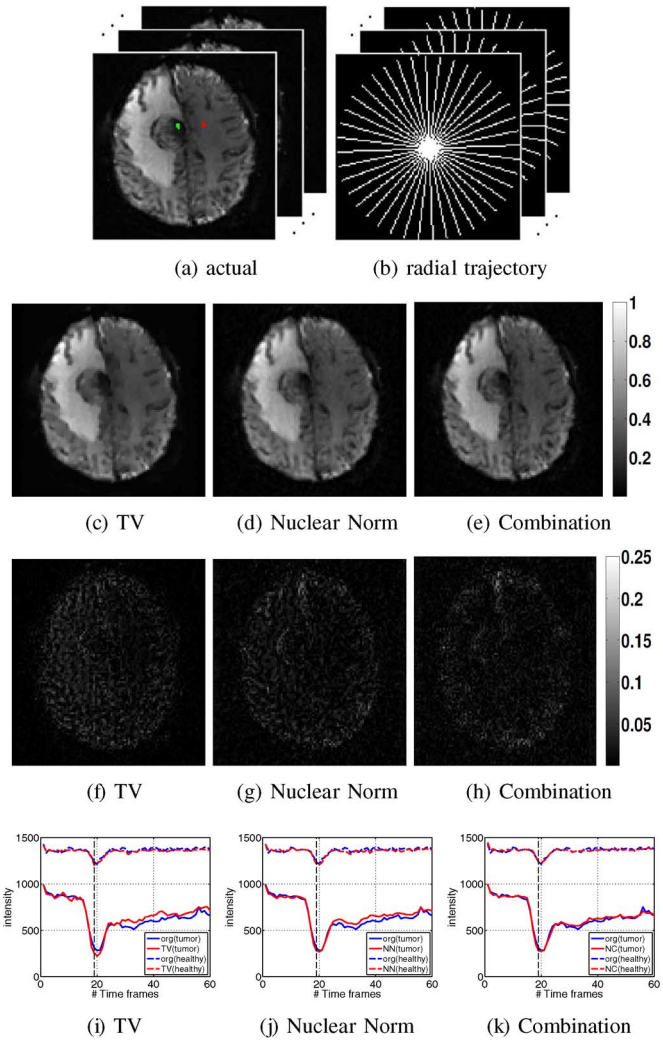


Fig. 10. Reconstruction of dynamic MRI data from undersampled Fourier samples. We consider 20 radial lines per image, which corresponds to an acceleration factor of $A = 7$. The images corresponding to (dotted line in the bottom row) the peak of the bolus, which are recovered using TV ($p_2 = 1$), nuclear norm ($p_1 = 1$), and the combination of nonconvex penalties ($p_1 = p_2 = 0.5$), are shown in (c)–(e). The corresponding error images are shown in (f)–(h), respectively. [(e), (h), and (k)] Reconstructions using the proposed combination of nonconvex penalties. The SNRs of the reconstructions using TV, nuclear norm, and the combination of nonconvex penalties are 23.81, 26.77, and 27.92 dB, respectively. Note that the proposed scheme provides a 1- to 4-dB improvement in performance and considerably better image quality compared with the classical schemes. (a) Actual. (b) Radial trajectory. (c) TV. (d) Nuclear norm. (e) Combination. (f) TV. (g) Nuclear norm. (h) Combination. (i) TV. (j) Nuclear norm. (k) Combination.

in the image quality is quite remarkable, particularly since we are only assuming a single channel acquisition; we expect to further improve the signal quality and/or acceleration using 12- or 32-channel head arrays that are now available.

VI. Conclusion

We have introduced a novel MM algorithm to recover sparse and low-rank matrices from its noisy and undersampled measurements. We have majorized the nonconvex spectral and sparsity penalties in the cost function using quadratic matrix functions, resulting in an iterative three-step alternating minimization scheme. Since each of the steps in the algorithm has computationally efficient implementations, the algorithm provides fast convergence. We have verified the utility of the combination of nonconvex spectral and sparsity penalties to significantly reduce the number of measurements required for perfect recovery in dynamic MRI data, as compared with current matrix recovery schemes.

Appendix I

Here, we derive the conjugate of function μ_g, specified by (29). Specifically, μ_g^* is defined as

$$\mu_g^*(w) = \max_x (wx - \mu_g(x)) = \max\left(\max_{x;x<\frac{1}{\beta}} wx, \max_{x;x>\frac{1}{\beta}} \left(wx - \frac{1}{2}(x-1/\beta)^2\right)\right).$$

The maximum value of the first term $(wx; x < 1/\beta)$ is given by w/β. The second term inside the bracket is true if $x = w + 1/\beta$, when the value of the function is given by $w^2/2 + w/\beta$. Since the second term is always greater than the first, we obtain

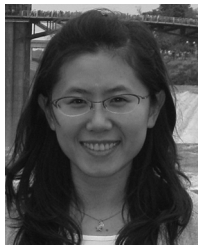
$$\mu_g^*(w) = \frac{w^2}{2} + \frac{w}{\beta}. \tag{50}$$

Since $\mu_\eta(w) = \beta(\mu_g(w) - w^2/2)$, we have $\mu_\eta(w) = w$.

References

[1] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.
[2] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Aug. 2010.
[3] Z. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. ISBI*, 2007, pp. 181–182.
[4] C. Brinegar, Y. Wu, L. Foley, T. Hitchens, Q. Ye, C. Ho, and Z. Liang, "Real-time cardiac MRI without triggering, gating, or breath holding," in *Proc. IEEE EMBC*, 2008, pp. 3381–3384.
[5] S. Goud, Y. Hu, E. Di Bella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.
[6] S. Goud, Y. Hu, and M. Jacob, "Real-time cardiac MRI using low-rank and sparsity penalties," in *Proc. ISBI*, 2010, pp. 988–991.
[7] 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.
[8] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.
[9] E. Candes, X. Li, Y. Ma, and J. Wright, Robust principal component analysis Stanford Univ., Stanford, CA, Tech. Rep. 2009-13, 2009.
[10] B. Zhao, J. Haldar, and Z. Liang, "Psf model-based reconstruction with sparsity constraint: Algorithm and application to real-time cardiac mri," in *Proc. IEEE EMBC*, 2010, pp. 3390–3393.
[11] B. Zhao, J. Haldar, A. Christodoulou, and Z. Liang, "Further development of image reconstruction with joint partial separability and sparsity constraints," in *Proc. IEEE ISBI*, 2011, pp. 1593–1596.
[12] J. Trzasko and A. Manduca, "Relaxed conditions for sparse signal recovery with general concave priors," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4347–4354, Nov. 2009.
[13] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.
[14] D. Wipf and S. Nagarajan, "Iterative reweighted l1 and l2 methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

[15] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
[16] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Programm.*, vol. 128, no. 1/2, pp. 321–353, Jun. 2011.
[17] K. Toh and S. Yun, "An accelerated proximal gradient method for nuclear norm regularized least squares," , 2009, to be published.
[18] J. Huang, S. Zhang, and D. Metaxas, "Efficient MR image reconstruction for compressed MR imaging," in *Proc. MICCAI*, 2010, pp. 135–142.
[19] R. T. Rockafellar and R.-B. Wets, *Variational Analysis*. New York: Springer-Verlag, 2004.
[20] M. Ehler and S. Geisel, "Arbitrary shrinkage rules for approximation schemes with sparsity constraints," in *Proc. Schloss Dagstuhl Semin. –Structured Decompositions Efficient Algorithms*, Wadern, Germany, 2009, no. 08492.
[21] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ell_0-minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, Jan. 2009.
[22] A. S. Lewis, "The convex analysis of unitarily invariant matrix functions," *J. Convex Anal.*, vol. 2, no. 1/2, pp. 173–183, 1995.
[23] J. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using power factorization," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 584–587, Jul. 2009.
[24] R. Chatrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *Proc. ISBI*, 2009, pp. 262–265.
[25] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.
[26] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, "Robust principal component analysis based four-dimensional computed tomography," *Phys. Med. Biol.*, vol. 56, no. 11, pp. 3181–3198, Jun. 2011.
[27] H. Jung, J. Park, J. Yoo, and J. C. Ye, "Radial k-t focus for high-resolution cardiac cine MRI," *Magn. Reson. Med.*, vol. 63, no. 1, pp. 68–78, Jan. 2010.
[28] M. Lustig, J. Santos, D. Donoho, and J. Pauly, "kt SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. 13th Annu. Meeting ISMRM*, Seattle, WA, 2006, p. 2420.
[29] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, no. 2, pp. 365–373, Feb. 2008.
[30] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Jul. 2008.
[31] Legendre Transformation [Online]. Available: http://en.wikipedia.org/wiki/Legendre_transformation
[32] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.
[33] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.
[34] E. Esser, Applications of Lagrangian-based alternating direction methods and connections to split Bregman UCLA, Los Angeles, CA, CAM Rep. TR09-31, 2009.
[35] T. Goldstein and S. Osher, The split Bregman algorithm for l1 regularized problems UCLA, Los Angeles, CA, UCLA CAM Rep. 08-29, 2008.
[36] S. Goud, Y. Hu, E. DiBella, and M. Jacob, "Accelerated first pass cardiac perfusion mri using improved k-t SLR," in *Proc. ISBI*, 2011, pp. 1280–1283.
[37] 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.
[38] M. Han, B. L. Daniel, and B. A. Hargreaves, "Accelerated bilateral dynamic contrast-enhanced 3-D spiral breast MRI using tsense," *J. Magn. Reson. Imag.*, vol. 28, no. 6, pp. 1425–1434, Dec. 2008.
[39] Jacket: Accelerating Matlab Using GPU [Online]. Available: http://www.accelereyes.com/products/jacket
[40] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo Sure: A black-box optimization of regularization parameters," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.
[41] P. Hansen and D. OLeary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, p. 1487, 1993.
[42] O. Scherzer, "The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems," *Computing*, vol. 51, no. 1, pp. 45–60, Mar. 1993.



Yue Hu (S'11) received the B.S. degree in 2008 from Harbin Institute of Technology, Harbin, China, and the M.S. degree in 2011 from the University of Rochester, Rochester, NY, where she is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering.

Her research interests include compressed sensing, image reconstruction with application in magnetic resonance imaging.

Sajan Goud Lingala (S'10) received the B.S. degree from Osmania University, Hyderabad, India, and the M.Tech. degree from the Indian Institute of Technology (IIT) Bombay, Mumbai, India, both in biomedical engineering (BME). He is currently working toward the Ph.D. degree in BME at the University of Iowa, Iowa City.

His research interests include multidimensional magnetic resonance image acquisition and reconstruction with focus on cardiac imaging.

Mr. Lingala is the recipient of the Best Undergraduate Thesis Award during his bachelors and the Nitish Thakor Award for excellence in M.Tech BME, IIT Bombay.

Mathews Jacob (M'11) received the B.Tech. in electronics and communication engineering from the National Institute of Technology, Calicut, India, in 1996, the M.E. degree in signal processing from the Indian Institute of Science, Bangalore, India, in 1999, and the Ph.D. degree from the Swiss Federal Institute of Technology, Zurich, Switzerland, in 2003.

Between 2003 and 2006, he was a Beckman Postdoctoral Fellow with the University of Illinois at Urbana Champaign, Urbana. He is an Assistant Professor with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA. His research interests include image reconstruction, image analysis and quantification in the context of a range of modalities including magnetic resonance imaging, near-infrared spectroscopic imaging, and electron microscopy.

Dr. Jacob is currently the Associate Editor of the IEEE TRANSACTIONS ON MEDICAL IMAGING. He is the recipient of the CAREER award from the National Science Foundation in 2009 and the Research Scholar Grant from the American Cancer Society in 2011.