Electronic Transactions on Numerical Analysis.
Volume 53, pp. 481–499, 2020.

ISSN 1068–9613.
DOI: 10.1553/etna_vol53s481

ETNA
Kent State University and
Johann Radon Institute (RICAM)

MATRIX COMPLETION FOR MATRICES WITH LOW-RANK DISPLACEMENT*

DAMIANA LAZZARO† AND SERENA MORIGI†

Abstract. The matrix completion problem consists in the recovery of a low-rank or approximately low-rank matrix from a sampling of its entries. The solution rank is typically unknown, and this makes the problem even more challenging. However, for a broad class of interesting matrices with so-called displacement structure, the originally ill-posed completion problem can find an acceptable solution by exploiting the knowledge of the associated displacement rank. The goal of this paper is to propose a variational non-convex formulation for the low-rank matrix completion problem with low-rank displacement and to apply it to important classes of medium-large scale structured matrices. Experimental results show the effectiveness and efficiency of the proposed approach for Toeplitz and Hankel matrix completion problems.

Key words. matrix completion, low-rank matrices, displacement rank, non-convex optimization

AMS subject classifications. 65K10, 65F22, 15A29

1. Introduction.

We consider the problem of the recovery of a low-rank structured matrix characterized by a low displacement rank from its undersampled/incomplete entries. This particular class of matrices with structure can be identified in a surprising variety of applications in engineering, mathematics, and physics. The displacement of a matrix X is the image $\mathcal{L}(X)$ of an appropriate linear displacement operator $\mathcal{L}$ applied to the matrix X and is revealing its structure. Structured matrices are naturally associated with linear displacement operators $\mathcal{L}$ (typically incorporating the operators of shift and scaling) and can be recovered easily from their displacement matrix $\mathcal{L}(X)$, which has a smaller and known rank r with respect to the rank of X [1, 20, 33]. Matrices like Cauchy, Vandermonde, Polynomial Vandermonde, Chebyshev Vandermonde, Toeplitz, Hankel, and others belong to this special class of matrices, and they only depend on $O(n)$ parameters instead of n^2.

Certain families of such particular class of structured matrices with low displacement rank are particularly relevant and commonly studied because they arise in a wide range of applications. Let us consider for example the standard image deconvolution problem, which aims to recover an image f given an observed image g, degraded by a convolution with a point spread function (PSF) h. The linear blurring process with the PSF h is represented by a matrix X, which is often a block matrix whose blocks take the form of a Toeplitz matrix. In this particular case, not only does X have Toeplitz blocks, but it is also block Toeplitz, that is, its blocks are aligned in a Toeplitz fashion [18]. If the PSF h is not known, then the problem becomes one of blind deconvolution, sometimes called inaccurate deconvolution if X is partially known [8] and must be recovered. The image blind deconvolution problem remains challenging and hard to solve.

The structure-preserving property is important also in the recovery of spectrally sparse signals from a random subset of samples, as illustrated in [10, 12], where the low-rank Hankel property has been exploited to recover the sparse signal, as well as in the restoration of signals in MRI imaging [16].

The more general problem to recover a rectangular matrix from a sampling of its entries is known as the Matrix Completion (MC) problem, which is highly ill-posed since with fewer samples than entries, we have infinitely many completions. A possible solution candidate can be obtained when the matrix that we wish to recover has low-rank or approximately low rank.

*Received December 29, 2019. Accepted June 23, 2020. Published online on July 13, 2020. Recommended by Marco Donatelli.

†Department of Mathematics, University of Bologna, Bologna, Italy
({damiana.lazzaro,serena.morigi}@unibo.it).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Considering this additional information, the general formulation of the MC problem for the recovery of a low-rank matrix M reads as

$$X^* = \underset{X\in\mathbb{R}^{m\times n}}{\operatorname{argmin}} \ \operatorname{rank}(X) \qquad \text{s.t. } P_\Omega(X) = P_\Omega(M), \tag{1.1}$$

where only p sampled entries of a matrix $M \in \mathbb{R}^{m\times n}$ are given,

$$\{M_{i,j} : (i,j) \in \Omega \subset \{1,\dots,m\} \times \{1,\dots,n\}\},$$

and P_Ω represents the projection operator on a random subset Ω of cardinality p,

$$P_\Omega(X) := \begin{cases} X_{i,j} & \text{if } (i,j) \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

The matrix rank minimization problem (1.1) is NP-hard in general due to the highly non-convex and combinatorial nature of the rank function [11]. To overcome such a computational difficulty, the authors in [5, 6, 37] introduced a convex relaxation of (1.1) which relies on the nuclear norm, which is the convex envelope of $\operatorname{rank}(X)$, namely $\|X\|_* = \sum_{i=1}^{\kappa} \sigma_i(X)$, where $\sigma_i(X)$, $i = 1,\dots,\kappa, \kappa \leq min(m,n)$, are the singular values of X. This convex relaxation results in the following nuclear norm minimization problem, which is the tightest convex relaxation to (1.1):

$$X^* = \underset{X\in\mathbb{R}^{m\times n}}{\operatorname{argmin}} \ \|X\|_* \qquad \text{s.t. } P_\Omega(X) = P_\Omega(M). \tag{1.2}$$

Candès and Recht in [5] showed that applying the convex optimization model (1.2) in case of MC problems with square matrices of dimension n, most low-rank matrices can be recovered exactly provided that the number of samples obeys

$$p \geq Cn^{1.2}\kappa \log n \tag{1.3}$$

for some positive numerical constant C. The condition above assumes that the rank κ is not too large. Similar results hold for arbitrary rectangular matrices as well.

Many algorithms have been designed to solve the optimization problem (1.2). For example, the singular value thresholding (SVT) method [4, 45], the accelerated proximal gradient (APG) method [41], the augmented Lagrange multiplier (ALM) method [26], as well as its variants [43, 44]. In the latter, the authors focused on the specific Toeplitz matrix completion problem, which has attracted a lot of attention in recent years. Many researchers have studied the recovery of Toeplitz and Hankel matrices [10, 12]. Toeplitz and Hankel matrices belong to a broader class of interesting matrices with displacement structure for which the originally ill-posed completion problem can find an acceptable solution by exploiting the knowledge of the associated displacement rank.

In this work we address the specific MC problem for the recovery of a low-rank structured matrix with low $\mathcal{L}$-displacement rank, which is a special case of the matrix completion problem (1.1). To this aim, we propose a constrained non-convex variational formulation which includes a low-rank promoting penalty function. This is directed to overcome the limits of the nuclear norm, whose performance for sparse regularization of the singular values is only suboptimal [31]. It is well known, in fact, that suitable non-convex penalty functions induce sparsity of the singular values more effectively than the nuclear norm [9, 25, 27, 29, 34].

For the recovery of a structured matrix M of low-rank κ with low displacement rank r, we propose to solve the following sparse low-rank minimization problem

$$X^* = \underset{X\in\mathbb{R}^{m\times n}}{\operatorname{argmin}} \sum_{i=1}^{\kappa} \phi\left(\sigma_i(X); a\right) \quad \text{s.t.} \begin{cases} P_\Omega(X) = P_\Omega(M), \\ \mathcal{L}(X) \in \mathcal{M}_r, \end{cases} \tag{1.4}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

with $\mathcal{M}_r$ the non-convex closed set of all matrices of size $m \times n$ with rank up to r defined as

$$\mathcal{M}_r = \left\{ X \in \mathbb{R}^{m \times n} \ : \ \operatorname{rank}(X) \leq r \right\},$$

and $\phi : \mathbb{R} \to \mathbb{R}$ a parameterized, non-convex penalty function that will be discussed in Section 3.

The use of non-convex penalty functions, however, generally leads to non-convex optimization problems, which suffer from numerous issues such as spurious local minima and a sensitivity to changes in the initialization. To avoid the intrinsic difficulties related to non-convex optimization, we present a simple forward-backward splitting method to solve the minimization problem (1.4), which relies on the Convex Non-Convex (CNC) strategy and a projection step [23, 24, 39].

In the reported experiments we aim to demonstrate that the additional constraint on the associated displacement rank, together with the non-convexity of the penalty function for the nuclear norm, allow us to obtain a very robust algorithm that is able to recover low-rank matrices starting from a set of sampled data with cardinality p much smaller than the estimated value in (1.3).

The paper is organized as follows: In Section 2 we introduce some basic definitions and concepts of structured matrices with low-rank displacement structure. In Section 3 we define the non-convex sparsity promoting penalty ϕ used in our model. The proposed variational MC model is presented in Section 4 together with the sketch of the forward-backward method applied to minimization. In Section 5 we address the details of the implemented algorithm, and in Section 6 we present two applications of the algorithm to the recovery of Toeplitz and Hankel matrices. Finally, conclusions are drawn in Section 7.

2. Structured matrices and their low-rank displacement structure. The concept of displacement structure was first introduced in [21]. Formally, we can associate structured matrices $X \in \mathbb{R}^{m \times n}$ with linear operators $\mathcal{L} : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ of Sylvester or Stein types, defined as follows:

DEFINITION 2.1. *Given two matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$, the* displacement operator *$\mathcal{L} : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ with respect to the displacement matrix pair (U, V) is defined for every matrix $X \in \mathbb{R}^{m \times n}$ as*

$$\mathcal{L}(X) := \nabla_{U,V}(X) = X - UXV \qquad \textit{Stein type} \tag{2.1}$$

or

$$\mathcal{L}(X) := \Delta_{U,V}(X) = UX - XV \qquad \textit{Sylvester type}. \tag{2.2}$$

The operators of Sylvester or Stein types can be transformed easily into each other if at least one of the two associated operator matrices is non-singular. The relation between the two different displacements of Sylvester type (2.2) and Stein type (2.1) is then given in the following proposition.

PROPOSITION 2.2. *If the operator matrix U is non-singular, then we have the identity $\nabla_{U,V} = U\Delta_{U^{-1},V}$; in case the operator matrix V is non-singular, then $\nabla_{U,V} = -\Delta_{U,V^{-1}}V$.*

In this work, without loss of generality, we will consider only displacements of Stein type. The matrix $X \in \mathbb{R}^{m \times n}$ is considered to possess a displacement structure with respect to U and V whenever the rank of $\mathcal{L}(X)$, the so-called $\mathcal{L}$-displacement rank, is low compared to its sizes m and n and it remains relatively small even for increasing dimensions of X. For example, if X is a Toeplitz matrix, then its $\mathcal{L}$-displacement rank is at most 2.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 2.1
Displacement rank for some structured matrices.

U	V	Class of structured matrices	rank of $\nabla_{U,V}$
Z_0	Z_0^T	Toeplitz and its inverse	$r <= 2$
Z_0	Z_0	Hankel	$r <= 2$
$Z_0 + Z_0^T$	$Z_0 + Z_0^T$	Toeplitz+Hankel	$r <= 4$

The displacement matrix $\mathcal{L}(X)$ can be represented by a pair of matrices $G \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{n\times r}$ called generator matrices (generators, for short) such that

$$\mathcal{L}(X) = GH^T = \sum_{k=1}^{r} g_k h_k^T, \tag{2.3}$$

where g_j and h_j represent the jth columns of G and H, respectively, and r denotes the $\mathcal{L}$-displacement rank of the matrix X.

In practice, the generator matrices G and H can be efficiently obtained by the singular value decomposition of the displacement matrix $\mathcal{L}(X)$ having mn entries and a reduced rank r. In particular, let U_X, S, V_X be the matrices obtained by the singular value decomposition of $\mathcal{L}(X)$. Then the orthogonal generators are given by

$$G = U_X\sqrt{S}, \qquad H = V_X\sqrt{S}. \tag{2.4}$$

For the specific classes of structured matrices of interest to us, the operator matrices U, V assume a particular structure Z_f, where Z_f is the unit f-circulant matrix and f is any scalar value:

$$Z_f = \begin{bmatrix} 0 & & & & f \\ 1 & 0 & & & \\ & 1 & 0 & & \\ & & \ddots & & \\ & & & 1 & 0 \end{bmatrix}.$$

In Table 2.1, matrices U and V are defined for the Stein operator for the case of the class of matrices of our interest.

Provided $\mathcal{L}$ is invertible, when r is small, the pair of generators G, H in (2.3) can play the role of a succinct data structure to represent the matrix X. More precisely, a matrix X can be recovered easily from its displacement $\mathcal{L}(X)$ by applying the inverse of the linear operator $\mathcal{L}$, denoted by $\mathcal{L}^{-1}$, defined according to [33], in the case of a Stein-type displacement ∇_{Z_e,Z_f^T} with $ef \neq 1$ as

$$(1-ef)X = \mathcal{L}^{-1}(GH^T) := \sum_{j=1}^{r}\left(\sum_{i=0}^{n-1}(Z_e^i g_j)(Z_f^i h_j)^T\right) \tag{2.5}$$

and in the case of a Stein-type displacement ∇_{Z_e,Z_f} with $ef \neq 1$ as

$$(1-ef)X = \mathcal{L}^{-1}(GH^T) := \sum_{j=1}^{r}\left(\sum_{i=0}^{n-1}(Z_e^i g_j)h_j^T Z_f^i\right), \tag{2.6}$$

which correspond to the first and second row in Table 2.1, respectively. Due to the simplified circulant structure of Z_e and Z_f, the computational effort required for the recovery of X by

ETNA
Kent State University and
Johann Radon Institute (RICAM)

applying (2.5) or (2.6) is nearly linear in time. In particular, for a square matrix of dimension n and rank r, the asymptotic complexity is $O(rn \log n)$; see [33] for details.

These results allow us to characterize a wide class of matrices with low displacement rank, which includes the well-known classes of Toeplitz and Hankel matrices as well as other classes of matrices which play an important role in some applications [1]. A Toeplitz matrix is a structured matrix of the form

$$T \in \mathbb{R}^{n\times n} = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-2} & t_{n-1} \\ t_{-1} & t_0 & \cdots & t_{n-3} & t_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{-n+2} & t_{-n+3} & \ddots & t_0 & t_1 \\ t_{-n+1} & t_{-n+2} & \cdots & t_{-1} & t_0 \end{bmatrix},$$

that is, its entries are constant along every diagonal. Thus, it is completely determined by the entries in the first row and first column. A matrix whose entries are constant along every antidiagonal is called a Hankel matrix and reads as

$$H \in \mathbb{R}^{n\times n} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_{n-1} \\ h_1 & h_2 & h_3 & \cdots & h_n \\ \vdots & \vdots & & \vdots & \vdots \\ h_{n-2} & h_{n-1} & \cdots & h_{2n-4} & h_{2n-3} \\ h_{n-1} & h_n & \cdots & h_{2n-3} & h_{2n-2} \end{bmatrix},$$

which is completely determined by the entries in the first column and last row.

3. Low-rank inducing penalty functions. In this section we formally introduce the class of non-convex penalty functions ϕ used as penalty term in the proposed variational model (1.4), which will be discussed in detail in Section 4. In the following, we denote the sets of non-negative and positive real numbers as $\mathbb{R}_+ := \{\, t \in \mathbb{R} : t \geq 0\}$ and $\mathbb{R}_+^* := \{\, t \in \mathbb{R} : t > 0\}$, respectively.

We consider parameterized penalty functions $\phi(t;a) : \mathbb{R} \to \mathbb{R}$ such that for any value of the parameter $a \in \mathbb{R}_+^*$, the following conditions are satisfied:

A1) $\phi(t;a) \in \mathcal{C}^2(\mathbb{R}_+^*),\ \phi(t;a) \in \mathcal{C}^0(\mathbb{R}_+)$,
A2) $\phi'(t;a) > 0, \qquad \forall\, t \in \mathbb{R}_+^*$,
A3) $\phi''(t;a) \leq 0, \qquad \forall\, t \in \mathbb{R}_+^*$,
A4) $\sup_{t\in\mathbb{R}_+^*} \phi'(t;a) = 1,\ \inf_{t\in\mathbb{R}_+^*} \phi''(t;a) = \phi''(0^+;a) = -a$.

Assumptions A1)–A3) are quite standard and encompass a wide class of non-convex sparsity-promoting penalty functions. Assumption A4), which corresponds to imposing boundedness (from above/below) of the first- and second-order derivatives of the penalty function ϕ, is mandatory when constructing CNC functionals; see, e.g., [23, 24, 34, 36, 39]. In fact, if the second-order derivative of the penalty function tends to $-\infty$ at any point in the domain, then there is no possibility to compensate it by the positive but bounded second-order derivatives of the convex quadratic fidelity term.

Without loss of generality, in this work we consider the non-convex parameterized penalty function

$$\phi(t;a) = \frac{1}{a}\log(1 + a|t|)\ , \tag{3.1}$$

where the parameter $a > 0$ controls the degree of non-convexity of the penalty function. This function satisfies the conditions A1)–A4).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

The proximity operator $\Theta : \mathbb{R} \to \mathbb{R}$ associated with the non-convex function $\phi(t; a)$ reads as

$$\Theta(y; \lambda, a) = \text{prox}_{\lambda\phi}(y; \lambda, a) = \underset{t \in \mathbb{R}}{\text{argmin}} \left\{ f(t) = \frac{1}{2}(y - t)^2 + \lambda\phi(t; a) \right\}. \tag{3.2}$$

In [35, 39, 40], the authors prove that for $0 \leq a < \frac{1}{\lambda}$, the function f in (3.2) is strictly convex. Therefore, the proximity operator admits a unique solution of the convex minimization problem (3.2).

The proximity operator associated with the logarithmic penalty (3.1) is a continuous non-linear *threshold function* with λ as threshold value [39], namely

$$\text{prox}_{\lambda\phi}(y; \lambda, a) = 0, \qquad \forall |y| < \lambda,$$

and is given by:

$$\text{prox}_{\lambda\phi}(y; \lambda, a) = \begin{cases} \left[\frac{|y|}{2} - \frac{1}{2a} + \sqrt{\left(\frac{|y|}{2} + \frac{1}{2a} \right)^2 - \frac{\lambda}{a}} \right] & |y| \geq \lambda, \\ 0 & |y| \leq \lambda. \end{cases}$$

4. Variational matrix completion model. We are concerned with the recovery of special matrices $X \in \mathbb{R}^{m \times n}$ with displacement structure that are characterized by a low-rank displacement matrix $\mathcal{L}(X) \in \mathcal{M}_r$. The rank r, $0 < r \ll \min(m, n)$, is chosen according to the class of matrices to be recovered; see Table 2.1. To this aim, we assume that p sampled entries $\{M_{ij} : (i, j) \in \Omega\}$ of an incomplete matrix M are available, with Ω a random subset of cardinality p. By relaxing the fidelity constraint in formulation (1.4), M can be recovered by solving the constrained optimization problem:

$$X^* \in \underset{\mathcal{L}(X) \in \mathcal{M}_r}{\text{argmin}} \left\{ \mathcal{J}(X; \lambda) = \frac{1}{2} \|P_\Omega(X) - P_\Omega(M)\|_F^2 + \lambda \sum_{i=1}^{\kappa} \phi\left(\sigma_i(X); a\right) \right\}, \tag{4.1}$$

where $\kappa < \min(m, n)$, $\sigma_i(X)$ is the ith singular value of the matrix X, $\|\cdot\|_F$ denotes the Frobenius norm, and $\phi : \mathbb{R} \to \mathbb{R}$, defined in (3.1), is a sparsity-inducing regularizer, possibly non-convex, and parameterized by a positive value a.

In order to solve the minimization problem (4.1), we consider the following unconstrained version with a functional given as the sum of a smooth convex function $F : \mathbb{R}^{m \times n} \to \mathbb{R}$ and a function G. The latter gathers the two non-convex non-smooth functions $R : \mathbb{R}^{m \times n} \to \mathbb{R}$, which induces a low-rank on the recovered matrix X, and $\iota_{\mathcal{M}_r}$, which is the indicator function of the non-convex set $\mathcal{M}_r$:

$$X^* \in \underset{X \in \mathbb{R}^{m \times n}}{\text{argmin}} \left\{ \mathcal{J}(X; \lambda) = \frac{1}{2} F(X) + \underbrace{\lambda R(X) + \iota_{\mathcal{M}_r}(\mathcal{L}(X))}_{G(X)} \right\}. \tag{4.2}$$

The indicator function $\iota_{\mathcal{M}_r} : \mathbb{R}^{m \times n} \to (-\infty, \infty]$ of the nonempty closed set $\mathcal{M}_r$ is a proper, lower semicontinuous, extended real-valued function defined by

$$\iota_{\mathcal{M}_r}(Y) := \begin{cases} 0 & \text{if } Y \in \mathcal{M}_r, \\ +\infty & \text{otherwise.} \end{cases}$$

Since the set $\mathcal{M}_r$ is non-convex, $\iota_{\mathcal{M}_r}(X)$ is a non-convex function.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

We propose to compute approximate solutions X^* of the minimization problem in (4.2) by means of the Forward-Backward (FB) iterative scheme, outlined in Algorithm 1, applied to the sum $F(X) + G(X)$. In particular, the forward step generates a sequence $\{Z_k\}$ via

$$Z_k = X_{k-1} - \frac{1}{\beta}\nabla F(X_{k-1}),$$

which relies on the gradient of F given by

$$\nabla F(X_{k-1}) = P_\Omega(X_{k-1}) - P_\Omega(M). \tag{4.3}$$

We adopt a constant stepsize β, which is an upper bound for the Lipschitz constant of ∇F, which can be seen to satisfy $L_F = 1$ as discussed in Section 4.1. The backward step requires the solution of the following optimization problem

$$X_k = P_{\mathcal{M}_r} \circ \operatorname{prox}_{\frac{\lambda}{\beta}R}(Z_k),$$

where

$$\tilde{X}_k = \operatorname{prox}_{\frac{\lambda}{\beta}R}(Z_k) = \operatorname*{argmin}_{X \in \mathbb{R}^{m\times n}} \left\{ \frac{1}{2}\|X - Z_k\|_F^2 + \frac{\lambda}{\beta}\sum_{i=1}^{\kappa} \phi(\sigma_i(X), a) \right\}. \tag{4.4}$$

The proximal map of the indicator function reduces to the projection operator onto the set, defined as $P_{\mathcal{M}_r}(X)$, where the projection $P_{\mathcal{M}_r} : \mathbb{R}^{m\times n} \rightrightarrows \mathbb{R}^{m\times n}$ has nonempty values and defines in general a multi-valued map, as opposed to the convex case, where orthogonal projections are guaranteed to be single-valued. However, the projection on the set $\mathcal{M}_r$ has a unique closed form computed by a truncated SVD of $\mathcal{L}(\tilde{X}_k)$ and the inversion formula (2.5) or (2.6), with r given as input according to the class of matrices to be recovered; see Table 2.1.

In particular, this projection can be efficiently computed by the singular value decomposition to reduce $\mathcal{L}(\tilde{X}_k)$ to a r-rank displacement matrix and to obtain the generators G and H. The matrix X_k is then built by applying the inverse operator $\mathcal{L}^{-1}$ to GH^T, and it is taken as input for the next FB splitting iteration k. The projection step is called Displacement Upgrade step in Algorithm 1.

Algorithm 1 Forward-Backward Low Displacement Rank (FB-LDR)

Given $X_0 = \mathcal{P}_{\mathbf{\Omega}}(M)$ and the parameters $\beta > 1, \lambda, r > 0$.
repeat
  Forward Step
    $Z_k = X_{k-1} - \frac{1}{\beta}\nabla F(X_{k-1})$
  Backward Step
    $\tilde{X}_k = \operatorname{prox}_{\frac{\lambda}{\beta}R}(Z_k)$
    Displacement Upgrade step
    $[U, S, V] = svd(\mathcal{L}(\tilde{X}_k), r)$
    $X_k = \mathcal{L}^{-1}((U\sqrt{S})(V\sqrt{S})^T)$
until convergence

The following Theorem 4.1, proved in [34], defines a closed-form solution to the minimization problem (4.4) in the backward step, which involves thresholding of the singular values of the matrix Z in (4.4), where the subscript index k is omitted. The result in Theorem 4.1 relies on the convexity condition of the cost function in (4.4), imposed on the concavity

ETNA
Kent State University and
Johann Radon Institute (RICAM)

parameter a. This can be easily derived from those of the scalar problem (3.2) as stated in [39]. Consequently, the objective function in (4.4) is strictly convex if

$$0 \leq a < \frac{\beta}{\lambda}, \tag{4.5}$$

hence its minimizer is unique.

THEOREM 4.1. *Let $Z = U\Sigma V^T$ be the SVD of a matrix Z and $\phi(t;a) : \mathbb{R} \to \mathbb{R}$ be a non-convex penalty function satisfying conditions A1)–A4). If the parameter a satisfies conditions (4.5), then the global minimizer of (4.4) is*

$$\tilde{X} = U \cdot \Theta(\Sigma; \lambda, a) \cdot V^T, \tag{4.6}$$

where Θ, defined in (3.2), is the threshold function associated with the non-convex penalty function ϕ.

This result allows the use of non-convex penalty functions ϕ while maintaining convexity of the objective function (4.4).

4.1. Basic convergence properties. In this section we first analyze the existence of minimizers for the proposed variational problem (4.2). Then some properties of the convergence of the iterative FB-LDR algorithm will be established in order to guarantee that the proposed iterative algorithm, which will be introduced in Section 5, does not break down.

PROPOSITION 4.2. *The objective function $\mathcal{J}(X;\lambda)$ in (4.2) is proper, non-convex, lower semicontinuous, coercive, and bounded from below by zero. Therefore, the set of global minimizers of problem (4.2) is non-empty.*

The gradient of the continuously differentiable function F given by (4.3) is a L_F-Lipschitz continuous function with constant $L_F = 1$. It easy to see from (4.3) that, for every $X, Y \in \mathbb{R}^{m\times n}$,

$$\|\nabla F(Y) - \nabla F(X)\| = \|P_\Omega(Y) - P_\Omega(X)\| \leq L_F \|X - Y\|.$$

If $X \in \Omega$ and $Y \in \Omega$, then $L_F = 1$. If $X \notin \Omega$ and $Y \notin \Omega$, then $L_F\|X - Y\| \geq 0$, for any constant L_F. Finally, when $X \in \Omega$ and $Y \notin \Omega$ (or vice versa), then $\|X\| \leq \|X\|$ (or $\|Y\| \leq \|Y\|$), which leads to $L_F = 1$.

This gradient-Lipschitz property implies the decrease of the objective function.

PROPOSITION 4.3 (Descent Lemma). *Let $F : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}$ be a continuously differentiable function with gradient ∇F assumed L_F-Lipschitz continuous. Then, for any $L \geq L_F$,*

$$F(X) \leq F(Y) + \langle X - Y, \nabla F(Y)\rangle + \frac{L}{2}\|X - Y\|^2, \qquad \forall\, X, Y \in \mathbb{R}^{m\times n}.$$

We adopt the following quadratic approximation model for F, for any $L > 0$:

$$Q_L(X, Y) := F(Y) + \langle X - Y, \nabla F(Y)\rangle + \frac{L}{2}\|X - Y\|^2 + G(X), \qquad \forall X, Y \in \mathbb{R}^{m\times n},$$

with prox-grad map given by

$$p_L(Y) \in \operatorname*{argmin}_X\{Q_L(X, Y)\} = \operatorname*{argmin}_X\{\langle X - Y, \nabla F(Y)\rangle + \frac{L}{2}\|X - Y\|^2 + G(X)\},$$

that is,

$$p_L(Y) \in \operatorname*{argmin}_X\{Q_L(X, Y)\} = \operatorname{prox}_{1/L}(G)(Y - \frac{1}{L}\nabla F(Y)). \tag{4.7}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

PROPOSITION 4.4 (Sufficient decrease property). *Let $F : \mathbb{R}^{m\times n} \to \mathbb{R}$ be a continuously differentiable function with gradient ∇F assumed L_F-Lipschitz continuous, and let $G(X)$ be a proper, extended-valued function bounded from below by zero. Fix any $L > L_F$. Then, for any $X \in \operatorname{dom} G$ and any $p_L(Y) \in \mathbb{R}^{m\times n}$ defined as in (4.7), we have*

$$\mathcal{J}(X) - \mathcal{J}(p_L(Y)) \geq \frac{1}{2}(L - L_F)\|p_L(Y) - X\|^2. \tag{4.8}$$

Proof. Since $p_L(Y)$ minimizes $Q_L(p_L(Y), Y)$, it holds that $Q_L(p_L(Y), Y) \leq Q_L(Y, Y)$. Hence, taking $X = p_L(Y)$ in $Q_L(X, Y)$ in (4.7), we obtain

$$\langle p_L(Y) - Y, \nabla F(Y)\rangle + \frac{L}{2}\|p_L(Y) - Y\|^2 + G(p_L(Y)) \leq G(Y). \tag{4.9}$$

Invoking the descent lemma (see Lemma 4.3) for F and using then inequality (4.9), we get

$$\begin{aligned} F(p_L(Y)) + G(p_L(Y)) &\leq F(Y) + \langle p_L(Y) - Y, \nabla F(Y)\rangle + \frac{L}{2}\|p_L(Y) - Y\|^2 + G(p_L(Y)) \\ &\leq F(Y) + \frac{L_F}{2}\|p_L(Y) - Y\|^2 + G(Y) - \frac{L}{2}\|p_L(Y) - Y\|^2 \\ &= F(Y) + G(Y) - \frac{1}{2}(L - L_F)\|p_L(Y) - Y\|^2. \end{aligned}$$

This proves the inequality (4.8). ∎

We now can establish basic convergence properties of the iteration sequence generated by Algorithm 1.

PROPOSITION 4.5 (Basic convergence properties). *Let $\{X_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1, and let $L > L_F$. Assume that the function $\mathcal{J}$ is coercive and bounded from below ($\inf \mathcal{J} > -\infty$). Then the following holds:*

(i) The sequence $\{\mathcal{J}(X_k)\}_{k\in\mathbb{N}}$ is nonincreasing, and in particular,

$$\mathcal{J}(X_k) - \mathcal{J}(X_{k+1}) \geq \rho\|X_{k+1} - X_k\|_F^2, \qquad \forall k = 0, 1, \ldots, \tag{4.10}$$

with $\rho = \frac{1}{2}(L - L_F) > 0$.

(ii) We have

$$\sum_{k=1}^{\infty} \|X_{k+1} - X_k\|_F^2 < +\infty,$$

which implies $\lim_{k\to\infty}(X_{k+1} - X_k) = 0$.

Proof. (i) For each $X \in \{X_k\}_{k\in\mathbb{N}}$, a generic iterate generated from Algorithm 1, the estimate (4.8) holds. Since $L > L_F$, it follows that $\rho > 0$, and thus $\{\mathcal{J}(X_k)\}_{k\in\mathbb{N}}$ is monotonically nonincreasing.

(ii) Let N be a positive integer. Summing up (4.10) from $k = 0$ to $N - 1$, we obtain that

$$\sum_{k=0}^{N-1} \|X_{k+1} - X_k\|_F^2 \leq \rho(\mathcal{J}(X_0) - \mathcal{J}(X_N)) \leq \rho(\mathcal{J}(X_0) - \inf \mathcal{J}).$$

Taking $N \to \infty$, the series is finite, and we get assertion (ii). ∎

Moreover, the sequence $\{X_k\}_{k\in\mathbb{N}}$ is bounded and has at least one accumulation point. Finally, one needs to prove that the set of limit points is a subset of the critical points for $\mathcal{J}$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

on which $\mathcal{J}$ is constant to establish global convergence. However, this would require a quite standard and essential requirement, i.e., to establish a subgradient lower bound for the iteration gap, which, in this case, depends on the Lipschitz continuity of ∇F but also on the subgradient of $G(X)$, which complicates the convergence analysis. Furthermore, an additional assumption on the class of non-convex functions $\mathcal{J}$ is required. In particular, $\mathcal{J}$ in (4.2) has the Kurdyka-Łojasiewicz property if the indicator function of the closed set $\mathcal{M}_r$ is semi-algebraic and $R(X)$ can be seen as the sum of semi-algebraic functions [32]. We will investigate these requirements in a future work.

5. Algorithm FB-LDR with continuation. The proposed FB-LDR Algorithm solves the minimization problem (4.1), where the choice of the regularization parameter λ represents a well-known crucial issue. With the aim to facilitate the choice of the effective λ parameter, we insert the FB-LDR algorithm in a continuation scheme (OUTER loop), which, starting from a reasonable value for λ^0, progressively diminishes the penalization parameter λ according to the decreasing of the objective functional, that is,

$$\lambda^i = c_\lambda \cdot \min(\lambda^{i-1}, \mathcal{J}\left(X^i; \lambda^{i-1}\right)), \qquad 0 < c_\lambda < 1. \tag{5.1}$$

For each outer iteration i, the minimization of (4.1) is then computed by the FB-LDR strategy (INNER loop) using a fixed λ^i value. The precision of the solution of the inner loop depends on the value of λ^i, which decreases at each iteration, and on a factor γ, which is taken small (0.01) or larger (0.1) depending on the freedom ratio of the sought-for solution as will be discussed in Section 6. The idea is to make the stopping criterion more stringent for harder MC problems.

This strategy, often successfully used under the name of *continuation* (see [13, 17, 30]), does not increase much the computing time of FB-LDR since the inner loop stopping criterion depends on λ^i, and each outer iteration uses as starting point the previous iterate (the so called *warm starting strategy*).

In order to accelerate the forward-backward inner iterative strategy, we used the variant of the *Fast Iterative Soft Thresholding Algorithm* (FISTA) introduced in [2], which has been proposed in [7] in order to achieve a convergent acceleration. In particular, at each inner iteration k, the weights α_k are computed as in [7], namely,

$$\alpha_k = \frac{t_k - 1}{t_{k+1}}, \qquad t_k = \frac{k + c - 1}{c}, \qquad c > 2 \qquad \forall k. \tag{5.2}$$

In our experiments the value for c is set to be 3. The stopping criterion for the inner loop is determined by the relative error:

$$err_{k+1} = \frac{|\mathcal{J}\left(X^i_{k+1}; \lambda^i\right) - \mathcal{J}\left(X^i_k; \lambda^i\right)|}{|\mathcal{J}\left(X^i_{k+1}; \lambda^i\right)|}. \tag{5.3}$$

In Algorithm 2 we report the main computational steps of the proposed implementation, which include the automatic choice of the regularization parameter λ in the cost function. The algorithm iterates over i in the outer loop and over k in the inner loop, and it is considered convergent for a given tolerance tol as soon as the following stopping criterion is satisfied:

$$\frac{\left|\sum_{s=1}^{\kappa} \phi\left(\sigma_s(X^{i+1}); a\right) - \sum_{s=1}^{\kappa} \phi\left(\sigma_s(X^{i}); a\right)\right|}{\sum_{s=1}^{\kappa} \phi\left(\sigma_s(X^{i+1}); a\right)} \leq \text{tol}. \tag{5.4}$$

The recovered low-rank matrix X^* will belong to the class of structured matrices (Toeplitz, Hankel, ...) as required by imposing the displacement rank r according to Table 2.1.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

The complexity of Algorithm 2 basically is determined by twice computing the singular value decompositions in the inner loop. The first SVD operator is applied to low-rank matrices Z_k since the number of sampled entries is typically much lower than the number of entries in the unknown matrix M. The second SVD operator is applied to the displacement matrices $\mathcal{L}(X)$, whose rank is very low; see Table 2.1. Consequently, we are interested in numerical methods for computing approximate SVD factorizations of large-scale low-rank matrices. Our implementation applies the routine *lansvd()* of the PROPACK library, which relies on Lanczos bidiagonalization with partial reorthogonalization [22]. Finally, the construction of the inverse displacement operator implies $O(rn \log n)$ operations. This allows us to classify the proposed algorithm as suitable for a class of medium-large scale MC problems.

Algorithm 2 FB-LDR-C.

inputs: $X_0 = \mathcal{P}_{\mathbf{\Omega}}(M)$, sampled matrix
outputs: X^* low-rank recovered matrix
parameters: $\lambda^0 > 0, \beta > 1, r > 0, \gamma > 0$

OUTER LOOP (OVER i): CONTINUATION
repeat
  INNER LOOP (OVER k): FB
  repeat
  Forward Step
    $Z_k = X^i_{k-1} - \frac{1}{\beta}(\mathcal{P}_{\mathbf{\Omega}}(X^i_{k-1}) - \mathcal{P}_{\mathbf{\Omega}}(M))$
  Backward Step
    $[U_Z, S_Z, V_Z] = svd(Z_k)$
    $\tilde{X}^{i+1}_k = U_Z \Theta(S_k; \lambda^i, a) V_Z^T$ (Θ defined in (4.6))
  Acceleration Strategy
    $t_{k+1} = \frac{k+c-1}{c}$, $\alpha_k = \frac{t_k - 1}{t_{k+1}}$
    $\hat{X}^{i+1}_k = \tilde{X}^{i+1}_k + \alpha_k(\tilde{X}^{i+1}_k - \tilde{X}^i_k)$
  Displacement Upgrade Step
    $[U, S, V] = svd(\mathcal{L}(\hat{X}^{i+1}_k), r)$
    $X^{i+1}_k = \mathcal{L}^{-1}(GH^T)$ (G, H defined in (2.4))
  until $err_{k+1} \le \gamma \cdot \lambda^i$ (err_{k+1} defined in (5.3))
  EXIT INNER LOOP
  Parameter update:
  $\lambda^{i+1} = c_\lambda \min(\lambda^i, \mathcal{J}\left(X^{i+1}_k; \lambda^i\right))$
until (5.4) is satisfied
$X^* = X^{i+1}$

The descent property of the sequence of values $\mathcal{J}(X^i; \lambda^i)$, for all i, is described by the following result, which guarantees that the reduction rule (5.1) generates a monotonically decreasing sequence of parameter values, and this ensures the convergence of the continuation strategy.

PROPOSITION 5.1. *Assume that the conditions of Proposition 4.5 hold. Let λ^i and λ^{i+1} be values of the penalization parameters at two successive iterative steps such that $\lambda^{i+1} < \lambda^i$, and X^i, X^{i+1} be the corresponding minimizers of the optimization problems (4.2). Then, the*

ETNA
Kent State University and
Johann Radon Institute (RICAM)

functional $\mathcal{J}(X;\lambda)$ *satisfies*

$$\mathcal{J}\left(X^{i+1};\lambda^{i+1}\right) < \mathcal{J}\left(X^{i};\lambda^{i}\right).$$

Proof. From the definition of $\mathcal{J}\left(X^{i};\lambda^{i}\right)$ in (4.1) and the assumption on X^i and X^{i+1} as minimizers, we have

$$\begin{aligned}\mathcal{J}\left(X^{i};\lambda^{i}\right) =& \frac{1}{2}\left\|P_\Omega(X^i)-P_\Omega(M)\right\|_F^2+\lambda^i R(X^i)\\ &> \frac{1}{2}\left\|P_\Omega(X^i)-P_\Omega(M)\right\|_F^2+\lambda^{i+1} R(X^i) = \mathcal{J}\left(X;\lambda^{i+1}\right)|_{X^i}\\ &\geq \frac{1}{2}\left\|P_\Omega(X^{i+1})-P_\Omega(M)\right\|_F^2+\lambda^{i+1} R(X^{i+1}) = \mathcal{J}(X^{i+1};\lambda^{i+1}),\end{aligned}$$

and this proves the result. ∎

6. Numerical experiments. In the following experiments we aim to demonstrate that, exploiting the non-convexity of the penalty function for the nuclear norm together with the rank constraint on the displacement structure, the proposed algorithm is able to recover low-rank matrices belonging to a given class of structured matrices (Toeplitz, Hankel,...), starting from a number of sampled data p much less than the estimated value in (1.3). We present two applications of the proposed method to the completion of low-rank Toeplitz matrices (Section 6.1) and to the spectrally sparse signal reconstruction via low-rank Hankel matrix completion (Section 6.2). The numerical experiments presented here aim to evaluate the proposed Algorithm FB-LDR-C and to compare its performance with some state-of-the-art methods proposed for similar problems. The initial regularization parameter λ^0 is set to be 0.1, the damping parameter c_λ defined in (5.1) is 10^{-5}, and the tolerance in (5.4) is 10^{-9}.

We will denote by FB-C the simplified version of Algorithm FB-LDR-C, where the displacement upgrade step is omitted. Hence, FB-C computes approximate solutions to the convex, unconstrained version of the low-rank matrix completion model (4.1).

The experiments were executed using Matlab 2019 on a PC with a 3.2GHz Intel i7-8700 CPU and 32 GB memory. The algorithms have been evaluated with respect to successful recovery rates, computational efficiency, robustness and are capability of handling particularly difficult recovery problems.

In the remainder of this section, before presenting some numerical results, we provide a brief summary of the evaluation metrics used throughout the reported experiments to measure the quality of the presented variational method as well as of the state-of-the-art algorithms considered for comparison.

Let $M \in \mathbb{R}^{m\times n}$ be a κ-rank matrix to be reconstructed. The set of matrices $\mathcal{M}_\kappa$ of rank up to κ is an algebraic variety of dimension $\kappa(m+n-\kappa)$.

We denote the *sampling ratio* as $SR = p/(mn)$, i.e., the number of measurements divided by the number of entries of the matrix, and we denote the degree of *freedom ratio* as $FR = \kappa(m+n-\kappa)/p$, i.e., the dimension of the set of rank-κ matrices divided by the number of measurements.

For the particular cases of Toeplitz and Hankel matrices $M \in \mathbb{R}^{n\times n}$, characterized by a string of $2n-1$ entries, the sampling ratio reduces to

$$SR = p/(2n-1), \tag{6.1}$$

where p is related to the string, i.e., $0 \leq p \leq 2n-1$, and the degree of freedom becomes

$$FR = \kappa(2n-\kappa)/p, \tag{6.2}$$

where p is here related to the full matrix M.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

The indicators (6.1) and (6.2) help to quantify the difficulty of a recovery problem. In particular, if $FR > 1$, then there is always an infinite number of matrices with rank κ with the given entries, so we cannot hope to recover the matrix M under this data condition. The largest rank such that $FR \leq 1$, denoted by κ_{max}, is given by

$$\kappa_{max} = (2n - \sqrt{(2n)^2 - 4p})/2. \tag{6.3}$$

We conducted several tests concerned with both easy matrix recovery, when $FR < 0.4$, and harder recovery problems, characterized by $FR \geq 0.4$. An MC problem is considered "easy" when the matrices are of very low-rank compared to the matrix size and the number of samples, and hence they are easy to recover. On the other hand, the âĂIJhardâĂİ problems involve matrices that are not of very low-rank and for which a very limited number of entries were sampled, thus they are considered very challenging problems; see [29] for more details.

6.1. Numerical results for Toeplitz matrix completion. In our tests, M denotes a real Toeplitz matrix of predefined rank κ. Random Toeplitz matrices of rank κ were created by the MATLAB function `M = gallery('toeppd',n,κ,w,theta)`, which returns an $n \times n$ symmetric, positive semi-definite (PSD) Toeplitz matrix composed of the sum of two PSD Toeplitz matrices of rank κ (or, for certain theta values, rank 1).

We compared the proposed Algorithm FB-LDR-C for structured MC problem with methods from the family of Augmented Lagrange Multipliers (ALM) modified to address the MC problem for Toeplitz matrices. In particular, we consider ALM [26], Smoothed ALM (SALM) [44], and Mean ALM (MV) [43], which recover Toeplitz matrices by solving convex optimization problems with a penalty given by the nuclear norm. The compared algorithms are optimized as proposed by the authors using the suggested parameter values. The performance of the algorithms for the MC problems is evaluated by the following error metric:

$$Err := \frac{\|X - M\|_F}{\|M\|_F}, \tag{6.4}$$

where X is the recovered matrix obtained from p samples of the full κ-rank matrix M.

We conducted two sets of MC tests for Toeplitz matrices; the first, reported in Table 6.1, is concerned with easy matrix recovery cases, where $FR < 0.4$, while the second, reported in Table 6.2, considers the hard recovery problems, characterized by $FR \geq 0.4$, where FR is defined by (6.2).

In Tables 6.1 and 6.2 we report the comparisons of the results obtained for 10 conducted random recovery tests in terms of average of execution time ($Time$) and error (Err), as defined in (6.4). In the first columns we report the dimension of the problem (n), the unknown rank of the matrix to be recovered (κ), the sampling ratio (SR), the freedom ratio (FR), and the algorithm applied. All times reported are in seconds.

For a given SR value, the higher the rank of the matrix to be reconstructed, the more difficult the MC problem is—compare, i.e., the first block of Table 6.1 with the second block of Table 6.2. Trivially, according to formulas (6.1) and (6.2), the higher the SR values, the easier the MC reconstruction problem. In our experiments we reconstruct Toeplitz matrices with a rank κ closest to the maximum rank κ_{max} in (6.3), such that the MC problem admits a unique solution starting from the given samples.

The proposed algorithm for the "easy" cases reported in Table 6.1 always obtained the exact reconstruction of the Toeplitz matrices, thanks to the improvement induced by the constraint on the low displacement rank matrix, at the expense of a slightly increased computational effort. When slightly lower accuracy is required, the unconstrained FB-C method shows very good performance in less computational time.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 6.1
Completion of Toeplitz matrices, easy cases ($FR < 0.4$).

n	κ	SR	FR	ALGORITHM	Time (sec)	Err
500	10	0.1	0.40			
				ALM	9.73	2.89e-07
				MV	**6.78**	**1.31e-10**
				SALM	8.46	1.13e-09
				FB-C	11.54	2.67e-04
				FB-LDR	9.18	2.27e-10
500	20	0.2	0.39			
				ALM	11.38	8.99e-02
				MV	15.58	8.38e-02
				SALM	13.41	5.06e-02
				FB-C	**3.80**	1.55e-06
				FB-LDR	6.42	**1.04e-10**
1000	20	0.1	0.39			
				ALM	33.27	1.80e-01
				MV	40.21	1.85e-01
				SALM	40.68	1.93e-01
				FB-C	**18.53**	4.54e-06
				FB-LDR	31.97	**1.45e-09**
2000	20	0.1	0.20			
				ALM	176.19	1.74e-01
				MV	212.39	1.92e-01
				SALM	217.39	1.92e-01
				FB-C	**125.94**	1.34e-04
				FB-LDR	167.99	**9.04e-09**

The real efficacy of the proposed FB-LDR-C algorithm is most encountered in Table 6.2 for the "difficult" cases, where all the compared methods fail to recover the Toeplitz matrices with an acceptable accuracy, while our proposal demonstrates robustness for high FR values, namely, the number of given samples is relatively small with respect to the dimension of the set of rank-k matrices.

6.2. Numerical results for Hankel matrix completion. The problem of spectrally sparse signal recovering from a random subset of regular time domain samples can be reformulated as a low-rank Hankel matrix completion problem [10, 12].

Let $x(t)$ be a continuous one-dimensional spectrally κ-sparse signal; that is, $x(t)$ is a weighted superposition of κ complex sinusoids

$$x(t) = \sum_{j=1}^{\kappa} c_j e^{(2\pi f_j - d_j)t}, \tag{6.5}$$

where c_j, f_j, and d_j represent the non-zero complex amplitude, the normalized frequency, and the damping factor of the jth sinusoid, respectively. Furthermore, let the column vector $x = [x(0), x(1), .., x(n-1)]^T$ denote the discrete samples of $x(t)$. In some circumstances, the measurements of the signal x are incomplete due to high experimental cost, hardware limitation, or other reasons. The spectrally sparse signal, as defined in (6.5), appears in a wide range of applications, for instance, in analog-to-digital conversion [42], nuclear magnetic resonance spectroscopy [14], and fluorescence microscopy [38].

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 6.2
Completion of Toeplitz matrices, difficult cases ($FR \geq 0.4$).

n	κ	SR	FR	ALGORITHM	Time (sec)	Err
500	10	0.05	0.76			
				ALM	10.47	7.85e-01
				MV	28.82	3.30e-01
				SALM	33.03	4.98e-01
				FB-C	**10.65**	4.76e-01
				FB-LDR-C	28.07	**4.50e-05**
500	20	0.1	0.78			
				ALM	13.05	4.20e-01
				MV	23.05	2.20e-01
				SALM	22.31	2.66e-01
				FB-C	**12.01**	2.04e-03
				FB-LDR-C	21.4	**3.99e-09**
500	20	0.08	0.96			
				ALM	10.24	6.47e-01
				MV	36.85	4.72e-01
				SALM	34.84	5.31e-01
				FB-C	**11.85**	9.81e-01
				FB-LDR-C	29.75	**1.72e-04**
1000	20	0.05	0.77			
				ALM	33.77	5.92e-01
				MV	97.20	4.39e-01
				SALM	144.73	4.61e-01
				FB-C	**34.98**	2.61e-01
				FB-LDR-C	109.28	**4.76e-05**
1000	30	0.1	0.59			
				ALM	81.3	9.18e-02
				MV	88.11	9.00e-02
				SALM	97.33	9.84e-02
				FB-C	**24.18**	9.90e-03
				FB-LDR-C	62.11	**3.80e-09**
2000	20	0.025	0.77			
				ALM	129.6	1.03
				MV	731.8	6.07e-01
				SALM	573.67	6.28e-01
				FB-C	**117.4**	1.82e-01
				FB-LDR-C	636.99	**1.03e-05**

In this experiment, we aim to recover $x(t)$ from partial measurements x_j, $j \in \Omega$, where Ω is a random set of indices of the observed entries, a subset of the complete index set $0, \ldots, 2n-1$, with $|\Omega| = p$, $p < 2n-1$. Recent results, such as those documented in [10, 12], reformulated the signal reconstruction problem within the framework of Low Rank Hankel Matrix Completion (LRHMC), which is the minimization problem (1.1) with $X = H(x)$, where $H(x)$ denotes a Hankel matrix arranged from the vector x.

In our experiment, the spectrally sparse signals of length n with κ frequency components are formed in the following way: each frequency f_j is uniformly sampled from $[0, 1)$, and the argument of each complex coefficient d_j is uniformly sampled from $[0, 2\pi)$, while the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

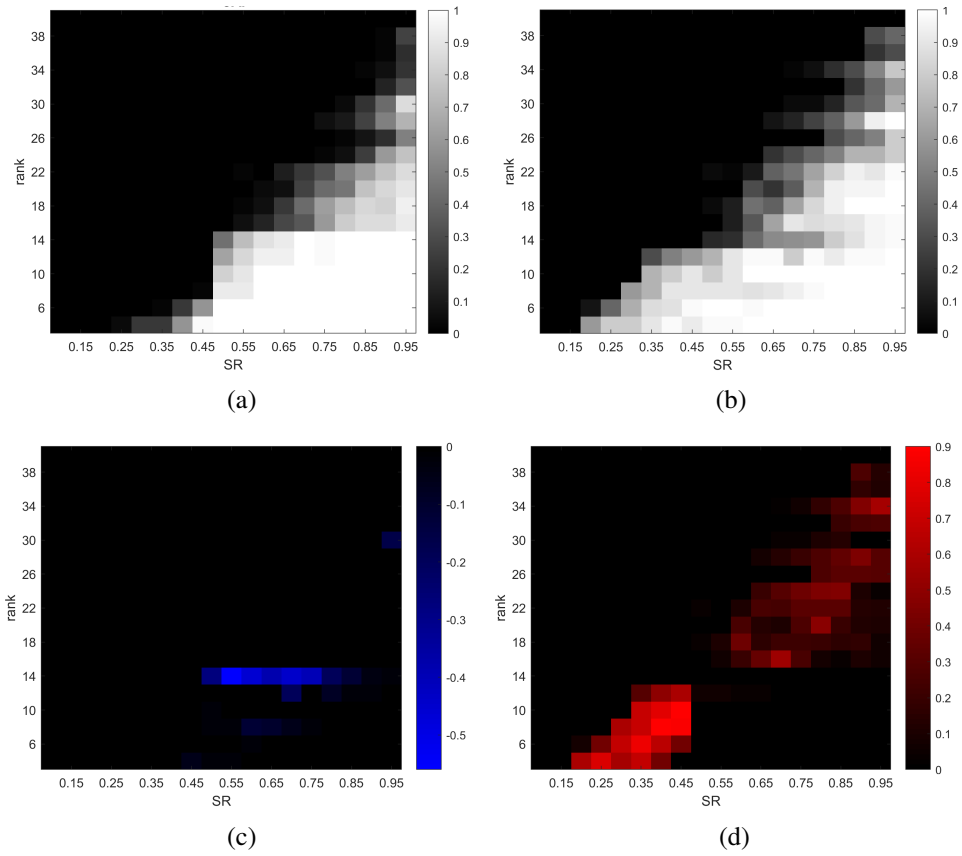

FIG. 6.1. *Phase transition recovery: (a) IHT algorithm; (b) FB-LDR-C algorithm; (c) IHT successes over FB-LDR-C; (d) FB-LDR-C successes over IHT.*

amplitude is selected to be $1 + 10^{0.5c_j}$ with c_j being uniformly distributed on $[0, 1]$. Then p entries of the test signals are uniformly randomly sampled. For a given triple (n, κ, p), 50 random tests were conducted. A successful reconstruction x_{rec} of a test signal x satisfies the relative error

$$\|x_{rec} - x\|_2 / \|x\|_2 \leq 10^{-3}. \tag{6.6}$$

The tests were conducted with $n = 127$ and $SR = p/n$ taking 18 equispaced values from 0.1 to 0.95. We compare our algorithm with the proposal in [3], named IHT.

The empirical phase transitions for the two tested algorithms are illustrated in Figure 6.1, first row, where each column corresponds to a different SR value, and each row corresponds to a different rank κ. Each pixel of the image is white colored when the algorithm recover all of the 50 random test signals, satisfying (6.6), while it is black colored if the algorithm fails to recover each of the randomly generated signals. Grayscale color indicates partial recover of the entire set of 50 signals.

The performance of IHT degrades severely when the sampling ratio SR is lower than 0.45, while FB-LDR-C can still achieve good performance even for the recovery of matrices with rank greater than 30. This is better highlighted observing Figure 6.1 (second row), where the differences between the performance in Figure 6.1(b) and Figure 6.1(a) are illustrated. In

ETNA
Kent State University and
Johann Radon Institute (RICAM)

particular, Figure 6.1(c) represents blue-colored the cases where IHT outperforms FB-LDR-C, while Figure 6.1(d) denotes in red color the cases where FB-LDR-C outperforms IHT.

7. Conclusion and research directions. This paper introduced a novel variational approach for matrix completion and related spectrally sparse signal recovery problems which involves matrices with low-rank displacement structure. We proposed a forward-backward strategy with projection for the minimization of the optimization problem, which leads to an easy to implement and surprisingly effective algorithm both in terms of computational cost and capability of handling particularly difficult recovery problems characterized by the freedom ratio $FR > 0.4$. Our algorithm exploits the constraint on the rank of the associated displacement matrix, which makes the original formulation (1.4) non-convex. Convexifying the set $\mathcal{M}_k$ is not of any interest since the convex hull is the whole space of rank-κ matrices in $\mathbb{R}^{m\times n}$; we could instead investigate the use of some relaxed forms of the rank function or the quasiconvex hull of the rank function [19].

Finally, a natural research direction will be to extend our optimization framework to include the recovery of other important classes of matrices with displacement structure such as the well-known classes of Vandermonde and Cauchy matrices, which play an important role in several applications.

Acknowledgements. We would like to thank the referees for comments that lead to improvements of the presentation. This research was supported in part by the National Group for Scientific Computation (GNCS-INDAM), Research Projects 2020.

REFERENCES

[1] M. BARNABEI AND L. B. MONTEFUSCO, *On a class of matrices with low displacement rank*, Linear Algebra Appl., 325 (2001), pp. 161–176.
[2] A. BECK AND M. TEBOULLE, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process., 18 (2009), pp. 2419–2434.
[3] J.-F. CAI, T. WANG, AND K. WEI, *Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank Hankel matrix completion*, Appl. Comput. Harmon. Anal., 46 (2019), pp. 94–121.
[4] J.-F. CAI, E. J. CANDÈS, AND Z. SHEN, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20 (2010), pp. 1956–1982.
[5] E. J. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.
[6] E. J. CANDÈS AND T. TAO, *The power of convex relaxation: near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2053–2080.
[7] A. CHAMBOLLE AND C. DOSSAL, *On the convergence of the iterates of the fast iterative shrinkage/thresholding algorithm*, J. Optim. Theory Appl., 166 (2015), pp. 968–982.
[8] R. H. CHAN, D. LAZZARO, S. MORIGI, AND F. SGALLARI, *A non-convex nonseparable approach to single-molecule localization microscopy*, in Scale Space and Variational Methods in Computer Vision, J. Lellmann, M. Burger, and J. Modersitzki, eds., Lecture Notes in Comput. Sci., vol. 11603, Springer, Cham, 2019, pp. 498–509.
[9] R. CHARTRAND, *Nonconvex splitting for regularized low-rank + sparse decomposition*, IEEE Trans. Signal Process., 60 (2012), pp. 5810–5819.
[10] Y. CHEN AND Y. CHI, *Robust spectral compressed sensing via structured matrix completion*, IEEE Trans. Inform. Theory, 60 (2014), pp. 6576–6601.
[11] A. L. CHISTOV AND D. Y. GRIGORIEV, *Complexity of quantifier elimination in the theory of algebraically closed fields*, in Mathematical Foundations of Computer Science, 1984, M. P. Chytil and V. Konbek, eds., Lecture Notes in Comput. Sci., vol. 176, Springer, Berlin, 1984, pp. 17–31.
[12] M. FAZEL, T. K. PONG, D. SUN, AND P. TSENG, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 946–977.
[13] M. FIGUEIREDO, R. NOWAK, AND S. WRIGHT, *Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems*, IEEE J. Sel. Topics Signal Proc., 1 (2007), pp. 586–598.
[14] D. GUO, H. LU, AND X. QU, *A fast low rank Hankel matrix factorization reconstruction method for non-uniformly sampled magnetic resonance spectroscopy*, IEEE Access, 5 (2017), pp. 16033–16039,

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[15] S. GU, L. ZHANG, W. ZUO, AND X. FENG, *Weighted nuclear norm minimization with application to image denoising*, in IEEE Conference on Computer Vision and Pattern Recognition, 2014, IEEE Conference Proceedings, Los Alamitos, 2014, pp. 2862–2869.
[16] J. P. HALDAR, *Low-rank modeling of local k-space neighborhood (LORAKS) for constrained MRI*, IEEE Trans. Med. Imag., 33 (2014), pp. 668–681.
[17] E. HALE, W. YIN, AND Y. ZHANG, *Fixed-point continuation for ℓ_1-minimization: methodology and convergence*, SIAM J. Optim., 19 (2008), pp. 1107–1130.
[18] P. C. HANSEN, *Deconvolution and regularization with Toeplitz matrices*, Numer. Algorithms, 29 (2002), pp. 323–378.
[19] J.-B. HIRIART-URRUTY AND H. Y. LE, *Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function*, Optim. Lett., 6 (2010), pp. 841–849.
[20] T. KAILATH AND A. H. SAYED, *Displacement structure: theory and applications*, SIAM Rev., 37 (1995), pp. 297–386.
[21] T. KAILATH, S. KUNG, AND M. MORF, *Displacement ranks of matrices and linear equations*, J. Math. Anal. Appl., 68 (1979), pp. 395–407.
[22] R. M. LARSEN, *Lanczos Bidiagonalization with Partial Reorthogonalization*, Ph.D. Thesis, Dept. Computer Science, University of Aarhus, Aarhus, 1998.
Code available at http://soi.stanford.edu/~rmunk/PROPACK
[23] A. LANZA, S. MORIGI, AND F. SGALLARI, *Convex image denoising via non-convex regularization with parameter selection*, J. Math. Imaging Vision, 56 (2016), pp. 195–220.
[24] A. LANZA, S. MORIGI, I. SELESNICK, AND F. SGALLARI, *Nonconvex nonsmooth optimization via convex-nonconvex majorization-minimization*, Numer. Math., 136 (2017), pp. 343–381.
[25] D. LAZZARO, *A nonconvex approach to low-rank matrix completion using convex optimization*, Numer. Linear Algebra Appl., 23 (2016), pp. 801–824.
[26] Z. LIN, M. CHEN, L. WU, AND Y. MA, *The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices*, Tech. Rep. UIUL-ENG-09-2214, University of Illinois at Urbana-Champaign, Champaign, 2010.
[27] C. LU, C. ZHU, C. XU, S. YAN, AND Z. LIN, *Generalized singular value thresholding*, in Proc. of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15), AAAI Press, Palo Alto, 2015, pp. 1805–1811.
[28] S. MA, D. GOLDFARB, AND L. CHEN, *Fixed point and Bregman iterative methods for matrix rank minimization*, Math. Program. 128 (2011), pp. 321–353.
[29] K. MOHAN AND M. FAZEL, *Iterative reweighted algorithms for matrix rank minimization*, J. Mach. Learn. Res., 13 (2012), pp. 3441–3473.
[30] L .B. MONTEFUSCO AND D. LAZZARO, *An iterative L_1-based image restoration algorithm with an adaptive parameter estimation*, IEEE Trans. Image Process., 21 (2012), pp. 1676–1686.
[31] R. R. NADAKUDITI, *OptShrink: an algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage*, IEEE Trans. Inform. Theory, 60 (2014), pp. 3002–3018.
[32] P. OCHS, A. DOSOVITSKIY, T. BROX, AND T. POCK, *On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision*, SIAM J. Imaging Sci., 8 (2015), pp. 331–372.
[33] V. Y. PAN, *Structured Matrices and Polynomials*, Springer, New York, 2001.
[34] A. PAREKH AND I. W. SELESNICK, *Enhanced low-rank matrix approximation*, IEEE Signal Proc. Lett., 23 (2016), pp. 493–497.
[35] ———*Convex denoising using non-convex tight frame regularization*, IEEE Signal Proc. Lett., 22 (2015), pp. 1786–1790.
[36] ———, *Improved sparse low-rank matrix estimation*, Signal Proc., 139 (2017), pp. 62–69.
[37] B. RECHT, M. FAZEL, AND P. A. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
[38] L. SCHERMELLEH, R. HEINTZMANN, AND H. LEONHARDT, *A guide to super-resolution fluorescence microscopy*, J. Cell Biol., 190 (2010), pp. 165–175.
[39] I. W. SELESNICK AND İ. BAYRAM, *Sparse signal estimation by maximally sparse convex optimization*, IEEE Trans. Signal Process., 62 (2014), pp. 1078–1092.
[40] I.W. SELESNICK, A. PAREKH, AND İ. BAYRAM, *Convex 1-D total variation denoising with non-convex regularization*, IEEE Signal Proc. Lett., 22 (2015), pp. 141–144.
[41] K. C. TOH AND S. YUN, *An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems*, Pac. J. Optim., 6 (2010), pp. 615–640.
[42] J. A. TROPP, J. N. LASKA, M. F. DUARTE, J. K. ROMBERG, AND R. G. BARANIUK, *Beyond Nyquist: efficient sampling of sparse bandlimited signals*, IEEE Trans. Inform. Theory, 56 (2010), pp. 520–544.
[43] C. L. WANG AND C. LI, *A mean value algorithm for Toeplitz matrix completion*, Appl. Math. Lett., 41 (2015), pp. 35–40.
[44] R. P. WEN, S. Z. LI, AND F. ZHOU, *Toeplitz matrix completion via smoothing augmented Lagrange multiplier algorithm*, App. Math. Comput., 355 (2019), pp. 299–310.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[45] R. P. WEN AND X.H. YAN, *A new gradient projection method for matrix completion*, Appl. Math. Comput., 258 (2015), pp. 537–544.