A nonconvex approach to low-rank matrix completion using convex optimization

Published Version:
A nonconvex approach to low-rank matrix completion using convex optimization / Lazzaro, Damiana. - In: NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS. - ISSN 1070-5325. - ELETTRONICO. - 23:5(2016), pp. 801-824. [10.1002/nla.2055]

Availability:
This version is available at: https://hdl.handle.net/11585/565705 since: 2016-10-25

Published:
DOI: http://doi.org/10.1002/nla.2055

NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS

Numer. Linear Algebra Appl. 0000; **00**:1–36

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla

A Nonconvex Approach to Low-Rank Matrix Completion using Convex Optimization

Damiana Lazzaro*

Department of Mathematics, University of Bologna, Piazza di Porta S. Donato, 40123 Bologna (Italy)

SUMMARY

This paper deals with the problem of recovering an unknown low-rank matrix from a sampling of its entries. For its solution we consider a nonconvex approach based on the minimization of a nonconvex functional that is the sum of a convex fidelity term and a nonconvex, nonsmooth relaxation of the rank function. We show that by a suitable choice of this nonconvex penalty it is possible, under mild assumptions, to use also in this matrix setting the iterative Forward-Backward splitting method. Specifically, we propose the use of certain parameter dependent nonconvex penalties that with a good choice of the parameter value allow us to solve in the backward step a convex minimization problem and we exploit this result to prove the convergence of the iterative Forward-Backward splitting algorithm. Based on the theoretical results, we develop for the solution of the matrix completion problem the efficient iterative Improved Matrix Completion Forward-Backward (IFBMC) Algorithm, which exhibits lower computing times and improved recovery performance when compared to the best state-of-the-art algorithms for matrix completion.

Received . . .

KEY WORDS: Matrix Completion, Nonconvex Minimization, Forward-Backward Splitting, Proximity Operator, Matrix Setting, Convex Optimization.

1. INTRODUCTION

The problem of recovering an unknown low-rank matrix from a sampling of its entries has recently become an important research area. The popularity and primary motivation of the matrix completion problem comes from recommendation systems, where users submit ratings only on a subset of entries of a low-rank user-item ratings matrix and the task is to complete the matrix from a small number of its entries. Many other real world models can be categorized as matrix completion problems, such as video denoising [1], model reduction [2], computer vision [3–5], multiclass learning in data analysis [6, 7] and so on.

The general form of the matrix completion problem is:

$$\min_{X \in \mathbb{R}^{m \times n}} \text{rank}(X), \quad \text{s. t.} \quad X_{ij} = M_{ij}, (i,j) \in \Omega, \tag{1}$$

where we are given only p sampled entries $\{M_{i,j} : (i,j) \in \Omega\}$, where Ω is a random subset of cardinality p.

Problem (1) is an NP-hard optimization problem, due to the nonconvexity and combinatorial nature of the rank function [8]. To overcome such a difficulty in [9–11] the authors prove that the convex envelope of rank(X) on the set $\left\{X \in \mathbb{R}^{m \times n} : \|X\|_2 \leq 1\right\}$ is the nuclear norm$\|X\|_*$, namely $\|X\|_* = \sum_{i=1}^{r} \sigma_i(X)$, where $\sigma_i(X),\ i = 1, ..., r, r \leq \min(m,n)$, are the singular values of X. They use, therefore, the nuclear norm as a convex relaxation for the rank function. Moreover, they prove that a random low-rank matrix can be recovered exactly with high probability from a rather small random sample of its entries by solving the convex minimization problem:

$$\min_{X \in \mathbb{R}^{m \times n}} \{||X||_* : X_{ij} = M_{ij}, (i,j) \in \Omega\}. \tag{2}$$

By defining

$$\mathcal{P}_{\mathbf{\Omega}}(X) = \begin{cases} \text{X}_{ij} & \text{if } (i,j) \in \Omega \\ 0 & \text{if } (i,j) \notin \Omega \end{cases},$$

problem (2) can be recast as:

$$\min_{X \in \mathbb{R}^{m \times n}} \{||X||_* : \mathcal{P}_{\mathbf{\Omega}}(X) = \mathcal{P}_{\mathbf{\Omega}}(M)\}. \tag{3}$$

1.1. Existing methods

Many existing algorithms rely on the convex relaxation (3). For example in [12, 13] it is shown that problem (3) can be solved making use of efficient semidefinite programming solvers, such as SeDuMi, SDPT3. Nevertheless the high computational cost and the memory demanding of these algorithms make them unsuitable for handling large size problems. Different approaches, more suited for large-scale matrix completion, solve the Lagrangian version of (3) i.e.

$$\min_{X \in \mathbb{R}^{m \times n}} \left\{ \frac{1}{2} ||\mathcal{P}_{\mathbf{\Omega}}(X) - \mathcal{P}_{\mathbf{\Omega}}(M)||_F^2 + \lambda ||X||_* \right\}, \tag{4}$$

where $\|\cdot\|_F$ is the Frobenius norm and λ is a suitable penalization parameter.

Since this formulation is similar to the well known and widely studied compressed sensing reconstruction problem [14], where the sparsity inducing ℓ_1-norm is used as convex relaxation of the ℓ_0-norm, most of the algorithms proposed for matrix completion are inspired by recent works in the area of sparse reconstruction. In particular, we refer to the singular values thresholding methods, such as the SVT-method proposed in [15], the alternating direction augmented Lagrangian methods (ALM) [16–18], the fixed point and Bregman iterative method (FPCA) [19] and the accelerated proximal gradient algorithm (APGL) [20]. All these algorithms are based on iterative soft thresholding of the singular values of certain matrices and are designed to obtain reconstructed matrices with the lowest nuclear norm. Even if these algorithms largely outperform the performance of SeDuMi and SDPT3, the high computational cost of the singular value decomposition greatly reduces their use in case of large dimension matrices. A remarkable improvement can be obtained by the use of truncated SVD algorithms, such as Linear-Time SVD [21] and randomized SVD [24], but this implies an efficient identification of a suitable truncation level.

Another strategy to solve the low-rank matrix completion problem is to apply low-rank matrix factorization methods [25, 26]: the aim is to find a matrix that is as close as possible to the data, according to a certain fidelity function, and that is, at the same time, factorisable into the product of two low-rank matrices. The corresponding algorithms are fast since they avoid the use of the SVD, but they are effective only for Easy problems (see Section 7.1 for definition of Easy and Hard problems). Their weakness derives from their need to know a good approximation of the unknown rank of the original matrix. When this approximation is not available, the methods based on the convex relaxation (2), (4) are preferable since they solve Easy and Hard problems with similar performance; but they still present some unresolved problems such as their inability, in certain situations, to retrieve the true relevant variables. In fact, the soft thresholding operator associated with the nuclear norm [15], shrinks the singular values of the same amount , not taking into account that the larger singular values are usually associated with the major projection orientations and therefore they should be penalized less, to preserve important information. To overcome such problems, other types of penalty that relax the convexity property have been considered in problem (4) that becomes:

$$\min_{X \in \mathbf{R}^{m \times n}} \{F(X) = H(X) + \lambda R(X)\} \tag{5}$$

where $H(X) = \frac{1}{2}\|\mathcal{P}_{\mathbf{\Omega}}(X) - \mathcal{P}_{\mathbf{\Omega}}(M)\|_F^2$ and $R(X)$ is a nonsmooth, nonconvex functional that approximates the rank function better than the nuclear norm. While being theoretically appealing, the nonconvex relaxation approach leads to a more challenging minimization problem, since the solution of problem (5) is usually not unique and algorithms performing nonconvex minimization may get trapped in bad local optimal solutions.

A widely used strategy to solve problem (5) is to replace it with a sequence of approximated convex subproblems. Specifically we refer to the iterative reweighted ℓ_2/ℓ_1 algorithms IRLS, t-IrucLq-M, IRNN, sIRLS0, proposed in [27–30], respectively. They solve a Weighted Singular value thresholding problem iteratively, where the weight vector is given by the gradient of the nonconvex penalty evaluated at the previous iterate. These algorithms exhibit superior performance when compared to the existing nuclear norm minimization algorithms, but they are more computationally demanding. A different approach is proposed in [31], where a low-rank solution is obtained iteratively by estimating an approximated matrix rank and shrinking only non-dominant singular values. The corresponding Iterative Partial Matrix Shrinkage algorithm (IPMS) has the same performance to solve hard problems as sIRLS0, but it takes less computing time.

Another line of research for solving the nonconvex optimization problem (5), where $H(X)$ is differentiable with a Lipschitz-continuous gradient, is to apply the Forward-Backward (FB) scheme [32–34] well known and largely used in the context of sparse reconstruction both for convex and nonconvex relaxation of the ℓ_0-norm [34–37]. This scheme consists in iteratively alternating between a descent step along the gradient of $H(X)$ (forward step) and a minimization step, involving the proximity operator of the functional $R(X)$ (backward step). While the convergence of this iterative method is well assessed for nonsmooth but convex $R(X)$ both in the vector and in matrix setting [20, 34], its behavior is not clear for nonconvex $R(X)$, since the corresponding proximity operator is not unique, in general, and its closed form may be not available. This represents a great limit to the applicability of this scheme to the nonconvex case.

1.2. Contribution and paper organization

Motivated by the theoretical problems related to the convergence of the FB scheme in the case of nonconvex $R(X)$, in this paper we focus on the application of this iterative splitting method for the solution of the nonconvex relaxation of the matrix completion problem (5). Our theoretical contribution is twofold: first we show that the use of certain parameter dependent nonconvex penalties, with a suitable choice of the parameter value, allows us to solve in the backward step a convex optimization problem whose corresponding proximity operator is unique and has a simple closed form. Based on this result, then we use the *KL-property* [38] of the cost functional $F(X)$ to give a rigorous proof of the convergence of the FB splitting method for matrix completion problem (FBMC) to a critical point of $F(X)$. Moreover, in the IMCBF algorithm we present an efficiently improved implementation of the proposed approach that includes an incremental strategy for matrix rank estimation and a continuation approach for the evaluation of the penalization parameter λ. Extensive numerical experiments and comparisons with the most efficient state-of-the-art algorithms for matrix completion highlight the effectiveness of the proposed algorithm, which outperforms the others both in terms of reconstruction capabilities and computing time.

The paper is organized as follows: in the next Section we briefly recall the Iterative Forward-Backward Splitting scheme for low-rank matrix recovery. In Section 3 the FB scheme is applied to the matrix completion problem and we propose a new strategy to deal with a convex optimization problem in spite of the nonconvexity of the penalty function. Section 4 is completely devoted to prove the convergence of the iterative FB scheme when applied to matrix completion and Section 5 our theoretical results are used to develop the efficient iterative algorithm IFBMC. The performance of the proposed algorithm is assessed in Section 6 by presenting comparisons with the best state-of-the-art algorithms on simulated and real recovery problems. A short conclusion in Section 7 summarizes the main contributions of the paper.

2. THE ITERATIVE FORWARD-BACKWARD SCHEME FOR LOW-RANK MATRIX RECOVERY

In this Section we briefly recall the derivation of the FB scheme for the solution of the following nonsmooth, nonconvex minimization problem

$$\min_{X \in \mathcal{M}_r} F(X) = H(X) + \lambda R(X), \tag{6}$$

where $\lambda \geq 0$, $\mathcal{M}_r$ is the set of real $m \times n$ matrices with rank less or equal to $r \leq \min(m, n)$, $R : \mathcal{M}_r \to \mathbb{R}$ is a nonsmooth, nonconvex relaxation of the rank function that enforces low-rank on the recovered matrix, and the functional H, that represents a constraint to be satisfied by the low-rank matrix we want to recover, is such that holds

A1- $H : \mathcal{M}_r \to \mathbb{R}$ is a smooth convex continuously differentiable functional with $L-$ Lipschitz continuous gradient:

$$\|\nabla H(X) - \nabla H(Y)\|_F \leq L\|X - Y\|_F, \quad X, Y \in \mathcal{M}_r.$$

For the solution of (6), for any matrix $Y \in \mathcal{M}_r$ and $\mu > 0$, we consider the following quadratic approximation of the objective functional $F(X)$ at Y:

$$Q(X, Y) = H(Y) + < X - Y, \nabla H(Y) > + \frac{\mu}{2}\|X - Y\|_F^2 + \lambda R(X) \tag{7}$$

that can be written as

$$Q(X, Y) = H(Y) + \frac{\mu}{2}\left\| X - \left(Y - \frac{1}{\mu}\nabla H(Y)\right)\right\|_F^2 - \frac{1}{2\mu}\|\nabla H(Y)\|_F^2 + \lambda R(X), \tag{8}$$

and we evaluate a minimizer of (8) iteratively by setting $Y = X_{k-1}$, and ignoring $X-$ independent terms. We obtain therefore that:

$$X_k = \min_{X \in \mathcal{M}_r} Q(X, X_{k-1}) = \text{prox}_{\frac{\lambda}{\mu} R}(Z_k), \tag{9}$$

where

$$Z_k = X_{k-1} - \frac{1}{\mu} \nabla H(X_{k-1}) \tag{10}$$

and the proximity operator of the functional $R(X)$ is defined for any $Z_k \in \mathcal{M}_r$ as

$$\text{prox}_{\frac{\lambda}{\mu} R}(Z_k) = \arg \min_{X \in \mathcal{M}_r} \left\{ \frac{\mu}{2} \|X - Z_k\|_F^2 + \lambda R(X) \right\}. \tag{11}$$

This leads to the following *Forward Backward Scheme for Low-Rank Matrix Recovery* (FBLRM)

Algorithm 1 (FBLRM)

Given $X_0 \in \mathbb{R}^{m \times n}$, $R(X)$, $H(X)$ and for fixed μ,$\lambda > 0$,

for k=1,2,...do

Forward Step

$\text{Z}_k = X_{k-1} - \frac{1}{\mu} \nabla H(X_{k-1})$

Backward Step

$\text{X}_k = \text{prox}_{\frac{\lambda}{\mu} R}(Z_k)$

until convergence

(12)

We remark that the FB iterative scheme is well known in the vector setting and its convergence has been proved in [32,34] in the case of R convex. In the nonconvex case, the problem is more involved because the evaluation of the proximity operator may be difficult and the result not unique, since they may be several local minima. This case is studied in [38] where the authors prove, under some restrictive hypotheses, a convergence result to a critical point of the cost functional.

The extension of the iterative FB scheme to the matrix setting, as done in FBLRM presents further unresolved problems: the first is the existence of a unique proximity operator of the nonconvex functional $R(X)$, or correspondingly the evaluation of a global minimum in the Backward step, necessary to obtain the monotonically decreasing of the objective function, last, but not least, is the proof of the convergence of the sequence of the iterates X_k to a critical point of $F(X)$. In the next two Sections we consider the above issues in the context of the nonconvex matrix completion problem and we propose a satisfactory and efficient solution.

3. THE NONCONVEX MATRIX COMPLETION PROBLEM: CONVEX MINIMIZATION USING NONCONVEX PENALTIES

In this Section we apply the FBLRM algorithm to the nonconvex matrix completion problem. More in detail, we consider

$$H(X) = \frac{1}{2}||\mathcal{P}_{\mathbf{\Omega}}(X) - \mathcal{P}_{\mathbf{\Omega}}(M)||_F^2, \tag{13}$$

that satisfies condition A1 and we choose $R(X)$ such that it holds:

A2- $R(X) : \mathcal{M}_r \to \mathbb{R}$, is a proper lower semicontinuous nonconvex functional defined as

$$R(X) = \sum_{i=1}^{r} \phi(\sigma_i(X), a), \tag{14}$$

where $\phi(x, a) : \mathbb{R} \to \mathbb{R}$ is a parametrized continuous, symmetric, twice differentiable on $\mathbb{R} \setminus \{0\}$, increasing and nonconvex function, $\sigma_i(X)$,$i = 1, .., r$, are the singular values of the matrix $X \in \mathcal{M}_r$, and $a > 0$.

The nonconvex matrix completion problem can be, then, formulated as:

$$\min_{X \in \mathcal{M}_r} \left\{ F(X) = \frac{1}{2}||\mathcal{P}_{\mathbf{\Omega}}(X) - \mathcal{P}_{\mathbf{\Omega}}(M)||_F^2 + \lambda \sum_i \phi(\sigma_i(X), a) \right\}. \tag{15}$$

By applying in this case the FBLRM algorithm, by taking into account (9),(10),(11) the Forward step becomes:

$$Z_k = X_{k-1} - \frac{1}{\mu}(\mathcal{P}_{\mathbf{\Omega}}(X_{k-1}) - \mathcal{P}_{\mathbf{\Omega}}(M)) \tag{16}$$

and the Backward step

$$X_k = \arg \min_{X \in \mathcal{M}_r} \left\{ \frac{\mu}{2}\|X - Z_k\|_F^2 + \lambda \sum_{i=1}^{r} \phi(\sigma_i(X), a) \right\} = \mathrm{prox}_{\frac{\lambda}{\mu} R}(Z_k) \tag{17}$$

Algorithm 1 in the case of the Matrix Completion Problem can be rewritten as:

Algorithm 2: Forward Backward Algorithm for Matrix Completion (FBMC)

Input: $X_0 = \mathcal{P}_{\mathbf{\Omega}}(M)$, $\lambda > 0$, $\mu > L$, $F(X_0)$, $\gamma < 1$, $r \leq \min(m, n)$

k=1

while $condition > \gamma \cdot \lambda$

Forward Updating Step

$Z_k = X_{k-1} - \frac{1}{\mu}(\mathcal{P}_{\mathbf{\Omega}}(X_{k-1}) - \mathcal{P}_{\mathbf{\Omega}}(M))$

Backward Minimization Step

$$\mathrm{X}_k = \arg\min_{X \in \mathcal{M}_r} \left\{ \frac{\mu}{2}\|X - Z_k\|_F^2 + \lambda \sum_{i=1}^{r} \phi(\sigma_i(X), a) \right\} = \ \mathrm{prox}_{\frac{\lambda}{\mu}}(Z_k) \quad (18)$$

$condition = \frac{F(X_k) - F(X_{k-1})}{F(X_k)}$

k=k+1

end

Output: $\hat{X} = X_{k-1}$ as low-rank solution

Finding an unique solution of the nonconvex minimization (18) is a really challenging problem. Despite that, we found a way to succeed by choosing the parametrized nonconvex penalty function $\phi(x, a)$ according to the recent results of the nonconvex sparse signal processing literature given [40–42].

3.1. Proximity operators of Parametrized Nonconvex penalty functions

Let $\phi(x, a)$ be the nonconvex parametrized penalty function given by

$$\phi(x, a) = \frac{1}{a} \log(1 + a|x|) \ , a > 0. \quad (19)$$

that satisfies regularity conditions similar to those given in condition A2 and where the parameter a controls the degree of nonconvexity of the penalty function. In [40–42] the authors prove that for $a \leq \frac{1}{\lambda}$ the function f in the definition of the proximity operator of such nonconvex $\phi(x, a)$

$$\mathrm{prox}_{\lambda\phi}(y, a) = \arg\min_{x \in \mathbb{R}} \left\{ f(x) = \frac{1}{2} |y - x|^2 + \lambda\phi(x, a) \right\} \quad (20)$$

is strict convex. The proximity operator is therefore the unique solution of the convex minimization problem (20) and it is a continuous non-linear *threshold function* with λ as threshold value, namely

$$\text{prox}_{\lambda\phi}(y,a) = 0, \, \forall \, |y| < \lambda. \tag{21}$$

It is given by:

$$\text{prox}_{\lambda\phi}(y,a) = \begin{cases} \left[\frac{|y|}{2} - \frac{1}{2a} + \sqrt{\left(\frac{|y|}{2} + \frac{1}{2a}\right)^2 - \frac{\lambda}{a}} \right] & |y| \geq \lambda \\ 0 & |y| \leq \lambda \end{cases} \tag{22}$$

While the soft thresholding function, that is the proximity operator of the absolute value, shrinks all arguments of the same amount, regardless of their values, the threshold function of the considered nonconvex penalties approaches identity asymptotically and in this way large values are not underestimated.

3.2. Proximity Operator in the matrix setting

In this subsection we use the previous results in order to prove that in the matrix setting it is possible to give an unique and simple expression of the proximity operator

$$\text{prox}_{\lambda R}(G) = \min_{X \in \mathcal{M}_r} \left\{ \frac{1}{2} ||X - G||_F^2 + \lambda \sum_i \phi(\sigma_i(X), a) \right\}. \tag{23}$$

where $\phi(x,a)$ is the nonconvex parametrized penalty function defined in (19) and $a \leq \frac{1}{\lambda}$.

Since in the matrix completion problem the rank revealing singular values are positive, from now on we restrict the function domain to be $\mathbb{R}^+$.

Theorem 3.1

Let $\phi(x,a) : \mathbb{R}^+ \to \mathbb{R}^+$ be the nonconvex parametrized penalty function given by (19) and that satisfies condition A2 with $a \leq \frac{1}{\lambda}$. Let $G = U\mathrm{D}iag(\sigma(G))V^T$ be the Singular Value Decomposition of $G \in \mathcal{M}_r$. The optimal solution to

$$\text{prox}_{\lambda R}(G) = \min_{X \in \mathcal{M}_r} \left\{ \frac{1}{2} ||X - G||_F^2 + \lambda \sum_i \phi(\sigma_i(X), a) \right\}, \tag{24}$$

is given by

$$X^* = U\mathrm{D}iag(\sigma^*)V^T, \tag{25}$$

where

$$\sigma_i^* = \mathrm{prox}_{\lambda\phi}(\sigma_i(G), a) = \arg\min_{\sigma_i \geq 0} \left\{ \frac{1}{2}(\sigma_i - \sigma_i(G))^2 + \lambda\phi_\lambda(\sigma_i, a) \right\}, \quad i = 1, ..., r$$

with $\sigma_1^* \geq \sigma_2^* \geq ... \geq \sigma_m^*, i = 1, ..., r$.

In order to prove Theorem (3.1), we need of the following results:

Proposition 3.1

Let $\phi(x, a)$ be the nonconvex parametrized penalty function given by (19) and that satisfies condition A2 with $a \leq \frac{1}{\lambda}$, then the proximity operator $\mathrm{prox}_{\lambda\phi}(\cdot, a)$ is strictly monotonically increasing, namely if $y_2 > y_1$, we have

$$\mathrm{prox}_{\lambda\phi}(y_2, a) > \mathrm{p}rox_{\lambda\phi}(y_1, a). \tag{26}$$

Proof: The proof immediately follows from the fact that under our hypotheses the function f in the definition of the proximity operator of $\phi(x, a)$ (20) is strict convex. □

Lemma 3.1

For any matrices $A, B \in \mathbb{R}^{m\times n}(m \leq n)$, the von Neumann'trace inequality [43, 44] holds:

$$Tr(A^TB) \leq \sum_{i=1}^{m} \sigma_i(A)\sigma_i(B) \tag{27}$$

where $\sigma_1(A) \geq \sigma_2(A) \geq ... \geq 0$ and $\sigma_1(B) \geq \sigma_2(B) \geq ... \geq 0$ are the singular values of A and B, respectively. Equality holds if A and B share the same right and left singular vectors, namely $A = UDiag(\sigma(A))V^T$ and $B = UDiag(\sigma(B))V^T$.

We are now in position to prove Theorem 3.1:

Proof of Theorem 3.1: By Lemma (3.1) we have:

$$\|X-G\|_F^2 = Tr((X-G)^T(X-G)) = \text{Tr}(X^TX) - 2Tr(X^TG) + Tr(G^TG)$$
$$= \sum_{i=1}^r \sigma_i^2(X) - 2Tr(X^TG) + \sum_{i=1}^r \sigma_i^2(G)$$
$$\geq \sum_{i=1}^r \sigma_i^2(X) - 2\sum_{i=1}^r \sigma_i(X)\sigma_i(G) + \sum_{i=1}^r \sigma_i^2(G)$$

Since equality holds when X and G share the same right and left singular vectors, we assume that $X = U\text{Diag}(\sigma(X))V^T$ and $G = U\text{Diag}(\sigma(G))V^T$. Therefore it holds that:

$$\|X-G\|_F^2 = \sum_{i=1}^r (\sigma_i(X) - \sigma_i(G))^2 .$$

The optimization problem (24) reduces now to

$$\arg \min_{\sigma_1(X)\geq\sigma_2(X)\geq\ldots\geq\sigma_r(X)\geq 0} \sum_{i=1}^r \left\{ \frac{1}{2}(\sigma_i(X) - \sigma_i(G))^2 + \lambda\phi(\sigma_i(X), a) \right\} . \tag{28}$$

The objective function in (28) is separable and each term is strict convex for $a \leq \frac{1}{\lambda}$: for each $\sigma_i(G)$ there exists a unique minimum

$$\sigma_i^* = \text{prox}_{\lambda\phi}(\sigma_i(G), a), \tag{29}$$

that is a thresholded version of $\sigma_i(G)$. Moreover, by Proposition (3.1), we have that $\sigma_1^* \geq \sigma_2^* \geq ... \geq \sigma_r^*$. Therefore it follows that $X^* = U\text{Diag}(\sigma^*)V^T$ is a global optimal solution to (24). □

We now define the *matrix thresholding operator* $\Theta_\lambda(G)$ as follows:

$$\Theta_\lambda(G) = U\text{Diag}(\sigma^*)V^T \tag{30}$$

where $\sigma_i^* = \text{prox}_{\lambda\phi}(\sigma_i(G), a), \quad i = 1, ...r, \quad r \leq \min(m, n)$, and U and V are the singular vectors of G, namely $G = U\text{Diag}(\sigma(G))V^T$.

Relation (18) in Algorithm 2, can be then rewritten as:

$$X_k = \min_{X\in\mathcal{M}_r} \left\{ \frac{\mu}{2}||X - Z_k||_F^2 + \lambda \sum_i \phi(\sigma_i(X), a) \right\} = \Theta_{\frac{\lambda}{\mu}}(Z_k) \tag{31}$$

4. CONVERGENCE ANALYSIS

In this Section we prove in our context the convergence of the sequence X_k generated by FBMC to a critical point of the objective functional $F(X)$ given in (15), when the penalty $\phi(x, a)$ is given by

(19) with $a \leq \frac{1}{\lambda}$. We suppose that Assumptions 1 and 2 are verified and that the following holds:

A3- $F(X) : \mathcal{M}_r \to \mathbb{R}$ is an analytical functional, bounded from below and coercive.

We remark that the choice of $\phi(x, a)$ as in (19), with $a \leq \frac{1}{\lambda}$, and the restriction of its domain to $\mathbb{R}^+$ ensures that A3 is satisfied.

Since the proof holds in a general context, provided that assumptions A1-A3 are verified, we do not give an explicit form for the functional $H(X)$.

To reach our aim we follow the steps listed below and we show that:

I- the sequence $\{F(X_k)\}$ is nonincreasing, namely satisfies the *Sufficient decrease condition*

$$F(X_k) + C_1\|X_k - X_{k-1}\|_F^2 \leq F(X_{k-1}), \quad C_1 > 0, \tag{32}$$

and the sequence $\{X_k\}$ has at least an accumulation point (Proposition 4.1);

II- the sequence $\{X_k\}$ satisfies the *Relative Error Condition*

$$\|W_{k+1}\|_F \leq C_2\|X_{k+1} - X_k\|_F, \tag{33}$$

where $W_{k+1} \in \partial F(X_{k+1})$ and $C_2 > 0$ (Proposition 4.2).

III- the sequence $\{X_k\}$ satisfies the *Continuity Condition*, (Lemma 4.2), namely, there exists a subsequence (X_{k_q}) and X^* such that

$$X_{k_q} \to X^* \text{ and } F(X_{k_q}) \to F(X^*), \quad \text{as } q \to \infty. \tag{34}$$

IV- the functional $F(X)$ possesses the *Kurdyka-Lojasiewicz* property (Lemma 4.3).

By exploiting these four results, we prove that any accumulation point of the sequence $\{X_k\}$ is a stationary point of (6) (Theorem 4.1).

For the proof of the point I, we need the following Lemma, which is the natural extension to functionals $H : \mathcal{M}_r \to \mathbb{R}$ of the descent lemma [35], and whose proof follows immediately from Assumption A1.

Lemma 4.1

Let $H : \mathcal{M}_r \to \mathbb{R}$ be a continuously differentiable function with L-Lipschitz continuous gradient.

Then for any $X, Y \in \mathcal{M}_r$

$$H(X) \leq H(Y) + < X - Y, \nabla H(Y) > + \frac{L}{2}\|X - Y\|_F^2. \qquad (35)$$

We can now state the following Proposition, whose detailed proof is given in Appendix A.

Proposition 4.1

Let $\{X_k\}$ be the sequence generated by the FBMC algorithm, and let $\mu > L$. By Assumption A3, the sequence $\{F(X_k)\}$ is nonincreasing, specifically

$$F(X_k) + C_1\|X_k - X_{k-1}\|_F^2 \leq F(X_{k-1}), \qquad C_1 = \frac{\mu - L}{2} > 0 \qquad (36)$$

and

$$\sum_{k=0}^{\infty} \|X_k - X_{k-1}\|_F^2 < \infty, \qquad (37)$$

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

Moreover, the sequence $\{X_k\}$ is bounded and has at least one accumulation point.

Before to starting with the proof of point II we recall some definitions that we use in the following.

Definition 4.1

Let $\psi(x) : \mathbb{R}^n \to]-\infty, +\infty]$ be a proper lower semicontinuous function.

1. Given $x \in$ domψ, the Fréchet subdifferential of ψ at x, written as $\hat{\partial}\psi(x)$ is defined as:

$$\hat{\partial}\psi(x) := \left\{ v \in \mathbb{R}^n : \liminf_{\substack{y \neq x \\ y \to x}} \frac{1}{\|x - y\|} [\psi(y) - \psi(x) - \langle v, y - x \rangle] \geq 0, \forall y \in \mathbb{R}^n \right\}$$

2. The limiting-subdifferential, or simply the subdifferential of $\psi(x)$ at $x \in$ domψ, is defined as:

$$\partial\psi(x) := \left\{ v \in \mathbb{R}^n : \exists x^k \to x, \psi(x^k) \to \psi(x) \text{ and } v^k \in \hat{\partial}\psi(x^k) \to v \text{ as } k \to \infty \right\}$$

   and each $v \in \partial\psi(x)$ is called a subgradient of ψ. Moreover if ψ is continuous differentiable, $\partial\psi(x) = \{\nabla\psi(x)\}$. A critical point or stationary point of ψ is a point $x_0 \in$ domψ, satisfying $0 \in \partial\psi(x_0)$. The set of critical points of ψ is denoted by crit ψ.

To give the definition of the subdifferential of $R(X)$ in the matrix setting, we formulate in the present context Corollary 2.5 given in [45]:

Lemma 4.2

Let $R(X)$ be defined as $R(X) = \sum_i \phi(\sigma_i(X))$, where $\sigma(X) : \mathcal{M}_r \to R^+$ has components $\sigma_1(X) \geq \sigma_2(X) \geq ...\sigma_q(X)$, $q = \min(m, n)$. By Corollary 2.5 in [45], it holds

$$\partial R(X) = \left\{UDiag(\mu)V^T : \mu \in \partial\phi(\sigma(X)), X = UDiag(\sigma(X))V^T\right\} \tag{38}$$

We now recall the definition of the subdifferential of the sum of two functionals, one of which is continuous and derivable and the other is nonsmooth [46].

Definition 4.2

Let $F(X) : \mathcal{M}_r \to R$ be defined as $F(X) = H(X) + \lambda R(X)$, where $H(X)$ satisfies the assumption A1 and $R(X) = \sum_i \phi(\sigma_i(X)$ where $\phi(x)$ is a nonsmooth, nonconvex function, then

$$\partial F(X) = \nabla H(X) + \lambda\partial R(X). \tag{39}$$

We are now in position to demonstrate the *Relative Condition Error*:

Proposition 4.2

Let $\{X_k\}$ be the sequence generated by FBMC algorithm, and let $\mu > L$. The sequence $\{X_k\}$ satisfies the following *Relative Error Condition*:

$$\|W_k\|_F \leq C_2\|X_k - X_{k-1}\|_F, \tag{40}$$

where $W_k \in \partial F(X_k)$ and $C_2 > 0$.

Proof: Since X_k is a global minimum to problem (7), where $Y = X_{k-1}$, writing down the optimality condition yields

$$\nabla H(X_{k-1}) + \mu(X_k - X_{k-1}) + V_k = 0 \tag{41}$$

where

$$V_k \in \partial(R(X_k)). \tag{42}$$

It follows that:

$$\|V_k + \nabla H(X_{k-1})\|_F = \mu\|X_k - X_{k-1}\|_F \leq \bar{\mu}\|X_k - X_{k-1}\|_F \tag{43}$$

where $\bar{\mu} \geq \mu$.

We define, now,

$$W_k = V_k + \nabla H(X_k).$$

From Corollary 4.2 and relation (42), it follows that

$$W_k \in \partial F(X_k).$$

Moreover, making use of relation (43) and of the triangular inequality, we have that

$$\begin{aligned} \|W_k\|_F &\leq \|V_k + \nabla H(X_{k-1})\|_F + \|\nabla H(X_k) - \nabla H(X_{k-1})\|_F \\ &\leq (\bar{\mu} + L)\|X_k - X_{k-1}\|_F = C_2\|X_k - X_{k-1}\|_F \end{aligned}$$

where $C_2 = (\bar{\mu} + L)$. □

In order to prove point III, we first define the limit point set of the sequence $\{X_k\}$ generated by FBMC algorithm from a starting point X_0, $\omega(X_0)$, as follows:

$$\omega(X_0) = \left\{X^* \in \mathcal{M}_r : \exists \quad \text{an increasing sequence of integer } \{k_l\}_{l \in \mathbb{N}}, \text{such that} X_{k_l} \to X^* \text{ as } l \to \infty\right\}.$$

We can now state the following lemma proving that the sequence $\{X_k\}$ satisfies the Continuity Condition (34) and that all points which belong to $\omega(X_0)$ are critical points of F.

Lemma 4.3

Let $\{X_k\}$ be the sequence generated by FBMC algorithm, and let $\mu > L$. Let $\omega(X_0)$ denote the set of its limit points. Considering that X_k is the global minimum of the corresponding minimization problem (9), the following assertions hold:

1. There exists a subsequence (X_{k_q}) and X^* such that

$$X_{k_j} \to X^* \quad \text{and} \quad F(X_{k_j}) \to F(X^*), \quad \text{as} \quad j \to \infty. \tag{44}$$

2. All points which belong to $\omega(X_0)$ are critical points of $F(X)$, namely

$$\emptyset \neq \omega(X_0) \subset \text{ crit}F. \tag{45}$$

Please refer to Appendix B for the detailed proof.

The proof of point IV follows immediately from Assumption A3, that allows us to use the following characterization of the *Kurdyka-Lojasiewicz* property in the matrix setting ([47],Theorem 3.8):

Lemma 4.4

Let $D \subseteq R^{m\times n}$ be open, $\mathcal{M}_r \subset D$ and $F : D \to R$ be real-analytic. Then function F is said to have the *Kurdyka-Lojasiewicz property* (and is called *KL function*) at each point $X^* \in \text{dom}\partial F = \{X \in \mathcal{M}_r : \partial F(X) < +\infty\}$, namely, there exist $\eta \in (0, +\infty]$, a neighborhood U of X^*, and a continuous concave function $\varphi : [0, \eta) \to \mathbb{R}_+$, called *desingularizing function* such that

(i) $\varphi(0) = 0$,

(ii) φ is C^1 on $(0, \eta)$,

(iii) for all $s \in (0, \eta)$,$\varphi'(s) > 0$

(iv) for all $X \in U \cap \{X \in \mathcal{M}_r : F(X^*) < F(X) < F(X^*) + \eta\}$, the *Kurdyka-Lojasiewicz property* holds:

$$\varphi'(F(X) - F(X^*)) \text{ dist}(0, \partial F(X)) \geq 1. \tag{46}$$

where dist(0,$\partial F(X)) = \inf\{\|V\|_F, V \in \partial F(X)\}$.

If $\varphi(s) = s^{1-\theta}, \theta = [1/2, 1)$, equation (46) is equivalent to Lojasiewicz inequality

$$|F(X) - F(X^*)|^\theta \leq \Lambda\|T\|_F \tag{47}$$

where $\Lambda > 0$ and $T \in \partial F(X)$.

Simply speaking, a function that possesses the KL-property in a critical point X^* admits locally a sharp reparametrization. Moreover, as we will show in the following Lemma, a sequence (X_k) that starts in the neighborhood of a point X^* (as in (49)) and that does not improve $F(X^*)$ (as in (48)) converges to a critical point near X^*.

Lemma 4.5

Let $F(X) = H(X) + \lambda R(X) : \mathcal{M}_r \to \mathbb{R}$ be a proper lower semicontinuous functional that possesses the KL property at each point $X^* \in \text{dom}\partial(F)$. Let $U \subset \mathcal{M}_r$ be a neighborhood of X^*, $\eta \in (0, +\infty]$ and $\varphi : [0, \eta) \to \mathbb{R}_+$ a continuous concave function that satisfies the properties (i),(ii),(iii),(iv). Let $\rho > 0$ be such that $B(X^*, \rho) = \{X \in \mathcal{M}_r \, : \, \|X - X^*\|_F \leq \rho.\} \subset U$. Let $(X_k)_{k \in \mathbb{N}}$ the sequence generated by FBMC algorithm which satisfies condition (32), (33) and (34) with X_0 as an initial point. Let us assume that

$$F(X^*) < F(X_k) < F(X^*) + \eta, \quad k \geq 0, \tag{48}$$

and

$$M\varphi(F(X_0) - F(X^*)) + \sqrt{\frac{2}{\mu - L}}\sqrt{F(X_0) - F(X^*)} + \|X^* - X_0\|_F < \rho \tag{49}$$

where $M = 2\frac{\bar{\mu}+L}{\mu-L}$, then the sequence $(X_k)_{k \in \mathbb{N}}$ satisfies:

$$\forall k \in \mathbb{N}, X_k \in B(X^*, \rho), \tag{50}$$

$$\sum_{k=0}^{\infty} \|X_{k+1} - X_k\|_F < +\infty \tag{51}$$

$$F(X_k) \to F(X^*), \quad \text{as} \quad k \to \infty. \tag{52}$$

and converges to a critical point of $F(X)$.

Please refer to Appendix C for the detailed proof.

Now we have all the tools to prove our main convergence result, which extends to the matrix setting the convergence property of the FB splitting algorithm proved for the vector setting in the nonconvex case, [38], Theorem 2.9.

Theorem 4.1

Let $F(X) = H(X) + \lambda R(X) : \mathcal{M}_r \to \mathbb{R}$ be a proper lower semicontinuous functional that satisfies Assumption A3 and let $\{X_k\}$ be the sequence generated by FBMC that satisfies (32), (33) and (34). If F possesses the KL-property at the cluster point $X^* = \lim_{j \to \infty} X_{k_j}$ then the sequence $\{X_k\}$ converges to X^* as k goes to ∞ and X^* is a critical point of the minimization problem (15).

Proof: Let X^* be a cluster point of $(X_k)_{k\in\mathbb{N}}$. By (34) there exists a subsequence $X_{k_q} \to X^*$ and $F(X_{k_q}) \to F(X^*)$. By (32) the sequence $F(X_k)$ is nonincreasing, it therefore follows that $F(X_k) \to F(X^*)$ and $F(X_k) \geq F(X^*)$ for all integers k. The functional F possesses the KL-property in X^*, hence, by Lemma (4.4), there exists a neighborhood of X^*, $U \subset \mathcal{M}_r$, a constant $\eta \in (0, +\infty]$ and a continuous concave function $\varphi : [0, \eta) \to \mathbb{R}_+$ such that property (46) holds. By using the continuity property of φ, we obtain the existence of an integer k_0 such that $F(X^*) \leq F(X_k) \leq F(X^*) + \eta$ for all $k \geq k_0$ and

$$M\varphi(F(X_{k_0}) - F(X^*)) + \sqrt{\frac{2}{\mu - L}}\sqrt{F(X_{k_0}) - F(X^*)} + \|X^* - X_{k_0}\|_F < \rho \tag{53}$$

The conclusion follows by applying Lemma (4.5) to the sequence $(Y_k)_{k\in\mathbb{N}}$ defined as $Y_k = X_{k_0+k}$ for all integers k. □

5. THE PROPOSED ALGORITHM: IFBMC

We give in the following some details of the proposed version of the FBMC algorithm. Since in the form presented in Algorithm 2 FBMC is very computational demanding, especially by the increase of the matrix dimensions, we have added some strategies to improve the computational time, and named IFBMC this improved version that is displayed in Table I.

The improvements are:

S0- *Continuation for the optimal value of* λ*.* FBMC solves the minimization problem (15). However, since the original problem was the constrained minimization problem

$$\min_{X\in\mathcal{M}_r} \sum_{i=1}^{r} \phi(\sigma_i(X), a) \qquad \text{such that} \qquad \mathcal{P}_{\mathbf{\Omega}}(X) = \mathcal{P}_{\mathbf{\Omega}}(M),$$

the parameter λ must be carefully chosen. To this aim we start with a reasonable value for λ_0 and we decrease this value by a fixed amount by adding in the algorithm an outer iteration loop. This strategy, often successfully used under the name of *continuation* [22, 23], does not increase much the computing time of IFBMC since the precision of the solution of the inner loop depends on the value of λ_i, namely $prec = \gamma \cdot \lambda_i$, and each outer iteration uses as starting point the previous iterate (*warm starting*). For the first value of λ_k the precision is not high, then it decreases according to the decrease of λ_k. Moreover, it is experimentally shown that changing the value of λ_i helps the algorithm to avoid getting stuck in unwanted local minima.

S1- *Rank estimation phase.* Since the main computational cost of the algorithm lies in computing at each inner iteration the r singular values of the update approximation Z_i, with $r \leq \min(m, n)$, to maximally reduce the computing time, we have divided the algorithm in two different phases: the first devoted to obtain an approximate estimation of the true matrix rank using an incremental strategy, and the second that uses and improves this estimate using the doubly iterative procedure described above. In the first phase, we fix the value of γ, $\lambda = \lambda_0$ and $r = r_e$, with r_e given by a small number, and we run FBMC algorithm until convergence. The output value r_c of the matrix rank, calculated by the algorithm after thresholding operations, is less or equal r_e. If $r_c = r_e$, we increase r_e, namely we set $r_e = r_c + q$, where q is given by a fixed percentage of r_c and we run again FBMC by repeating this procedure until $r_c < r_e$. At this point we accept r_c as our initial ”good” rank estimation and we start with the doubly iterative algorithm: if r_c is not the exact rank, r_c evolves throughout the IFBMC iterations and it converges to the exact rank of the matrix to be recover.

S2- *Improvements of the speed-up of SVD.* In order to speed-up the computing time of the SVD, we use the randomized truncated SVD (RSVD [24]), and since it is based on the power method we use the eigenvectors of the previous iteration for its initialization. This initial approximation is revealed to be very efficient and, with this choice, the power method usually converges in only one iteration.

S3- *Acceleration strategy.* Another technique to accelerate the doubly iterative algorithm consists in the use of the FISTA approach [35]. Specifically, FISTA defines a sequence of positive numbers $\{\alpha_k\}$, (see Table I for their definition), and, at each iteration step, improves the estimate yielded by FBMC performing a specific linear combination with the unmodified previous iterate. This preserves the computational simplicity of our scheme, but significantly improves its convergence speed. Since in general FISTA does not maintain the descent

Table I: **Improved Algorithm IFBMC**

```
Inputs: X_{0,-1} = P_Ω(M), X_{0,0} = 0, λ_0 > 0, τ < 1 μ > L, γ > 0
```

FIRST PHASE: RANK ESTIMATION PHASE

$(r_c, \hat{X})\leftarrow$ rank estimated *as suggested in S1* and corresponding recovery matrix $\hat{X}$

$X_{-1,0} = \hat{X}$

SECOND PHASE: RECONSTRUCTION PHASE

I. Outer continuation Loop

for i=0,1,... do

$k = 0$

$t_k = 1$

II. Inner Forward-Backward Loop

1. Determination of the step $\frac{1}{\bar{\mu}_k}$ *as suggested in S4.*

2. Forward Updating Solution

$Z_{k,i+1} = X_{k-1,i} - \frac{1}{\bar{\mu}_k}(\mathcal{P}_{\mathbf{\Omega}}(X_{k-1,i}) - \mathcal{P}_{\mathbf{\Omega}}(M))$

3. Backward Minimization Step

$\tilde{X}_{k,i+1} = \Theta_{\frac{\lambda_i}{\bar{\mu}_k}}(Z_{k,i+1})$ where Θ is defined as in (30) and $a \leq \frac{\bar{\mu}_i}{\lambda_i}$

if *the descent property is satisfied*

4. FISTA of the solution as suggested in S3.

$t_{k+1} = \frac{1+\sqrt{1+4tk^2}}{2}, \alpha_k = \frac{t_k - 1}{t_{k+1}}$

$X_{k,i+1} = \tilde{X}_{k,i+1} + \alpha_k(\tilde{X}_{k,i+1} - \tilde{X}_{k,i})$

, *else*

$X_{k,i+1} = \tilde{X}_{k,i+1}$

end

if $\frac{|F(\lambda_i, X_{k,i+1}) - F(\lambda_i, X_{k,i})|}{|F(\lambda_i, X_{k,i+1})|} > \gamma \cdot \lambda_i, \quad k = k+1$

goto 1

else

$\lambda_{i+1} = \tau \cdot \lambda_i$ *Updating of the penalization parameter*

$X_{0,i+1} = X_{k,i+1}$ *Warm Starting*

until convergence

Output: $\hat{X} = X_{0,i+1}$ low-rank solution

property of the functional $F(X)$, we insert a test to accept the FISTA accelerate value only if the value of the functional F(X) still decreases.

S4- *Choice of the step-size.* Our experimentation has highlighted that the condition $\mu > L$, that is necessary for the convergence of the algorithm in the nonconvex case, is usually too conservative. It was therefore possible to further speed-up IFBMC in the Easy problems by incorporating at each iteration of IFBMC a linesearch-like acceleration strategy for the step-size $\frac{1}{\mu}$ starting from $\mu > L/2$, as proposed in [20]. On the contrary, in the Hard cases, we use the line-search strategy starting from $\mu > L$ in order to achieve convergence.

5.1. *Choice of the parameter values*

To run IFBMC is necessary to initialize the value of the penalization parameter λ_0, its reduction rate τ, and the constant γ of the inner loop exit test. Guided by a large number of numerical experiments, we made the following choices for the above parameters. We have chosen λ_0 as a suitable fraction of the greatest singular value of $\mathcal{P}_\Omega(M)$, i.e. $\lambda_0 = cost_\lambda \cdot \|\sigma(\mathcal{P}_\Omega(M))\|_\infty$, where the value of $cost_\lambda$ ranges in $[0.001, 0.3]$ according to the difficulty of the problem ($cost_\lambda < 0.2$ for Hard problems and $cost_\lambda > 0.2$ for Easy problems). Concerning the reduction strategy, we have used $\lambda_{i+1} = \tau\lambda_i$, with $\tau = 0.8$ in all the considered experiments. The parameter γ is very important since it influences the precision requested in the inner loop which is $\gamma \cdot \lambda_i$. In fact we have experimentally observed that, when solving Easy and not too Hard problems, at the algorithm beginning, i. e. for large values of λ_i, is not necessary to solve the inner problems with high accuracy, while accuracy must increase by decreasing λ_i. On the contrary, in the very Hard cases it is necessary to solve inner problems with high accuracy already from beginning, and this can be obtained by tuning the value of γ suitably. In our experimentation we have chosen γ ranging form from 10^{-2} to 10^{-5}, according to the difficulty of the minimization problem.

6. EXPERIMENTS

In this Section, to validate the effectiveness of the proposed algorithm, we show the results of several experiments, both on synthetic and real data, and we compare the recovery performance of IFBMC with those of the best state-of-art algorithms. Specifically, we consider FPCA [19] (available from http://www1.se.cuhk.edu.hk/˜sqma/FPCA.html), APGL [20] (available from http://www.math.nus.edu.sg/˜mattohkc/NNLS.html), sIRLS-0 [30] (available from https://faculty.washington.edu/mfazel/https://faculty.washington.edu/mfazel/IRLS_final.zip), IPMSr [31] (that we have implemented according to the specifications of the authors), t-IRucLq-M [28] (available from http://www.caam.rice.edu/˜optimization/codes/), IRNN [29] (available from https://sites.google.com/site/canyilu/). We also point out that in the considered packages some acceleration strategies are implemented. In particular, APGL uses continuation, line-search, Lanczos method for computation of singular values, and FISTA acceleration. FPCA uses continuation, matlab SVD or Linear Time SVD in the case of hard and easy problems respectively, sIRLS-0 uses randomized SVD, iPSMr uses randomized SVD and continuation, IRNN uses continuation and economized version of matlab SVD.

All the numerical computations are conducted on an Intel i7-3770 CPU with 16 GB of RAM. The supporting software is MATLAB R2012b.

6.1. Experiments on synthetic data

In this subsection we experiment IFBMC for the completion of randomly generated square matrices of size n and rank r. In all experiments the matrix $M \in \mathcal{M}_r$ is generated as a product $M_L M_R^T$, where both matrices $M_L \in \mathbb{R}^{n\times r}$ and $M_R \in \mathbb{R}^{n\times r}$ have i.i.d. Gaussian entries. The index set Ω of the given p entries is sampled uniformly at random. We evaluate the recovery performance by the Relative Error, defined as

$$Relative\,Error = ||\widehat{X} - M||_F/||M||_F,$$

where $\widehat{X}$ is the recovered matrix. For each experiment we run several trials and we use the Probability of Perfect Reconstruction (*PPR*) as performance index, defined as the ratio of the number of the successful trials over the total number of trials. A trial is considered successful when its Relative Error is smaller than 10^{-3}. This value plus a maximum iteration number, *maxit=3000* represents an exit test for the algorithm. The sampling ratio, namely the percentage of known data, is measured by $sr = p/n^2$, and, according to the current literature, to classify the numerical experiments into two categories, we use the *degree of freedom ratio*, that, for an $m \times n$ matrix, is defined as $fr = \frac{r\cdot(m+n-r)}{p}$. If fr is small (≤ 0.4) recovering M it is not difficult and the problem is considered Easy. If fr is large (> 0.4), recovering M becomes harder and the problem is classified as Hard.

We recall that for given *m,n* matrix dimensions and p sampled entries, the largest rank for which the matrix completion problem has an unique solution is the rank $r_{\max}$ for which $fr = 1$, namely $r_{\max} = \lfloor (m + n - \sqrt{(m+n)^2 - 4p})/2 \rfloor$. We therefore consider $r_{\max}$ as the maximum possible

Table II: Execution times of FPCA, APGL, IFBMC, IPMSr, sIRLS-0 for Easy Problems: $FR < 0.4$

Problem				FPCA	APGL	IFBMC	IPMSr	sIRLS-0
m=n	r	sr	FR	Time	Time	Time	Time	Time
500	10	0.2	0.2	1.22	0.98	**0.47**	1.27	3.80
500	10	0.12	0.33	1.91	**1.12**	1.19	3.75	6.86
1000	10	0.12	0.17	4.52	**1.47**	2.24	5.35	16.59
1000	20	0.12	0.33	7.14	**3.14**	4.05	10.97	32.85
2000	20	0.12	0.17	20.56	**5.84**	8.26	20.75	71.10
2000	40	0.12	0.33	32.34	**11.73**	15.61	46.36	156.32
5000	10	0.15	0.026	114.92	**19.10**	26.74	45.02	921.93
5000	50	0.12	0.17	275.22	**41.88**	52.52	162.1	2220.08
10000	10	0.15	0.013	737.49	**78.68**	87.3612	159.66	2915.92
10000	50	0.12	0.083	817.84	**131.02**	185.0632	470.39	5433.05

rank, once fixed m, n, p, for the matrix completion problem.

Experiment 1: Comparisons on Easy problems

The aim of this experiment is to show the performance of IFBMC on Easy problems, where the use of nonconvex penalties would be not strictly necessary. We make the same tests as in [31], by repeating the experiments 10 times. In Table II we compare the execution times of IFBMC with those of some of the best convex and nonconvex reconstruction algorithms of the recent literature. More precisely, we consider the convex APGL, FPCA and the nonconvex sIRLS-0, IPMSr methods. The results highlight that, while all algorithms solve successfully the matrix completion problem in all cases, the computing time of IFBMC is much lower than that used by nonconvex algorithms and comparable with that of the best state-of-art convex algorithms for Easy problems.

Experiment 2: Comparisons on Hard problems

To show the performance of IFBMC on Hard problems, we consider the same tests again as in [31], where the authors show that IPSMr in Hard cases is more efficient than both sIRLS-0 and FPCA. Furthermore, because it is known from the literature that the nuclear norm algorithm APGL fails in the case of hard problems, we compare IFBMC only with IPSMr. In this experiment we do not consider t-IrucLq-M and IRNN because these algorithms for n large are not very efficient. In fact, their authors in [28] and [29], respectively, present experiments only for small problems. Table III shows that, for Hard problems, IFBMC is approximately 5 times faster than IPSMr.

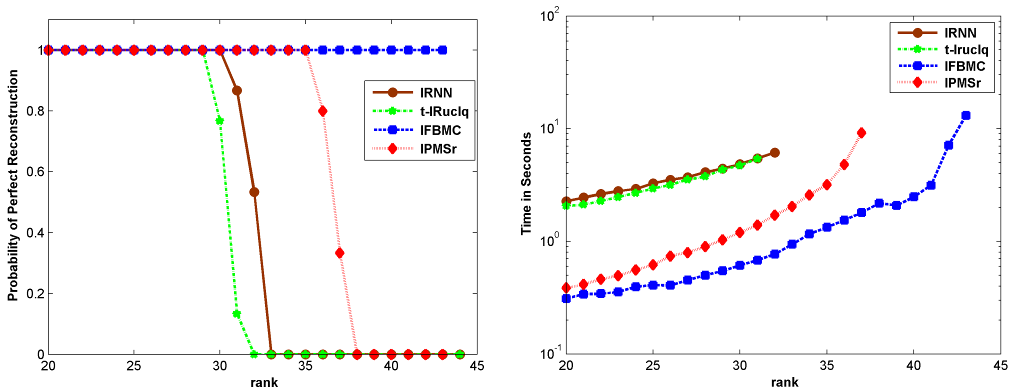
Table III: Execution times of IFBMC and IPMSr for Hard Problems: $FR \geq 0.4$

Problem				IPMSr	IFBMC
m=n	r	sr	fr	Time	Time
40	9	0.50	0.80	0.66	**0.18**
500	20	0.15	0.52	7.09	**2.18**
500	20	0.10	0.78	66.86	**6.81**
1000	20	0.10	0.40	18.83	**5.95**
1000	20	0.06	0.66	141.19	**16.23**
1000	30	0.10	0.59	47.71	**10.07**
2000	40	0.10	0.40	65.39	**22.97**
2000	40	0.06	0.66	386.06	**59.77**
10000	300	0.10	0.59	9629.51	**1430.00**

Experiment 3: Scalability respect to r

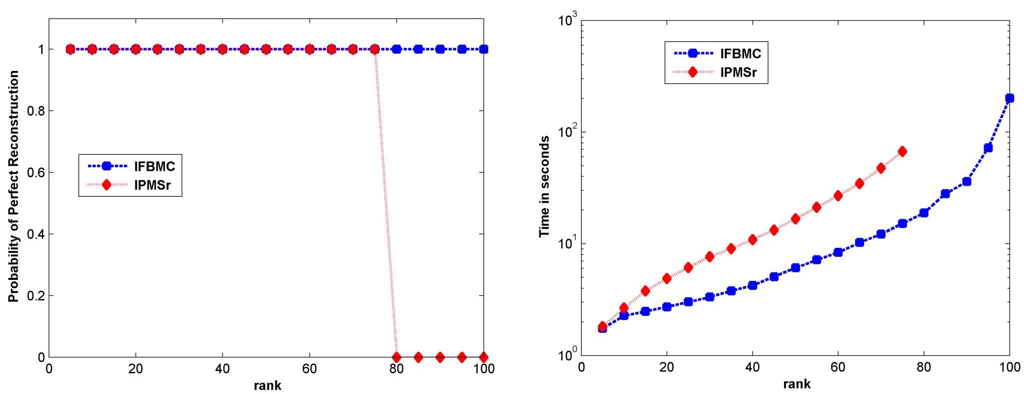
In this experiment we fix the matrix order n, the sampling ratio of observed entries sr, and we test the performance of IFBMC by varying the rank r. We consider two problems. In the first, similarly to the experiments presented in [29], we set $n = 150$, $sr = 0.5$; since in this case the largest rank for which the matrix completion problem has an unique solution, corresponding to $fr = 1$, is $r_{\max} = 43$, we test the performance of the reconstruction algorithms IFBMC, IRNN, t-IRlucq-M and IPMSr by repeating the experiment 30 times for rank r varying from 20 to 43, ($fr = 0.9823$). The results, depicted in Fig. 1, show that IFBMC outperforms in all the experiments IRNN, t-IRlucq-M and IPMSr both in terms of Probability of Perfect Reconstruction and execution time. We observe that IRNN achieves $PPR = 1$ up to matrices of rank $r = 30$, t-IRlucq-M up to matrices of rank $r = 29$ and IPMSr up to matrices of rank $r = 35$, while IFBMC obtains $PPR = 1$ up to matrices of rank $r = 43$. Moreover the running times of IRNN and t-IRlucq-M are approximately 9 higher times than those of IFBMC, while the running times of IPMSr are approximately 5 higher times than those of IFBMC, especially in the Hard problems.

In the second problem we set $n = 1000$, $sr = 0.2$, and we repeat the experiment 10 times with the rank r varying from 5 to 100, largest rank for which the matrix completion problem has an unique solution, and we compare the results of IFBMC with those of IPSM-r. As shown in Fig. 2a, IFBMC reaches for all rank values $PPR = 1$, while IPSMr obtains $PPR = 1$ only for $5 \leq r \leq 75$. Fig. 2b



(a) (b)

Figure 1. a) Experiment 3: Case 1: m=n=150, sr=0.5, $r \in [20, 44]$. Comparison between reconstruction results obtained by IFBMC and those obtained by t-IRuclq-M and IRNN-log. (a)Probability of Perfect Reconstruction as a function of the rank r. (b)The average perfect reconstruction time in seconds.



(a) (b)

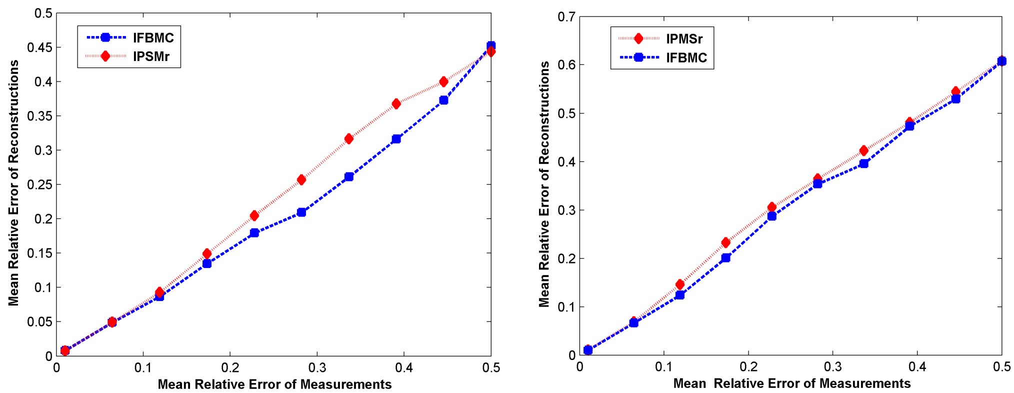
Figure 2. a) Experiment 3: Case 2: m=n=1000, sr=0.2, $r \in [5, 100]$. Comparison between reconstruction results obtained by IFBMC and those obtained by IPMSr. (a)Probability of Perfect Reconstruction as a function of the rank r. (b)The average perfect reconstruction time in seconds.

highlights the major efficiency in terms of computational cost of IFBMC respect to IPSMr especially for $r \geq 15$.

Experiment 4: Robustness to noise

This experiment demonstrates the robustness of the proposed algorithm to different noise levels. We consider the following noisy matrix completion problem

$$\min_{X \in \mathcal{M}_r} \left\{ \frac{1}{2} ||\mathcal{P}_{\mathbf{\Omega}}(X) - \mathcal{P}_{\mathbf{\Omega}}(B)||_F^2 + \lambda \sum_i \phi(\sigma_i(X)) \right\}. \tag{54}$$



(a) (b)

Figure 3. Experiment 4: (a) m=n=500 sr=0.12, $r = 10$, $fr = 0.33$; (b) m=n=500, sr=0.15, $r = 20$, $fr = 0.52$. Measurements corrupted by withe noise with standard deviation $\sigma \approx \epsilon\sqrt{r}$, $\epsilon \in [10^{-2}, 0.5]$. Mean Relative Error of recovery as a function of Mean Relative Error of measurements.

where $\mathcal{P}_{\mathbf{\Omega}}(B) = \mathcal{P}_{\mathbf{\Omega}}(M) + \mathcal{P}_{\mathbf{\Omega}}(Z)$ and $\mathcal{P}_{\mathbf{\Omega}}(Z)$ is the noise on the measurements, with $Z_{i,j}$ i.i.d Gaussian random variables with normal distribution with 0-mean and variance σ^2. As in [15], we consider a perturbation such that $\|\mathcal{P}_{\mathbf{\Omega}}(Z)\|_F \leq \epsilon \|\mathcal{P}_{\mathbf{\Omega}}(M)\|_F$. This is true if $\sigma \approx \epsilon\sqrt{r}$. We consider two cases: an Easy problem with $N = 500$, $r = 10$ and $sr = 0.12$, corresponding to an $fr = 0.33$ and an Hard problem with $N = 500$, $r = 20$ and $sr = 0.15$, corresponding to an $fr = 0.52$. In both problems we choose ϵ ranging in interval $[0.01, 0.5]$. We repeat the two sets of experiments 5 times. We stop the algorithm when $\|\mathcal{P}_{\mathbf{\Omega}}(\widehat{X}) - \mathcal{P}_{\mathbf{\Omega}}(B)\|_F / \|\mathcal{P}_{\mathbf{\Omega}}(B)\|_F \leq \epsilon \cdot \eta$, with $\eta = 0.8$. In Fig. 3a and Fig. 3b we display the behavior of the mean Relative Error of the reconstructions obtained both by IPSMr and IFBMC, as a function of the mean Relative Error of the measurements, for Easy and Hard problems, respectively. An analysis of the results highlights that IFBMC outperforms IPSMr in terms of stability in particular in Easy cases.

6.2. Experiments on real data

In this section, we consider matrix completion problems based on the real MovieLens data set [48]. The dataset is collected by the GroupLens Research Project at the University of Minnesota and contains 100,000 rating information from 943 users on 1,682 movies. The data has been cleaned up such that users who had less than 20 ratings were removed. So each user in the data has rated at least 20 movies. The ratings are from 1 (strongly unsatisfactory) to 5 (strongly satisfactory). In the

Table IV: Comparison of Normalized Mean Absolute Error NMAE of s-IRLS0, IPSMr and IFBMC for different splits of the 100k movie-lens dataset

Method	Split 1	Split 2	Split 3	Split 4
s-IRLS0	0.1919	0.1878	0.1870	0.1899
IPSMr	0.1904	0.1873	0.1863	0.1863
IFBMC	0.1814	0.1795	0.1786	0.1792

recommendation situation, the data matrix is highly sparse (only about 6.3% entries are known). In order to test the proposed algorithm, we split the ratings into training and test sets. In our experiment, 80,000 ratings (80%) are randomly chosen to be the training set and the test set contains the remaining 20,000 ratings (20%) In particular, for our numerical experiments, we consider as in [31] four different splits of the 100k ratings into (training set, test set): (u1.base,u1.test), (u2.base,u2.test), (u3.base,u3.test), (u4.base,u4.test) for our numerical experiments. Any given set of ratings (e.g., from a data split) can be represented as a matrix. This matrix has rows representing the users and columns representing the movies and an entry (i, j) of the matrix is non-zero if we know the rating of user i for movie j. Thus estimating the remaining ratings in the matrix corresponds to a matrix completion problem. For each data split, we train our algorithms on the training sets, 80,0000 ratings (namely, u1.base,u2.base,u3.base,u4.base) to estimate the 20,000 ratings in the test sets (namely, u1.test, u2.test, u3.test, u4.test). The performance metric here is Normalized Mean Absolute Error or NMAE defined as follows. Let M be the matrix representation corresponding to the actual test ratings and X be the ratings matrix output by an algorithm when input the training set. Then

$$NMAE = \frac{1}{(rt_{max} - rt_{min})|\Omega|} \left(\sum_{i,j \in |\Omega|} |M_{ij} - X_{i,j}| \right)$$

where rt_{min} and rt_{max} are the minimum and maximum possible value of movie ratings. For the MovieLens data set, we have that $rt_{min} = 1$ and $rt_{max} = 5$

Table IV shows that the IFBMC has a better NMAE than IPSMr and s-IRLS0 across different splits of the data.

6.3. Application to image recovery

In this subsection we evaluate the effectiveness of IFBMC for low-rank image recovery and we compare our results with those of IPSMr. For this aim we consider the 512×512 Boat image, full of fine details and edges (Fig 4.a). Because the real images are not low-rank, but the top singular values maintain the main information, we apply SVD to the original image and truncate this decomposition to get the 50-rank image M_{50} (Fig. 4.b). We selected 30 % samples randomly from M_{50} (Fig. 4.c) and we solve the recovery problem with IFBMC and IPSMr. To measure the quality of the reconstruction in this case, in addition to the *Relative Error*, that uses M_{50}, we use the *Peak Signal to-Noise Ratio*, defined as:

$$PSNR = 20 \log_{10} \frac{\max(M)}{rmse} \qquad rmse = \sqrt{\frac{\sum_i \sum_j (\widehat{X}_{ij} - M_{ij})^2}{m \cdot n}}$$

where M is the original full rank image.

The reconstructions obtained by both the algorithms, depicted in Fig. 4d, have the same $Relative\,Error = 9.90E - 011$ and $PSNR = 50.19$, but the execution time of IFBMC (25.79 sec) is lower than that of IPSMr (136.41 sec).

7. CONCLUSIONS

We considered a nonconvex approach to the matrix completion problem. We showed that also in this nonconvex, nonsmooth matrix setting it is possible, under certain hypotheses, to use the iterative Forward-Backward splitting algorithm and we proved its convergence to a critical point of the nonconvex objective functional. Since the most difficult request for the convergence proof is that in the backward step the solution of the nonconvex minimization is a global minimizer, we showed that by using certain parameter dependent nonconvex, nonsmooth relaxations of the rank function and by a suitable choice of the parameter value it is possible to solve in the backward step a convex minimization problem that has a closed form solution. We used our theoretical results to solve the matrix completion problem and we developed the efficient, iterative algorithm IFBMC, that presents better recovery performances and lower computing times when compared, on simulated and real recovery problems, with the best state-of-the-art algorithm for matrix completion. The performance of the proposed IFBMC algorithm is illustrated on problem of real and synthetic low-rank matrix completion and image recovery.

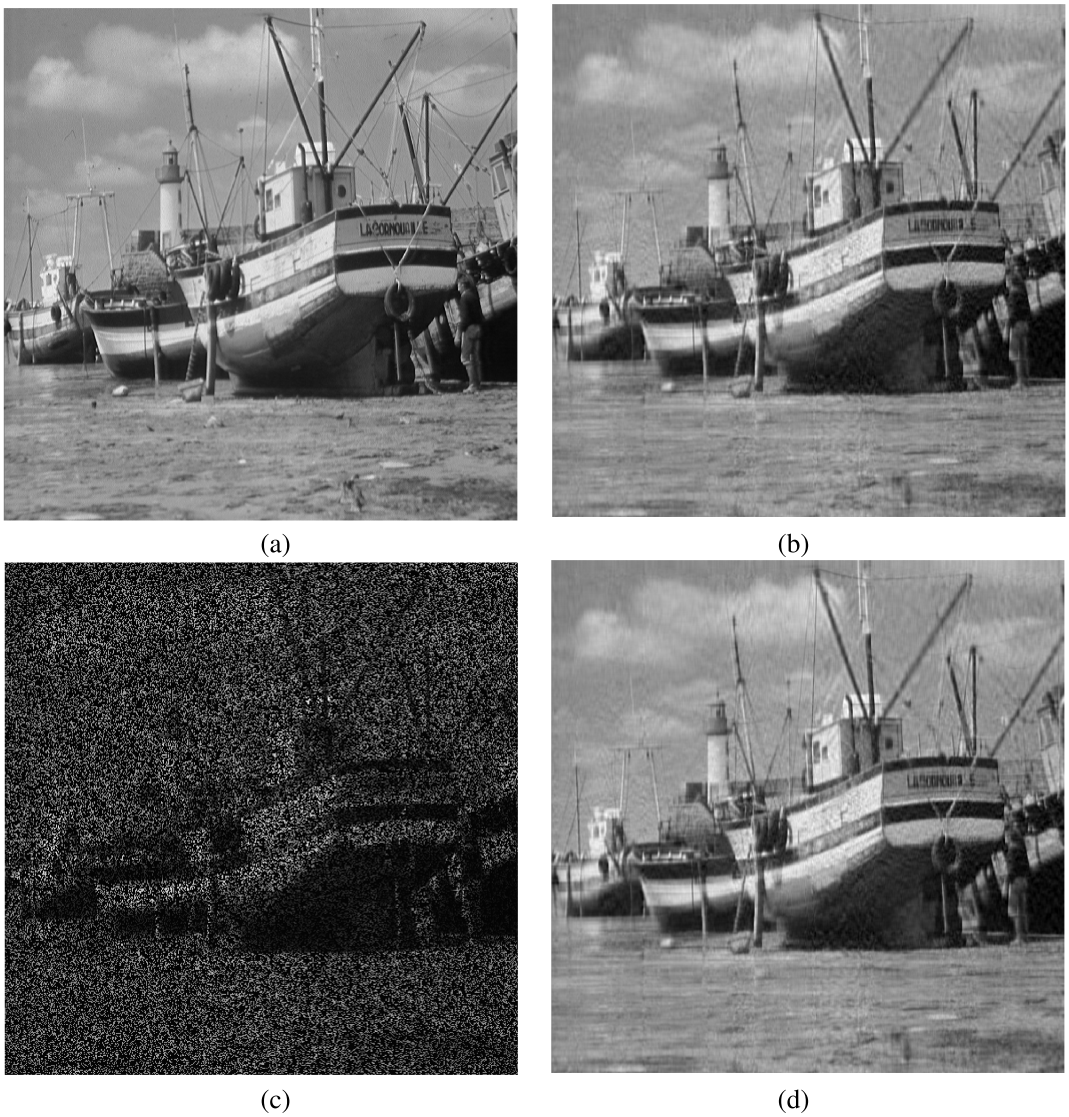


Figure 4. a) Original 512× 512 Boat image with full rank. b) 50-rank approximation: M_{50}. c) Starting data for the reconstruction: only 30% of M_{50}. d) Recovered Images by both IFBMC and IPSMr (the two reconstructions are almost identical)

A. PROOF PROPOSITION 4.1

Proof: Since X_k is a global minimum to problem (9), it holds that

$$Q(X_k, X_{k-1}) \leq Q(X_{k-1}, X_{k-1}),$$

namely

$$\lambda R(X_k) + H(X_{k-1}) + \langle \nabla H(X_{k-1}), X_k - X_{k-1} \rangle + \frac{\mu}{2}\|X_k - X_{k-1}\|_F^2 \leq$$
$$\lambda R(X_{k-1}) + H(X_{k-1}).$$

then,

$$\lambda(R(X_{k-1}) - R(X_k)) \geq \langle \nabla H(X_{k-1}), X_k - X_{k-1} \rangle + \frac{\mu}{2}\|X_k - X_{k-1}\|_F^2. \tag{55}$$

Using Lemma 4.1 with X=X$_k$ and $Y = X_{k-1}$, we obtain:

$$H(X_{k-1}) - H(X_k) \geq -\langle \nabla H(X_{k-1}), X_k - X_{k-1} \rangle - \frac{L}{2}\|X_k - X_{k-1}\|_F^2. \tag{56}$$

Since $F(X_{k-1}) = H(X_{k-1}) + \lambda R(X_{k-1})$ and $F(X_k) = H(X_k) + \lambda R(X_k)$, by using (55) and (56) it follows that

$$\begin{aligned} F(X_{k-1}) - F(X_k) = H(X_{k-1}) - H(X_k) + \lambda(R(X_{k-1})) - \lambda R(X_k)) \geq \\ \geq \frac{\mu - L}{2}\|X_k - X_{k-1}\|_F^2, \end{aligned} \tag{57}$$

Since $\mu > L$, it follows that $F(X_k)$ is monotonically nonincreasing and (36) holds.

In order to prove relation (37), we rewrite (57) for $k = 1$,2,..., and by back-substitution and summing for $k \geq 1$, we obtain

$$F(X_0) \geq \sum_{k=1}^{+\infty} \|X_k - X_{k-1}\|_F^2 \frac{\mu - L}{2}$$

namely

$$\sum_{k=1}^{+\infty} \|X_k - X_{k-1}\|_F^2 \leq \frac{2}{\mu - L} F(X_0).$$

The series $\sum_{k=1}^{+\infty} \|X_k - X_{k-1}\|_F^2$ is, therefore, finite, it holds then that

$$\lim_{k\to\infty} \|X_k - X_{k-1}\|_F = 0. \tag{58}$$

Since the sequence $\{F(X_k)\}$ is monotonically nonincreasing, the boundness of $\{(X_k)\}$ follows from the coercivity and from boundness from below of F.

By the Bolzano-Weirstrass theorem, there exists a matrix X^* and a subsequence $\{X_{k_j}\}$ such that $\lim_{k\to\infty} X_{k_j} = X^*$. □

B. PROOF LEMMA 4.2

Proof: 1. Since X_k is a global minimum to problem (9), we have

$$F(X_k) + \frac{\mu - L}{2}\|X_k - X_{k-1}\|_F^2 \le F(\Xi) + \frac{\mu - L}{2}\|\Xi - X_{k-1}\|_F^2, \quad \Xi \in \mathcal{M}_r$$

namely

$$H(X_k) + \lambda R(X_k) + \frac{\mu - L}{2}\|X_k - X_{k-1}\|_F^2 \le H(\Xi) + \lambda R(\Xi) + \frac{\mu - L}{2}\|\Xi - X_{k-1}\|_F^2, \quad \Xi \in \mathcal{M}_r \tag{59}$$

Let X^* be a point in $\omega(X_0)$; there exists a subsequence X_{k_q} of X_k that converges to X^* as $q \to \infty$. Because by (58) $\|X_k - X_{k-1}\|_F \to 0$, by Assumption A1 $H(X)$ is continuous, and $R(X)$ is lower semicontinuos functional, from (59) we deduce that:

$$\limsup \lambda R(X_{k_q}) + H(X^*) \le H(\Xi) + \lambda R(\Xi) + \frac{\mu - L}{2}\|\Xi - X^*\|_F^2, \quad \Xi \in \mathcal{M}_r$$

In particular, for $\Xi = X^*$ we obtain

$$\limsup_{q \to \infty} R(X_{k_q}) \le R(X^*)$$

and since $R(X)$ is a lower semicontinuous functional we can assume that $R(X_{k_q}) \to R(X^*)$, as $q \to \infty$.

It follows therefore that

$$\lim_{q\to\infty} F(X_{k_q}) = \lim_{q\to\infty} \left\{H(X_{k_q}) + \lambda R(X_{k_q})\right\} = H(X^*) + \lambda R(X^*) = F(X^*),$$

and thus the Continuity Condition (34) results satisfied.

2. Since by Lemma 4.1 X_k is a bounded sequence , the set $\omega(X_0)$ is nonempty. Moreover, by (40) we have that $\|W_k\|_F \le (\bar{\mu} + L)\|X_k - X_{k-1}\|_F, \;\; W_k \in \partial F(X_k)$. By (58) it follows that $W_k \to 0$ as $k \to \infty$. Therefore $0 \in \partial F(X^*)$. This proves that X^* is a critical point of F.

□

C. PROOF LEMMA 4.4

Proof: Without loss of generality, we assume that $F(X^*) = 0$, (otherwise we can consider $F(X) = F(X) - F(X^*)$).

By Proposition (32) it holds that:

$$F(X_i) - F(X_{i+1}) \geq \frac{\mu - L}{2} \|X_{i+1} - X_i\|_F^2. \tag{60}$$

We observe that $\varphi'(F(X_i))$ makes sense since $0 < F(X_i) < \eta$ for hypothesis (48) (being $F(X^*) = 0$), and $\varphi'(F(X_i)) > 0$. Hence, by multiplying both sides of (60) for $\varphi'(F(X_i))$ we have

$$\varphi'(F(X_i))(F(X_i - F(X_{i+1})) \geq \varphi'(F(X_i))\frac{\mu - L}{2} \|X_{i+1} - X_i\|_F^2 \tag{61}$$

Due to the concavity of φ and by using (61) it follows that

$$\varphi(F(X_i)) - \varphi(F(X_{i+1})) \geq \varphi'(F(X_i))(F(X_i) - F(X_{i+1})) \geq \varphi'(F(X_i))\frac{\mu - L}{2} \|X_{i+1} - X_i\|_F^2. \tag{62}$$

We now prove by induction relation (50). Let us first check (50) for $k = 0$ and $k = 1$.

From hypothesis (49) it follows that $X_0 \in B(X^*, \rho)$.

As to X_1, relation (60) yields, in particular, for $i = 0$,

$$\frac{\mu - L}{2} \|X_1 - X_0\|_F^2 \leq F(X_0) - F(X_1) \leq F(X_0) \tag{63}$$

Hence

$$\|X_1 - X^*\|_F \leq \|X_1 - X_0\|_F + \|X_0 - X^*\|_F \leq \sqrt{\frac{2}{\mu - L}} \sqrt{F(X_0)} + \|X_0 - X^*\|_F \tag{64}$$

Hence, from hypothesis (49) it follows that $X_1 \in B(X^*, \rho)$.

By induction, we now prove that $X_k \in B(X^*, \rho)$ for all $k \geq 0$. This being true for $k \in \{0, 1\}$, let us assume that it holds up some $k \geq 1$.

For $0 \leq i \leq k$, because $X_i \in B(X^*, \rho)$, and $0 < F(X_i) < \eta$, we can write the Kurdyka-Loyasiewicz inequality at X_i

$$\varphi'(F(X_i)) dist(0, \partial F(X_i)) \geq 1 \tag{65}$$

By (33), we have that

$$\|W_i\|_F \leq (\bar{\mu} + L)\|X_i - X_{i-1}\|_F \tag{66}$$

with $W_i \in \partial F(X_i)$. We can then write:

$$\varphi'(F(X_i)) \geq \frac{1}{\|W_i\|_F} \geq \frac{1}{(\bar{\mu} + L)\|X_i - X_{i-1}\|_F} \tag{67}$$

By combining (62) and (67), we have

$$M\left(\varphi(F(X_i)) - \varphi(F(X_{i+1}))\right) \geq \frac{\|X_{i+1} - X_i\|_F^2}{\|X_i - X_{i-1}\|_F} \tag{68}$$

where $M = 2\frac{\bar{\mu}+L}{\mu-L}$.

Inequality (68) can be rewritten as:

$$M(\varphi(F(X_i)) - \varphi(F(X_{i+1})))\|X_i - X_{i-1}\|_F \geq \|X_{i+1} - X_i\|_F^2 \tag{69}$$

Taking the square root of both sides of the relation (69)

$$\sqrt{\|X_i - X_{i-1}\|_F}\,(M(\varphi(F(X_i)) - \varphi(F(X_{i+1}))))^{1/2} \geq \|X_{i+1} - X_i\|_F \tag{70}$$

and by recalling that $\alpha\beta \leq (\alpha^2 + \beta^2)/2$, relation (70) can be rewritten as:

$$\|X_i - X_{i-1}\|_F + M(\varphi(F(X_i)) - \varphi(F(X_{i+1})) \geq 2\|X_{i+1} - X_i\|_F \tag{71}$$

This inequality holds for $1 \leq i \leq k$; let us sum over i:

$$\|X_1 - X_0\|_F + M(\varphi(F(X_1)) - \varphi(F(X_{k+1})) \geq \sum_{i=1}^{k} \|X_{i+1} - X_i\|_F + \|X_{k+1} - X_k\|_F \tag{72}$$

Due to the monotonicity of φ and $F(X_k)$, we can write

$$\|X_1 - X_0\|_F + M\varphi(F(X_0)) \geq \sum_{i=1}^{k} \|X_{i+1} - X_i\|_F \tag{73}$$

By considering that

$$X_{k+1} - X^* = X_2 - X_1 + X_3 - X_2 + \cdots + X_{k+1} - X_k + X_1 - X^*,$$

and by using triangular inequality and relation (73), we get

$$\|X_{k+1} - X^*\|_F \leq \sum_{i=1}^{k} \|X_{i+1} - X_i\|_F + \|X_1 - X^*\|_F \leq M\varphi(F(X_0)) + \|X_1 - X_0\|_F + \|X_1 - X^*\|_F$$

that, by hypothesis (49) and relations (63) and (64), implies $X_{k+1} \in B(X^*, \rho)$. It results, thus, proved (50).

Indeed inequality (71) holds for $i \geq 1$; let us sum it for i running from some k to some $K > k$

$$\|X_k - X_{k-1}\|_F + M(\varphi(F(X_k)) - \varphi(F(X_{K+1})) \geq \sum_{i=k}^{K} \|X_{i+1} - X_i\|_F + \|X_{K+1} - X_K\|_F \quad (75)$$

Hence

$$\|X_k - X_{k-1}\|_F + M\varphi(F(X_k)) \geq \sum_{i=k}^{K} \|X_{i+1} - X_i\|_F \quad (76)$$

Letting $K \to \infty$ and by using (60), we conclude that

$$\sum_{i=k}^{\infty} \|X_{i+1} - X_i\|_F < +\infty, \quad (77)$$

which implies that the sequence $(X_k)_{k\in\mathbb{N}}$ is a convergent Cauchy sequence and, by (44), converges to X^*. Moreover, by (45), it follows that X^* is a critical point of $F(X)$. □

ACKNOWLEDGMENTS

This work was supported by Miur, R.F.O. project.

REFERENCES

1. Ji H, Liu C. Shen Z., Xu Y. Robust video denoising using low-rank matrix completion, *CVPR* **2010**: 1791–1798.
2. Fazel M, Hindi H, Boyd S. A rank minimization heuristic with application to minimum order system approximation, *Proceedings of the American Control Conference Arlington, VA* **2001**: 4734–4739.
3. Chen, P, Suter, D. Recovering the missing components in a large noisy low-rank matrix: application to SMF, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2004**; **26(2)**:1051–1063.
4. Elad, M, Aharon, M. Image Denoising via sparse and redundant representations over learned dictionaries, *IEEE Transaction on Image Processing* **2006**; **15(12)**: 3736–3745.
5. Zeng B, Fu JJ. Directional discrete cosine transforms: a new framework for image coding, *IEEE Transactions on Circuits and Systems for Video Technology* **2008**; **18(3)**: 305–313.
6. Amit Y, Fink M, Srebro N, Ullman S. Uncovering shared structures in multiclass classification, in *in Proc. 24th International Conference on Machine Learning* **2007**.
7. Argyriou A, Evgeniou T, Pontil M. Multi-task feature learning, in *Neural Information Processing Systems* **2007**.
8. Chistov AL, Grigoriev DY. Complexity of quantifier elimination in the theory of algebraically closed fields, *Mathematical Foundations of Computer Science*, Springer Verlag **1984**; **176**: 17–31.
9. Recht B, Fazel M, Parrilo PA. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Review* **2010**; **52(3)**: 471–501.
10. Candés EJ, Recht B. Exact matrix completion via convex optimization, *Foundations of Computational Mathematics* **2009**; **9(6)**: 717–772.
11. Candés EJ, Tao T. The power of convex relaxation: near-optimal matrix completion, *IEEE Transactions on Information Theory* **2010**; **56(5)**: 2053–2080.
12. Sturm JF. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones, *Optimization Methods and Software* **1999**; **11**: 625–653.
13. Tütüncü RH, Toh KC, Todd MJ. Solving semidefinite-quadratic-linear programs using SDPT3, *Mathematical Programming* **2003**; **95(2)**: 189–217.
14. Candés EJ, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics* **2003**; **59(8)**: 12071223.

15. Cai JF, Candés EJ, Shen Z. A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal of Optimization* **2010**; **20(4)**: 1956–1982.
16. Goldfarb D, Ma S, Wen Z. Solving low-rank matrix completion problems efficiently. *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing, Allerton* **2009**: 1013-1020.
17. Yang, J, Yuan, X. An inexact alternating direction method for trace norm regularized least squares Problem. Tech. Rep., Dept. of Mathematics. Nanjing University **2010**.
18. Yuan, X, Yang, J. Sparse and low-rank matrix decomposition via alternating direction methods. Tech. Rep., Dept. of Mathematics, Hong Kong Baptist University **2009**.
19. Ma S, Goldfarb D, Chen L. Fixed point and Bregman iterative methods for matrix rank minimization, *Mathematical Programming Series A.* **2011**; **128(1)**: 321–353.
20. Toh KC, Yun S. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems, *Pacific Journal of Optimization* **2010**; **6**: 615 – 640.
21. Drineas H, Kannan R, Mahoney MW. Fast Monte Carlo algorithms for matrices II: computing low-rank approximations to a matrix, *SIAM Journal on Computing* **2006**; **36**: 132–157.
22. E. Hale, W. Yin, and Y. Zhang, Fixed-point continuation for 1- minimization: Methodology and convergence, *SIAM Journal on Optimization* **2008**; **19(3)**, 1107–1130.
23. M. Figueiredo, R. Nowak, and S. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE Journal of Selected Topics in Signal Processing* **2007** ,**1(4)**, 586–598.
24. Halko N, Martinsson PG, Tropp JA. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Review* **2011**; **53(2)**: 217–288.
25. Keshavan RH, Montanari A, Oh S. Matrix Completion from a few entries, *IEEE Transactions on Information Theory* **2010**; **56(6)**: 2980-2998.
26. Zaiwen W, Yin W, Zhang Y. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm, *Mathematical Programming Computation* **2012**; **4**: 333-361.
27. Fornasier M, Rauhut H, Ward R. Low-rank matrix recovery via iteratively reweighted least squares minimization, *SIAM Journal on Optimization* **2011**; **2(4)**: 1614–1640.
28. Lai MJ, Xu Y, Yin W. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization, *SIAM Journal on Numerical Analysis* **2013**; **51(2)**: 927–957.
29. Lu C, Tang J, Yan S, Lin Z. Generalized nonconvex nonsmooth low-rank minimization, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* **2014**: 4130–4137.
30. Mohan K, Fazel M. Iterative reweighted algorithms for matrix rank minimization, *Journal of Machine Learning Research* **2012**; **13**: 3441–3473.
31. Konishi K, Uruma K, Takahashi T, Furukawa T. Iterative partial matrix shrinkage algorithm for matrix rank minimization, *Signal Processing* **2014**; **100**: 124–131.
32. Combettes PL, Pesquet JC. Proximal splitting methods in signal processing, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H.H. Bauschke, Ed. et al. New York, NY, USA: Springer Verlag **2011**: 185–2012.
33. Combettes PL, Pesquet JC. Proximal thresholding algorithm for minimization over orthonormal bases, *SIAM Journal on Optimization* **2008**; **18(4)**: 13511376.
34. Combettes PL, Wajs VR. Signal recovery by proximal forward-backward splitting, *SIAM Journal on Multiscale Modelling and Simulation* **2005**; **4(4)**: 1168–1200.
35. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal Imaging Science* **2009**; **2(1)**: 183–202.
36. Montefusco LB, Lazzaro D, Papi S. Fast sparse image reconstruction using adaptive nonlinear filtering, *IEEE Transaction on Image Processing* **2011**; **20(2)**: 534–544.
37. Montefusco LB, Lazzaro D, Papi S. A fast algorithm for nonconvex approaches to sparse recovery problems, *Signal Processing* **2013**; **93**: 2636–2647.
38. Attouch H, Bolte J, Svaiter BF. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forwardbackward splitting, and regularized GaussSeidel methods, *Mathematical Programming, Series A* **2013**; **137**: 91-129.
39. R.T. Rockafellar and R.J-B. Wets, Variational analysis, Springer (1998).
40. A. Parekh and I. Selesnick, Convex Denoising using Non-Convex Tight Frame Regularization, *IEEE Signal Processing Letters* **2015***;* **22(10)***: 1786–1790*
41. *Selesnick IW, Parekh A, Bayram I. Convex 1-D Total variation denoising with non-convex regularization* IEEE Signal Processing Letters **2015***;* **22(2)***: 141–144.*
42. *Selesnick IW, Bayram I. Sparse signal estimation by maximally sparse convex optimization,* IEEE Transactions on Signal Processing **2014***;* **62(5)***: 1078–1092.*
43. *Mirsky L. A trace Inequality of John von Neumann,* Monatshefte fur Mathematik **1975***;* **79***: 303–306.*
44. *Von Neumann J. Some matrix inequalities and metrization of matrix-space, in* Tomsk Polytechnic University Review **1937***;* **1***: 286–300.*
45. *Lewis AS. The convex analysis of unitarily invariant matrix norms, in* J. Convex Analysis **1995***;* **2(1/2)***: 173–183.*
46. *Kruger AY. On Fréchet Subdifferentials,* Journal of Mathematical Science **2003***;* **116(3)***:3325–3358.*
47. *Schneider R, Uschmajew A. Convergence results for projected line-search methods on varieties of low-rank matrices via Lojasiewicz inequality,* SIAM Journal on Optimization **2015***;* **25(1)***: 622–646, 2015.*
48. *GroupLens. MovieLens. http://www.grouplens.org/taxonomy/term/14,* **2009**.