Robust Recovery of Low-Rank Matrices using Multi-Penalty Regularization

Massimo Fornasier massimo.fornasier@ma.tum.de
Technical University of Munich, Germany
Johannes Maly johannes.maly@ma.tum.de
Technical University of Munich, Germany
Valeriya Naumova valeriya@simula.no
Simula Research Laboratory, Norway

Abstract

We propose a novel approach to robust and efficient recovery of low-rank sparse matrices from few noisy compressive measurements. Our approach is based on minimization of a multi-penalty functional with a sparsity-promoting term on right singular vectors. We propose and analyze an iterative alternating algorithm for minimizing the functional. The main virtue of the proposed algorithm is that each singular vector pair is updated iteratively rather than a whole matrix. This modification allows not only to achieve near optimal performance guarantees at low computational costs but also to remove assumptions on the matrix entries distribution, required in other methods. The theoretical results are exemplified by numerical experiments, demonstrating state of the art performance.

1 Introduction

In many data acquisition and reconstruction applications, the data or signal being acquired is assumed to have sparse representations with respect to suitable bases. Classical compressed sensing results guarantee recovery of a sparse signal from a small number of random linear measurements, under certain assumptions on the measurement matrix.

In this paper, we consider a generalization of classical compressed sensing towards the recovery of low-rank sparse matrices. We provide below the precise notion of matrix sparsity, which we consider in this paper. In particular, we consider the general problem of recovering a low-rank sparse matrix $X \in \mathbb{R}^{n_1 \times n_2}$ from the noisy measurements y collected by a linear map $\mathcal{A} \colon \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ with $m \ll n_1 n_2$:

$$y = \mathcal{A}(X) + \eta, \text{ where } \operatorname{rank}(X) \leq R \ll \min\{n_1, n_2\} \text{ and } X \text{ is sparse}$$

where η is some additive noise. This problem formulation is common for many real-life applications in signal and image processing such as hypperspectral image recovery [Golbabaee and Vandergheynst(2012)] or in inverse problems such as blind deconvolution [Lee et al.(2013)Lee, Wu, and Bresler].

Related work The recovery from linear measurements of low-rank matrices without sparsity constraints has been well studied as an extension of classical compressed sensing theory [Candes and Plan(2011), Recht et al.(2010)Recht, Fazel, and Parrilo]. When the unknown matrix is assumed to have both low-rankness and sparsity, the number of compressed sensing measurements required to its successful recovery can be further reduced, see [Jain et al.(2013)Jain, Netrapalli, and Sanghavi] and references therein. The conceptually closest work [Lee et al.(2013)Lee, Wu, and Bresler] to ours proposes an alternating minimization algorithm, the so-called Sparse Power Factorization (SPF), for recovering low-rank sparse matrices from compressed measurements. The authors represent the unknown matrix X as product of two matrices $X = UV^T$ and then apply alternating minimization based on Hard Thresholding Pursuit (HTP), which enforces additional sparsity on the columns of U and/or V. The authors showed that with a suitable initialization and for small noise levels, SPF provides a stable recovery of a rank-R matrix with s_1-sparse columns and s_2-sparse rows whenever the number of measurements is $O(R(s_1 + s_2)\log(\max\{en_1/s_1, en_2/s_2\}))$. At the same time, to achieve these results, the authors require that all columns (resp. rows) of X share a common support, which is not always practical and/or feasible. Moreover, convergence heavily relies on $\mathcal{A}$ fulfilling a certain Restricted Isometry Property (RIP) which can only be guaranteed for a small class of randomly generated measurement ensembles.

OPTML 2017: 10th NIPS Workshop on *Optimization for Machine Learning* (NIPS 2017).

Contribution We present a new alternating iterative algorithm for efficient and robust low-rank and sparse matrix recovery with theoretical guarantees. The algorithm is based on minimization of a non-smooth multi-penalty functional with sparsity promoting terms for the right singular vectors. The main virtue of the algorithm is the given explicit formulas for computation of the successive iterations, resulting in low computational complexity. We also illustrate numerically that by using convex relaxation instead of solving a non-convex problem as in [Jain et al.(2013)Jain, Netrapalli, and Sanghavi] we can achieve better approximation properties, esp. in the presence of noise, and show convergence results without any conditions on the measurement operator $\mathcal{A}$ or assumptions on the support distribution in X. Additionally, we generalize the concept of alternating minimization by not only alternating between two matrices but between R pairs of vectors. This enables us to drop the assumption of a common support for all columns (resp. rows).

The paper is organized as follows: in Section 2 we present and discuss the iterative alternating algorithm for low-rank matrix recovery. The main results are presented in Section 3, whereas numerical experiments are provided in Section 4. Proofs and further results are provided in [Fornasier et al.(2017)Fornasier, Maly, and Naumova].

2 Algorithm

Let us first define our notion of matrix sparsity: We are interested in recovering the unknown low-rank sparse matrix X, which can be represented as $X = \sum_{r=1}^{R} u^r (v^r)^T$ such that $\|u^r\|_2 = 1$ and $\|v^r\|_2 = \sigma_r$ where R is the rank of the matrix; $\sigma_1, \dots \sigma_R$ are the positive singular values, and $U = [u^1, \dots, u^R], V = [v^1, \dots, v^R]$ are the matrices of left- and right-singular vectors. We consider the case when the right singular vectors v^r are $s-$sparse, i.e., v^r has only s non-zero coefficients. We take measurements of X using a linear map $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ such that

$$y = \mathcal{A}(X) + \eta = \begin{pmatrix} \langle A_1, X \rangle_F \\ \vdots \\ \langle A_m, X \rangle_F \end{pmatrix} + \eta. \tag{1}$$

Motivated by recent works, providing theoretical and numerical evidences of superior performance of multi-penalty regularization, see [Naumova and Peter(2014), Grasmair and Naumova(2016), Daubechies et al.(2016)Daubechies, Defrise, and De Mol] and references therein, we propose to recover X by minimizing the following multi-penalty functional:

$$J_{\alpha,\beta}^R(u^1, \dots, u^R, v^1, \dots, v^R) := \left\| y - \mathcal{A}\left(\sum_{r=1}^{R} u^r (v^r)^T \right) \right\|_2^2 + \alpha \sum_{r=1}^{R} \|u^r\|_2^2 + \beta \sum_{r=1}^{R} \|v^r\|_1, \tag{2}$$

where α, β are regularization parameters. We will denote the global minimizer of (2) by $(u_{\alpha,\beta}^1, \dots, u_{\alpha,\beta}^R, v_{\alpha,\beta}^1, \dots, v_{\alpha,\beta}^R)$. Note that $J_{\alpha,\beta}^R$ also applies to matrices by viewing each $2R$-uple $(u^1, \dots, u^R, v^1, \dots, v^R)$ as the matrix $\sum_{r=1}^{R} u^r (v^r)^T$, and we denote $X_{\alpha,\beta}$ the one corresponding to $(u_{\alpha,\beta}^1, \dots, u_{\alpha,\beta}^R, v_{\alpha,\beta}^1, \dots, v_{\alpha,\beta}^R)$. The functional $J_{\alpha,\beta}^R$ has a restricted domain (the decomposition can only consist of R vector pairs) to enforce low-rankness of X and uses a non-smooth term $\|\cdot\|_1$ to promote sparsity in right singular vectors of X.

The minimization of functional (2) is performed by using the alternating algorithm based on simple iterative soft-thresholding, the so-called **A**lternating **T**ikhonov regularization and **Las**so (ATLAS), see (3). As most of the non-convex minimization algorithms, empirical performance of ATLAS depends heavily on a proper initialization $(u_0^1, \dots, u_0^R, v_0^1, \dots, v_0^R)$. We show in Section 4 that initialisation by the leading right singular values of $\mathcal{A}^*(y)$ can provide a stable recovery. After the initialization step, ATLAS iteratively updates the components pairwise $\{(u_k^r, v_k^r)\}, r = 1 \dots R$. The algorithm is efficient as it reduces the minimization of $J_{\alpha,\beta}^R$ to solving a sequence of least-squares and ℓ_1-minimization problems. In the following section, we show that under general assumptions algorithm (3) provides stable recovery of the low-rank sparse matrix at a low-computational cost.

$$\begin{cases} u_{k+1}^1 & = \arg\min_u \left\| \left(y - \mathcal{A}\left(\sum_{r=2}^R u_k^r (v_k^r)^T \right) \right) - \mathcal{A}(u(v_k^1)^T) \right\|_2^2 + \alpha \|u\|_2^2 + \frac{1}{2\lambda_k^1} \|u - u_k^1\|_2^2, \\ v_{k+1}^1 & \in \arg\min_v \left\| \left(y - \mathcal{A}\left(\sum_{r=2}^R u_k^r (v_k^r)^T \right) \right) - \mathcal{A}(u_{k+1}^1 v^T) \right\|_2^2 + \beta \|v\|_1 + \frac{1}{2\mu_k^1} \|v - v_k^1\|_2^2, \\ & \vdots \\ u_{k+1}^R & = \arg\min_u \left\| \left(y - \mathcal{A}\left(\sum_{r=1}^{R-1} u_{k+1}^r (v_{k+1}^r)^T \right) \right) - \mathcal{A}(u(v_k^R)^T) \right\|_2^2 + \alpha \|u\|_2^2 + + \frac{1}{2\lambda_k^R} \|u - u_k^R\|_2^2, \\ v_{k+1}^R & \in \arg\min_v \left\| \left(y - \mathcal{A}\left(\sum_{r=1}^{R-1} u_{k+1}^r (v_{k+1}^r)^T \right) \right) - \mathcal{A}(u_{k+1}^R v^T) \right\|_2^2 + \beta \|v\|_1, + \frac{1}{2\mu_k^R} \|v - v_k^R\|_2^2 \end{cases} \tag{3}$$

3 Main Results

Before stating our main results, we define a set of low-rank matrices with approximately sparse singular vectors and a corresponding RIP which is necessary for approximation guarantees. We call a decomposition $Z = UV^T = \sum_{r=1}^R u^r (v^r)^T$ an approximate Sparse Decomposition (SD) of Z if $U \in \mathbb{R}^{n_1 \times R}$ and $V \in \mathbb{R}^{n_2 \times R}$ have columns which are approximately sparse, i.e. $\|u^r\|_1 / \|u^r\|_2 \le \sqrt{s_1}$ (resp. $\|v^r\|_1 / \|v^r\|_2 \le \sqrt{s_2}$) for all $r \in [R]$; see [Plan and Vershynin(2013)] for more details on approximate sparsity. The SD is not unique and the SVD of Z is not necessarily an SD. We define the sets of approximately sparse matrices as

$$K_{s_1,s_2}^R = \{Z \in \mathbb{R}^{n_1 \times n_2} \colon Z \text{ possesses approximate SD}\}.$$

Note that a set of sparse matrices is a subset of K_{s_1,s_2}^R. Contrary to [Lee et al.(2013)Lee, Wu, and Bresler], we do not require neither columns to share a common support, nor orthogonality of U and V.

Definition 3.1 (Rank-R and (s_1, s_2)-sparse RIP,[Fornasier et al.(2017)Fornasier, Maly, and Naumova]) *A linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ satisfies the rank-R and approximately (s_1, s_2)-sparse RIP with isometry constant $0 < \gamma < 1$ if for all $Z \in K_{s_1,s_2}^R$*

$$(1 - \gamma)\|Z\|_F \le \|\mathcal{A}(Z)\|_2 \le (1 + \gamma)\|Z\|_F, \tag{4}$$

The next Lemma shows that up to log-factors $m \approx \mathcal{O}\left(R^3(s_1 + s_2)\right)$ Gaussian measurements are sufficient to preserve the norms within K_{s_1,s_2}^R.

Lemma 3.2 (RIP for Gaussian Operators, [Fornasier et al.(2017)Fornasier, Maly, and Naumova]) *Let $\mathcal{A} \colon \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be the linear measurement operator of form (1). Assume, all A_i for $1 \le i \le m$ have i.i.d. Gaussian entries. If $m \gtrsim \gamma^{-4} R^3 (s_1 + s_2) \log^3 \left(\max\{n_1, n_2\}\right)$, $\mathcal{A}$ has the rank-R and approximately (s_1, s_2)-sparse RIP with probability at least $1 - 2\exp(-C'\sqrt{m})$, $C' = C'(R, \log(n_1), \log(n_2), s_1, s_2)$.*

Our main result demonstrates that under suitable RIP condition the approximation error is linear in noise level with the slope depending on sparsity level and RIP constant.

Theorem 3.3 ([Fornasier et al.(2017)Fornasier, Maly, and Naumova]) *Let X fulfils the noisy measurements $y = \mathcal{A}(X) + \eta$ with $\hat{v}^r$ s-sparse for all $r \in [R]$ and let $\alpha = \beta = \|\eta\|_2^2 / \|X\|_{\frac{2}{3}}^{\frac{2}{3}} < 1$. Assume $\mathcal{A}$ has the rank-$2R$ approximately $\left(n_1, \max\{s, C^2(\|X\|_{\frac{2}{3}}^{\frac{2}{3}} / \|\eta\|_2^2)^2\}\right)$-sparse RIP with some RIP-constant $0 < \gamma < 1$, $C = C(\|X\|_{\frac{2}{3}}, \eta, s)$. Then, for $X_{\alpha,\beta}$ with $\|v_{\alpha,\beta}^r\|_2 \ge \|y\|_2 / C$, $r = 1, \dots, R$, we have*

$$\|X - X_{\alpha,\beta}\|_F \le \frac{2\sqrt[6]{s} + 2}{1 - \gamma} \|\eta\|_2.$$

As the above results apply only to global minimizers of $J_{\alpha,\beta}^R$, an important question whether the sequences produced by (3) converge to it. Adapting the results [Attouch et al.(2010)Attouch, Bolte, Redont, and Soubeyran], we can show the local convergence to a global minimizer. The current results do not provide proof for any initialization to fulfill the above requirements, which remains to be an open problem for future research.

Theorem 3.4 ([Fornasier et al.(2017)Fornasier, Maly, and Naumova]) *ATLAS converges to a stationary point of $J^R_{\alpha,\beta}$. Let $(u^1_{\alpha,\beta}, ..., v^R_{\alpha,\beta})$ be a global minimiser of $J^R_{\alpha,\beta}$. There exist $\varepsilon, \eta > 0$ such that*

$$\|(u^1_0, \ldots, v^R_0) - (u^1_{\alpha,\beta}, \ldots, v^R_{\alpha,\beta})\|_2 < \varepsilon, \quad \min J^R_{\alpha,\beta} < \min J^R_{\alpha,\beta}(u^1_0 \ldots v^R_0) < \min J^R_{\alpha,\beta} + \eta,$$

implies the sequence $(u^1_k, \ldots, v^R_k)$ generated by ATLAS converges to $(u^1_{\alpha,\beta}, ..., v^R_{\alpha,\beta})$.

4 Numerical Experiments

We compare the performance of ATLAS to the state-of-the art method SPF [Jain et al.(2013)Jain, Netrapalli, and Sanghavi] for the low-rank sparse matrix recovery. Actually, SPF can be considered as a benchmark, as it has been shown to outperform most of the popular recovery algorithms based on convex relaxation [Jain et al.(2013)Jain, Netrapalli, and Sanghavi]. We compare the performance of the algorithms in terms of recovery probability and mean approximation error for 30 experiments with Gaussian random sampling operator $\mathcal{A}$. In each experiment, we randomly generate $X \in \mathbb{R}^{16\times 100}$ with $\|X\|_F = 10$ and 10-sparse right singular vectors. We consider the experiments with noisy measurements, and $\eta = 0.3\|X\|_F$. We fix both regularization parameters at $\alpha = \beta = 0.5$. As initialisation, we use leading singular vectors of $\mathcal{A}^*(y)$ for both algorithms which is a sub-optimal choice but sufficient for illustration. We count the recovery as successful if $\|X - X_{\text{appr}}\|_F / \|X\|_F \leq 0.4$. All computations reported in this paper are performed in Matlab using standard toolboxes.

The comparison results are displayed in Fig. 1 and Fig 2. As can be seen from the results, ATLAS shows a higher level of robustness w.r.t. noise, in contrast to SPF that has a good performance in noise-free cases. Additionally, ATLAS outperforms SPF as soon as singular vectors of X do not share common support, which is a quite restrictive assumption for SPF to be successful.

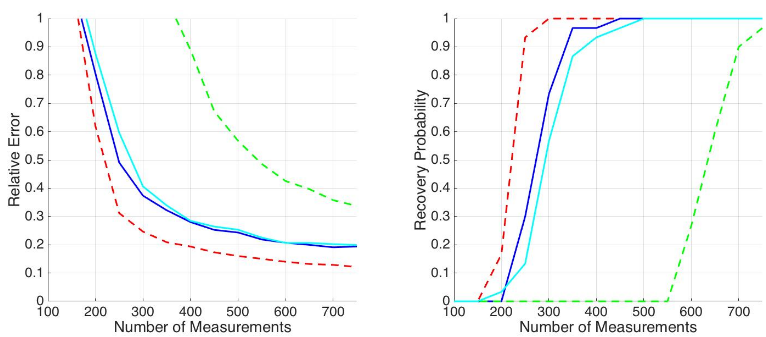
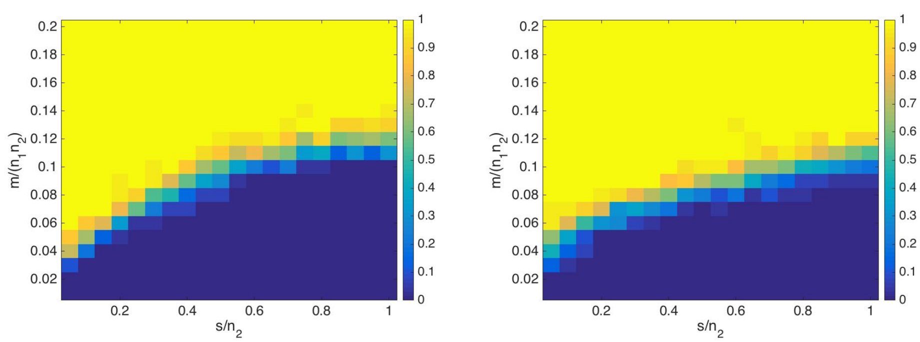


Figure 1: Comparison of SPF and ATLAS with and without common support for $R = 5$. Depicted are average approximation error relative to $\|X\|_F$ and recovery probabilities of SPF (dashed) and ATLAS (solid). Common Support: SPF (red) vs ATLAS (blue). Arbitrary Support: SPF (green) vs ATLAS (cyan).



(a) SPF (b) ATTILA

Figure 2: Comparison of SPF and ATLAS with noise on the measurements for $R = 1$. Recovery probability is depicted by color from zero (blue) to one (yellow).

Acknowledgments

The first two authors acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) in context of CoSIP SPP 1798. The third author acknowledges the support of project No 251149/O70 'Function-driven Data Learning in High Dimension' (FunDaHD) funded by the Research Council of Norway.

References

[Attouch et al.(2010)Attouch, Bolte, Redont, and Soubeyran] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[Candes and Plan(2011)] E. J. Candes and Y. Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57(4):2342–2359, 2011.

[Daubechies et al.(2016)Daubechies, Defrise, and De Mol] I. Daubechies, M. Defrise, and C. De Mol. Sparsity-enforcing regularisation and ISTA revisited. *Inverse Problems*, 32(10):104001, 2016.

[Fornasier et al.(2017)Fornasier, Maly, and Naumova] M. Fornasier, J. Maly, and V. Naumova. Atlas: A multi-penalty approach to compressed sensing of low-rank matrices with sparse decomposition. *In preparation*, 2017.

[Golbabaee and Vandergheynst(2012)] M. Golbabaee and P. Vandergheynst. Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery. *Proceedings of ICASSP 2012*, 2012.

[Grasmair and Naumova(2016)] M. Grasmair and V. Naumova. Conditions on optimal support recovery in unmixing problems by means of multi-penalty regularization. *Inverse Problems*, 32(10):104007, 2016.

[Jain et al.(2013)Jain, Netrapalli, and Sanghavi] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. pages 665–674, 2013.

[Lee et al.(2013)Lee, Wu, and Bresler] K. Lee, Y. Wu, and Y. Bresler. Near optimal compressed sensing of a class of sparse low-rank matrices via sparse power factorization. *arXiv preprint arXiv:1312.0525*, 2013.

[Naumova and Peter(2014)] V. Naumova and S. Peter. Minimization of multi-penalty functionals by alternating iterative thresholding and optimal parameter choices. *Inverse Problems*, 30(12):125003, 2014.

[Plan and Vershynin(2013)] Y. Plan and R. Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66(8):1275–1297, Aug. 2013.

[Recht et al.(2010)Recht, Fazel, and Parrilo] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.