Signal Processing 139 (2017) 62–69

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Improved sparse low-rank matrix estimation

Ankit Parekh[a,*], Ivan W. Selesnick[b]

[a] Dept. of Mathematics, Tandon School of Engineering, New York University, USA
[b] Dept. of Electrical and Computer Engineering, Tandon School of Engineering, New York University, USA

ARTICLE INFO

Article history:
Received 30 January 2017
Revised 9 April 2017
Accepted 11 April 2017
Available online 12 April 2017

Keywords:
Low-rank matrix
Sparse matrix
Speech denoising
Non-convex regularization
Convex optimization

ABSTRACT

We address the problem of estimating a sparse low-rank matrix from its noisy observation. We propose an objective function consisting of a data-fidelity term and two parameterized non-convex penalty functions. Further, we show how to set the parameters of the non-convex penalty functions, in order to ensure that the objective function is strictly convex. The proposed objective function better estimates sparse low-rank matrices than a convex method which utilizes the sum of the nuclear norm and the ℓ_1 norm. We derive an algorithm (as an instance of ADMM) to solve the proposed problem, and guarantee its convergence provided the scalar augmented Lagrangian parameter is set appropriately. We demonstrate the proposed method for denoising an audio signal and an adjacency matrix representing protein interactions in the 'Escherichia coli' bacteria.

1. Introduction

We aim to estimate a sparse low-rank matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ from its noisy observation $\mathbf{Y} \in \mathbb{R}^{m\times n}$, i.e.,

$$\mathbf{Y} = \mathbf{X} + \mathbf{W}, \qquad \mathbf{W} \in \mathbb{R}^{m\times n}, \tag{1}$$

where $\mathbf{W}$ represents additive white Gaussian noise (AWGN) matrix. The estimation of sparse low-rank matrices has been studied [7] and used for various applications such as covariance matrix estimation [4,21,54], subspace clustering [27], biclustering [34], sparse reduced rank regression [8,15], graph denoising and link prediction [46,47], image classification [53] and hyperspectral unmixing [24].

In order to estimate the sparse low-rank matrix $\mathbf{X}$, it has been proposed [47] to solve the following optimization problem

$$\arg\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \left\{ \frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2 + \lambda_0\|\mathbf{X}\|_* + \lambda_1\|\mathbf{X}\|_1 \right\}, \tag{2}$$

where $\|\cdot\|_*$ is the nuclear norm, $\|\cdot\|_1$ is the entry-wise ℓ_1 norm and $\lambda_i \geq 0$ are the regularization parameters. The nuclear norm induces sparsity of the singular values of the matrix $\mathbf{X}$, while the entry-wise ℓ_1 norm induces sparsity of the elements of $\mathbf{X}$.

The nuclear norm and the ℓ_1 norm are convex relaxations of the non-convex rank and sparsity constraints, respectively. The nuclear norm can be considered as the ℓ_1 norm applied to the singular values of the matrix. It is known that the ℓ_1 norm underestimates non-zero signal values, when used as a sparsity-inducing regularizer. As a result, the sparse low-rank (SLR) problem in (2) can be considered, in general, to be over-relaxed [32]. Further, it is known that the performance of nuclear norm for sparse regularization of the singular values is sub-optimal [39].

In order to estimate the non-zero signal values more accurately, non-convex regularization has been favored over convex regularization [13,43,45,51,52]. Furthermore, it has been shown that non-convex penalty functions can induce sparsity of the singular values more effectively than the nuclear norm [14,26,28,37,40]. Indeed, it was shown that non-convex regularizers are better able to estimate simultaneously sparse and low-rank matrices in the context of spectral unmixing for hyperspectral images [24]. The use of non-convex regularizers (penalty functions), however, generally leads to non-convex optimization problems. The non-convex optimization problems suffer from numerous issues (sub-optimal local minima, sensitivity to changes in the input data and the regularization parameters, non-convergence, etc.).

In this paper, we avoid the issues of non-convexity by using parameterized penalty functions, which aid in ensuring the strict convexity of the proposed objective function. We propose to solve the following improved sparse low-rank (ISLR) formulation

$$\arg\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \Bigg\{ F(\mathbf{X}) := \frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2 + \lambda_0 \sum_{i=1}^{k} \phi(\sigma_i(\mathbf{X}); a_0) + \lambda_1 \sum_{i=1}^{m}\sum_{j=1}^{n} \phi(\mathbf{X}_{i,j}; a_1) \Bigg\}, \tag{3}$$

* Corresponding author.
E-mail address: ankit.parekh@nyu.edu (A. Parekh).
URL: https://github.com/aparek/ISLRMatrix (A. Parekh)

http://dx.doi.org/10.1016/j.sigpro.2017.04.011
0165-1684/

where $k = \min(m, n)$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a parameterized non-convex penalty function (see Section 2.1). Note that, if $\lambda_1 = 0$, then the ISLR formulation reduces to the generalized nuclear norm minimization problem [36,40]. Further, if $\lambda_1 = 0$ and $\phi(x; a) = |x|$, then the ISLR problem (3) reduces to the singular value thresholding (SVT) problem.

The contributions of this paper are two-fold. First, we show how to set the parameters a_0 and a_1 to ensure that the function F in (3) is strictly convex. Second, we provide an ADMM based algorithm to solve (3), which utilizes single variable-splitting compared to two variable-splitting as in [53]. We guarantee the convergence of ADMM, provided the scalar augmented Lagrangian parameter μ, satisfies $\mu > 1$.

1.1. Related work

The parameterized non-convex penalty functions used in this paper have designated non-convexity, which enables the overall objective function F in (3) to be strictly convex. In particular, if the parameters a_0 and a_1 exceed their critical value, then the function F in (3) is non-convex. A similar framework of convex objective functions with non-convex regularization was studied for several signal processing applications (see for eg., [18,33,42,50] and the references therein). It was reported that non-convex regularization outperformed convex regularization methods for these applications.

The sparse low-rank (SLR) formulation in (2) is different from the low-rank + sparse decomposition [9], also known as the robust principal component analysis (RPCA). Both the SLR and the RPCA formulations utilize the nuclear norm and the ℓ_1 norm as sparsity-inducing regularizers [55,56]. The RPCA formulation aims to estimate the matrix, which is the sum of a low-rank and a sparse matrix. Note that, in the case of RPCA, the matrix to be estimated is itself neither sparse or low-rank [11,12]. In contrast, the SLR problem (2), and the one proposed in this paper, considers the case wherein the matrix to be estimated is simultaneously sparse and low-rank (similar to [24]).

Several well-studied convex optimization algorithms, such as ADMM [1,25], ISTA/FISTA [2,22], and proximal gradient methods [17] can be applied to solve convex objective functions of the type (2). The SLR objective function (2), has been solved using Generalized Forward-Backward [44], Incremental Proximal Descent [47] (introduced in [3]), Majorization-Minimization [31], and the Inexact Augmented Lagrangian Multiplier (IALM) method [35]. The IALM method can also be used to solve the SLR problem, although with a different data-fidelity term [53].

2. Preliminaries

We denote vectors and matrices by lower and upper case letters respectively. For a matrix $\mathbf{Y}$, we use the following entry-wise norms,

$$\|\mathbf{Y}\|_F^2 := \sum_{i,j} |\mathbf{Y}_{i,j}|^2, \quad \|\mathbf{Y}\|_1 := \sum_{i,j} |\mathbf{Y}_{i,j}|. \tag{4}$$

Further, we use the nuclear norm (also called the 'Schatten-1' norm) defined as

$$\|\mathbf{Y}\|_* := \sum_{i=1}^{k} \sigma_i(\mathbf{Y}), \tag{5}$$

where $\sigma_i(\mathbf{Y})$ represent the singular values of the matrix $\mathbf{Y} \in \mathbb{R}^{m\times n}$ and $k = \min(m, n)$.

2.1. Parameterized non-convex penalty functions

We propose to use non-convex penalty functions $\phi(x;\ a)$ parameterized by the parameter $a \geq 0$. The value of a provides the degree of non-convexity of the penalty functions. Below we define such non-convex penalty functions and list their properties.

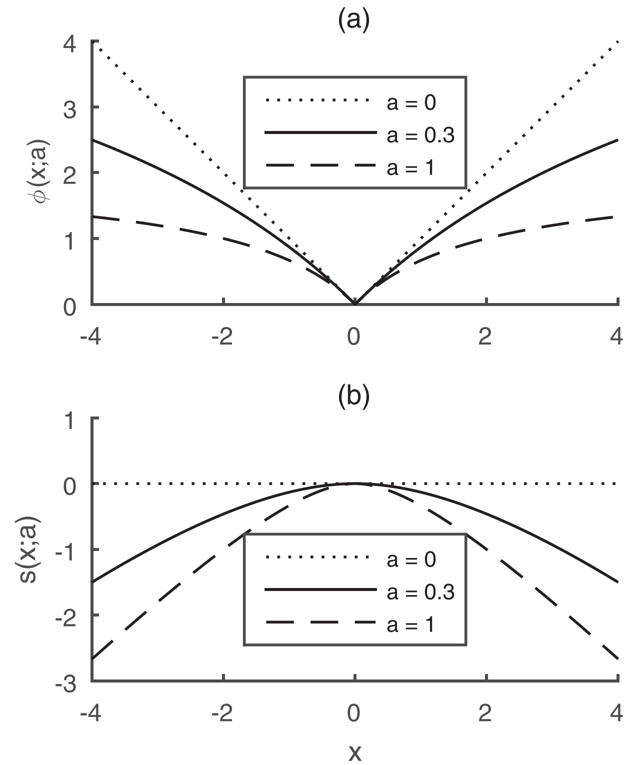

Fig. 1. (a) Non-convex penalty function ϕ in (6) for three values of a. (b) The twice continuously differentiable concave function $s(x; a) = \phi(x; a) - |x|$ in (8) for the corresponding values of a.

Assumption 1. The non-convex penalty function $\phi : \mathbb{R} \to \mathbb{R}$ satisfies the following

1. ϕ is continuous on $\mathbb{R}$, twice differentiable on $\mathbb{R}\backslash\{0\}$ and symmetric, i.e., $\phi(-x; a) = \phi(x; a)$
2. $\phi'(x) > 0,\ x > 0$
3. $\phi''(x) \leq 0,\ x > 0$
4. $\phi'(0^+) = 1$
5. $\inf_{x\neq 0} \phi''(x; a) = \phi''(0^+; a) = -a$

An example of a non-convex penalty function satisfying Assumption 1 is the rational penalty function [23] defined as

$$\phi(x; a) := \frac{|x|}{1 + a|x|/2}, \qquad a \geqslant 0. \tag{6}$$

The ℓ_1 norm is recovered as a special case of the non-convex rational penalty function (i.e., if $a = 0$, then $\phi(x; 0) = |x|$). Fig. 1(a) shows the rational penalty function (6) for different values of a. Other examples of penalty functions satisfying Assumption 1 are the logarithmic penalty [10,38], arctangent penalty [49] and the Laplace penalty [51].

The proximity operator of ϕ [16], $\mathrm{prox}_\phi : \mathbb{R} \to \mathbb{R}$, is defined as

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

The proximity operator associated with the function $\phi(x;\ a)$, satisfying Assumption 1, is continuous with

$$\mathrm{prox}_\phi(y; \lambda, a) = 0, \forall |y| < \lambda, \tag{7}$$

if $0 \leq a < 1/\lambda$. The proximity operators associated with the arctangent and the logarithmic penalty are provided in [49]. Note that for $a = 0$, the proximity operator is the soft-threshold function [19].