ARTICLE IN PRESS
JID: NEUCOM [m5G;February 28, 2018;10:4]
Neurocomputing 000 (2018) 1–13
Contents lists available at ScienceDirect
Neurocomputing
journal homepage: www.elsevier.com/locate/neucom

Bayesian inference for adaptive low rank and sparse matrix estimation

Xixi Jia[a], Xiangchu Feng[a,*], Weiwei Wang[a], Chen Xu[b], Lei Zhang[c]

[a] School of Mathematics and Statistics, Xidian University, Xi'an 710126, China
[b] Shenzhen University, Guangdong, Shenzhen 518060, China
[c] The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 2 July 2017
Revised 31 January 2018
Accepted 12 February 2018
Available online xxx

Communicated by Xin Luo

Keywords:
MAP estimation
ADMM algorithm
Image denoising
Low rank and sparse decomposition

ABSTRACT

Low rank and sparse matrix estimation has been attracting significant interest in recent years. Generally, such a problem is modeled by imposing the l_1-norm to pursuit a sparse and low rank matrix decomposition. However, the l_1-norm is only a conservative sparse regularizer which leads to over-penalty. To remedy this issue, this paper presents an adaptive regularizer learning strategy to provide advanced low rank solution and avoid over-penalty. The new method is termed ARLLR. In the Bayesian inference, the prior distribution of the singular values is assumed to be Laplacian with hyper scale parameters. With the help of full *Maximize A Posterior (MAP)*, we learn the optimal scale parameters by revealing its correlation to the inherent variables. We indicate that the adaptively estimated regularizer corresponds to the *log* function and the global minimum is given for the proposed non-convex problem. Furthermore, by employing the adaptive regularizer on the sparse part, a double *log* regularized low rank and sparse matrix decomposition model which is denoted by ARLLRE, is proposed. The ADMM algorithm is utilized to solve the ARLLRE problem, and the convergence of the algorithm is proved. In experiment, we use ARLLR for image denoising and ARLLRE for foreground and background extraction, respectively. Experimental results show that ARLLR enhances image denoising performance compared with the state-of-the-art image denoising algorithms in both quantity value and visual quality. Meanwhile, ARLLRE delivers excellent results in foreground and background extraction.

1. Introduction

The low rank structure of a given matrix has been extensively exploited in many signal processing applications, such as subspace clustering [1–4] compressed sensing [5], background modeling [6,7], [8–12] and especially for image restoration [13–15] and recommendation system [16–19]. It is observed that the video sequences when vectorized as a matrix contain a low rank background [6] and a sparse foreground movements. In image restoration, the matrix formed by non-local similar image patches is usually assumed to be a low rank matrix and such low rank prior gains significant improvements for image denoising problem. In factor models [20], each factor is a preference vector, and a user's preferences correspond to a linear combination of these factor vectors, with user-specific coefficients, training such a linear factor model amounts to approximating the empirical preferences with a low-rank matrix.

Exploiting the low rank structure of a matrix leads one to deal with the low rank matrix approximation problem (LRMA). Generally, minimizing the rank function in LRMA is ill-posed and NP-hard for some specific problems [21]. A tractable solution is to replace the rank function by the convex the nuclear norm, which gives the well known nuclear norm minimization (NNM) problem [22] defined as

$$\min_{\boldsymbol{X}} \frac{1}{2\tau}\|\boldsymbol{Y}-\boldsymbol{X}\|_F^2+\lambda\|\boldsymbol{X}\|_* \qquad (1)$$

where $\|\boldsymbol{X}\|_* = \sum_{i=1}^r |\sigma_i|$ and σ_i is the ith singular values of the matrix $\boldsymbol{X}$, τ is a parameter that correlated with the variance of the data term and λ is the regularization parameter. The nuclear norm has been widely studied as a surrogate of the rank function, and theoretical analysis [6] guarantee that under some certain conditions the nuclear norm can accurately recover the low rank structure of a matrix. Cai *et al.*, [22] proved that the NNM problem (1) can be solved in closed form and proposed the well known singular value thresholding method.

Albeit its success, the NNM (1) is still sub-optimal for low rank approximation, since the nuclear norm is only a loose approximation of the rank function. Meanwhile, the l_1-norm may cause over-

* Corresponding author.
E-mail addresses: xixijia@stu.xidian.edu.cn (X. Jia), xcfeng@mail.xidian.edu.cn (X. Feng), wwwang@mail.xidian.edu.cn (W. Wang), cslzhang@comp.polyu.edu.hk (L. Zhang).

https://doi.org/10.1016/j.neucom.2018.02.057
0925-2312/

penalty and result in a biased solution. To alleviate the problem, some non-convex regularizers have been proposed, which mainly include Smoothly Clipped Absolute Deviation(SCAD) [23], Minimax Concave Penalty (MCP) [24], Capped L1 [25], Exponential-Type Penalty (ETP) [26], Geman [27], Laplace [28] and correntropy induced metric(CIM) [29]. The non-convex regularization often performs better than the l_1-norm [30] and they are proposed to approximate the rank function when applied on the singular values of the matrix, such as the Schatten-p norm $(0 < p < 1)$ [31]. Yet, they are hand-crafted functions with some tunable parameters and lack of statistical reasonability. Moreover, it still remains unclear what non-convex function can better fits the problem. In the Bayesian perspective, the regularizer term in (1) is closely related to the prior distribution of the inherent variables. Proper regularizer should fits well the prior knowledge, and accompany with the likelihood to compose an advanced model.

In this paper, by taking full advantage of the Bayesian inference, we propose to learn an adaptive regularizer. The new regularizer adaptively fits the prior distribution of the singular values and provides a more accurate estimation for solving low rank matrix approximation problem (ARLLR). Generally, the singular values of a matrix follows Laplacian distribution with a hyper scale parameters. In this paper, rather than using hand-tuned scale parameters or given regularizers, and by maximizing the *full MAP*, we show that the prior distribution of the inherent variable corresponds to the *log* function as regularizer. Based on the analysis, we make an instructive conclusion that the log function is statistically more suitable for low rank approximation among the non-convex functions. Moreover, we prove that the ARLLR (*log* regularizer applied on singular values) has closed form global optimum solution, although the function is non-convex.

In experiments, we find that the *log* regularized LRMA method well preserves the large singular values and penalizes more on the small ones. In this scene, the regularizer is go by the name of adaptive regularizer, which has significant meanings in real applications since large singular values of a matrix are associated with the major projection directions in a low dimensional space, which should be preserved and the small ones is associated with noise which should be discarded [32]. The adaptivity is not only meaningful for the singular values of a low rank matrix, but also appropriate for a sparse matrix. To take one step forward, we propose a double *log* regularized model for low rank and sparse matrix decomposition (known as RPCA [6]), which we term ARLLRE. The ADMM algorithm is used to solve the ARLLRE problem. We prove that our algorithm for solving the two variables non-convex problem converges to a stationary point of the ARLLRE problem. From the process of ADMM algorithm, it can be seen that the low rank parts and the sparse parts are successively separated.

In application, ARLLR is used for image denoising. It is observed that the matrix constructed by the vectorization of nonlocal similar image patches is of low rank and various low rank approximation methods have been used for image denoising [13,15,33–37]. The denoising results of ARLLR are compared with the state-of-art image denoising methods. Meanwhile, ARLLRE is used for foreground and background extraction and the results are compared with [6,32,38–40]. Extensive experimental results verified the superiority of our method over the compared methods.

This paper is an substantial extension of our pioneer conference paper [41], compared to [41], we make following improvements:

- We present detail analysis of the global optimal of the proposed non-convex model, and show that the global optimal can be easily obtained.
- The non-convex adaptive regularizer is extended to low rank and sparse matrix decomposition, and a double log regularized model ARLLRE is proposed;
- The ADMM algorithm is applied to solve the ARLLRE model, and the convergence is proved;
- The ARLLRE model is employed for texture removal and foreground background extraction. The implementation delivers excellent results and confirms the validity of this new model.

The rest of this paper is organized as follows. Section 2 provides a brief survey of the related works. Section 3 first presents the Bayesian inference of our ARLLR model and then give the corresponding regularization. In Section 4, the ARLLR model is extended to low rank and sparse matrix decomposition ARLLRE. In Section 5, extensive experiments are conducted to evaluate the performance of ARLLR and ARLLRE. Finally, several concluding remarks are given in Section 6.

2. Related works and preliminary

In this section, we briefly review some related works and preliminaries which help strengthen our presentation.

2.1. Applications of the low rank matrix approximation

The low rank matrix approximation problem has been extensively used in the past decades, with applications ranging from computer vision, scientific computing and machine learning [1,6,20,32]. One of the most representative application is to solve the Recommendation system problem, such as [16–19] use nonnegative low rank matrix factorization to estimate the low rank structure of the data for recommendation system. In image denoising, the matrix constructed by the similar image patches is assumed to be a low rank matrix and nuclear norm minimization algorithm has been successfully used for image denoising [32,35,42]. In video background extraction, the background of the video sequence which is captured by a surveillance camera and taken under a static scenario naturally share the low rank property, algorithms such as: RPCA [6,32,40], GoDec [10,11] are able to accurately extract the background. The recent work [12] improved the GoDec [10] by introducing correntropy and it is robust to varies noise and outliers compared to GoDec [10]. Most of these low rank approximation algorithms are based on nuclear norm minimization (NNM) problem.

2.2. NNM to generalized low rank approximation

The closed form solution of the NNM problem (1) is given by singular value thresholding [22] as:

$$\boldsymbol{X} = \boldsymbol{U}\mathcal{S}_{\lambda,\tau}(\boldsymbol{\Sigma})\boldsymbol{V}^T, \tag{2}$$

where $\mathcal{S}_{\lambda,\tau}(\boldsymbol{\Sigma})$ is the soft thresholding function operating on the diagonal matrix $\boldsymbol{\Sigma}$ with parameters λ and τ. For each diagonal elements $\boldsymbol{\Sigma}_{ii}$ in $\boldsymbol{\Sigma}$, the function $\mathcal{S}_{\lambda,\tau}$ is defined by:

$$\mathcal{S}_{\lambda,\tau}(\boldsymbol{\Sigma})_{ii} = max(|\boldsymbol{\Sigma}_{ii}| - \lambda\tau, 0). \tag{3}$$

In Eq. (3), the parameter τ is correlated with the noise level, and the parameter λ comes from the prior knowledge of the inherent variable which is crucial for characterizing the prior distribution of the data. Usually, the parameter λ is set uniformly for all singular values, composing an easily solved convex objective function (1). However, as argued in [32], assigning uniform regularization parameter λ on the singular values is inferior. Gu *et al.*, [32] proposed to assign different weights on the singular values and achieved the state-of-the-art image denoising performance.

Some recent works [30,43] have shown that non-convex sparse promoting regularizer perform better than the l_1-norm, and many non-convex models have been proposed such as: l_p-norm $(0 < p < 1)$ [44], Smoothly Clipped Absolute Deviation (SCAD) [23],