Pattern Recognition Letters 71 (2016) 31–37

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A local convex method for rank-sparsity factorization[☆]

Jiayun Mao, Zhenyue Zhang[*]

Department of Mathematics and State Key Laboratory of CAD & CG, Zhejiang University, Yuquan Campus, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 23 February 2015
Available online 18 December 2015

Keywords:
Low-rank matrices
Sparsity
Convex relaxation
Matrix factorization

ABSTRACT

A novel method is proposed for recovering low-rank component and sparsity component of noisy observations, using a local convex envelope of the matrix cardinality function over a local box. Two local relaxation models combined with implicit or explicit rank restriction are proposed for solving the rank-sparsity factorization. An iterative approach of the local relaxation and a post-processing refinement are also given to further improve the factorization, together with updating rules of the local box. Numerical examples show the efficiency of the proposed methods in two applications.

1. Introduction

Low-rank or sparse approximation is a main task in many applications such as image processing. For example, in photometric stereo, a sequence of face images of the same person form a matrix $A \in \mathcal{R}^{m\times n}$, each column of which is a transformed image. Because a face image is nearly convex and almost Lambertian [3,9], if these images are taken under various lighting conditions, A can be roughly represented as $A = L_0 + S_0$ with a low-rank matrix L_0 from a clean face image by removing shadows or specularities and a sparse matrix S_0 of the polluted parts. This factorization also occurs in other applications such as video surveillance [15], rigid SFM [18], nonrigid SFM [5], Gaussian graphical model [14] and matrix rigidity [20].

In the early work for example [19] and [1], the sparsity component is regarded as intra-sample outlier for estimating the low-rank component by a weighted ℓ_2-norm minimization, although the sparsity component is also meaningful in some applications. In [6] and [11], the sparsity component is assumed to have an independent Laplacian distribution. The factorization is then estimated by the solution of the ℓ_1-minimization with an estimated rank r,

$$\min_{S,\, \mathrm{rank}(L)\le r} \|S\|_1, \quad \text{s.t.} \quad A = L + S, \tag{1}$$

where $\|S\|_1 = \sum_{ij} |s_{ij}|$ is the ℓ_1-norm. Due to the explicit rank-restriction, this problem is non-convex. In [21], the convex nuclear norm function $\|X\|_*$ of matrix, the sum of all the singular values of X, is used for penalizing rank or imposing an implicitly rank restriction, yielding a convex problem

$$\min_{S,L} \|L\|_* + \mu\|S\|_1, \quad \text{s.t.} \quad A = L + S \tag{2}$$

with a suitably set parameter $\mu > 0$. It can be solved cheaply, using a shrink operator as in [7]. The implicit rank-constrained model is more suitable for noisy observed data A,

$$A = L_0 + S_0 + N_0. \tag{3}$$

Assuming that we have an estimate for the noise matrix N_0 as its Frobenius norm $\|N_0\|_F \le \varepsilon$, (2) can be extended to

$$\min_{S,L} \|L\|_* + \mu\|S\|_1, \quad \text{s.t.} \quad \|A - L - S\|_F \le \varepsilon, \tag{4}$$

as did in [2] and [17].

There are other relaxation approaches for low-rank/sparsity factorization. In [16], a so-called capped-norm is used to relax the rank function rank(L) and the cardinality function $\|S\|_0$. In [12,13], the minimization problems of $\|L - A\|_F^2 + \lambda\, \mathrm{rank}(L)$ and $\|L - A\|_F^2 + \lambda \max\{r, \mathrm{rank}(L)\}$ for low-rank approximation with implicit or explicit rank-restriction are relaxed by the convex envelopes of the objective functions. It is a bit similar to [10] in which the objective $\|x\|_2^2 + \lambda\|x\|_0$ is relaxed by its local convex envelope over $\{x\colon \|x\|_\infty \le 1\}$.

In applications like video surveillance, the sparsity component in the factorization (3) is meaningful. It is known that the ℓ_1-minimization may give inferior results with small entries in the sparsity component. These undesirable small components can be reduced much if we look for an S as sparse as possible.

In this paper, we propose a new method for the low-rank-sparsity factorization (3), focusing on the sparsity component. The main contribution consists of three parts. First, we propose to use a local convex envelope of the ideal but non-convex and noncontinuous cardinality function $\|S\|_0$ over a local box around a center

[☆] This paper has bee recommended for acceptance by Prof. A. Heyden.

[*] Corresponding author. Tel.: +86 571 87951609/8307.
E-mail address: zyzhang@zju.edu.cn (Z. Zhang).

http://dx.doi.org/10.1016/j.patrec.2015.12.002
0167-8655/

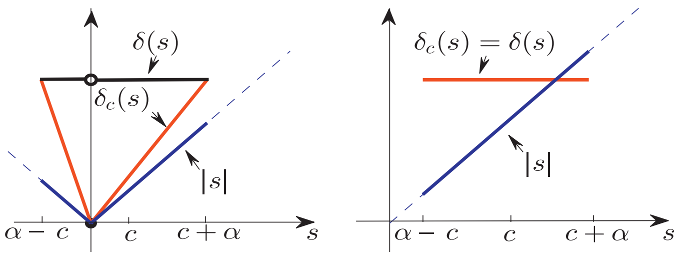

Fig. 1. The convex envelope $\delta_c(s)$ of $\delta(s)$ in $[c-\alpha, c+\alpha]$ (red lines) and the ℓ_1-function $|s|$ (blue lines), when $|c| \le \alpha$ (left) or $|c| > \alpha$ (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

matrix. Combining the local convex envelope on the sparse matrix S with the nuclear norm on the implicit low-rank matrix L or an explicit low-rank factorization of L, we give two optimization formulations to compute the factorization. Because the envelope is closer to $\|S\|_0$ than $\|S\|_1$ in the local box if the box size is suitable, it is expectable to obtain a sparser solution than the ℓ_1-model, if the local box contains the optimum. Second, the local relaxation is flexible. In practice, it can be also used to improve solutions from other algorithms if necessary. We give an outer loop to iteratively improve the location of the local box and improve the solution if the current local box does contain the optimum. Third, we give a post-processing refinement to further improve the sparsity of S, while the implicit low-rank of L is at last preserved.

We will also show some algorithms for solving the proposed local optimization problems, including convergence analysis.

2. The local convex method

Generally, by searching in a local domain via a local convex envelope, one can get a better solution than that via an envelope over a larger or the whole domain. In this section, we consider a restriction box around a center matrix $C = (c_{ij})$ with radius α,

$$\mathcal{B}(C,\alpha) = \left\{ S = (s_{ij}) : \ |s_{ij} - c_{ij}| \le \alpha, \forall (i,j) \right\}$$

and the convex envelope of the cardinality function $\|S\|_0 = \sum_{ij} \delta(s_{ij})$ over the box, where $\delta(s) = 1$ if $s \ne 0$, and $\delta(0) = 0$. Two local relaxation formulations will be considered, combining the envelope with the nuclear norm for implicit rank-restriction or an explicit rank factorization, respectively.

2.1. The local convex envelope

The convex envelope of a function f in a closed subdomain is the largest convex function less than or equal to f. In our case, both the function $\|S\|_0$ and the subdomain $\mathcal{B}(C,\alpha)$ are separable with respect to variables $\{s_{ij}\}$. The convex envelope is then a sum of the convex envelopes of the δ-function over each interval $[c_{ij}-\alpha, c_{ij}+\alpha]$ for the single variable s_{ij}.

It is easy to verify that the convex envelope of δ-function $\delta(s)$ over $[c-\alpha, c+\alpha]$ is given by $\delta_c(s) = \frac{|s|}{\alpha + \text{sign}(s)c}$ if $|c| \le \alpha$, or $\delta_c(s) \equiv 1$ if $|c| > \alpha$. Hence, the convex envelope of $\|S\|_0$ is

$$\phi_{C,\alpha}(S) = k_{C,\alpha} + \sum_{|c_{ij}| \le \alpha} \frac{|s_{ij}|}{\alpha + \text{sign}(s_{ij})c_{ij}}, \quad S \in \mathcal{B}(C,\alpha), \tag{5}$$

where $k_{C,\,\alpha}$ is the number of $|c_{ij}|$'s larger than α.

As illustrated in Fig. 1, the convex envelope $\delta_c(s)$ is a better approximation to $\delta(s)$ than $|s|$ in the interval $[c-\alpha, c+\alpha]$ if $|c| > \alpha$ or $|c| \le \alpha < 1 - |c|$. If α is suitably chosen for a given center matrix C, for example, $\alpha = \max\{|c_{ij}| : |c_{ij}| < 1/2\}$, $\phi_{C,\,\alpha}(S)$ should be closer to $\|S\|_0$ than $\|S\|_1$ for any $S \in \mathcal{B}(C,\alpha)$:

$$\|S\|_1 < \phi_{C,\alpha}(S) \le \|S\|_0, \quad \forall S \in \mathcal{B}(C,\alpha).$$

2.2. Local relaxation models

We consider two kinds of local minimizations for the low rank sparsity factorization, utilizing the convex relaxation $\phi_{C,\,\alpha}(S)$ of $\|S\|_0$ over the box $\mathcal{B}(C,\alpha)$. One is that

$$\min_{S \in \mathcal{B}(C,\alpha), L} \ \phi_{C,\alpha}(S) + \lambda \|L\|_*, \quad \text{s.t.} \quad \|A - L - S\|_F \le \varepsilon, \tag{6}$$

which implicitly imposes a low-rank restriction on L.

It is known that an iterative method for minimizing a function with the nuclear norm $\|\cdot\|_*$ may require to compute singular value decomposition (SVD) repeatedly. Because an SVD costs much for matrices in a large scale, such an approach may be only suitable for problems in a small scale. In that case, an explicit rank-restriction model may be a good choice for the problem in a large scale since no SVDs are required.

Our second formulation adopts an explicit rank-restriction $\text{rank}(L) \le r$, where r is an estimated rank of the optimum L_0.

$$\min_{S \in \mathcal{B}(C,\alpha),\ \text{rank}(L) \le r} \ \phi_{C,\alpha}(S), \quad \text{s.t.} \quad \|A - L - S\|_F \le \varepsilon \tag{7}$$

One may rewrite L as $L = HF^T$ with two matrices H and F; each is of r columns. A further restriction on the factors H and F can be imposed. For example, one of them is orthonormal, or both are nonnegative, and so on. However, such a restriction may increase the complicity in solving.

Different from the implicit rank-restriction model, the minimization problem (7) is no longer convex because of the nonconvexity of the set of low-rank matrices, though the feasible set $S \in \mathcal{B}(C,\alpha)$ is convex. However, this problem becomes convex when H or F is fixed. Hence, (7) can be solved by alternatively minimizing the two convex subproblems.

3. The algorithms

We will slightly modify the Alternating Direction Method of Multipliers (ADMM) [4] for solving (6). Each ADMM iteration can be implemented simply. The alternative method mentioned above will be discussed, focusing on analysis of its convergence. A modified ADMM method will be given for solving the subproblems involved in the alternating iterations, together with its convergence.

3.1. ADMM algorithm for solving (6)

The problem (6) can be equivalently represented as

$$\min_{S \in \mathcal{B}(C,\alpha), L, \|N\|_F \le \varepsilon} \ \phi_{C,\alpha}(S) + \lambda \|L\|_*, \quad \text{s.t. } A = L + S + N. \tag{8}$$

Its augmented Lagrangian function with a dual variable Y is

$$\mathcal{L}(L,S,N,Y) = \phi_{C,\alpha}(S) + \lambda \|L\|_* + \frac{\rho}{2}\|A - L - S - N + \rho^{-1}Y\|_F^2$$

with a tunable constant $\rho > 0$.

The basic iteration of the ADMM method applied on the above problem consists of four steps. The first three steps minimize $\mathcal{L}$ with one of the primal variables L, S, N over the feasible set, and the last one modifies the dual variable Y. Below we show the procedure in details with $\hat{A} = A + \rho^{-1}Y$.

Step 1. Find the nearest neighbor $\hat{N}$ of $W_N = \hat{A} - L - S$ in the ε-ball $\{\tilde{N} : \|\tilde{N}\|_F \le \varepsilon\}$, which gives the solution

$$\hat{N} = \min\left\{1, \varepsilon / \|W_N\|_F\right\} W_N. \tag{9}$$

Step 2. Let $W_S = \hat{A} - L - \hat{N}$. Update S by the minimizer of $\phi_{C,\alpha}(\tilde{S}) + \frac{\rho}{2}\|\tilde{S} - W_S\|_F^2$ over $\mathcal{B}(C,\alpha)$. Because the objective is separable on the entries $\tilde{s}_{ij}$ of $\tilde{S}$, the minimizer $\hat{S}$ has entries