Neurocomputing 172 (2016) 253–261

Contents lists available at ScienceDirect

Neurocomputing

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

Convex optimization based low-rank matrix decomposition for image restoration

Ning He [a,*], Jin-Bao Wang [a], Lu-Lu Zhang [a], Ke Lu [b]

[a] College of Information Technology, Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 28 November 2013
Received in revised form
30 October 2014
Accepted 1 November 2014
Available online 9 May 2015

Keywords:
Image restoration
Low-rank matrix
Principal component pursuit
Singular value thresholding

ABSTRACT

This paper addresses the problem of image denoising in the presence of significant corruption. Our method seeks an optimal set of image domain transformations such that the matrix of transformed images can be decomposed as the sum of a sparse matrix of errors and a low-rank matrix of recovered denoised images. We reduce this optimization problem to a sequence of convex programs minimizing the sum of the ℓ_1-norm and the nuclear norm of the two component matrices, which can be solved efficiently using scalable convex optimization techniques. We verify the efficacy of the proposed image denoising algorithm through extensive experiments on both numerical simulations and different types of images, demonstrating its highly competent objective performance compared with several state-of-the-art methods for matrix decomposition and image denoising. Our subjective quality results compare favorably with those obtained by existing techniques, especially at high noise levels and with a large amount of missing data.

1. Introduction

In recent years, image restoration has been an important, yet challenging problem widely studied in computer vision and image processing. The purpose of image restoration is to "compensate for" or "undo" defects that degrade an image. Degradation comes in many forms, including motion blur, noise, and camera misfocus. In cases like motion blur, it is possible to obtain a very good estimate of the actual blurring function and "undo" the blur to restore the original image. However, in cases where the image is corrupted by noise, the best we can hope to achieve is to compensate for the resulting degradation. Owing to the ill-posed nature of image restoration, an image restoration solution is generally not unique. To find a better solution, prior knowledge of images can be used to regularize the image restoration problem. One of the most commonly used regularization models is the total variation (TV) model [1,2]. Since the TV model favors piecewise constant image structures, it tends to smooth out the finer details of an image. To better preserve the image edges, algorithms have subsequently been developed to improve the TV models [3,4,5].

The success of TV regularization validates the importance of good image prior models in solving image restoration problems. In wavelet based image denoising [6], researchers have found that the sparsity of wavelet coefficients can serve as a good prior. This implies that many types of signals, e.g., natural images, can be sparsely represented using a dictionary of atoms, such as discrete cosine transforms (DCT) or wavelet bases. In addition, recent studies have shown that iteratively reweighting the ℓ_1-norm sparsity regularization term can lead to better image restoration results [7]. Sparse representation has been successfully used in various image processing applications [8,9,10]. However, sparse decomposition over a highly redundant dictionary is potentially unstable and tends to generate visual artifacts [11,12]. Recently, a representative study introduced low-rank matrix recovery (LR) theory [13] into image restoration. Existing LR-based image denoising models share a common assumption that an image can be represented as a highly redundant information part (e.g., background regions) plus a main part (e.g., the foreground object) including several homogeneous regions. The redundant information part usually lies in a low dimensional feature subspace, which can be approximated as a low-rank feature matrix, whereas the main part can be viewed as a sparse sensory matrix. In this paper, we introduce a new method, the matrix rank minimization image restoration algorithm. Our solution builds on recent advances in rank minimization and formulates the image restoration problem as a solution that connects low-rank methods with simultaneous sparse coding. We utilize the low-rank matrix convex optimization scheme to estimate the local sparsity of the image and adjust the

* Corresponding author.
E-mail addresses: xxthening@buu.edu.cn (N. He), luk@ucas.ac.cn (K. Lu).

http://dx.doi.org/10.1016/j.neucom.2014.11.090
0925-2312/

sparsity regularization parameters. Extensive experiments on image denoising show that the proposed approach can effectively reconstruct the image details.

The remainder of this paper is organized as follows: In Section 2, we introduce the related matrix rank optimization technique and an analysis of existing problems. In Section 3, we introduce matrix rank as a measure of image similarity and reformulate the image denoising problem as one of matrix rank minimization. In addition, we propose an efficient algorithm to solve the rank minimization problem by iterative convex optimization. Experimental results showing the efficacy of our method on numerical simulations and different types of images are presented in Section 4.

2. Related work

In recent years, the search for more scalable algorithms for high-dimensional convex optimization problems has prompted a return to first-order methods. Principal component analysis (PCA) is a popular tool for high-dimensional data analysis, with applications ranging across a wide variety of scientific and engineering fields. It relies on the basic assumption that the given high-dimensional data lie in a much lower-dimensional linear subspace. Correctly estimating this subspace is crucial for reducing the dimension of the data and facilitating tasks such as processing [23,29], analyzing [22,28], compressing, or visualizing the data [14,15]. Suppose that the given data are arranged as columns in a large matrix, $D \in R^{m\times n}$. Classical PCA assumes that this data matrix was generated by perturbing a matrix, $A \in R^{m\times n}$, whose columns lie in a subspace of dimension $r < < \min(m,n)$. In other words, $D = A + E$, where A is a rank-r matrix and E is a matrix whose entries are independent and identically distributed (i.i.d.) Gaussian random variables. In this setting, PCA seeks an optimal estimate of A, by the following constrained optimization:

$$\min_{A,E} \|E\|_F, \text{s.t. } \mathrm{rank}(A) \le r, D = A + E \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm. It is well-known that this problem can be solved efficiently by simply computing the singular value decomposition (SVD) of D. The optimal estimate of low-rank matrix A is simply the projection of the columns of D onto the subspace spanned by the r principal left singular vectors of D [16].

Although PCA offers an optimal estimate of the subspace for data corrupted by small amounts of Gaussian noise, it breaks down under large corruption, even if that corruption affects only a few of the observations. This undesirable behavior motivated the study of the problem of recovering a low-rank matrix A from a corrupted data matrix $D = A + E$, where some entries of E may be of arbitrarily large magnitude.

Recently, the authors in [17] showed that under surprisingly broad conditions, one can recover low-rank matrix A exactly from $D = A + E$ with gross but sparse errors E, by solving the following convex optimization problem:

$$\min_{A,E} \|A\|_* + \lambda \|E\|_1 \quad \text{s.t.} D = A + E \tag{2}$$

where the nuclear norm $\|\cdot\|_*$ (sum of the singular values of a matrix) is a convex relaxation of the matrix rank function, $\|\cdot\|_1$ denotes l_1-norm, which promotes sparsity, and parameter $\lambda > 0$ is a trade-off between the two items. In [18], this optimization is robust PCA (RPCA), because it enables one to correctly recover the underlying low-rank structure in the data, even in the presence of gross errors or outlying observations. This optimization can easily be reformulated as a semi-definite program and solved by an off-the-shelf interior point solver. However, although interior point methods offer superior convergence rates, the complexity of computing the step direction is $O(n^6)$, and thus they do not scale well with the size of the matrix.

One striking example of this is the current popularity of iterative thresholding algorithms for ℓ_1-norm minimization problems arising in compressed sensing [19,20]. Similar iterative thresholding techniques [21,24] can be applied to the problem of recovering a low-rank matrix from an incomplete subset of its entries [25,26,27]. This optimization is closely related to the RPCA problem, and the convergence proof extends quite naturally to RPCA. However, the iterative thresholding scheme proposed in [30] exhibits extremely slow convergence; solving one instance requires about 10^4 iterations, each of which has the same cost as one SVD. Hence, even for matrix sizes as small as 800×800, the algorithm requires more than 8 h on a typical PC.

In this paper, our goal is to develop faster and more scalable algorithms, by further studying the convex optimization problem in Eq. (2) associated with RPCA and applied to the image restoration problem.

3. Image restoration by matrix rank minimization

In this section, we formulate image restoration as a search for a set of transformations minimizing the rank of the transformed images, viewed as the columns of a matrix. We discuss why rank is a natural measure of image similarity, and how this conceptual framework can be made robust to gross errors due to corruption or occlusion.

3.1. Low-rank matrix structure and iterative decomposition

Measuring the degree of similarity within a set of images is a fundamental problem in computer vision and image processing. Consider matrix $X \in R^{m\times n}$ constructed by stacking all the vectorized images, denoted by $\mathrm{vec}(I_k)$, as $X = [\mathrm{vec}(I_1)|\cdots|\mathrm{vec}(I_n)]$, where $\mathrm{vec}(I_j) = [I_j(1), \cdots, I_j(m)]^T$ for $j = 1, \cdots, n$. It follows that X can be factorized as $X = NL$, where $N = [\rho_1 n_1 | \cdots | \rho_m n_m]^T \in R^{m\times 3}$ and $L = [l_1 | \cdots | l_n] \in R^{3\times n}$. Suppose that the number of images is $n \ge 3$. Irrespective of the size of n, the rank of matrix X is clearly at most 3.

Notions: Let $X = (x_1, \cdots, x_n)$ be an $m \times n$ matrix, $\Omega \subset \{1, \cdots, m\} \times \{1, \cdots, n\}$ denote the indices of the observed entries of X, and Ω^c denote the indices of the missing entries. The Frobenius norm of X is defined as $\|X\|_F^2 = \sqrt{\sum_{(i,j)} X_{ij}^2}$. Let P_Ω be the orthogonal projection operator onto the span of matrices vanishing outside of Ω so that the $(i,j) -$ th component of $P_\Omega(X)$ is equal to X_{ij} when $(i,j) \in \Omega$, and zero otherwise. Let $X = U\Sigma V^T$ be the SVD of X, where $\Sigma = \mathrm{diag}(\sigma_i)$, $1 < i < \min\{m,n\}$ and σ_i the $i -$ th largest singular value of X. The shrinkage operator $D_T(X)$ is defined [17] as $D_T(X) = U\Sigma_T V^T$, where $\Sigma_T = \mathrm{diag}(\max\{\sigma_i - T, 0\})$, $1 \le i \le \min\{m,n\}$. We summarize the main result below.

Lemma 1. For a given vector $y \in R^n$ and the thresholding weight vector $w \in R^n_{++}$, the non-uniform singular value operator $S_w[y]$ satisfies:

$$S_w[y] = \arg\min_x \left(\frac{\mu}{2}\|x - y\|_2^2 + \|w \odot x\|_1\right), \tag{3}$$

where $\odot$ is the vector of corresponding matrix multiplication operators, and $\|\cdot\|_2$ and $\|\cdot\|_1$ are the ℓ_2- and ℓ_1-norms, respectively. $\mu > 0$is the penalty factor. As μ approaches 0, any solution to Eq. (3) approaches the solution set of Eq. (2). In other words, the non-uniform singular value operator satisfies:

$$S_w[y] = S_{w0}[z] = \arg\min_x \left(\frac{\mu}{2}\|x - z\|_2^2 + w_0\|x\|_1\right), \tag{4}$$

where $z = \mathrm{sign}(y) \odot (|y| + w_0\mathbf{1} - w)$, $w_0 = \max(\{w_i\})$, $|y| = \mathrm{sign}(y) \odot y$, and vector $\mathbf{1} \in R^n$ with all elements equal to 1.

Proof. For any element of vector y,

$$S_{w_i}[y_i] = \arg\min_{x_i} \left(\frac{\mu}{2}\|x_i - y_i\|_2^2 + w_i\|x_i\|_1\right), \tag{5}$$