Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Incoherent dictionary learning with log-regularizer based on proximal operators

Zhenni Li [a,*], Shuxue Ding [a], Takafumi Hayashi [b], Yujie Li [c]

[a] School of Computer Science and Engineering, The University of Aizu, Aizu-Wakamatsu City, Fukushima 965-8580, Japan
[b] Graduate School of Science and Technology, Niigata University, Niigata City, Niigata, 950-2181, Japan
[c] Artificial Intelligence Center, AIST, Tsukuba, Ibaraki 305-8560, Japan

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Dictionary learning
Coherence
log-regularizer
Proximal operator
Sparse representation

ABSTRACT

In this study, we propose a novel dictionary learning algorithm with the *log*-regularizer and simultaneously with the coherence penalty based on proximal operators. Our proposed algorithm simply employs a decomposition scheme and alternating optimization, which transforms the overall problem into a set of single-vector variable subproblems, with either one dictionary atom or one coefficient vector. Although the subproblems are still nonsmooth and even nonconvex, remarkably they can be solved by proximal operators, and the closed-form solutions of the dictionary atoms and the coefficient vectors are obtained directly and explicitly. To the best of our knowledge, no previous studies of dictionary learning have applied proximal operators to sparse coding with the *log*-regularizer and simultaneously to dictionary updating with the coherence penalty. According to our analysis and simulation study, the main advantages of the proposed algorithm are its greater ability of recovering the dictionary and its faster convergence for reaching the values of the dictionary recovery ratios than state-of-the-art algorithms. In addition, for real-world applications, our proposed algorithm can obtain good performances on audio data and image classification.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Sparse representation is a significantly important model for signal and image processing, which has been studied theoretically and practically [1–5]. A basic assumption to apply this model is that a natural signal $\mathbf{y} \in \mathbb{R}^m$ can be sparsely represented or approximated by a linear combination of a small number of elementary components, called atoms, which are chosen from an overcomplete dictionary $\mathbf{W} \in \mathbb{R}^{m \times r}$. Here overcomplete means that the number r of atoms is greater than the dimension m of the atom. The model can be described as:

$$\mathbf{y} \approx \mathbf{W}\mathbf{h}, \quad \mathbf{h} \text{ with } s \text{ nonzero elements}, \tag{1}$$

where $\mathbf{h} \in \mathbb{R}^r$ is a sparse representation of $\mathbf{y}$ and s ($s < r$) is the sparsity of $\mathbf{h}$. Each column of $\mathbf{W}$ is one atom. The atoms corresponding to the s nonzero elements of $\mathbf{h}$ are used to synthesize the signal $\mathbf{y}$ via their linear combination.

A key problem related to sparse representation is the choice of the dictionary employed to decompose the signals of interest in an efficient manner. A simple approach is to consider predefined dictionaries, such as the discrete cosine transform (DCT) and wavelets [6]. Another approach is to use an adaptive dictionary which is learned from the signals that need to be represented, resulting in better matches to the contents of the signals. This problem leads to the issue of dictionary learning. Solving a dictionary learning problem is usually established as an optimization process involving a serious of iterations between two alternative stages: sparse coding and dictionary update as the following.

Input: the training sample data $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_L]$, where $\mathbf{y} \in \mathbb{R}^m$; initial dictionary $\mathbf{W}^{(0)} \in \mathbb{R}^{m \times r}$.
Procedure: Initialize $t = 0$, and repeat until convergence:

1) *Sparse coding stage*: The sparse representation of $\mathbf{y}$ are often found by,

$$\min_{\mathbf{h}} \phi(\mathbf{h}) = \|\mathbf{y} - \mathbf{W}^{(t)}\mathbf{h}\|_2^2 \quad \text{s.t. } \|\mathbf{h}\|_p \le s, \tag{2}$$

where $\|\cdot\|_p$ is so called ℓ^p-norm (typically $p \in [0, 1]$), which constrains the sparsity. When $p = 0$, $\|\cdot\|_0$ is the so-called ℓ^0-norm, which counts the number of nonzero elements. When $p = 1$, $\|\cdot\|_1$ is the so-called ℓ^1-norm. The aforementioned optimization task can be easily transformed as follows,

$$\min_{\mathbf{h}} \phi(\mathbf{h}) = \|\mathbf{y} - \mathbf{W}^{(t)}\mathbf{h}\|_2^2 + \lambda \|\mathbf{h}\|_p^p, \tag{3}$$

* Corresponding author.
E-mail addresses: lizhenni2012@gmail.com (Z. Li), sding@u-aizu.ac.jp (S. Ding), takafumi@ie.niigata-u.ac.jp (T. Hayashi), yujie-li@aist.go.jp (Y. Li).

http://dx.doi.org/10.1016/j.dsp.2016.12.014
1051-2004/© 2017 Elsevier Inc. All rights reserved.

UNCORRECTED PROOF

where $\lambda > 0$ is a regularization parameter. Thus, the solution to sparse coding becomes the ℓ^p-regularized optimization problem. An alternative to minimizing (3) individually on each vector is to find a joint sparse representation of the matrix $\mathbf{Y}$ by employing a sparsity measure in the matrix form. Then the optimization problem can be formulated as,

$$\min_{\mathbf{H}} \phi(\mathbf{H}) = \|\mathbf{Y} - \mathbf{W}^{(t)}\mathbf{H}\|_F^2 + \lambda \| \mathbf{H} \|_p^p, \tag{4}$$

where $\mathbf{H} \in \mathbb{R}^{r \times L}$ is also called the coefficient matrix. $\| \cdot \|_F$ is the Frobenius norm.

2) *Dictionary update stage*: For the obtained $\mathbf{h}^{(t)}$ or $\mathbf{H}^{(t)}$, update $\mathbf{W}^{(t)}$ such that

$$\min_{\mathbf{W}} \phi(\mathbf{W}) = \|\mathbf{y} - \mathbf{W}\mathbf{h}^{(t)}\|_2^2, \tag{5}$$

or

$$\min_{\mathbf{W}} \phi(\mathbf{W}) = \|\mathbf{Y} - \mathbf{W}\mathbf{H}^{(t)}\|_F^2. \tag{6}$$

The constraints might be imposed on $\mathbf{W}$ to avoid a trivial result or scale ambiguity. Dictionaries with fixed column norm or fixed Frobenius norm have been used in different papers [7,8].

3) *increment* $t = t + 1$.

Many approaches have been proposed for dictionary learning [2,8–13]. K-SVD [8] and MOD [9] are classical and successful dictionary learning approaches that use the ℓ^0-norm as sparsity constraint. They are both solved using orthogonal matching pursuit (OMP) [14] for the ℓ^0-norm. OMP is a greedy algorithm that can find a sufficiently sparse representation, but this algorithm is not suitable for high-dimensional problems. The difference between K-SVD and MOD occurs in the dictionary update stage. MOD updates the overall set of atoms simultaneously by optimizing a least squares problem, but the convergence could not be guaranteed. By contrast, K-SVD employs singular value decomposition (SVD) to update the atoms one by one, which can learn a better dictionary and improve the convergence. However, SVD is computationally expensive. Bao et al. [15] have proposed to use the proximal gradient method for dictionary learning with the ℓ^0-norm as sparsity regularization, which improves the computational complexity compared to K-SVD. However, the proximal gradient method is essentially a gradient-based method which requires more iterations to converge.

As a convex relaxation of the ℓ^0-norm, the ℓ^1-norm has been widely used in many dictionary learning methods to improve the computational feasibility and efficiency of sparse coding. Yaghoobi [7] has proposed the dictionary learning algorithm with the ℓ^1-norm based on the majorization method, which has been proven to be faster than K-SVD. Rakotomamonjy [12] has developed an algorithm called Dir with the ℓ^1-norm as the sparsity regularizer, which simultaneously optimizes the dictionary and the coefficient matrix in one stage based on a nonconvex proximal splitting framework, instead of alternating optimization. However, Dir is an extension of the gradient-based method which has slow convergence.

As far as we know, the good performance of sparse representation in various recognition tasks requires imposing some additional constraints on the dictionary [16–18]. One of such essential dictionary properties is the coherence between atoms. The coherence defined as the penalty term of $\| \mathbf{W}^T\mathbf{W} - \mathbf{I} \|_F$ imposed on the approximation error is studied in [16]. However, the dictionary is updated by the SVD method, which is costly. In addition, INK-SVD [17] uses the peak coherence to control the maximal correlation of different atoms in the dictionary. However, the corresponding problem is difficult to be optimized directly, since it becomes a min-max problem. An additional step after updating the dictionary is required to constrain the peak coherence of the dictionary, leading to increasing the complexity of the algorithm.

1.1. Motivation and contributions

In general, the ℓ^0-norm and ℓ^1-norm are used as sparsity constraints. The ℓ^1-norm has special significance because it is the convex proxy for sparsity and thus the corresponding optimization is relatively easy to be solved. Recently, the *log*-regularizer, which promotes sparsity more strongly and yields more accurate results in many sparse signal estimation and reconstruction problems than the ℓ^1-norm, has been proposed as sparsity constraint [19–21]. Our previous work [22] has studied dictionary learning with the *log*-regularizer for sparsity. The proximal operator was employed to obtain the closed-form solution of the coefficient vector directly and exactly. While the subproblems with respects to the dictionary atoms are quadratic and univariate functions, which are easy to be solved in the closed form. The log-regularized problem is nonconvex and nonsmooth. The corresponding optimal problem is complex. However, it can lead to further sparsity compared to the ℓ^1-norm, thereby resulting in superior performance, and it ensures that a lower approximation error is obtained, which is why it is attracting much attention.

In this study, we address the problem of learning an overcomplete dictionary with the *log*-regularizer for sparsity. Moreover, we propose to impose the ℓ^1-norm based term to the cost function as a penalty to control the degree of coherence in the dictionary, which is different from that of state-of-the-art algorithms. Hence, we propose a novel cost function, which is constructed by minimizing the approximation error with the *log*-regularizer on the coefficient matrix and with the coherence penalty on the dictionary atoms. Unfortunately, the overall problem is not only nonconvex but also nonsmooth due to the *log*-regularizer. State-of-the-art algorithms [8,9,12,15], different sparsity constraints imposed, are not suitable for our problem.

To solve the problem efficiently and explicitly, we propose a novel and fast algorithm which is effective as well as efficient, where it is crucial to handle nonconvexity and nonsmoothness of the problem in the procedure. We focus on single-vector variable problems instead of the overall problem with respect to the whole matrix. Hence, the overall problem can be transformed into a series of subproblems, each of which is with respect to either one dictionary atom or one coefficient vector. However, the subproblem with respect to the coefficient vector is nonsmooth and still nonconvex due to the *log*-regularizer. To our best knowledge, the proximal operator [23,24] can be treated as a powerful tool for solving nonsmooth and even nonconvex problem. We employ proximal operator to solve the subproblems with respect to the coefficient vectors and moreover the closed-form solutions of the subproblems can be obtained directly. In addition, the subproblems with respect to the dictionary atoms are also nonsmooth due to the coherence penalty. We can also apply proximal operator [13] to these problems and thus the closed-form solutions of the dictionary atoms are obtained explicitly. Hence, the overall problem can be solved efficiently and rapidly based on proximal operators, which we refer to as the proximal dictionary learning algorithm (PDLA). PDLA forces the coefficients to be as sparse as possible while simultaneously forcing the atoms of the dictionary to be as incoherent as possible.

We summarize the main contributions of this study as follows.

- We use the *log*-regularizer for sparsity in dictionary learning because it can induce sparsity strongly and ensure obtaining the low approximation errors. The proximal operator has been used to solve the signals recovery problem with the *log*-regularizer, where there is one unknown variable. To the best of our knowledge, there have been no similar treatments for dictionary learning problem where there are two unknown variables. Based on a decomposition scheme and alternating optimization, we transform the overall problem into a set of minimizations of the single-vector variable

UNCORRECTED PROOF

Please cite this article in press as: Z. Li et al., Incoherent dictionary learning with log-regularizer based on proximal operators, Digit. Signal Process. (2017), http://dx.doi.org/10.1016/j.dsp.2016.12.014

Download English Version:

https://daneshyari.com/en/article/4973821

Download Persian Version:

https://daneshyari.com/article/4973821

Daneshyari.com