Pattern Recognition Letters 80 (2016) 64–69

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A family of the subgradient algorithm with several cosparsity inducing functions to the cosparse recovery problem ☆

Guinan Wang [a], Hongjuan Zhang [a,*], Shiwei Yu [a], Shuxue Ding [b,**]

[a] Department of Mathematics, Shanghai University, Shanghai 200444, PR China
[b] Department of Computer Software, The University of Aizu, Tsuruga, Ikki-Machi, Aizu-Wakamatsu, Fukushima 965-8580, Japan

ARTICLE INFO

Article history:
Received 30 September 2015
Available online 27 May 2016

Keywords:
Cosparse analysis model
Cosparsity inducing function
Local linear approximation

ABSTRACT

In the past decade, there has been a great interest in the sparse synthesis model for signal. The researchers have obtained a series of achievements about the sparse representation. The cosparse analysis model as the corresponding version of the sparse synthesis model has drawn much attention in recent years. Many approaches have been proposed to solve this model. In some conventional general, these methods usually relaxed l_0-norm to l_1-norm or l_2-norm to represent the cospasity of signal, from which some reasonable algorithms have been developed. Furthermore, this work will present a new alternative way to replace the l_0-norm based on the cosparsity inducing function, which is closer to l_0-norm than l_1-norm and l_2-norm. Based on this function, we firstly construct the objective function and give a constrained optimal model of the cosparse recovery problem. Then we propose a subgradient algorithm – cosparsity inducing function (CIF) algorithm, which belongs to a two-layer optimization algorithm. Specifically, through converting the constrained optimal problem into the unconstrained case, we firstly obtain a temporary optimal variable, in which the cosparsity inducing function is approximated using its local linear approximation in order to avoid its nonconvex property. Secondly, a new cosupport is given by projecting the temporary optimal variable into the cosparse subspace and then keeping the l smallest elements. Besides, the desired signal is estimated using a conjugate gradient algorithm on the new cosupport. Moreover, we study the relative theoretical analysis about CIF algorithm. Simulations on the recovering of the unknown signal in the cosparse analysis model indicate its better performance at last.

1. Introduction

In recent years, signal models have drawn much attention and been successfully used for a variety of signal processing tasks such as denoising, deblurring and compressing sensing. To obtain the sparse representation for signals, the researchers are interested in the following problem: the signal of interest $\mathbf{x} \in \mathbb{R}^d$ is observable only through a set of linear measurements $\mathbf{y} \in \mathbb{R}^m$ and the observation matrix $\mathbf{M} \in \mathbb{R}^{m\times d} (m < d)$

$$\mathbf{y} = \mathbf{Mx} + \mathbf{e}, \tag{1}$$

where $\mathbf{e}$ is the additive noise that satisfies $\|\mathbf{e}\|_2 \le \varepsilon$. When $\mathbf{e} = 0$, Eq. (1) becomes $\mathbf{y} = \mathbf{Mx}$, it is the noiseless case. The aim of solving the problem (1) is to recover or approximate $\mathbf{x}$ from $\mathbf{y}$. Since this problem is a linear equation system with more unknowns than equations, i.e., under complete, generally, it is impossible. If $\mathbf{x}$ is known to be sparse in prior, the problem has been shown solvable. So far, there are two important signal models to solve the problem (1): the sparse synthesis model and the cosparse analysis model.

In the sparse synthesis model, the main optimization problem about the sparse representation is

$$\hat{\mathbf{x}} = \mathbf{D}\hat{\alpha} \text{ and } \hat{\alpha} = \underset{\alpha\in\mathbb{R}^n}{\operatorname{argmin}} \|\alpha\|_0, \text{ subject to } \|\mathbf{y} - \mathbf{Mx}\|_2 \leqslant \varepsilon, \tag{2}$$

where the signal $\mathbf{x} \in \mathbb{R}^d$ is assumed to be composed as linear combinations of a few atoms from a given dictionary $\mathbf{D} \in \mathbb{R}^{d\times n}$, which is overcomplete, i.e., $n > d$, such as $\mathbf{x} = \mathbf{D}\alpha$. The vector $\alpha \in \mathbb{R}^n$ is the sparse representation of $\mathbf{x}$, that is to say, $\boldsymbol{\alpha}$ contains few nonzeros elements, and the sparsity k is the number of nonzero elements in $\boldsymbol{\alpha}$, i.e. $\|\alpha\|_0 = k \ll d$ [1–3].

In the cosparse analysis model, the researchers often consider the following optimization problem to recover $\mathbf{x}$ from $\mathbf{y}$

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^d}{\operatorname{argmin}} \|\boldsymbol{\Omega}\mathbf{x}\|_0, \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{Mx}\|_2 \leqslant \varepsilon, \tag{3}$$

where $\boldsymbol{\Omega} \in \mathbb{R}^{p\times d} (p > d)$ is a fixed analysis operator. We can find that the goal of the problem (3) is to make the cosparse

☆ This paper has been recommended for acceptance by Dr. J. Yang.

* Corresponding author. Tel.: +86 2166135676.
** Corresponding author.
E-mail addresses: zhanghongjuan@shu.edu.cn (H. Zhang), sding@u-aizu.ac.jp (S. Ding).

http://dx.doi.org/10.1016/j.patrec.2016.05.012

representation vector $\mathbf{\Omega x}$ sparse. In other words, $\mathbf{\Omega x}$ contains many zeros. The cosparsity l is the number of zeros in $\mathbf{\Omega x}$, i.e. $l = p - \|\mathbf{\Omega x}\|_0$ $(0 \leqslant l \leqslant d)$, where $\|\mathbf{\Omega x}\|_0 \geqslant p - d$ [4].

Certainly, there are some connections and differences between the sparse synthesis model and the cosparse analysis model that has proved in some literatures [2,5,6]. Specially, these two models may become equivalent in the general case, when $\mathbf{D}$ is a square and invertible matrix, i.e. $\mathbf{D} = \mathbf{\Omega}^{-1}$, and $\mathbf{D} \in \mathbb{R}^{m\times n}$ is a full-rank matrix with $n < m$ for $\mathbf{D} = \mathbf{\Omega}^{+} = (\mathbf{\Omega}^T\mathbf{\Omega})^{-1}\mathbf{\Omega}^T$. While this seems like a perfect transfer from the analysis model to the synthesis model, it is in fact missing a key element. It simply states that $\mathbf{x}$ must reside in the range of $\mathbf{\Omega}$, i.e. $\mathbf{\Omega\Omega}^{+}\mathbf{x} = \mathbf{\Omega}(\mathbf{\Omega}^T\mathbf{\Omega})^{-1}\mathbf{\Omega}^T\mathbf{x} = \mathbf{x}$. Adding this as a constraint to the synthesis model, we get an exact equivalence, and otherwise, the synthesis model gets a larger number of degrees of freedom, and thus its minimum is deeper. In this work, we will concentrate on the cosparse analysis model.

As we all know, l_0-norm problem is generally NP-hard. From the previous works, we can find that there are some alternative ways to replace the l_0-norm. For example, the l_1-norm or l_2-norm due to their convexity [2,6,7]

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^d}{\operatorname{argmin}} \|\mathbf{\Omega x}\|_1, \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{Mx}\|_2 \leqslant \varepsilon \tag{4}$$

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^d}{\operatorname{argmin}} \|\mathbf{\Omega x}\|_2, \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{Mx}\|_2 \leqslant \varepsilon \tag{5}$$

At present, for these problems in the cosparse analysis model, the main works are not only the related algorithms to estimate or approximate the sparse representation of the possibly observed signal, but also the theoretical success guarantees for such algorithms [4,6,8].

Specifically, for the optimal model (4), Cai et al. introduce a split Bregman method which solves this model prevailingly using the Bregman iteration and provide the detailed convergence analysis [6]. The Bregman iteration converges very quickly when applied to certain types of objective functions, especially for problems involving an l_1 regularization term. So the Split Bregman method takes only a few steps of iterations to give good results for the cosparse analysis model.

In order to solve the optimal model (5), literature [4] gives a greedy algorithm termed “Greedy Analysis Pursuit” (GAP), which is an effective pursuit methods and similar to the Orthogonal Matching Pursuit (OMP)[2]. This approach updates the cosupports of the cosparse signals in a greedy fashion to find approximate solution under the noiseless condition. GAP recovers the signal perfectly in the relevant experiments. Besides, another kind of the greedy algorithm is also used to solve this model. The work presented in [8] describes a new family of greedy-like methods for the cosparse analysis model, including Analysis IHT (AIHT), Analysis HTP (AHTP), Analysis CoSaMP (ACoSaMP) and Analysis SP (ASP). These algorithms are the analysis versions of the synthesis counterpart approaches, i.e. Iterative Hard Thresholding (IHT), Hard Thresholding Pursuit (HTP), Compressive Sampling Matching Pursuit (CoSaMP), Subspace Pursuit (SP) [9–12]. When $\mathbf{x}$ is a low dimensional signal, the methods of AHTP, ACoSaMP, ASP need to solve the transformation of the problem (5), i.e. $\min_x \|\mathbf{y} - \mathbf{Mx}\|_2^2$ s.t.$\|\mathbf{\Omega}_\wedge \mathbf{x}\|_2^2 = 0$, where $\wedge$ is the cosupport. And for high dimensional signals, the model (5) is replaced by the following unconstrained minimization problem $\min_x \|\mathbf{y} - \mathbf{Mx}\|_2^2 + \lambda\|\mathbf{\Omega}_\wedge \mathbf{x}\|_2^2$ (λ is a relaxation constant). Meanwhile, literature [8] also provides performance grantees for these methods, which relied on a restricted isometry property (RIP) adapted to the context of the cosparse analysis model.

Although the l_1-norm and l_2-norm overcome the computational difficulty, the researchers still prefer to find the better substitutions than the popular l_1-norm and l_2-norm. For the problem (2), many of works related sparse recovery replaced the l_0-norm using l_q-norm $(0 < q < 1)$ and have proved that $l_q (0 < q < 1)$ minimization has better sparse recovery ability than l_1 minimization [13–21]. Besides, another novel idea of the sparsity inducing functions which used in the synthesis model (2) has proposed by Montefusco et al. [22], in which the sparse representation of recovering $\mathbf{x}$ from $\mathbf{y}$ under the noiseless situation be cast as

$$\min_{\mathbf{x}\in\mathbb{R}^d} \mathbf{\Phi}(\mathbf{x}), \quad \text{subject to} \ \ \mathbf{y} = \mathbf{Mx}, \tag{6}$$

where $\mathbf{\Phi}(\mathbf{x})$ is called the sparsity inducing function, which can be chosen l_q-norm$(0 < q < 1)$, atan function, log-sum function and so on. These sparsity inducing functions are more or less closely resembling the l_0-norm, which have been proved that the recovery effect is better by the related experiments in some other literatures [23–25]. Although these above functions are nonsmooth and nonconvex (concave), they also maintain some good properties of the l_1-norm, such as continuity and differentiability (for $\mathbf{x} \neq \mathbf{0}$).

In fact, there are some methods to resolve this nonsmooth and nonconvex optimization problem [22,26–30], which includes a first order approximation method, a neural network approach based on smoothing approximation and so on. Here, we mainly focus on the first order approximation method. As we all know, the good properties of the first order approximation method are easy to implement, and are easily obtained by exploiting the concavity of the function, which always lies below its tangent. And it is also shown that this method has ability to yield the best convex majorization of a concave objective function. At present, there are some representative works about the first order approximation method [22,26,27]. One of them is the local linear approximation (LLA) method proposed in [22,26], which is possible to transform the nonconvex constrained minimization problem into a convex unconstrained problem by inserting the local linear approximation in the context of a Lagrangian approach. In literature [27], there is another first order method named smoothing quadratic regularization (SQR) algorithm, which solves a strongly convex quadratic minimization problem with a diagonal Hessian matrix at each iteration.

Inspired by the literature [22,26,27], this work will concentrate on a more suitable relaxation of l_0-norm in the analysis model which may be differ from the l_1-norm and l_2-norm. One new substitution, named cosparsity inducing function, will be given to replace the l_0-norm, which is closer to it than l_1-norm and l_2-norm. The cosparsity inducing function includes the following forms, such as A-l_q, A-atan and A-log-sum function, whose details will be given in Section 2. Based on these functions, we firstly construct the objective function and give a constrained optimal model of the cosparse recovery problem. Then we propose a subgradient algorithm – cosparsity inducing function (CIF) algorithm, which belongs to a two-layer optimization algorithm. Specifically, through converting the constrained optimal problem into the unconstrained case, we firstly obtain a temporary optimal variable based on the gradient learning step, in which the cosparsity inducing function is approximated using its local linear approximation in order to avoid its nonconvex property. Secondly, a new cosupport is given by projecting the temporary optimal variable into the cosparse subspace and then keeping the l smallest elements. Finally, the desired signal is estimated using a conjugate gradient algorithm on the new cosupport. The CIF algorithm has a better recovery ability than some existing methods for solving the cosparse analysis problem, which will be indicated by the numerical experiments.

The manuscript is organized as follows. Section 2 will introduce the optimal model firstly, and give the cosparsity inducing functions including their expression, and their subgradients. Meanwhile, we will propose CIF algorithm and provide its main procedure. In Section 3, we will provide theoretical guarantees for the recovering performance of CIF algorithm. The numerical