Digital Signal Processing 26 (2014) 21–35

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Recovering non-negative and combined sparse representations

Karthikeyan Natesan Ramamurthy *, Jayaraman J. Thiagarajan, Andreas Spanias

SenSIP Center, School of ECEE, Arizona State University, Tempe, AZ 85287-5706, USA

ARTICLE INFO

Article history:
Available online 11 December 2013

Keywords:
Underdetermined linear system
Sparse representations
Non-negative representations
Orthogonal matching pursuit
Unique sparse solution

ABSTRACT

The non-negative solution to an underdetermined linear system can be uniquely recovered sometimes, even without imposing any additional sparsity constraints. In this paper, we derive conditions under which a unique non-negative solution for such a system can exist, based on the theory of polytopes. Furthermore, we develop the paradigm of combined sparse representations, where only a part of the coefficient vector is constrained to be non-negative, and the rest is unconstrained (general). We analyze the recovery of the unique, sparsest solution, for combined representations, under three different cases of coefficient support knowledge: (a) the non-zero supports of non-negative and general coefficients are known, (b) the non-zero support of general coefficients alone is known, and (c) both the non-zero supports are unknown. For case (c), we propose the combined orthogonal matching pursuit algorithm for coefficient recovery and derive the deterministic sparsity threshold under which recovery of the unique, sparsest coefficient vector is possible. We quantify the order complexity of the algorithms, and examine their performance in exact and approximate recovery of coefficients under various conditions of noise. Furthermore, we also obtain their empirical phase transition characteristics. We show that the basis pursuit algorithm, with partial non-negative constraints, and the proposed greedy algorithm perform better in recovering the unique sparse representation when compared to their unconstrained counterparts. Finally, we demonstrate the utility of the proposed methods in recovering images corrupted by saturation noise.

1. Introduction

We investigate the problem of recovering non-negative and combined sparse representations from underdetermined linear models. The system of linear equations with the constraint that the solution is non-negative can be expressed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha}, \quad \text{where } \boldsymbol{\alpha} \geqslant 0, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the data vector, $\boldsymbol{\alpha} \in \mathbb{R}^{K_x}$ is the non-negative solution (coefficient vector) and $\mathbf{X} \in \mathbb{R}^{M \times K_x}$ is the dictionary with $K_x > M$. When only a part of the solution is constrained to be non-negative and the rest is unconstrained (general), we obtain the combined representation model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{D}\boldsymbol{\beta}, \quad \text{where } \boldsymbol{\alpha} \geqslant 0. \tag{2}$$

Here, the coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^{K_d}$ is unconstrained, and $\mathbf{X} \in \mathbb{R}^{M \times K_x}$ and $\mathbf{D} \in \mathbb{R}^{M \times K_d}$ are the sub-dictionaries for the non-negative and general representations respectively. We denote the combined coefficient vector as $\boldsymbol{\delta} = [\boldsymbol{\alpha}^T \ \ \boldsymbol{\beta}^T]^T$, and the combined dictionary as $\mathbf{G} = [\mathbf{X} \ \mathbf{D}]$. We assume that $\mathbf{G}$ is overcomplete with $K_x + K_d > M$, and the columns of the dictionaries are normalized to have unit ℓ_2 norm. The sparsest solutions to (1) and (2) are obtained by minimizing the ℓ_0 norm, the number of non-zero elements, of the corresponding non-negative coefficient vector ($\boldsymbol{\alpha}$) or the combined coefficient vector ($\boldsymbol{\delta}$). In both the cases, the unique minimum ℓ_0 norm solution, when it exists, will be referred to as ML_0 solution. In this paper, we focus on obtaining deterministic guarantees for recovery of the ML_0 solutions to the linear systems (1) and (2), using both convex and greedy algorithms, based on the properties of the dictionaries.

1.1. Applications

Some of the applications of the non-negative representation model in (1), and the combined model in (2) are in image recovery [1], automatic speech recognition using exemplars [2], protein mass spectrometry [3], astronomical imaging [4], spectroscopy [5], source separation [6], clustering/semi-supervised learning of data [7,8], sparse portfolio optimization [9] to name a few. In particular, we will briefly mention two applications where the proposed combined model is directly relevant.

1.1.1. Signal/image recovery

Natural image patches can be sparsely represented using predefined and learned dictionaries and this property is used favorably

* Corresponding author. Fax: +1 (480) 965 8325.
E-mail address: knatesan@asu.edu (K. Natesan Ramamurthy).

1051-2004/$ – see front matter
http://dx.doi.org/10.1016/j.dsp.2013.11.003

in many image recovery applications such as denoising, inpainting, super-resolution and compressed sensing. When the representation of the image has two components, which are sparse in two distinct dictionaries, and when the sign of the coefficients in one of the dictionaries is known, the proposed combined sparse models can be used to recover the coefficients and hence the image itself. One such example application for the proposed model is demonstrated in Section 5.4, where we recover images corrupted by saturation noise. Another potential application is in compressed recovery of sparse signals, when the signs of a subset of the coefficient vector are known. The utility of the combined model in this application is illustrated in Sections 5.1, 5.2 and 5.3.

1.1.2. Sparse Markowitz portfolio optimization

In portfolio optimization, the goal is to select assets for a given capital that balances high returns with low risk. Recently, it has been proposed that this can be solved as a sparse optimization problem with appropriate constraints [9]. In this context, a negative coefficient corresponds to a *short-position* on the portfolio and a non-negative coefficient corresponds to a *no-short position*. When certain positions are mandated to be no-shorts because of possible government or market regulations, the combined model can be effectively used to select an optimal portfolio.

1.2. Prior work

For the non-negative representation model in (1), a sufficiently sparse ML_0 solution can be recovered by minimizing the ℓ_1 norm of $\boldsymbol{\alpha}$, using the non-negative version of the basis pursuit (BP) algorithm [10], which we refer to as NN-BP. The optimization program can be expressed as

$$\min_{\boldsymbol{\alpha}} \mathbf{1}^T\boldsymbol{\alpha} \quad \text{subject to } \mathbf{y} = \mathbf{X}\boldsymbol{\alpha},\ \boldsymbol{\alpha} \geqslant \mathbf{0}. \tag{3}$$

The conditions on $\mathbf{X}$ under which the recovery of ML_0 solution using (3) is possible have been derived based on the neighborliness of polytopes [11–13], and the non-negative null-space property [14]. A non-negative version of the greedy orthogonal matching pursuit (OMP) algorithm [15], which we will refer to as NN-OMP, for recovering the coefficients has also been proposed [16]. If the set

$$\{\boldsymbol{\alpha} | \mathbf{y} = \mathbf{X}\boldsymbol{\alpha},\ \boldsymbol{\alpha} \geqslant 0\} \tag{4}$$

contains only one solution, we can use any variational function instead of the ℓ_1 norm in order to obtain the unique non-negative solution [12,13,16]. In particular, the solution can be obtained by using the non-negative least squares (NNLS) algorithm [3,17].

A major part of our work investigates the combined sparse representation model introduced in (2), where only a part of the sparse coefficient vector is constrained to be non-negative. We consider the *deterministic sparsity thresholds*, i.e., the maximum number of non-zero coefficients possible in the ML_0 solution, such that the ML_0 solution can be uniquely recovered. To the best of our knowledge, such an investigation has not been reported so far in the literature. However, when both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unconstrained general sparse vectors, the sparsity thresholds for recovery of the ML_0 solution have been presented in [18,19]. By considering the coherence parameters of $\mathbf{X}$ and $\mathbf{D}$ separately, the authors in [18] show that an improvement up to a factor of two can be achieved in the deterministic sparsity threshold when compared to considering $\mathbf{X}$ and $\mathbf{D}$ together as a single dictionary. Note that deterministic sparsity thresholds provide guarantees that hold for all sparsity patterns and non-zero values in the coefficient vectors. *Probabilistic* or *robust* sparsity thresholds, that hold for most sparsity patterns and non-zero values in the coefficient vectors have also been derived in [18], again for the case where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are general sparse vectors. When this representation is approximately sparse and corrupted by additive noise, theory and algorithms for coefficient recovery are presented in [20].

1.3. Contributions

We present deterministic recovery guarantees for both the non-negative and the combined sparse representation models given by (1) and (2) respectively. Furthermore, we propose a greedy algorithm for performing coefficient recovery in combined representations and derive deterministic sparsity thresholds for unique recovery using ℓ_1 minimization and the proposed greedy algorithm.

For the non-negative model in (1), we derive the sufficient conditions for (4) to be singleton based on the neighborliness properties of the quotient polytope corresponding to the dictionary $\mathbf{X}$. Similar analyses reported in [12,13] assume that the dictionary $\mathbf{X}$ is obtained from a random ensemble and append a row of ones to it, such that the row span of $\mathbf{X}$ contains the vector $\mathbf{1}^T$. In contrast, we do not assume any randomness on $\mathbf{X}$ and only require that its row span intersects the positive orthant. We show that the sparsity threshold on $\boldsymbol{\alpha}$, for the set (4) to be singleton, is the same as the deterministic sparsity threshold for recovering the ML_0 solution of a general sparse representation. Whenever this threshold is satisfied, ℓ_1-norm regularization in (3) can be replaced with any variational function. Section 2 presents the analysis of the non-negative representation model.

For the combined model in (2), we propose a variant of the greedy OMP algorithm, the combined OMP (COMB-OMP) algorithm, for performing coefficient recovery. We also consider an ℓ_1 regularized convex algorithm, which we refer to as combined BP (COMB-BP). We derive the deterministic sparsity thresholds for recovering the ML_0 solution using both the COMB-BP and COMB-OMP algorithms. We show that a factor-of-two improvement in the sparsity threshold, observed when $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are general sparse vectors [18], holds for recovery using the COMB-BP also. We also show that such an improvement in the sparsity threshold cannot be observed using the COMB-OMP algorithm, because of the partial non-negativity constraint on the coefficient vector. Furthermore, we obtain the sparsity thresholds in the following cases of coefficient support knowledge: (a) the non-zero support of both $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ is known, and (b) non-zero support of $\boldsymbol{\beta}$ alone is known. When analyzing case (b), we factor out the contribution of the general representation component and arrive at conditions under which ℓ_1-norm regularization in the resulting optimization can be replaced with any variational function for the recovery of $\boldsymbol{\alpha}$. Section 3 presents all the details in the analysis of the combined representation model. As a final piece of our theoretical investigation, we present the computational complexities of OMP, COMB-OMP, BP and COMB-BP algorithms.

The performance of the COMB-BP and the COMB-OMP algorithms is also analyzed using simulations. The dictionary $\mathbf{G}$ is obtained from a Gaussian ensemble and the non-zero coefficients are obtained either from a uniform distribution or from a random sign distribution (± 1). It is shown that both COMB-BP and COMB-OMP respectively perform better than their unconstrained counterparts, the BP and the OMP, particularly as the K_x becomes larger. The algorithms show a similar behavior when recovering the sparse coefficients from noisy signals. Furthermore, the empirical phase transition characteristics of the proposed algorithms are provided and their utility in a real-world application of recovering images from saturation noise is demonstrated.

1.4. Notation

Lowercase boldface letters denote column vectors and uppercase boldface denote matrices, e.g., $\mathbf{a}$ and $\mathbf{A}$ denote a vector and