ARTICLE IN PRESS
JID:YDSPR AID:2224 /FLA [m5G; v1.224; Prn:8/11/2017; 13:18] P.1 (1-10)
Digital Signal Processing ••• (••••) •••–•••

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

On joint optimization of sensing matrix and sparsifying dictionary for robust compressed sensing systems ☆

Gang Li [a], Zhihui Zhu [b,*], Xinming Wu [c], Beiping Hou [a]

[a] School of Automation & Electrical Engineering, Zhejiang University of Science & Technology, Hangzhou, 310023, Zhejiang, PR China
[b] Department of Electrical Engineering, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401, USA
[c] University of Texas at Austin, Bureau of Economic Geology, Austin, TX, 78713-8924, USA

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
Maximum likelihood estimation
Compressed sensing
Sparse representation error
Signal and image compression

A B S T R A C T

This paper deals with joint design of sensing matrix and sparsifying dictionary for compressed sensing (CS) systems. Based on the maximum likelihood estimation (MLE) principle, a preconditioned signal recovery (PSR) scheme and a novel measure are proposed. Such a measure allows us to optimize the sensing matrix and dictionary jointly. An alternating minimization-based iterative algorithm is derived for solving the corresponding optimal design problem. Simulation and experiments, carried with synthetic data and real image signals, show that the PSR scheme and the CS system, obtained using the proposed approaches, outperform the prevailing ones in terms of reducing the effect of sparse representation errors.

1. Introduction

Compressed sensing (CS) has attracted a lot of attention in the signal processing community since its appearance in the early 2000s [1–3]. Mathematically, CS is a framework that involves compressing a signal vector $x \in \Re^{N\times 1}$ into a low dimensional one $y \in \Re^{M\times 1}$ $(M \ll N)$ via

$$y = \Phi x \tag{1}$$

and reconstructing the original signal x from the measurement y. The matrix $\Phi \in \Re^{M\times N}$ is called a *sensing/projection* matrix.

Let $x \in \Re^{N\times 1}$ be modeled as a linear combination of a set of vectors $\{\psi_l\}_{l=1}^{L}$ [1]:

$$x = \sum_{l=1}^{L} s(l)\psi_l \triangleq \Psi s \tag{2}$$

where the matrix $\Psi \in \Re^{N\times L}$ is usually called a *dictionary* and is said *over-complete* if $N < L$ and s is the coefficient vector. We say x is κ-sparse (in Ψ) if $\|s\|_0 \le \kappa$, where $\|s\|_0$ denotes the number of non-zero elements in s.

A CS system refers to equations (1)–(2), characterized with the sensing matrix Φ and dictionary Ψ. Its ultimate goal is to reconstruct the original signal x from y that senses the former via (1). Traditionally, the reconstructed signal is given by $\hat{x} = \Psi\hat{s}$ with $\hat{s}$ being a proper solution of the following problem

$$y = As \tag{3}$$

where $A \triangleq \Phi\Psi$ is called the *equivalent* dictionary.

As $M < N$, equation (3) has an infinite number of solutions. To make the above equation have a unique solution, extra properties of this linear system have to be enforced and the concept of *spark* is one of such properties. The spark of a matrix $A \in \Re^{M\times L}$, denoted as *spark*(A), is defined as the smallest number of columns in A that are linearly dependent. It was shown in [4] that as long as *spark*$(A) > 2\kappa$, any κ-sparse signal $x_0 = \Psi s_0$ can be exactly recovered from its measurement $y = \Phi x_0$ by solving

$$s_0 = \arg\min_{s} \|s\|_0 \quad \text{s.t.} \quad y = As \tag{4}$$

or

$$s_0 = \arg\min_{s} \|y - As\|_2^2 \quad \text{s.t.} \quad \|s\|_0 \le \kappa \tag{5}$$

where $\|.\|_p$ denotes the l_p-norm of vector $v \in \Re^{N\times 1}$ and is defined as

☆ This work was supported by the Grants of NSFC 61273195, 61304124, and 61473262.

* Corresponding author.
E-mail addresses: ieligang@zjut.edu.cn (G. Li), zzhu@mines.edu (Z. Zhu), xinming.wu@beg.utexas.edu (X. Wu).

[1] Throughout this paper, MATLAB notations are adopted: $Q(m,:)$, $Q(:,k)$ and $Q(i,j)$ denote the mth row, kth column, and (i,j)th entry of the matrix Q; $q(n)$ denotes the nth entry of the vector q.

https://doi.org/10.1016/j.dsp.2017.10.023
1051-2004/

Please cite this article in press as: G. Li et al., On joint optimization of sensing matrix and sparsifying dictionary for robust compressed sensing systems, Digit. Signal Process. (2017), https://doi.org/10.1016/j.dsp.2017.10.023

$$\|v\|_p \triangleq (\sum_{n=1}^{N} |v(n)|^p)^{1/p}, \quad p \geq 1 \tag{6}$$

In general, both of the problems (4) and (5) are NP-hard. The problem defined in (5) is practically addressed using greedy algorithms such as the orthogonal matching pursuit (OMP) technique [5–9]. Furthermore, it can be shown [4] that under some conditions, (4) is equivalent to the following l_1-based minimization

$$s_0 = \arg\min_s \|s\|_1 \quad \text{s.t.} \quad y = As \tag{7}$$

which can be solved efficiently using algorithms such as basis pursuit (BP) [3] and the l_1/l_2-based optimization techniques [10].

1.1. Related works

Designing optimal CS systems usually refers to determine a pair (Ψ, Φ) such that the corresponding CS system yields a desired performance in terms of signal compression and signal recovery. Such performance depends strongly on the properties of Ψ and Φ.

As mentioned above, the spark of the equivalent dictionary is one of the properties for exact reconstruction. The restricted isometry property (RIP) [2], [3] is another one. A matrix A is said (κ, δ)-RIP if there exists a δ with $0 \leq \delta < 1$ such that

$$(1-\delta)\|s\|_2^2 \leq \|As\|_2^2 \leq (1+\delta)\|s\|_2^2$$

holds for all s satisfying $\|s\|_0 \leq \kappa$. It has been shown that when $A = \Phi\Psi$ is $(2\kappa, \delta)$-RIP, a κ-sparse x in Ψ can be reconstructed exactly from its low dimensional measurement [2–4]. Furthermore, as shown in [4], [5], any κ-sparse coefficient vector s can be exactly obtained from $y = As$ as long as

$$\kappa < \frac{1}{2}[1 + \frac{1}{\mu(A)}] \tag{8}$$

where $\mu(A)$ is the mutual coherence of matrix A and is defined as

$$\mu(A) \triangleq \max_{1 \leq i \neq j \leq L} \{ \frac{|< A(:,i), A(:,j) >|}{\|A(:,i)\|_2 \|A(:,j)\|_2} \triangleq r_{ij} \} \tag{9}$$

where $< ., . >$ and r_{ij} denote the inner product and the cross-correlation factor between two vectors, respectively. Roughly speaking, $\mu(A)$ measures the maximum linear dependency possibly achieved by any two columns of matrix A.

The relation specified by (8) suggests that the equivalent dictionary with small mutual coherence can enlarge the signal space in which the coefficient vector s can be achieved exactly. The optimal sensing matrix design, initialized in [11], deals with how to design the sensing matrix Φ with a dictionary Ψ given such that the CS system yields an accurate reconstruction of signals. This can be achieved via choosing Φ to enhance the mutual coherence property of the equivalent dictionary A. A class of approaches under this framework can be unified as

$$\begin{aligned} \tilde{\Phi} &\triangleq \arg\min_{\Phi, H} \|H - A^{\mathcal{T}} A\|_F^2 \\ &\quad \text{s.t.} \ H \in \mathcal{S}_H, \ A = \Phi\Psi \end{aligned} \tag{10}$$

where $\|.\|_F$ denotes the *Frobenius* norm and $\mathcal{S}_H$ is a non-empty set of target Gram matrices of desired mutual coherence with $H(k,k) = 1, \forall\, k$. See [12–22].

In (10), the target Gram H has its diagonal elements all equal to one and the Gram $A^{\mathcal{T}} A$ is directly related to the coherence behavior of A if and only if $\|A(:,l)\|_2 = 1, \ \forall\, l$. In order to make the cost function have the designated physical meaning, (10) was studied recently in [23] with an additional constraint: $\|A(:,l)\|_2 = 1, \ \forall\, l$, and the corresponding problem was attacked using a gradient descent algorithm.

Dictionary design is to find a dictionary to represent a class of signals for a given sparsity level κ. Typical examples include the Fourier matrix for frequency-sparse signals, a multiband modulated Discrete Prolate Spheroidal Sequences (DPSS's) dictionary for sampled multiband signals [24,25], and learning a sparsifying dictionary from a training dataset. Let $X \in \Re^{N \times J}$ with $X(:,j) = x_j$ be the data matrix formed by a collection of training samples $\{x_j\}_{j=1}^J$ from a certain class of signals. The traditional dictionary learning is to solve the following problem

$$\begin{aligned} \{\tilde{\Psi}, \ \tilde{S}\} &\triangleq \arg\min_{\Psi, S} \|X - \Psi S\|_F^2 \\ \text{s.t.} \quad & \|S(:,j)\|_0 \leq \kappa, \ \forall\, j \\ & \|\Psi(:,l)\|_2 = 1, \ \forall\, l \end{aligned} \tag{11}$$

where the normalization constraint on dictionary is mainly for avoiding degenerate solutions as the solutions to minimizing $\|X - \Psi S\|_F^2$ *w.r.t.* Ψ and S are not unique. A practical approach used to attack such a highly non-convex problem is based on the alternating minimization strategy [26–33].

1.2. Problems to be investigated

It should be pointed out that the sensing matrices obtained using the classical approaches (10) do yield a very good performance when the signals to be compressed are exactly sparse in a dictionary Ψ. A more practical signal model is

$$x = \Psi s + e \tag{12}$$

where e is the representation error, which is practically not nil in general. In that case, the CS system using the sensing matrix designed based on (10) usually fails in resulting an accurate reconstruction. See Section 4.2.2 and also [35,36] for a typical example in image compression. The main reason for this phenomenon is due to the fact that the classical approaches (10) to optimal sensing matrix design do not take into account of priori information (such as sparse representation error) of the signals.

It has been noted that in most of the existing works on design of optimal CS systems, the sensing matrix and sparsifying dictionary are designed independently. From the reconstruction equation $y = As$, one can see that the performance of a CS system is determined by the properties of the equivalent dictionary $A = \Phi\Psi$ and hence can be enhanced by designing the two Φ and Ψ jointly in one and the same framework. As far as we know, there have been a few works reported on this topic. The very first piece of work closely related to this topic was perhaps given by Duarte-Cavajalino et al. [13]. An improved work was reported in [14], in which the same framework as that in [13] is used but both sensing matrix and dictionary are updated using analytical solutions. However, both approaches in [13] and [14] alternatively update the sensing matrix and the dictionary with different measures rather than under the same criterion.

The main problem to be considered in this paper is to investigate how to learn both sensing matrix and dictionary jointly in one and the same framework and measure (a.k.a. criterion) using a set of training samples. Our contributions in this paper are highlighted in next subsection.

1.3. Contributions

- Based on the maximum likelihood estimation (MLE) principle [34], an alternative signal reconstruction scheme, called pre-conditioned signal recovery (PSR), is derived and a new measure is proposed, which allows us to optimize the sensing matrix and the dictionary simultaneously. Unlike [13] and [14],