Signal Processing 143 (2018) 200–210

Contents lists available at ScienceDirect

Signal Processing

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

# An efficient method for robust projection matrix design

Tao Hong[a,*], Zhihui Zhu[b]

[a] Department of Computer Science, Technion - Israel Institute of Technology, Haifa, 32000, Israel
[b] Department of Electrical Engineering, Colorado School of Mines, Golden, CO 80401 USA

ARTICLE INFO

*Article history:*
Received 19 January 2017
Revised 3 September 2017
Accepted 6 September 2017
Available online 7 September 2017

*Keywords:*
Robust projection matrix
Sparse representation error (SRE)
High dimensional dictionary
Mutual coherence

ABSTRACT

Our objective is to efficiently design a robust projection matrix $\boldsymbol{\Phi}$ for the Compressive Sensing (CS) systems when applied to the signals that are not exactly sparse. The optimal projection matrix is obtained by mainly minimizing the average coherence of the equivalent dictionary. In order to drop the requirement of the sparse representation error (SRE) for a set of training data as in [15,16], we introduce a novel penalty function independent of a particular SRE matrix. Without requiring of training data, we can efficiently design the robust projection matrix and apply it for most of CS systems, like a CS system for image processing with a conventional wavelet dictionary in which the SRE matrix is generally not available. Simulation results demonstrate the efficiency and effectiveness of the proposed approach compared with the state-of-the-art methods. In addition, we experimentally demonstrate with natural images that under similar compression rate, a CS system with a learned dictionary in high dimensions outperforms the one in low dimensions in terms of reconstruction accuracy. This together with the fact that our proposed method can efficiently work in high dimension suggests that a CS system can be potentially implemented beyond the small patches in sparsity-based image processing.



## 1. Introduction

Since the beginning of this century, Compressive Sensing or Compressed Sensing (CS) has received a great deal of attention [1–6]. Generally speaking, CS is a mathematical framework that addresses accurate recovery of a signal vector $\boldsymbol{x} \in \Re^N$ from a set of linear measurements

$$\boldsymbol{y} = \boldsymbol{\Phi x} \in \Re^M \tag{1}$$

where $M \ll N$ and $\boldsymbol{\Phi} \in \Re^{M \times N}$ is referred to as the projection or sensing matrix. CS has found many applications in the areas such as image processing, machine learning, pattern recognition, signal detection/classification etc. We refer the reader to [5,6] and the references therein to find the related topics mentioned above.

Sparsity and coherence are two important concepts in CS theory. We say a signal $\boldsymbol{x}$ of interest approximately sparse (in some basis or dictionary) if we can approximately express it as a linear combination of few columns (also called atoms) from a well-chosen dictionary:

$$\boldsymbol{x} = \boldsymbol{\Psi\theta} + \boldsymbol{e} \tag{2}$$

where $\boldsymbol{\Psi} \in \Re^{N \times L}$ is the given or determined dictionary, $\boldsymbol{\theta} \in \Re^L$ is a sparse coefficient vector with few non-zero elements, and $\boldsymbol{e} \in \Re^N$ stands for the sparse representation error (SRE). In particular, the vector $\boldsymbol{x}$ is called (purely or exactly) $K$-sparse in $\boldsymbol{\Psi}$ if $\|\boldsymbol{\theta}\|_0 = K$ and $\boldsymbol{e} = \boldsymbol{0}$ and approximately $K$-sparse in $\boldsymbol{\Psi}$ if $\|\boldsymbol{\theta}\|_0 = K$ and $\boldsymbol{e}$ has relatively small energy. Here, $\|\boldsymbol{\theta}\|_0$ denotes the number of non-zero elements in $\boldsymbol{\theta}$ and $\boldsymbol{0}$ represents a vector whose entries are equivalent to 0. Through this paper, we say $\boldsymbol{\theta}$ is $K$-sparse if $\|\boldsymbol{\theta}\|_0 = K$ regardless whether $\boldsymbol{e} = \boldsymbol{0}$.

Substituting the sparse model (2) of $\boldsymbol{x}$ into (1) gives

$$\boldsymbol{y} = \boldsymbol{\Phi\Psi\theta} + \boldsymbol{\Phi e} \triangleq \boldsymbol{D\theta} + \boldsymbol{\Phi e} \tag{3}$$

where the matrix $\boldsymbol{D} = \boldsymbol{\Phi\Psi}$ is referred to as the equivalent dictionary of the CS system and $\boldsymbol{\epsilon} \triangleq \boldsymbol{\Phi e}$ denotes the projection noise caused by SRE. The goal of a CS system is to retrieve $\boldsymbol{\theta}$ (and hence $\boldsymbol{x}$) from the measurements $\boldsymbol{y}$. Due to the fact that $M \ll L$, solving $\boldsymbol{y} \approx \boldsymbol{D\theta}$ for $\boldsymbol{\theta}$ is an undetermined problem which has an infinite number of solutions. By utilizing the priori knowledge that $\boldsymbol{\theta}$ is sparse, a CS system typically attempts to recover $\boldsymbol{\theta}$ by solving the following problem:

$$\boldsymbol{\theta} = \arg\min_{\tilde{\boldsymbol{\theta}}} \|\tilde{\boldsymbol{\theta}}\|_0, \ \text{s.t.} \ \|\boldsymbol{y} - \boldsymbol{D}\tilde{\boldsymbol{\theta}}\|_2 \le \|\boldsymbol{\epsilon}\|_2 \tag{4}$$

which can be solved by many efficient numerical algorithms including basis pursuit (BP), orthogonal matching pursuit (OMP), least absolute shrinkage and selection operator (LASSO) etc. All of the methods can be found in [5,7] and the references therein.

To ensure exact recovery of $\boldsymbol{\theta}$ through (4), we need certain conditions on the equivalent dictionary $\boldsymbol{D}$. One of such conditions is

* Corresponding author.
*E-mail addresses:* hongtao@cs.technion.ac.il (T. Hong), zzhu@mines.edu (Z. Zhu).

http://dx.doi.org/10.1016/j.sigpro.2017.09.007
0165-1684/

related to the concept of *mutual coherence*. The mutual coherence of a matrix $\boldsymbol{D} \in \Re^{M\times L}$ is denoted by

$$\mu(\boldsymbol{D}) \triangleq \max_{1\le i\ne j\le L} |\bar{\boldsymbol{G}}(i,j)| \tag{5}$$

where $\bar{\boldsymbol{G}} = \bar{\boldsymbol{D}}^{\mathcal{T}}\bar{\boldsymbol{D}}$ is called the Gram matrix of $\bar{\boldsymbol{D}} = \boldsymbol{D}\boldsymbol{S}_{sc}$ with $\boldsymbol{S}_{sc}$ a diagonal scaling matrix such that each column of $\bar{\boldsymbol{D}}$ is of unit length. Here $^{\mathcal{T}}$ represents the transpose operator. It is known that $\mu(\boldsymbol{D})$ is lower bounded by the Welch bound $\underline{\mu}(\boldsymbol{D}) = \sqrt{\frac{L-M}{M(L-1)}}$, i.e., $\mu(\boldsymbol{D}) \in [\sqrt{\frac{L-M}{M(L-1)}}, 1]$. The mutual coherence $\mu(\boldsymbol{D})$ measures the worst-case coherence between any two columns of $\boldsymbol{D}$ and is one of the fundamental quantities associated with the CS theory. As shown in [5], when there is no projection noise (i.e., $\boldsymbol{\epsilon} = 0$), any $K$-sparse signal $\boldsymbol{\theta}$ can be exactly recovered by solving the linear system (4) as long as

$$K < \frac{1}{2}\left[1 + \frac{1}{\mu(\boldsymbol{D})}\right] \tag{6}$$

which indicates that a smaller $\mu(\boldsymbol{D})$ ensures a CS system to recover the signal with a larger $K$. Thus, Barchiesi et al. and the following workers [8,9] proposed methods to design a dictionary with small mutual coherence. For a given dictionary $\boldsymbol{\Psi}$, the mutual coherence of the equivalent dictionary is actually determined or controlled by the projection matrix $\boldsymbol{\Phi}$. So it would be of great interest to design $\boldsymbol{\Phi}$ such that $\mu(\boldsymbol{D})$ is minimized. Another similar indicator used to evaluate the average performance of a CS system is named *average* mutual coherence $\mu_{av}$. The definition of $\mu_{av}$ is given as follows:

$$\mu_{av}(\boldsymbol{D}) \triangleq \frac{\sum_{\forall(i,j)\in S_{av}} |\bar{\boldsymbol{G}}(i,j)|}{N_{av}}$$

where $S_{av} \triangleq \{(i,j) : \bar{\mu} \le |\bar{\boldsymbol{G}}(i,j)|\}$ with $0 \le \bar{\mu} < 1$ as a prescribed parameter and $N_{av}$ is the number of components in the index set $S_{av}$.

There has been much effort [10–14] devoted to designing an optimal $\boldsymbol{\Phi}$ that outperforms the widely used random matrix in terms of signal recovery accuracy (SRA). However, all these methods are based on the assumption that the signal is exactly sparse under a given dictionary, which is not true for practical applications. It is experimentally observed that the sensing matrix designed by Elad and the following workers [10–14] based on mutual coherence results in inferior performance for real images (which are generally approximately but not exactly sparse under a well-chosen dictionary). To address this issue, the recent work in [15,16] proposed novel methods to design a robust projection matrix when the SRE exists.[1] Through this paper, similar to what is used in [15,16], a robust projection (or sensing) matrix means it is designed with consideration of possible SRE and hence the corresponding CS system yields superior performance when the SRE $\boldsymbol{e}$ in (2) is not nil. However, the approaches in [15,16] need the explicit value of the SRE on the training dataset, making them inefficient in several aspects. First, many practical CS systems with predefined analytical dictionaries (e.g., the wavelet dictionary, and the modulated discrete prolate spheroidal sequences (DPSS) dictionary for sampled multiband signals [17]) actually do not involve any training dataset and hence no SRE available. In order to design the robust projection matrix for these CS systems using the framework presented in [15,16], one has to first construct plenty of extra representative dataset for the explicit SRE with the given dictionary, which limits the range of applications. Second, even for the CS system with a dictionary learned typically on a large-scale dataset, we need a lot of memories and computations to store and compute with the huge dataset as well its corresponding SRE for designing a robust sensing matrix. Moreover, if the CS system is applied to a dynamic dataset, e.g., video stream, it is practically impossible to store all the data and compute its corresponding SRE. Therefore, the requirement of the explicit value of SRE for the training dataset makes the methods in [15,16] limited and inefficient for all the cases discussed above.

In this paper, to drop the requirement of the training dataset as well as its SRE, we propose a novel robust projection matrix framework only involving a predefined dictionary. With this new framework, we can efficiently design projection matrices for the CS systems mentioned above. We stress that by efficient method for robust projection matrix design (which is the title of this paper), we are not providing an efficient method for solving the problems in [15,16]; instead we provide a new framework in which the training dataset and its corresponding SRE are not required any more. Experiments on synthetic data and real images demonstrate the proposed sensing matrix yields a comparable performance in terms of SRA compared with the ones obtained by Li and Hong et al. [15,16].

Before proceeding, we first briefly introduce some notation used throughout the paper. MATLAB notations are adopted in this paper. In this connection, for a vector, $\boldsymbol{v}(k)$ denotes the $k$th component of $\boldsymbol{v}$. For a matrix, $\boldsymbol{Q}(i,j)$ means the $(i,j)$th element of matrix $\boldsymbol{Q}$, while $\boldsymbol{Q}(k,:)$ and $\boldsymbol{Q}(:,k)$ indicate the $k$th row and column vector of $\boldsymbol{Q}$, respectively. We use $\boldsymbol{I}$ and $\boldsymbol{I}_L$ to denote an identity matrix with arbitrary and $L\times L$ dimension, respectively. The $k$th column of $\boldsymbol{Q}$ is also denoted by $\boldsymbol{q}_k$. trace($\boldsymbol{Q}$) denotes the calculation of the trace of $\boldsymbol{Q}$. The Frobenius norm of a given matrix $\boldsymbol{Q}$ is $\|\boldsymbol{Q}\|_F = \sqrt{\sum_{i,j}\|\boldsymbol{Q}(i,j)\|^2} = \sqrt{\text{trace}(\boldsymbol{Q}^{\mathcal{T}}\boldsymbol{Q})}$ where $\mathcal{T}$ represents the transpose operator. The definition of $l_p$ norm for a vector $\boldsymbol{v} \in \Re^N$ is $\|\boldsymbol{v}\|_p \triangleq (\sum_{k=1}^N |\boldsymbol{v}(k)|^p)^{\frac{1}{p}}, \quad p \ge 1$.

The remainder is arranged as follows. Some preliminaries are given in Section 2 to state the motivation of developing such a novel model. The proposed model which does not need the SRE is shown in Section 3 and the corresponding optimal sensing problem is solved in this section. The synthetic and real data experiments are carried out in Section 4 to demonstrate the efficiency and effectiveness of the proposed method. Some conclusions are given in Section 5 to end this paper.

## 2. Preliminaries

A sparsifying dictionary $\boldsymbol{\Psi}$ for a given dataset $\{\boldsymbol{x}_k\}_{k=1}^P$ is usually obtained by considering the following problem

$$\{\boldsymbol{\Psi}, \boldsymbol{\theta}_k\} = \arg\min_{\tilde{\boldsymbol{\Psi}},\tilde{\boldsymbol{\theta}}_k} \sum_{k=1}^P \|\boldsymbol{x}_k - \tilde{\boldsymbol{\Psi}}\tilde{\boldsymbol{\theta}}_k\|_2^2 \quad \text{s.t.} \quad \|\tilde{\boldsymbol{\theta}}_k\|_0 \le K \tag{7}$$

which can be addressed by some practical algorithms [18], among which the popularly utilized are the K-singular value decomposition (K-SVD) algorithm [19] and the method of optimal direction (MOD) [20]. As stated in the previous section, the SRE $\boldsymbol{e}_k = \boldsymbol{x}_k - \boldsymbol{\Psi}\boldsymbol{\theta}_k$ is generally not nil. We concatenate all the SRE $\{\boldsymbol{e}_k\}$ into an $N\times P$ matrix:

$$\boldsymbol{E} \triangleq \boldsymbol{X} - \boldsymbol{\Psi}\boldsymbol{\Theta}$$

which is referred to as the SRE matrix corresponding to the training dataset $\{\boldsymbol{x}_k\}$ and the learned dictionary $\boldsymbol{\Psi}$.

The recent work in [15,16] attempted to design a robust projection matrix with consideration of the SRE matrix $\boldsymbol{E}$ and proposed to solve

$$\boldsymbol{\Phi} = \arg\min_{\tilde{\boldsymbol{\Phi}}} \|\boldsymbol{I}_L - \boldsymbol{\Psi}^{\mathcal{T}}\tilde{\boldsymbol{\Phi}}^{\mathcal{T}}\tilde{\boldsymbol{\Phi}}\boldsymbol{\Psi}\|_F^2 + \lambda\|\tilde{\boldsymbol{\Phi}}\boldsymbol{E}\|_F^2 \tag{8}$$

or

$$\boldsymbol{\Phi} = \arg\min_{\tilde{\boldsymbol{\Phi}},\boldsymbol{G}\in H_\xi} \|\boldsymbol{G} - \boldsymbol{\Psi}^{\mathcal{T}}\tilde{\boldsymbol{\Phi}}^{\mathcal{T}}\tilde{\boldsymbol{\Phi}}\boldsymbol{\Psi}\|_F^2 + \lambda\|\tilde{\boldsymbol{\Phi}}\boldsymbol{E}\|_F^2 \tag{9}$$

[1] We note that the approaches considered in [15,16] share the same framework. The difference is that in [16] Hong et al. utilized an efficient iterative algorithm giving an approximate solution, while a closed form solution is derived in [15].