Digital Signal Processing 73 (2018) 62–71

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

ARTICLE INFO

Article history:
Available online 6 November 2017

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

ABSTRACT

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 $M < L$ is assumed, 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), houbeiping@126.com (B. Hou).

[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/

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

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

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

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

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

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],

where the sensing matrix and dictionary are updated alternatively with different measures, our proposed approach allows us to optimize the sensing matrix and the dictionary jointly with one and the same measure. As such a design procedure is data-driven and it takes the representation error into account, the obtained CS system is expected to be more robust against the sparse representation errors and hence yields a more accurate reconstruction.
- An alternating minimization-based iterative procedure is proposed, in which the sparse coefficients, dictionary and sensing matrix are updated one by one. A closed form solution is obtained for updating the sensing matrix. Analysis shows that our proposed algorithm is convergent under some mild condition, while this is not the case for the iterative procedures in [13], [14]. Further elaborations will be given on this issue in Section 2.2.

The outline of this paper is given as follows. Based on the maximum likelihood estimation principle, the PSR scheme is derived and a novel measure is proposed in Section 2, which allows us to investigate the joint optimization in the same framework. In Section 3, an alternating minimization-based algorithm is derived to solve the corresponding optimal design problem. Section 4 is devoted to presenting simulations and experimental results, which show that the proposed PSR is superior to the traditional reconstruction scheme and the CS system obtained using the proposed approach outperforms the prevailing ones in terms of reducing the effect of sparse representation errors. Some concluding remarks are given in Section 5 to end this paper.

2. An MLE-based framework for designing optimal CS systems

Let $y = \Phi x$ be the projection of x given by (12) and clearly,

$$y = \Phi\Psi s + \Phi e \triangleq As + \epsilon \tag{13}$$

where ϵ is the representation error of y in the equivalent dictionary A.

2.1. Pre-conditioned signal recovery (PSR) scheme

First of all, let us consider the signal reconstruction from a statistical point of view. Assume that e is a normal distribution $\mathcal{N}(\mu, C)$, that is the probability density function (PDF) of e obeys

$$f_e(\xi) = \frac{1}{\sqrt{2\pi \det(C)}} e^{-\frac{1}{2}(\xi-\mu)^{\mathcal{T}} C^{-1}(\xi-\mu)}$$

where det(.) denotes the determinant operator. In the sequel, $\mu = \mathbf{0}$, $C = \sigma_0^2 I_N$ is assumed, where I_N denotes the identity matrix of dimension N.

Let $x = \Psi s + e$ and the measurement y be given by $y = \Phi x$. What is the best estimate of s that can be achieved from the observation y?

Denote $P^{1/2}$ as a square root (symmetric) matrix of a (semi-) positive definite matrix P and assume that $\Phi\Phi^{\mathcal{T}}$ is positive definite. Note that the PDF of ϵ, which has a multivariate normal distribution with $\mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$, is given by

$$f_\epsilon(\xi) = \frac{e^{\left(-\frac{1}{2}\xi^{\mathcal{T}}\Sigma_\epsilon^{-1}\xi\right)}}{\sqrt{2\pi \det(\Sigma_\epsilon)}}$$

where $\Sigma_\epsilon = \sigma_0^2 \Phi\Phi^{\mathcal{T}}$. According to the maximum likelihood estimation (MLE) principle, which states that for a given sample of random variable, the best estimation of the parameter(s) characterizing the PDF of the random variable can be achieved by maximizing the PDF corresponding to the sample given [34], the best estimate $\hat{s}$ of the sparse coefficient vector s is the one that maximizes the PDF $f_\epsilon(\xi)$ at $\xi = y - As$. This leads to

$$\hat{s} \triangleq \arg\max_s -(y - As)^{\mathcal{T}}(\Phi\Phi^{\mathcal{T}})^{-1}(y - As) \quad \text{s.t.} \quad \|s\|_0 \le \kappa$$

which is equivalent to

$$\hat{s} \triangleq \arg\min_s \|(\Phi\Phi^{\mathcal{T}})^{-1/2}(y - As)\|_2^2 \quad \text{s.t.} \quad \|s\|_0 \le \kappa \tag{14}$$

Clearly, (14) is an alternative signal recovery scheme to the traditional one by (5). To see the connection between (14) and (5), we rewrite the objective function in (14) as

$$\|(\Phi\Phi^{\mathcal{T}})^{-1/2}(y - As)\|_2^2 = \|(\Phi\Phi^{\mathcal{T}})^{-1/2}y - (\Phi\Phi^{\mathcal{T}})^{-1/2}As\|_2^2$$

which has the same form as the objective function in (5) if we denote by $y_c = (\Phi\Phi^{\mathcal{T}})^{-1/2}y$ and $A_c = (\Phi\Phi^{\mathcal{T}})^{-1/2}A = \Phi_c\Psi$ with $\Phi_c = (\Phi\Phi^{\mathcal{T}})^{-1/2}\Phi$. Thus to recover the signal x that is not exactly sparse represented in Ψ (see (12)) from the measurement y as in (1), the MLE principle indicates that it is better to first precondition the measurement y and the sensing matrix Φ by $(\Phi\Phi^{\mathcal{T}})^{-1/2}$, and then solve the traditional sparse recovery problem (5). In the sequel, (14) is referred to as *pre-conditioned signal recovery* (PSR). As it will be seen in Section 4 that for a given projection matrix Φ, the proposed PSR scheme outperforms the traditional one (5) in terms of signal reconstruction accuracy.

2.2. Joint optimization of Ψ and Φ

Let $\{x_j\}_{j=1}^J$ be a set of signal samples obeying model (12); that is $x_j = \tilde{\Psi}\tilde{s}_j + e_j$. Under the assumption that e_j is a normal distribution $\mathcal{N}(\mathbf{0}, \sigma_0^2 I_N)$ and that all e_j are statistically independent, the MLE principle suggests that the best estimate of the true $(\tilde{\Psi}, \{\tilde{s}_j\})$ that can be obtained using $\{x_j\}_{j=1}^J$ is the one which minimizes

$$\varrho_1 \triangleq \sum_{j=1}^{J} \|x_j - \Psi s_j\|_2^2 \triangleq \|X - \Psi S\|_F^2 \tag{15}$$

subject to $\|s_j\|_0 \le \kappa, \forall\, j$, where $X(:, j) = x_j$ and $S(:, j) = s_j$. This coincides with the traditional dictionary learning (11).

It is easy to understand that for a given CS system (Φ, Ψ), the (column) sparse coefficient matrix $\tilde{S}$ can be optimally estimated from the low dimensional measurements $\{y_j\}_{j=1}^J$ using the proposed PSR by minimizing

$$\varrho_2 \triangleq \sum_{j=1}^{J} \|(\Phi\Phi^{\mathcal{T}})^{-1/2}(y_j - As_j)\|_2^2 \triangleq \|(\Phi\Phi^{\mathcal{T}})^{-1/2}\Phi(X - \Psi S)\|_F^2 \tag{16}$$

where $A = \Phi\Psi$.

Let Φ have an SVD of form

$$\Phi = U \begin{bmatrix} \Sigma & \mathbf{0} \end{bmatrix} V^{\mathcal{T}} \tag{17}$$

where $U \in \Re^{M\times M}$ and $V \in \Re^{N\times N}$ are orthonormal matrices, and Σ is an $M \times M$ diagonal matrix with diagonals being the singular values of Φ. It can be seen that the PSR specified by (14) is equivalent to

$$\hat{s} \triangleq \arg\min_s \|y_c - A_c s\|_2^2 \quad \text{s.t.} \quad \|s\|_0 \le \kappa \tag{18}$$

where $y_c \triangleq \Phi_c x$ with $A_c \triangleq (\Phi\Phi^{\mathcal{T}})^{-1}A = \Phi_c\Psi$ and

$$\Phi_c \triangleq (\Phi\Phi^{\mathcal{T}})^{-1/2}\Phi = U\begin{bmatrix} I_M & \mathbf{0}\end{bmatrix}V^{\mathcal{T}} \tag{19}$$

since $(\Phi\Phi^{\mathcal{T}})^{-1/2} = U\Sigma^{-1}U^{\mathcal{T}}$. This implies that *using the proposed PSR scheme, a CS system characterized by* (Φ, Ψ) *is totally equivalent to that by* (Φ_c, Ψ), where Φ and Φ_c are given by (17) and (19), respectively.

One notes that the set of sensing matrices of form Φ_c can be characterized with

$$\mathcal{S}_c \triangleq \{\Phi:\ \Phi\Phi^{\mathcal{T}} = I_M\} \tag{20}$$

In this paper, the optimal CS system design is considered with the sensing matrices belonging to the set characterized by (20).

As mentioned before, our objective in this paper is to design the sensing matrix and dictionary jointly in an optimal way. The first thing is to derive a proper measure. It is observed that ϱ_2 is a function of Φ, Ψ and S and hence such a measure would be a candidate of cost function for the joint optimization of sensing matrix and dictionary. However, due to the existence of the sensing matrix Φ in ϱ_2, the solution of minimizing ϱ_2 with respect to S is in general different from the best one that minimizes the sparse representation error ϱ_1. Therefore, for a given Φ the best (Ψ, S) should minimize ϱ_1 and ϱ_2 simultaneously in some manner.

What is the key factor that should be taken into account in designing the sensing matrix in an CS system? Let $X = \Psi S + E \triangleq X^* + E$, where $X^* = \Psi S$ is the matrix of the clean signals. Note that the measurements are obtained via $Y = \Phi X = \Phi X^* + \Phi E$ and that $\varrho_2 = \|\Phi E\|_F^2$ due to $\Phi \in \mathcal{S}_c$. It is understood that besides reducing ϱ_2, the projection matrix Φ is expected to sense the key information underlying X^* as much as possible. It can be shown that the best linear estimate of X^* with the (clean) measurements ΦX^* is $\Phi^{\mathcal{T}}\Phi X^*$. So, it is also desired to design the projection matrix Φ such that $\|X^* - \Phi^{\mathcal{T}}\Phi X^*\|_F^2 = \|(I_N - \Phi^{\mathcal{T}}\Phi)\Psi S\|_F^2$ is minimized

As seen, the joint optimization is a multi-objective optimization problem and to deal with it, we propose the following measure/cost function:

$$\varrho(\Phi, \Psi, S) \triangleq \varrho_1 + \alpha\,\varrho_2 + \beta\,\|(I - \Phi^{\mathcal{T}}\Phi)\Psi S\|_F^2 \tag{21}$$

where $\alpha \geq 0$ and $\beta \geq 0$ are weighting factors to balance the importance of each term. The three terms in the measure are related to the actual sparse representation error, the cost function of the PSR scheme, and the difference between the clean signal and its best estimate from its sensed version, respectively, all having the same physical dimension (meaning) – signal energy.

Generally speaking, the choice of the best weights α and β depends on specific applications. We provide a rough guide for choosing α and β. In terms of α, $\alpha < 1$ is suggested (similar to what is suggested in [13]) since for signal compression application we need to highlight ϱ_1 (the sparse representation error E). With respect to β, we suggest $\beta > \alpha$ since capturing most of the key information in ΨS is more important for a sensing matrix. For most applications in which the CS systems are designed off-line, we can determine the best α and β by searching. This will be demonstrated in Section 4.2.

The corresponding joint optimization problem is then formulated as

$$\begin{aligned}\{\tilde{\Phi}, \tilde{\Psi}, \tilde{S}\} \triangleq\ & \min_{\Phi,\Psi,S} \varrho(\Phi, \Psi, S)\\ \text{s.t.}\quad & \Phi\Phi^{\mathcal{T}} = I_M;\ \ \|S(:, j)\|_0 \leq \kappa,\ \forall\, j\\ & \|\Psi(:, l)\|_2 = 1,\ \forall\, l\end{aligned} \tag{22}$$

We stress that our interest is to provide a measure (as ϱ in (22)) such that both the sensing matrix and the dictionary are optimized by minimizing this objective function. This is totally different than the framework utilized in [13], [14]. To better illustrate it, we briefly summarize the framework in [13], [14] as follows:

$$\begin{aligned}&\min_{\Psi,S} f_1(\Phi, \Psi, S)\\ &\min_{\Phi} f_2(\Phi, \Psi)\end{aligned} \tag{23}$$

where $f_1 = \varrho_1 + \alpha\varrho_2$ and f_2 has different forms in [13] and [14] (we refer the reader to [13] and [14] for the explicit expressions for f_2). It is clear that (22) is different from (23). An iterative procedure is used in [13], [14] to attack (23), in which the dictionary and sensing matrix are alternatively updated by minimizing f_1 and f_2, respectively. Clearly, the convergence of such a procedure is questionable. With respect to our framework in (22) for jointly optimizing the sensing matrix and the dictionary, it is possible to come up with a numerical algorithm that solves (22) with guaranteed convergence. However, such convergence is not possible (or even has no sense) for the algorithms in [13] and [14] since once we update the sensing matrix by minimizing f_2, we are not sure how f_1 changes (and it is possible that f_1 has larger value).

3. The proposed algorithm

In this section, we will derive an algorithm for solving the joint optimization problem (22).

Keeping $\Phi \in \mathcal{S}_c$ in mind, one can see

$$\varrho(\Phi, \Psi, S) = \|A(\Phi) - B(\Phi)\Psi S\|_F^2 \tag{24}$$

where

$$A(\Phi) \triangleq \begin{bmatrix} X \\ \sqrt{\alpha}\Phi X \\ \mathbf{0}\end{bmatrix},\ B(\Phi) \triangleq \begin{bmatrix} I_N \\ \sqrt{\alpha}\Phi \\ \sqrt{\beta}(I_N - \Phi^{\mathcal{T}}\Phi)\end{bmatrix}$$

Here, we adopt the alternating minimization strategy to address (22). The basic idea is to construct a sequence $\{\Phi_k, \Psi_k, S_k\}$ such that $\{\varrho(\Phi_k, \Psi_k, S_k)\}$ is a descent sequence. The proposed algorithm is outlined below:

Algorithm 1.

Initialization: With a collection of training sample matrix $X \in \Re^{N\times J}$, dimensionality (M, L) and sparsity level κ given, design a dictionary $\Psi_0 \in \Re^{N\times L}$ (say using overcomplete discrete cosine transform (DCT)) and generate $\Phi_0 \in \Re^{M\times N}$ randomly;

Begin $k = 1, 2, \cdots, N_{ite}$

- Coarse update of S with $\Phi = \Phi_{k-1}, \Psi = \Psi_{k-1}$: Note that (22) becomes

$$\begin{aligned}\tilde{S}_k = \arg\min_S & \left\|A(\Phi_{k-1}) - B(\Phi_{k-1})\Psi_{k-1}S\right\|_F^2\\ \text{s.t. } & \|S(:, j)\|_0 \leq \kappa,\ \forall\, j\end{aligned} \tag{25}$$

  which can be solved using OMP.
- Coarse update Ψ with $\Phi = \Phi_{k-1}, S = \tilde{S}_k$: (22) leads to

$$\tilde{\Psi}_k = \arg\min_\Psi \left\|A(\Phi_{k-1}) - B(\Phi_{k-1})\Psi\tilde{S}_k\right\|_F^2 \tag{26}$$

  which can be solved with the results in [14].
- Update Ψ and S: let $D_{\tilde{\Psi}_k}$ be an $N \times N$ diagonal matrix with $D_{\tilde{\Psi}_k}(n, n) = \|\tilde{\Psi}_k(:, n)\|_2$. Set[2]

$$\Psi_k = \tilde{\Psi}_k D_{\tilde{\Psi}_k}^{-1},\ S_k = D_{\tilde{\Psi}_k}\tilde{S}_k \tag{27}$$

  As to be shown later in Lemma 1, this is the best normalized Ψ_k for S_k.

[2] We note that $\Psi_k = \tilde{\Psi}_k D_{\tilde{\Psi}_k}^{-1}$ is also the orthogonal projection of $\tilde{\Psi}_k$ onto the set $\{\Psi \in \Re^{N\times L} : \|\Psi(:, l)\|_2 = 1,\ \forall\, l\}$.

- Update Φ with $\Psi = \Psi_k$, $S = S_k$: The design problem (22) turns to

$$\Phi_k \triangleq \arg \min_{\Phi\Phi^{\mathcal{T}} = I_M} \varrho(\Phi, \Psi_k, S_k) \tag{28}$$

As guaranteed by Lemma 2, the optimal Φ_k is given by (31).

End
Output: $\tilde{\Phi} = \Phi_{N_{ite}}$ and $\tilde{\Psi} = \Psi_{N_{ite}}$.

Remark 3.1. Our proposed algorithm, MOD and KSVD are used to update dictionary and all require the same $\tilde{S}_k$ obtained in (25). With $\Phi = \Phi_{k-1}$, $S = \tilde{S}_k$, the dictionary is updated by solving the problem defined in (22). Such a constrained problem is difficult to solve and all the three algorithms just yield sub-optimal solutions.

- MOD updates (Ψ, S) with $(\Psi_k, \tilde{S}_k)$. By doing so, it can not ensure the cost function to decrease and hence its convergence is not guaranteed, while we update (Ψ, S) with (Ψ_k, S_k) given in (27). Such a update has a property specified in Lemma 1 to be given below and hence its convergence, as to be shown later, can be guaranteed.
- Our proposed algorithm differs from KSVD algorithm in that 1) we update the whole dictionary concurrently rather than column by column as in KSVD; 2) we update S with (27) by scaling $\tilde{S}_k$ only, while KSVD takes $\tilde{S}_k$ as initial condition and updates the coefficients S_k and the dictionary simultaneously. Like in our algorithm, the update (Ψ_k, S_k) in KSVD can also ensure the cost function to decrease. However, there is nothing to ensure that one is better than another.

To analyze the convergence behavior of our proposed algorithm, we need the following lemma.

Lemma 1. *Let $\tilde{S}_k$ and (Ψ_k, S_k) be defined respectively as in (25) and (27). Ψ_k is the optimal solution to*

$$\Psi_k = \arg \inf_{\Psi: \|\Psi(:,l)\|_2 = 1,\ \forall\, l} \varrho(\Phi_{k-1}, \Psi, S_k)$$

and

$$\varrho(\Phi_{k-1}, \Psi_k, S_k) \le \inf_{\Psi: \|\Psi(:,l)\|_2 = 1,\ \forall\, l} \varrho(\Phi_{k-1}, \Psi, \tilde{S}_k) \qquad \square \tag{29}$$

Proof. We prove the arguments by contradiction. Suppose there exists Ψ' with column-wise normalized such that $\varrho(\Phi_{k-1}, \Psi', S_k) < \varrho(\Phi_{k-1}, \Psi_k, S_k)$. This implies

$$\left\| A(\Phi_{k-1}) - B(\Phi_{k-1})\Psi' D_{\tilde{\Psi}_k} D^{-1}_{\tilde{\Psi}_k} S_k \right\|_F^2 < \left\| A(\Phi_{k-1}) - B(\Phi_{k-1})\tilde{\Psi}_k \tilde{S}_k \right\|_F^2,$$

which constricts to the fact by (26). Thus, we have $\varrho(\Phi_{k-1}, \Psi_k, S_k) \le \varrho(\Phi_{k-1}, \Psi, S_k)$ for all $\Psi \in \Re^{N\times L}$ with $\|\Psi(:,l)\|_2 = 1,\ \forall\, l$, which completes the first part of the proof.

Now, let us consider the proof for the 2nd part of the lemma. Suppose there exists Ψ' with column-wise normalized such that $\varrho(\Phi_{k-1}, \Psi', \tilde{S}_k) < \varrho(\Phi_{k-1}, \Psi_k, S_k)$. As $\varrho(\Phi_{k-1}, \Psi_k, S_k) = \varrho(\Phi_{k-1}, \tilde{\Psi}_k, \tilde{S}_k)$, the above inequality is equivalent to

$$\varrho(\Phi_{k-1}, \Psi', \tilde{S}_k) < \varrho(\Phi_{k-1}, \tilde{\Psi}_k, \tilde{S}_k)$$

which can not be held due to the fact that $\tilde{\Psi}_k$ is the solution of (26). $\square$

The following lemma yields the solution to (28), which is used for updating the sensing matrix.

Lemma 2. *Let $G(\Psi, S)$ be defined as*

$$G(\Psi, S) \triangleq \beta\Psi S(\Psi S)^{\mathcal{T}} - \alpha(X - \Psi S)(X - \Psi S)^{\mathcal{T}} \tag{30}$$

and $G(\Psi, S) = V_G \Lambda_G V_G^{\mathcal{T}}$ be an eigen-decomposition (ED) of $G(\Psi, S)$, where $\Lambda_G = diag(\lambda_1, \cdots, \lambda_n, \cdots, \lambda_N)$ with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_N$. Then the solution to (28) with $S = S_k$, $\Psi = \Psi_k$ is given by

$$\Phi_k = U \begin{bmatrix} I_M & \mathbf{0} \end{bmatrix} V_G^{\mathcal{T}} \tag{31}$$

where U is an arbitrary $M \times M$ orthonormal matrix. $\square$

Proof. First of all, one note that $\varrho(\Phi, \Psi, S)$ is equivalent to

$$\begin{aligned}\varrho(\Phi, \Psi, S) &= \alpha\|\Phi(X - \Psi S)\|_F^2 - \beta\|\Phi\Psi S\|_F^2 + C_1 \\ &= -\mathrm{tr}\left[\Phi^{\mathcal{T}}\Phi G(\Psi, S)\right] + C_2\end{aligned}$$

where tr[.] denotes the trace operator, C_1 and C_2 are some constants that are independent of Φ, and $G(\Psi, S)$ is defined above in the Lemma. As understood, (28) is equivalent to

$$\Phi_k = \arg \max_{\Phi\Phi^{\mathcal{T}} = I_M} \mathrm{tr}\left[\Phi^{\mathcal{T}}\Phi G(\Psi, S)\right]$$

It then follows from the ED of $G(\Psi, S) = V_G \Lambda_G V_G^{\mathcal{T}}$ that

$$\begin{aligned}\mathrm{tr}\left[\Phi^{\mathcal{T}}\Phi G(\Psi, S)\right] &= \mathrm{tr}\left[\Phi^{\mathcal{T}}\Phi V_G \Lambda_G V_G^{\mathcal{T}}\right] \\ &= \mathrm{tr}\left[\Lambda_G V_G^{\mathcal{T}} \Phi^{\mathcal{T}}\Phi V_G\right] \\ &= \sum_{n=1}^{N} \lambda_n (Q(:,n))^{\mathcal{T}} Q(:,n)\end{aligned}$$

where $Q \triangleq \Phi V_G$.

As $\Phi = U \begin{bmatrix} I_M & \mathbf{0} \end{bmatrix} V^{\mathcal{T}}$, one has $QQ^{\mathcal{T}} = I_M$, that is $\|Q(m,:)\|_2 = 1,\ \forall\, m$. Let $\tilde{V} = V_G^{\mathcal{T}} V$. We have

$$\begin{aligned}(Q(:,n))^{\mathcal{T}} Q(:,n) &= (\tilde{V}(n,:))^{\mathcal{T}} \begin{bmatrix} I_M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tilde{V}(n,:) \\ &\le \|\tilde{V}(n,:)\|^2 = 1.\end{aligned}$$

Also, note that

$$\sum_{n=1}^{N} (Q(:,n))^{\mathcal{T}} Q(:,n) = \|Q\|_F^2 = M.$$

Combining the above two results gives

$$\sum_{n=1}^{N} \lambda_n (Q(:,n))^{\mathcal{T}} Q(:,n) \le \sum_{n=1}^{M} \lambda_n$$

and that the equality holds if and only if

$$(Q(:,n))^{\mathcal{T}} Q(:,n) = 1,\ \forall\, n \le M \ \Leftrightarrow\ V_G^{\mathcal{T}} V = I_N$$

which leads to (31). $\square$

Now, let us consider the convergence analysis of the proposed algorithm. The proposed algorithm is an iterative procedure, generating a sequence $\{(\Phi_k, \Psi_k, S_k)\}$. The convergence analysis of such an algorithm has two lines. The first one involves the convergence of the iterates $\{(\Phi_k, \Psi_k, S_k)\}$ and the second one is the convergence of the cost function $\varrho(\Phi_k, \Psi_k, S_k)$. The analysis of convergence for the iterates generated by the alternating minimization approach is usually very difficult and is beyond the scope of this paper. In what it follows, we will look at the behavior of the proposed algorithm in terms of convergence of the cost function.

Suppose the sparse coding stage (25) can be performed perfectly and the best κ-sparse approximations to the signals in $A(\Phi_{k-1})$ are obtained. In this case, it is clear that

$$\varrho(\Phi_{k-1}, \Psi_{k-1}, S_{k-1}) \geq \varrho(\Phi_{k-1}, \Psi_{k-1}, \tilde{S}_k)$$

Ideally, suppose Ψ is updated using $\bar{\Psi}_k$:

$$\bar{\Psi}_k = \arg \min_{\Psi, \|\Psi(:,l)\|_2=1} \varrho(\Phi_{k-1}, \Psi, \tilde{S}_k) \quad (32)$$

Obviously, $\varrho(\Phi_{k-1}, \Psi_{k-1}, \tilde{S}_k) \geq \varrho(\Phi_{k-1}, \bar{\Psi}_{k-1}, \tilde{S}_k)$ and according to Lemma 1, one has

$$\varrho(\Phi_{k-1}, \Psi_k, S_k) \leq \varrho(\Phi_{k-1}, \bar{\Psi}_k, \tilde{S}_k)$$

and our proposed algorithm can ensure

$$\varrho(\Phi_{k-1}, \Psi_{k-1}, S_{k-1}) \geq \varrho(\Phi_{k-1}, \Psi_k, S_k)$$

Finally, since Φ is updated with Φ_k, which is the solution of minimizing $\varrho(\Phi, \Psi_k, S_k)$ and obtained analytically using Lemma 2, one has

$$\varrho(\Phi_k, \Psi_k, S_k) \leq \varrho(\Phi_{k-1}, \Psi_{k-1}, S_{k-1}) \quad (33)$$

which means that our proposed algorithm guarantees convergence of the cost function.

We conclude this section by discussing the computational complexity of the proposed algorithm and the two algorithms in [13] and [14]. Note the dimensionality in $\Phi \in \Re^{M\times N}, \Psi \in \Re^{N\times L}$ and $X \in \Re^{L\times J}$. Our algorithm is an iterative method and in each iteration, there are four step. They are sparse coding which requires $O(\kappa^2(L+N)J)$ operations, dictionary updating which solves (26) and requires $O(LJ^2 + (N+M)^2N)$ operations, the dictionary normalizing (includes computing the diagonal matrix $D_{\tilde{\Psi}_k}$) and sparse coefficients scaling which are very cheap and respectively require $O(NL)$ and $O(LJ)$ flops, and sensing matrix updating which requires computing $G(\Psi, S)$ in (30) with $O(NLJ)$ operations and ED of $G(\Psi, S)$ with $O(N^3)$ operations and totally $O(NLJ)$ operations. Thus, the total cost for Algorithm 1 is $O(LJ^2N_{iter})$ under the assumption of $N \leq L \ll J$.

As to the algorithm proposed in [14], in each iteration, it runs N_{sen} iterations to update the sensing matrix (with computational complexity of $O(NL^2)$ in each iteration) and N_{dic} iterations to update the dictionary and space coefficients (with computational complexity of $O(LJ^2)$ in each iteration). Thus the total cost for the algorithm in [14] is $O(LJ^2N_{iter}N_{dic})$. With respect to the algorithm proposed in [13], in each iteration, it performs sensing matrix updating which requires $O(MN^2)$, sparse coding which requires $O(\kappa^2(L+N)J)$ operations, and dictionary updating which utilizes KSVD algorithm and needs $O(LJ^2)$ flops. Thus, the total cost for the algorithm in [13] is also $O(LJ^2N_{iter})$.

4. Experiments

In this section, we will examine the performance of the proposed PSR and the CS system, denoted as CS_{New}, whose sensing matrix and dictionary are jointly optimized with (22) using Algorithm 1. Particularly, we will compare CS_{New} with the CS systems CS_{DCS} and CS_{BH}, which were respectively proposed in [13] and [14], and the ones defined below: For the same dictionary, say the oracle/true dictionary or Ψ_{KSVD} design using KSVD,

- CS_{Rdm} – where the sensing matrix is generated randomly.
- CS_{Eld} – where the sensing matrix is designed using the approach in [11].
- CS_{LZYCB} – where the sensing matrix is designed using the approach in [16].
- $CS_{Classic}$ – where the sensing matrix is the solution of

$$\min_{\Phi, H\in\mathcal{S}_H^{etf}} \|H - A^{\mathcal{T}}A\|_F^2 \ \text{ s.t. } A = \Phi\Psi_{KSVD}$$

  where $\mathcal{S}_H^{etf}$ is a relaxed equiangular tight frames (ETF) Gram set. The solution is solved using the alternating minimization with the sensing matrix updated using one of the algorithms in [35], [37]. This is actually the same approach used in [15] and [20].
- CS_{TKK} – where the sensing matrix is obtained using the algorithm proposed in [38].

In our experiments, both real images and synthetic data are used. For the latter, the original signal samples are generated in the following way. A set of J κ-sparse $L\times 1$ vectors $\{s_j\}_{j=1}^J$ is generated, where each non-zero element of s_j is randomly positioned with an Gaussian distribution of *i.i.d.* zero-mean and unit variance. The set of signal vectors $\{x_j\}_{j=1}^J$ is generated with $x_j = \tilde{\Psi}s_j + e_j \triangleq x_j^* + e_j,\ \forall\ j$, where $\tilde{\Psi}$ is given and e_j is produced with Gaussian distribution of zero-mean and variance σ_e^2. Denote σ_{snr} as the signal-to-noise ratio (in dB) of the signals. For a given sensing matrix Φ, the measurements are given by $y_j = \Phi x_j,\ \forall\ j$.

We use the mean squared error (MSE), denoted as σ_{mse}, to measure the signal recovery performance of a CS system, say CS_z:

$$\sigma_{mse} \triangleq \frac{1}{N\times J}\sum_{j=1}^{J}\|x_j^* - \hat{x}_j\|_2^2 \quad (34)$$

where N is the dimension of the signal (vector) x_j^*, and $\hat{x}_j$ is the recovered version of x_j^* obtained using $\hat{x}_j = \Psi\hat{s}_j$ with $\hat{s}_j$ being the solution of the traditional reconstruction problem (i.e., (5)):

$$\hat{s}_j = \arg\min_s \|y_j - As\|_2^2 \ \text{ s.t. } \|s\|_0 \leq \kappa, \forall j$$

which is attached using OMP.

We use CS_{P-z} to denote the corresponding CS system CS_z which uses the PSR scheme to find $\hat{s}_j$, that is

$$\hat{s}_j = \arg\min_s \|(\Phi\Phi^{\mathcal{T}})^{-1/2}(y_j - As)\|_2^2 \ \text{ s.t. } \|s\|_0 \leq \kappa, \forall j$$

which is also solved using OMP.

4.1. Demonstration of the PSR scheme

For this part, the set of signal vectors $\{x_j\}_{j=1}^J$ is generated in the way described previously with an $N\times L$ dictionary $\tilde{\Psi}$ of independent and identically distributed (i.i.d.) normally distributed entries.

First of all, let us examine the performance of CS_{Rdm} and CS_{P-Rdm}, where the sensing matrix Φ is generated with i.i.d. normally distributed entries. Fig. 1 shows the MSE performance versus different sparsity level κ for each of CS systems, where $M = 25, N = 60, L = 80, J = 2000$ and $\sigma_{snr} = 10$ dB are used. Figs. 2 and 3 show the same results but with $\sigma_{snr} = 20, +\infty$ dB, respectively.

Fig. 4 displays the MSE performance versus different signal-to-noise ratio level σ_{snr} for each of CS systems, where $M = 25, N = 60, L = 80, J = 2000$ and $\kappa = 4$ are used.

Remark 4.1. As seen, CS_{P-Rdm} yields a better performance than CS_{Rdm}. The same is also observed for the other CS systems when the SNR σ_{snr} is small. See Fig. 5. This has been further confirmed with real images, for which the sparse representation error is very large. See Table 1.

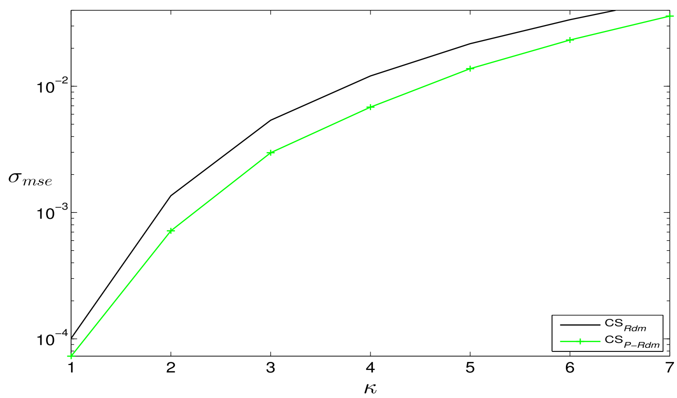

Fig. 1. Reconstruction error σ_{mse} versus signal sparsity κ for $\sigma_{snr} = 10$ dB.

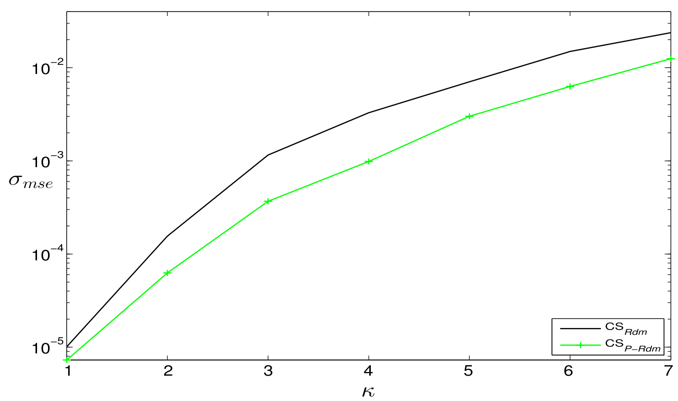

Fig. 2. Reconstruction error σ_{mse} versus signal sparsity κ for $\sigma_{snr} = 20$ dB.

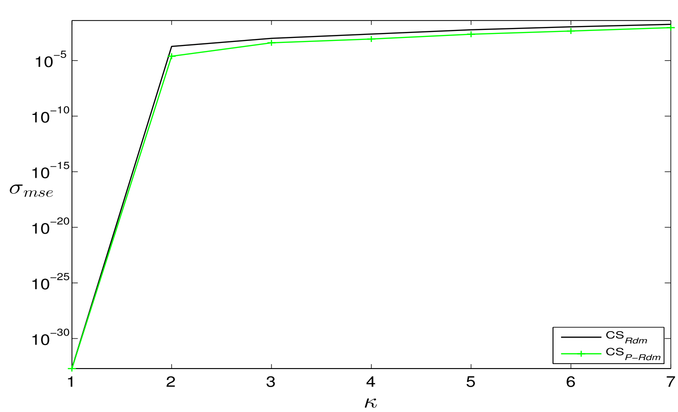

Fig. 3. Reconstruction error σ_{mse} versus signal sparsity κ for $\sigma_{snr} = +\infty$ dB.

4.2. Evaluation of CS systems

4.2.1. Synthetic data

First of all, for all the CS systems, including CS$_{DCS}$, CS$_{BH}$ and CS$_{New}$, we fix the dictionary Ψ with $\tilde{\Psi}$ – the oracle one utilized for generating the signal vectors $\{x_j\}_{j=1}^J$. This allows us to test the performance of the sensing matrix in each CS system. We separate the data into two sets: the training data $\{x_j\}_{j=1}^{J/2}$ and the testing data $\{x_j\}_{j=J/2+1}^{J}$. Denote $X_1 = \begin{bmatrix} x_1 & \cdots & x_{J/2} \end{bmatrix}$ and $S_1 = \begin{bmatrix} s_1 & \cdots & s_{J/2} \end{bmatrix}$. In this case, the sensing matrix Φ for CS$_{New}$ is simply obtained from (22) with $\Psi = \tilde{\Psi}$, which is equivalent to

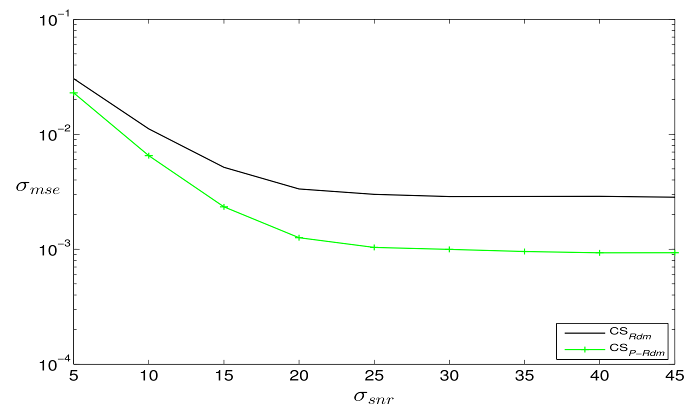

Fig. 4. Reconstruction error σ_{mse} versus σ_{snr}.

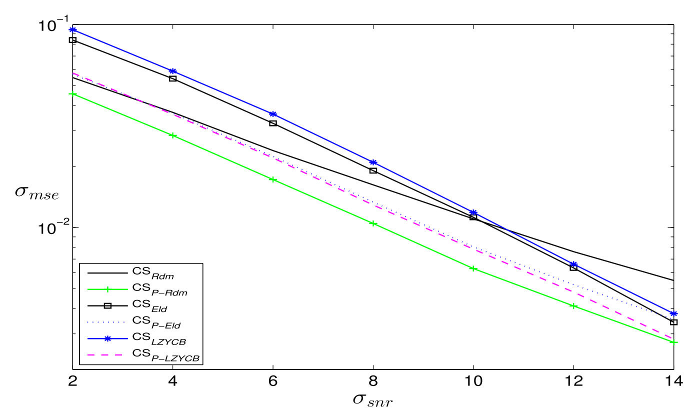

Fig. 5. Reconstruction error σ_{mse} versus σ_{snr}.

Table 1
Statistics of σ_{psnr}(dB) for images processed with $M = 20$, $N = 64$, $L = 100$, $\kappa = 4$.

	Lena	Couple	Elanie	Lake	Lax	Man	Plane	Baboon	Child	Fingerprint	Crowd	Boat
CS$_{Rdm}$	29.59	26.90	29.39	25.97	22.41	27.26	28.51	22.02	31.20	23.52	27.51	26.64
CS$_{P-Rdm}$	29.98	27.36	29.78	26.44	22.81	27.74	29.00	22.38	31.57	24.07	28.12	27.26
CS$_{Eld}$	17.68	15.21	10.30	11.71	7.25	14.76	19.45	7.95	20.88	16.52	18.56	13.33
CS$_{P-Eld}$	23.12	20.64	19.71	18.87	14.09	20.35	22.99	14.31	25.29	18.80	22.26	19.64
CS$_{LZYCB}$	13.03	10.75	7.56	7.02	2.57	10.19	14.17	3.44	15.81	11.00	13.44	8.17
CS$_{P-LZYCB}$	17.52	14.92	12.74	12.83	7.68	14.41	18.23	8.03	19.79	14.72	17.39	13.88
CS$_{DCS}$	32.77	30.07	32.24	29.01	25.40	30.36	31.60	24.94	34.17	26.56	30.71	29.87
CS$_{P-DCS}$	33.06	30.30	32.44	29.39	25.69	30.75	31.93	25.28	34.51	26.87	31.12	30.15
CS$_{Classic}$	12.97	10.49	6.68	7.32	2.62	9.80	14.72	3.03	11.58	26.77	13.65	8.34
CS$_{P-Classic}$	18.97	16.12	14.75	14.38	9.38	15.93	19.12	9.51	20.95	15.56	18.29	15.18
CS$_{TKK}$	15.20	12.90	8.41	9.14	4.78	12.21	17.07	5.65	18.54	14.20	16.04	10.90
CS$_{P-TKK}$	20.79	18.02	16.36	16.15	11.50	17.88	20.92	11.25	22.75	16.96	19.87	17.51
CS$_{BH}$	32.91	30.29	32.44	29.32	25.61	30.65	31.99	25.18	34.50	26.95	31.07	30.11
CS$_{P-BH}$	32.91	30.28	32.44	29.32	25.61	30.65	32.00	25.18	34.50	26.94	31.07	30.12
CS$_{New}$	33.17	30.44	32.62	29.52	25.69	30.78	32.20	25.33	34.75	27.30	31.35	30.26

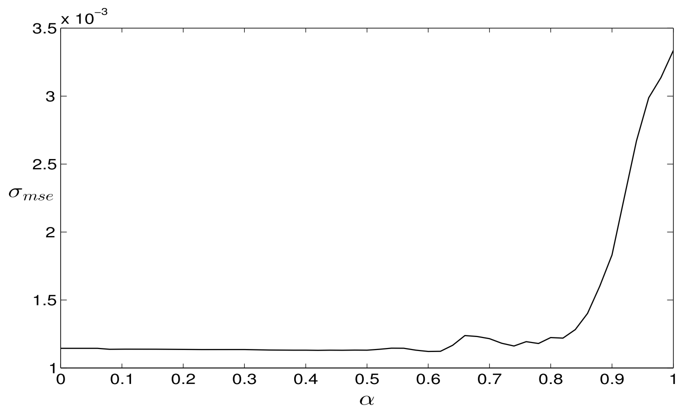

Fig. 6. Reconstruction error σ_{mse} versus α.

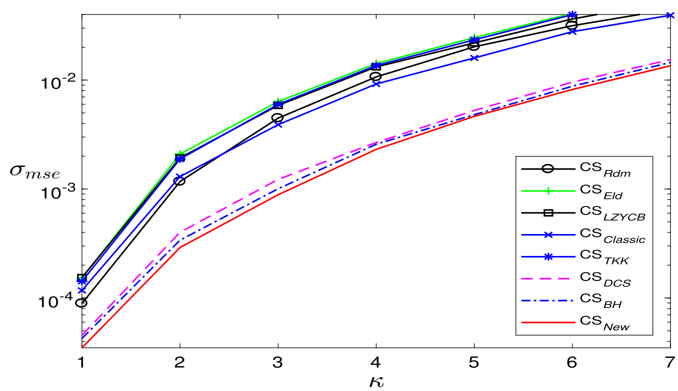

Fig. 7. Reconstruction error σ_{mse} versus signal sparsity κ for $\sigma_{snr} = 10$ dB.

$$\tilde{\Phi} = \arg \min_{\Phi\Phi^{\mathcal{T}}=I_M} \alpha \varrho_2(\Phi, \tilde{\Psi}, S_1) + \beta \|(I - \Phi^{\mathcal{T}}\Phi)\tilde{\Psi}S_1\|_F^2$$

where ϱ_2 is the cost function of the PSR scheme. The solution of the above problem can be obtained directly with (31).

Set $M = 25, N = 60, L = 80, J = 2000, \sigma_{snr} = 10$ dB and $\beta = 1 - \alpha$. Let us look at the effect of the parameter α on the MSE of signal reconstruction for CS_{New}. Fig. 6 depicts σ_{mse} versus the values of α.

Remark 4.2. Though there is no systematic way to find the best α, intensive simulations show that α taking value within $(0.4, 0.80)$ usually leads to a high performance sensing matrix. In the sequel, $\alpha = 0.45$ is used.

Fig. 7 shows the MSE performance versus different sparsity level κ for each of CS systems, where $M = 25, N = 60, L = 80, J = 2000$ and $\sigma_{snr} = 10$ dB are used. Fig. 8 shows the results for $\sigma_{snr} = 20$.

Remark 4.3. It is observed that CS_{New}, CS_{BH} and CS_{DSC} outperform the others, while our proposed CS_{New} is slightly superior over CS_{DCS} and CS_{BH} when the sparse representation error is high. This superiority is also demonstrated in the experiments to be presented below using real images, when the three systems are designed such that both the sensing matrix and the dictionary are simultaneously optimized.

4.2.2. Image compression

Now, we will examine these CS systems for the application of image compression. We randomly extract 15 non-overlapping patches (the dimension of each patch is 8×8) from each of 400 images in the LabelMe [39] training data set. Each patch of 8×8 is

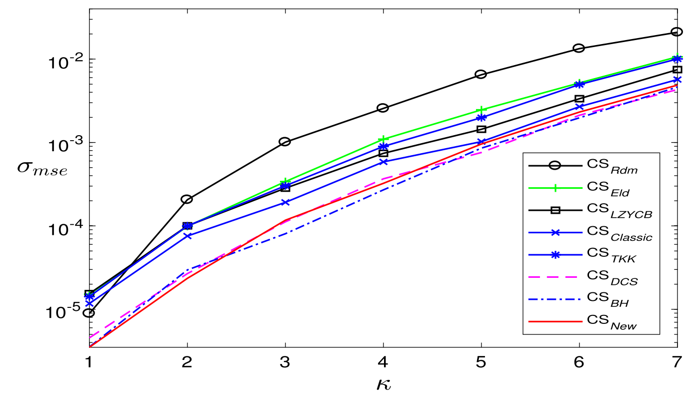

Fig. 8. Reconstruction error σ_{mse} versus signal sparsity κ for $\sigma_{snr} = 20$ dB.

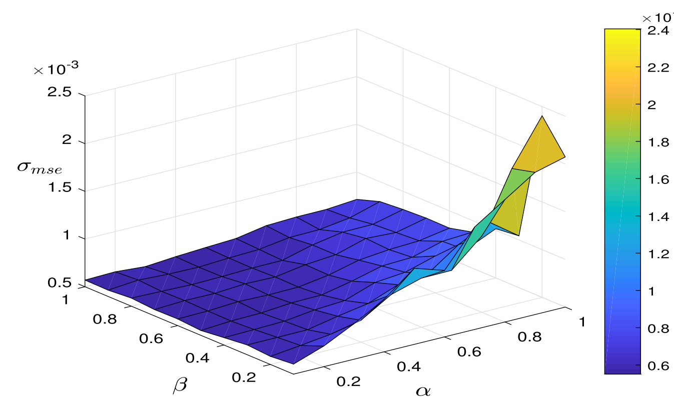

Fig. 9. Statistics of σ_{mse} versus α and β.

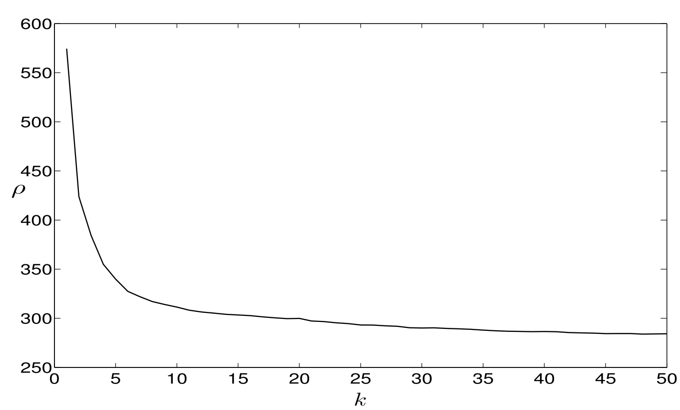

Fig. 10. Evolution of the cost function $\varrho(\Phi_k, \Psi_k, S_k)$.

then re-arranged as a vector of 64×1. With a set of 6000 training samples $\{x_j\}$, we apply the K-SVD algorithm with $N = 64, L = 100$ for a given sparsity level $\kappa = 4$ to obtain a sparsifying dictionary Ψ_{KSVD}. Based on the Ψ_{KSVD}, one can design the corresponding sensing matrix $\Phi \in \Re^{M\times N}$ for each of CS_{Eld}, CS_{LZYCB}, $\text{CS}_{Classic}$, and CS_{TKK}. The sensing matrix and dictionary for CS_{DCS} are simultaneously designed with this training data using the method in [13].[3]

To examine the effect of the parameters α and β on the CS system CS_{New}, we extract another 1000 samples $\{z_j\}$ (that are different than the training samples $\{x_j\}$) from the LabelMe [39] testing data set. Fig. 9 depicts σ_{mse} of CS_{New} for the samples $\{z_j\}$ versus the values of α and β. As observed from Fig. 9, $\alpha > 0.5$ and $\beta \leq 0.5$ yield poor performance. In the sequel, we perform Al-

[3] We set the coupling factor in CS_{DCS} to 0.5, which is suggested in [13].

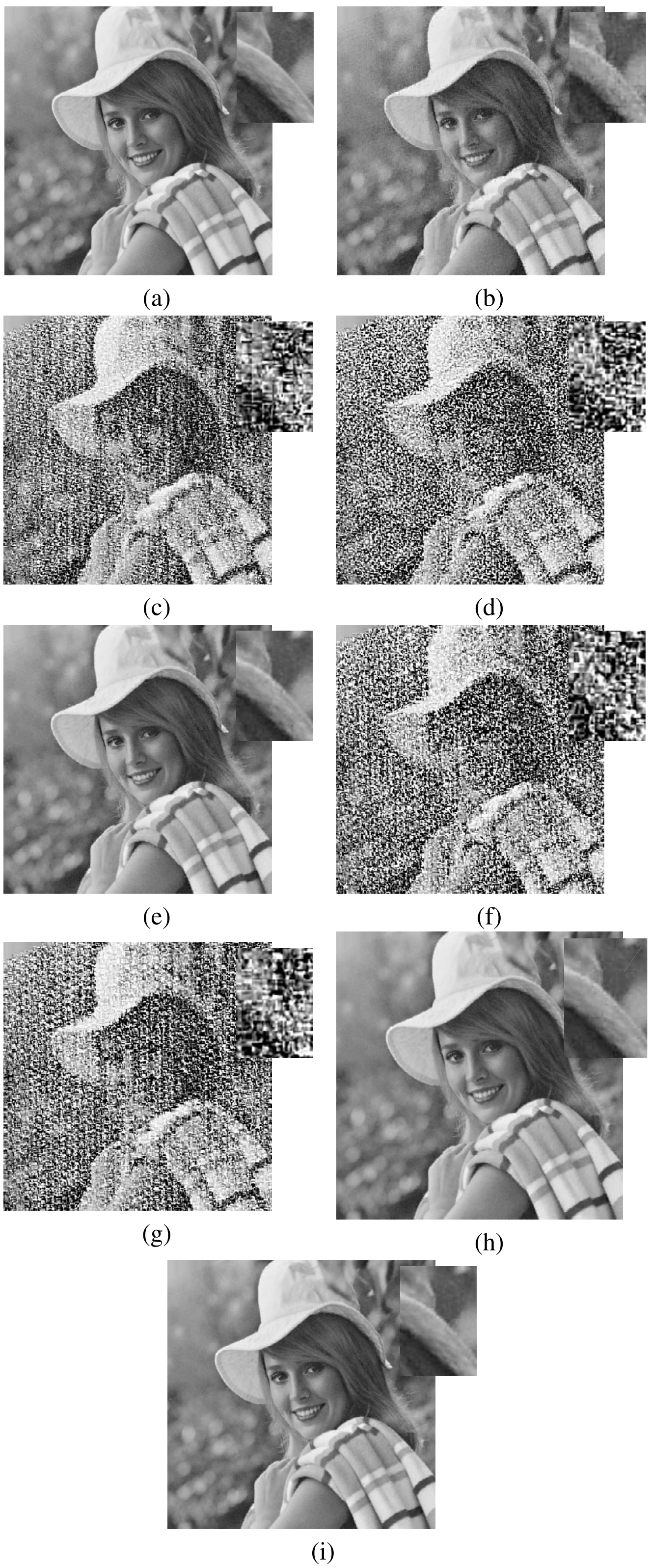

Fig. 11. The original 'Elanie' and reconstructed images from their CS samples with $M = 20$ and $\kappa = 4$. (a) The original; (b) CS_{Rdm}; (c) CS_{Eld}; (d) CS_{LZYCB}; (e) CS_{DCS}; (f) $CS_{Classic}$; (g) CS_{TKK}; (h) CS_{BH}; (i) CS_{New}.

gorithm 1 with 100 iterations and $\alpha = 0.4$, $\beta = 0.8$ to obtain the sensing matrix and sparsifying dictionary for CS_{New}.

Fig. 10 displays the evolution of the objective function $\varrho(\Phi_k, \Psi_k, S_k)$. As seen, the cost function $\varrho(\Phi_k, \Psi_k, S_k)$ decays steadily for Algorithm 1, which coincides with the analysis of convergence.

Twelve real images are processed with each CS system with compression and reconstruction. The reconstruction accuracy is evaluated in terms of peak signal-to-noise ratio (PSNR), defined as

$$\sigma_{psnr} \triangleq 10 \times \log_{10}\left(\frac{(2^r - 1)^2}{\sigma_{mse}}\right)$$

where $r = 8$ bits per pixel and σ_{mse} is defined in (34).

Table 1 provides σ_{psnr} for the CS systems tested with the twelve images. Fig. 11 displays the visual effects of 'Elanie'.

Remark 4.4.

- It is interesting to note from Table 1 that the CS systems with PSR are better than the original CS systems in terms of PSNR. We note that CS_{P-BH} and CS_{BH} have almost identical performance because the sensing matrix Φ obtained in CS_{BH} yields a $\Phi\Phi^{\mathcal{T}}$ very close to I_M.
- Note that any real image is in general not exactly sparse under any dictionary which is designed to minimize the sparse representation error. Table 1 indicates that CS_{TKK}, $CS_{Classic}$, CS_{LZYCB} and CS_{Eld}, designed without taking sparse representation errors into account, are very sensitive to the sparse representation errors. As shown in [3], the random sensing matrix is generally robust to the noise. This is confirmed with the performance of CS_{Rdm}.
- The superiority of the proposed CS_{New} over the others is clearly shown in Table 1. As can be observed, CS_{DCS} CS_{New} and CS_{BH} outperform greatly the others in terms of PSNR, which justifies the potentials of simultaneous design of sensing matrix and dictionary in improving the performance of CS systems, while our proposed CS_{New} is 0.4 dB–0.8 dB better than CS_{DCS} and 0.15 dB – 0.35 dB better than CS_{BH} due to the joint optimization of sensing matrix and dictionary.

5. Concluding remarks

This paper deals with designing optimal CS systems. Based on the maximum likelihood estimation principle, a preconditioned signal recovery scheme for CS has been developed and more importantly, a novel framework for CS system design has been proposed, which allows us to jointly optimize the sensing matrix and sparsifying dictionary with the same measure. An algorithm has been derived for solving the corresponding optimal design problem with ensured convergence in terms of reducing the cost function. Simulations have been carried out using synthetic data and real images, in which the superiority of the proposed approaches over the prevailing ones is clearly demonstrated.

Acknowledgments

The authors would like to thank the reviewers for their helpful comments and constructive suggestions that help improve the quality of this paper.

References

[1] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (Sept. 2006) 1289–1306.

[2] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (Feb. 2006) 489–509.

[3] E.J. Candès, M.B. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (Mar. 2008) 21–30.

[4] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthonormal) dictionaries via l_1 minimization, Proc. Natl. Acad. Sci. USA 100 (5) (Mar. 2003) 2197–2202.

[5] J.A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Inf. Theory 50 (10) (Sep. 2004) 2231–2242.

[6] J.A. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (Dec. 2007) 4655–4666.

[7] N.B. Karahanoglu, E. Hakan, A^* orthogonal matching pursuit: best-first search for compressed sensing signal recovery, Digit. Signal Process. 22 (4) (July 2012) 555–568.

[8] N.B. Karahanoglu, E. Hakan, Compressed sensing signal recovery via forward-backward pursuit, Digit. Signal Process. 23 (5) (Sep. 2013) 1539–1548.

[9] T. Liu, T. Qiu, R. Dai, J. Li, L. Chang, R. Li, Nonlinear regression A^* OMP for compressive sensing signal reconstruction, Digit. Signal Process. 69 (Oct. 2017) 11–21.

[10] M. Zibulevsky, M. Elad, L_1–L_2 optimization in signal and image processing, IEEE Signal Process. Mag. (May 2010) 76–88.

[11] M. Elad, Optimized projections for compressed sensing, IEEE Trans. Signal Process. 55 (12) (Dec. 2007) 5695–5702.

[12] W. Chen, M.R.D. Rodrigues, I.J. Wassell, Projection design for statistical compressive sensing: a tight frame based approach, IEEE Trans. Signal Process. 61 (8) (Apr. 2013) 2016–2029.

[13] J.M. Duarte-Carvajalino, G. Sapiro, Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization, IEEE Trans. Image Process. 18 (7) (July 2009) 1395–1408.

[14] H. Bai, G. Li, S. Li, Q. Li, Q. Jiang, L. Chang, Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing, IEEE Trans. Signal Process. 63 (6) (Mar. 2015) 1581–1594.

[15] V. Abolghasemi, S. Ferdowsi, S. Sanei, A gradient-based alternating minimization approach for optimization of the measurement matrix in compressive sensing, Signal Process. 92 (4) (Apr. 2012) 999–1009.

[16] G. Li, Z.H. Zhu, D.H. Yang, L.P. Chang, H. Bai, On projection matrix optimization for compressive sensing systems, IEEE Trans. Signal Process. 61 (11) (Jun. 2013) 2887–2898.

[17] N. Cleju, Optimized projections for compressed sensing via rank-constrained nearest correlation matrix, Appl. Comput. Harmon. Anal. 36 (3) (May 2014) 495–506.

[18] L. Zelnik-Manor, K. Rosenblum, Y.C. Eldar, Sensing matrix optimization for block-sparse decoding, IEEE Trans. Signal Process. 59 (9) (Sept. 2011) 4300–4312.

[19] M.P. Pereira, L. Lovisolo, D. Silva, M. Campos, On the design of maximally incoherent sensing matrices for compressed sensing using orthogonal bases and its extension for biorthogonal bases case, Digit. Signal Process. 27 (April 2014) 12–22.

[20] J. Xu, Y. Pi, Z. Cao, Optimized projection matrix for compressive sensing, EURASIP J. Adv. Signal Process. 43 (Feb. 2010) 1–8.

[21] H. Bai, Z. Zhu, G. Li, S. Li, Design of optimal measurement matrix for compressive detection, in: Proc. of the 10th International Symposium on Wireless Communication Systems, Ilmenau, Germany, Aug. 2013.

[22] B. Li, L. Zhang, T. Kirubarajan, S. Rajan, Projection matrix design using prior information in compressive sensing, Signal Process. 135 (June 2017) 36–47.

[23] X. Li, H. Bai, B. Hou, A gradient-based approach to optimization of compressed sensing systems, Signal Process. 139 (Oct. 2017) 49–61.

[24] Z. Zhu, M.B. Wakin, Approximating sampled sinusoids and multiband signals using multiband modulated DPSS dictionaries, J. Fourier Anal. Appl. (2017), https://doi.org/10.1007/s00041-016-9498-2, in press.

[25] Z. Zhu, M.B. Wakin, Wall clutter mitigation and target detection using discrete prolate spheroidal sequences, in: 3rd Int. Workshop Compressed Sensing Theory and Its Appl. to Radar, Sonar and Remote Sensing, CoSeRa, Pisa, Italy, June 2015.

[26] K. Engan, S.O. Aase, J.H. Hakon-Housoy, Method of optimal direction for frame design, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 5, March 1999, pp. 2443–2446.

[27] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (Nov. 2006) 4311–4322.

[28] M. Yaghoobi, L. Daudet, M.E. Davies, Parametric dictionary design for sparse coding, IEEE Trans. Signal Process. 57 (12) (Dec. 2009) 4800–4810.

[29] W. Chen, M.R.D. Rodrigues, Dictionary learning with optimized projection design for compressive sensing applications, IEEE Signal Process. Lett. 20 (10) (Oct. 2013) 992–995.

[30] L.N. Smith, M. Elad, Improving dictionary learning: multiple dictionary updates and coefficient reuse, IEEE Signal Process. Lett. 20 (1) (Jan. 2013) 79–82.

[31] M. Sadeghi, M. Babaie-Zadeh, C. Jutten, Dictionary learning for sparse representation: a novel approach, IEEE Signal Process. Lett. 20 (12) (Dec. 2013) 1195–1198.

[32] S.K. Sahoo, A. Makur, Dictionary training for sparse representation as generalization of K-means clustering, IEEE Signal Process. Lett. 20 (6) (June 2013) 587–590.

[33] G. Li, Z.H. Zhu, H. Bai, A.H. Yu, A new framework for designing incoherent sparsifying dictionaries, in: Proc. Int. Conf. Accoustic, Speech, and Signal Processing, ICASSP, New Orleans, USA, Mar. 2017.

[34] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall Inc., 1993.

[35] G. Li, X. Li, S. Li, H. Bai, Q.R. Jiang, X.X. He, Designing robust sensing matrix for image compression, IEEE Trans. Image Process. 24 (12) (Dec. 2015) 5389–5400.

[36] T. Hong, Z. Zhu, An efficient method for robust projection matrix design, Signal Process. 143 (Feb. 2018) 200–210.

[37] H. Tao, H. Bai, S. Li, Z. Zhu, An efficient algorithm for designing projection matrix in compressive sensing based on alternating optimization, Signal Process. 125 (Aug. 2016) 9–20.

[38] E.V. Tsiligianni, L.P. Kondi, A.K. Katsaggelos, Construction of incoherent unit norm tight frames with application to compressed sensing, IEEE Trans. Inf. Theory 60 (4) (Apr. 2014) 2319–2330.

[39] B. Russell, A. Torralba, W.T. Freeman, Labelme, the open annotation tool, MIT, online available: http://labelme2.csail.mit.edu/Release3.0/browserTools/php/dataset.php.

Gang Li received the B.Eng. degree in EE from BIT, China in 1982, and the M.Eng. and the Ph.D. both from Louvain University, Belgium, in 1988 and 1990, respectively.

He was a PDF from 1990 to 1992 at Louvain University. From May 1992 to Jan. 2007, he held a faculty position as an associate professor in the School of EEE, Nanyang Technological University (NTU), Singapore. From Feb. 2007 to Apr. 2015, he was with the Zhejiang University of Technology as a Qianjiang Chair Professor. He is now a professor with the Zhejiang University of Science & Technology in Hangzhou, China. His current researches focus on signal transforms, data compression, and optimization theory. He is a co-author of two books published by Springer-Verlag (London) and De Gruyter of Germany, respectively.

Zhihui Zhu obtained the B.Eng. degree in communications engineering from Zhejiang University of Technology in 2012. He is now working towards the Ph.D. degree in the Department of Electrical Engineering at the Colorado School of Mines. He was the meritorious winner of MCM/ICM in 2011. His research interests include exploiting inherent structures and applying optimization methods for signal processing and data analysis.

Xinming Wu received an engineering degree (2009) in Geophysics from Central South University, an M.Sc. (2012) in Geophysics from Tongji University, and a Ph.D. (2016) in Geophysics from the Colorado School of Mines where he was a member of the Center for Wave Phenomena. He interned twice at Transform Software and Services/DrillingInfo during the summer and winter of 2014. Since Aug. 2015, he has been working at Bureau of Economic Geology, University of Texas at Austin as a visiting scholar and postdoctoral fellow.

He received the award for Best Paper in Geophysics in 2016 (3D seismic image processing for faults). He also received the Shanghai excellent master thesis award in 2013 (Generating 3D seismic Wheeler volumes: methods and applications).

Beiping Hou received his Ph.D. in Control Science and Engineering from Zhejiang University in 2005. He is currently an Associate Professor at the School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou, China. His current research interests include machine vision and pattern recognition, texture analysis, and automatic control.