Measurement-Driven Framework With Simultaneous Sensing Matrix and Dictionary Optimization for Compressed Sensing

HUANG BAI, (Member, IEEE), AND XIUMEI LI, (Member, IEEE)
School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 311121, China
Corresponding author: Xiumei Li (xiumei_li@hotmail.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61801159, Grant 61571174, Grant 61503330, and Grant 61502133, and in part by the Start-Up Research Program from Hangzhou Normal University under Grant 2018QDL018.

ABSTRACT This paper presents a measurement-driven framework to deal with the compressed sensing (CS) system design problem. Under this novel framework, the sparse coefficient matrix is calculated according to the low-dimension measurements, rather than updated along with the dictionary as in the traditional cases. Moreover, a new cost function is proposed to simultaneously optimize the sensing matrix and dictionary. In order to minimize this cost function, an iterative algorithm is carried out. In every iteration, the solutions of the sensing matrix and dictionary are derived analytically. Experiments are executed with real images, especially medical images. The results demonstrate the superiority of the designed CS system composed of the optimized sensing matrix and dictionary with improved performance for image compression and reconstruction.

INDEX TERMS Compressed sensing, measurement-driven, sensing matrix, dictionary, iterative optimization.

I. INTRODUCTION

Compressed sensing (CS) has attracted vast attention since its introduction around 2006 [1]–[3] and it is a useful tool in many engineering fields, such as wireless sensor networks [4]–[6] and medical image processing [7], [8]. As the necessary condition of CS, signal sparse representation has a much longer history [9]. In this theory, the original signal \( x \in \mathbb{R}^{N \times 1} \) is assumed to be in the form of

\[
x \triangleq \Psi s
\]

(1)

where \( \Psi \in \mathbb{R}^{N \times L} \) is called a dictionary with the columns referred to as atoms and \( s \) is the sparse coefficient vector.

Let \( v \in \mathbb{R}^{L \times 1} \) with \( v(l) \) being its \( l \)-th element. The \( l_p \)-norm of vector \( v \) is defined as

\[
\|v\|_p \triangleq \left( \sum_{l=1}^{L} |v(l)|^p \right)^{1/p}, \quad p \geq 1
\]

The associate editor coordinating the review of this manuscript and approving it for publication was Kai Li.

Note that \( \|v\|_p \) is not a norm in a strict sense for \( 0 \leq p < 1 \). For convenience, \( \|v\|_0 \) is used to denote the number of non-zero elements in \( v \). Then, \( x \) given by (1) is said \( K \)-sparse in \( \Psi \) if \( \|s\|_0 = K \).

Sparse representation theory leads to a widely investigated problem named dictionary design. Dictionary design is to find a dictionary to represent a class of signals for a given sparsity level \( K \). Typical examples include the Fourier dictionary for frequency-sparse signals, a multiband modulated Discrete Prolate Spheroidal Sequences (DPSS) dictionary for sampled multiband signals [10], [11], and a dictionary learned from a training dataset [12], [13]. Considering learning a dictionary, let \( \{x_k\}_{k=1}^{P} \) be a set of training samples with \( x_k \in \mathbb{R}^{N \times 1} \). The basic problem, usually referred to as dictionary learning, can be formulated as

\[
\min_{\Psi, S} \|X - \Psi S\|_F^2 \quad \text{s.t.} \quad \|s_k\|_0 \leq K, \quad \forall k = 1, 2, \ldots, P \quad (2)
\]

with \( \|\cdot\|_F \) denoting the Frobenius norm,

\[
X \triangleq [x_1, \ldots, x_k, \ldots, x_P], \quad S \triangleq [s_1, \ldots, s_k, \ldots, s_P]
\]
(2) is difficult to be solved directly. A two-stage alternating minimization procedure is usually adopted. The first stage is sparse coding, aiming at finding the (column) sparse matrix \( S \) with a given \( \Psi \), i.e.,

\[
\min_{\mathbf{s}_k} \| \mathbf{s}_k \|_0 \quad \text{s.t.} \quad \mathbf{x}_k = \mathbf{s}_k \Psi, \quad \forall \ k = 1, 2, \ldots, P
\]  

(3) or

\[
\min_{\mathbf{s}_k} \| \mathbf{x}_k - \mathbf{s}_k \|_2^2 \quad \text{s.t.} \quad \| \mathbf{s}_k \|_0 \leq K, \quad \forall \ k = 1, 2, \ldots, P
\]  

(4)

This problem can be solved by orthogonal matching pursuit (OMP) based techniques [14]–[16]. Many algorithms differentiate each other mainly in the second stage named dictionary updating [12], [13], that is to update \( \Psi \) with the obtained \( S \) fixed. In the method of optimal direction (MOD) [12], the dictionary \( \Psi \) is simply taken as the solution of

\[
\tilde{\Psi} = \min_{\Psi} \| \mathbf{X} - \Psi^T \mathbf{s} \|_F^2 = \mathbf{X} (\mathbf{S} S^T)^{-1} \mathbf{T}
\]

followed by an extra normalization procedure. Here \( \mathbf{T} \) denotes the transpose operator. Taking the sparse structure of the given \( S \) into consideration, Aharon *et al.* proposed an algorithm, named K-singular value decomposition (K-SVD) [13], in which the atoms of the dictionary are updated one by one. Such an algorithm usually yields a better performance than the MOD as the non-zero elements in \( S \) are simultaneously updated. In general, the sparse coefficients are always updated along with the corresponding dictionary under the same model, such as (2) discussed above.

If signal sparse representation holds, CS can be formulated as a mathematical framework dealing with the accurate reconstruction of a high-dimension signal \( \mathbf{x} \in \mathbb{R}^{N \times 1} \) from its corresponding low-dimension measurement \( \mathbf{y} \in \mathbb{R}^{M \times 1} \) with \( M \ll N \). This measuring process is implemented by projecting the original signal via a well chosen sensing matrix \( \Phi \in \mathbb{R}^{M \times N} : \)

\[
\mathbf{y} = \Phi \mathbf{x}
\]  

(5)

Substitute \( \mathbf{x} \) in (5) by (1), then \( \mathbf{y} \) can be rewritten as

\[
\mathbf{y} = \Phi \Psi \mathbf{s} \triangleq D \mathbf{s}
\]  

(6)

\( D \triangleq \Phi \Psi \) is sometimes named equivalent dictionary. With both \( \mathbf{y} \) and \( D \) given, reconstruction is to find \( \mathbf{s} \) through

\[
\min_{\mathbf{s}} \| \mathbf{s} \|_0 \quad \text{s.t.} \quad \mathbf{y} = D \mathbf{s}
\]  

(7) or

\[
\min_{\mathbf{s}} \| \mathbf{y} - D \mathbf{s} \|_2^2 \quad \text{s.t.} \quad \| \mathbf{s} \|_0 \leq K
\]  

(8)

Similar with (3) and (4), such problem can also be addressed using the greedy algorithms, such as OMP based techniques [14]–[16]. The estimated signal is simply obtained with \( \mathbf{x} = \Psi \mathbf{s} \).

A key notion for signal reconstruction is the restricted isometry property (RIP) [1], [3]. A matrix \( D \) is said to be \((K, \delta)\)-RIP if there exists a \( \delta \) with \( 0 \leq \delta < 1 \) such that

\[
(1 - \delta) \| \mathbf{s} \|_2^2 \leq \| D \mathbf{s} \|_2^2 \leq (1 + \delta) \| \mathbf{s} \|_2^2
\]

holds for all \( K \)-sparse vectors \( \mathbf{s} \). Clearly, a small \( \delta \) can ensure that any subset of \( D \) with cardinality less than \( K \) are nearly orthogonal. It has been shown that when the equivalent dictionary \( D \) is \((2K, \delta)\)-RIP, a \( K \)-sparse \( X \) in \( \Psi \) can be reconstructed exactly from its low-dimension measurement [1]–[3], [9]. Another significant property for exact signal reconstruction is spark. The spark of a matrix \( D \) is defined as the smallest number of columns in \( D \) that are linearly dependent. It has been shown that as long as the spark of \( D \) is greater than \( 2K \), \( K \)-sparse signal \( s \) can be exactly recovered via (7) or (8) [9].

Since neither RIP nor spark is tractable, it is preferable to use alternative properties of \( D \) that can be easily manipulated. One of such properties is the mutual coherence [9], [17], which is defined as

\[
\mu(D) \triangleq \max_{1 \leq i \neq j \leq L} \frac{|D(:,i)^T D(:,j)|}{\| D(:,i) \|_2 \| D(:,j) \|_2}
\]  

(9)

\( \mu(D) \) measures the maximum linear dependency possibly achieved by any two columns of matrix \( D \). As shown in [9], a \( K \)-sparse signal can be exactly recovered from the measurement as long as

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

(10)

With the dictionary \( \Psi \) learned from the training data via (2), the property of the equivalent dictionary \( D \) depends only on the sensing matrix \( \Phi \). (10) implies that if \( \Phi \) is well designed such that \( \mu(D) \) is as small as possible, the CS system will allow a wider set of candidate signals to be successfully recovered.

Roughly speaking, many approaches for designing sensing matrix are based on a similar criterion:

\[
\min_{\Phi, \mathbf{G}} \| \mathbf{G}_f - \Psi^T \Phi \Psi \|_F^2
\]  

(11)

The dictionary \( \Psi \) is assumed to be given and \( S \) is a class of Gram matrices possessing certain properties, e.g., the Gram of equiangular tight frame leads to good coherence property, see [17], [18]. However, it should be pointed out that (10) is just a worst-case bound and the mutual coherence cannot reflect the actual performance of the system accurately. The sensing matrix designed based on coherence property as (11) yields impressive performance when the signal to be considered is exactly sparse in the dictionary, but this is not true when there exists sparse representation error [19]–[26].

Note that the sensing matrix and the dictionary in a CS system are coupled in pair, therefore it would be better to optimize these two matrices simultaneously [22]–[28]. Moreover, in the practical CS scenario, what we receive is the low-dimension measurement \( y \) rather than the original \( x \). So it is crucial to recover the sparse coefficient \( s \) accurately according to \( y \) and the well-designed sensing matrix \( \Phi \) and dictionary \( \Psi \).

The main objective of this paper is to propose a novel framework for CS system optimization and the contribution is three-fold:
A novel measurement-driven framework is presented to design the CS system in order to improve the reconstruction accuracy of the sparse coefficients from the low-dimension measurements. In such a design framework, sparse coefficients and dictionary are updated with different measures.

Under the measurement-driven framework, a new cost function is proposed for optimizing sensing matrix and dictionary simultaneously. Different from the work in [22] and [23], the sensing matrix and the dictionary can be designed jointly under this new cost function.

An iterative algorithm that updates sensing matrix and dictionary alternately is put forward to solve the design problem, in which the solutions of sensing matrix and dictionary are derived analytically, hence the convergence of the cost function can be ensured.

In addition, the potential of the designed system is demonstrated by a great many of experiments based on real images, especially medical images.

The remainder of this paper is arranged as follows. In Section II, some preliminaries are provided and the work in [22] and [23] is introduced. Section III is devoted to formulating problem of a novel measurement-driven CS system. Besides, a new cost function for simultaneous sensing matrix and dictionary optimization is presented. The main algorithm for updating sensing matrix and dictionary is investigated in Section IV, where in each iteration, the solutions of sensing matrix and dictionary are derived analytically. Simulations are carried out in Section V to examine the performance of the proposed system and to compare with the existing works. Some concluding remarks are given in Section VI.

In this paper, we use lowercase and capital bold characters to indicate vectors and matrices, respectively. All the parameters are real-valued.

II. PRELIMINARIES AND RELATED WORK

In this section, we review some preliminaries on sensing matrix and dictionary optimization based on the work in [22] and [23].

There are lots of work concerning designing optimal CS systems. In most of these papers, the sensing matrix and dictionary are optimized semi-independently, i.e., the dictionary is learned according to the signal data under the model (2) [12], [13], while the sensing matrix is designed to make the equivalent dictionary possess certain properties with corresponding dictionary fixed, such as the criterion (11) [17]–[21]. Seen from the reconstruction procedure (7) or (8), one can observe that the performance of a CS system is determined by the equivalent dictionary \( D = \Phi \Psi \). Therefore, it would be better to optimize these two matrices \( \Phi \) and \( \Psi \) simultaneously under a uniform framework.

To the best of our knowledge, the first work dealing with the simultaneous sensing matrix and dictionary optimization was probably [22]. In general, dictionary learning intends to design a dictionary \( \Psi \) along with the sparse coefficient matrix \( S \) such that the sparse representation error

\[
E \triangleq X - \Psi S
\]

is minimized. In practical applications, this error \( E \) is not nil, for example, in the case of image compression as we will show in Section V. In the design of CS system, besides \( \Psi \), one has to consider how to choose the sensing matrix \( \Phi \) such that the original signals can be reconstructed from the corresponding measurements \( Y = \Phi X = \Phi \Psi S + \Phi E \) with high accuracy. \( \Phi E \) is the amplified version of \( E \). This error may seriously degrade the reconstruction accuracy of \( S \) as \( Y \) may be far away from being sparse in \( \Phi \Psi \). This observation is the main motivation of [22] and [23] that takes the sensing matrix into account in dictionary learning and leads to the following optimal dictionary learning problem with sensing matrix embedded:

\[
\min_{\Phi, \Psi} \left\{ \alpha \|E\|_F^2 + (1 - \alpha)\|\Phi E\|_F^2 \right\}
\]

s.t. \( \|S(:, k)\|_0 \leq K, \ \forall \ k, \ \Phi = f(\Psi) \) (12)

The parameter \( \alpha \) is used to balance \( E \) and its amplified version \( \Phi E \). Note that the sensing matrix \( \Phi \) is a function of dictionary \( \Psi \). Such a function, denoted as \( \Phi = f(\Psi) \), is generally implicit and determined by the method used for sensing matrix design.

In [22], the sensing matrix was designed using (11) with \( G_t = I_L \) fixed, \( I_L \) denotes the identity matrix of dimension \( L \). With some algebraic operations, the design problem was converted into the following

\[
\min_{\Psi, \Psi^T} \left\| \Pi - \sum_{k=1}^{M} v_k v_k^T \right\|_F^2
\]

where \( \Pi \) is the diagonal matrix in the SVD of \( \Psi \Psi^T \). The optimal \( \{v_k\} \) were searched one by one using an iterative algorithm:

\[
\min_{\Psi, \Psi^T} \left\| (\Pi - \sum_{k \neq j} v_k v_k^T) - v_j v_j^T \right\|_F^2, \quad j = 1, 2, \ldots, M
\]

This algorithm is based on eigen-decomposition and it is computationally efficient but cannot guarantee the global solution.

In [23], the following modified cost function was adopted:

\[
\min_{\Phi} \left\{ (1 - \beta)\|\Psi^T \Psi - \Psi^T \Phi \Psi\|_F^2 + \beta\|G_t - \Psi^T \Phi \Psi\|_F^2 \right\}
\]

(13)

where \( G_t \) is a target Gram matrix possessing certain property, e.g., coherence property. One main contribution of [23] is to derive the closed-form solution of problem (13).

From then on, several works have been published for jointly optimizing sensing matrix and dictionary. For example, [24] extended the work of [23] to tensor compressive sensing and the experiments on multi-dimensional signals verified its superiority; [25] mainly concerned the normalization of the incoherent dictionary and the equivalent dictionary
for clear physical meaning when optimizing these two matrices. Although these works have tried to optimize the sensing matrix and dictionary of CS system in a joint way, it should be pointed out that the model of (12) is still far away from simultaneously optimizing Φ and Ψ. An iterative procedure is used to attack (12), in which the sensing matrix and dictionary are alternatively updated by minimizing different cost functions, i.e., when updating sensing matrix, the authors jump out of the problem (12) and turn to a new measure Φ = f(Ψ), e.g., (13) in [23]. The key to simultaneous optimization is to derive a cost function of Φ and Ψ that is tractable.

III. PROBLEM FORMULATION

In this section, the measurement-driven framework for designing CS system with simultaneous sensing matrix and dictionary optimization will be introduced in detail. Before that, a preprocessing procedure on sensing matrix is investigated, and the resulted constraint on sensing matrix will be applied in our method throughout this paper.

A. PREPROCESSING ON SENSING MATRIX

Let us take a close look at the signal reconstruction problem (8). Denote X − Ψs ∆ e and

\[ y − Ds = Φ(x − Ψs) = Φe ∆ e \]

From a statistical point of view, assume that e follows a normal distribution \( N(0, vI_N) \). So the probability density function (PDF) of e is

\[ f_ε(ξ) = \frac{1}{\sqrt{2π} \det(vI_N)} e^{-\frac{1}{2}ξ^T(vI_N)^{-1}ξ} \]

in which \( \det(\cdot) \) denotes the determinant operator [29]. e = Φe indicates that e has a multivariate normal distribution \( N(0, vΦΦ^T) \), which means that the PDF of e obeys

\[ f_ε(ξ) = \frac{1}{\sqrt{2π} \det(vΦΦ^T)} e^{-\frac{1}{2}ξ^T(vΦΦ^T)^{-1}ξ} \]

According to the maximum likelihood estimation (MLE) principle [26], [30], the best estimate of the sparse coefficient vector s is the one that maximizes the PDF \( f_ε(ξ) \) at \( ξ = y − Ds \), which leads to

\[
\max_s f_ε(ξ) = \max_s e^{-\frac{1}{2}ξ^T(vΦΦ^T)^{-1}ξ} = \min_s ξ^T(vΦΦ^T)^{-1}ξ = \min_s \|((ΦΦ^T)^{-1/2}(y − Ds))\|_2^2
\]

as in a statistical point of view, the variance \( v \geq 0 \) is independent from s.

With the above discussions, we modify (8) as

\[
\min_s \|((ΦΦ^T)^{-1/2}y − (ΦΦ^T)^{-1/2}Ds)\|^2_2 = \|Φx − Ψs\|^2_2
\]

s.t. \( \|s\|_0 \leq K \) (14)

where \( \hat{Φ} = (ΦΦ^T)^{-1/2}Φ \). When \( x − Ψs \neq 0 \), the MLE principle indicates that the reconstruction procedure carried out by (14) is better than (8) [26].

Let general \( \boldsymbol{M} × N \) full row rank matrix Φ has the following singular value decomposition (SVD)

\[ Φ = U \left[ \begin{array}{c} \Sigma \end{array} \right] V^T \]

with U and V two orthonormal matrices of dimensions \( \boldsymbol{M} \times \boldsymbol{M} \) and \( \boldsymbol{N} \times \boldsymbol{N} \), respectively, and \( \Sigma \) the diagonal singular value matrix. It is easy to show that

\[ \Phi = (ΦΦ^T)^{-1/2}Φ = U \left[ I_{M \times 0} \right] V^T \]

as \( (ΦΦ^T)^{-1/2} = UΣ^{-1}U^T \). Our model is depicted as (14), so for convenience, in the following, the optimal CS system design is considered with the sensing matrix Φ constrained by the form

\[ Φ = U \left[ I_{M \times 0} \right] V^T \] (15)

Remark 1: Hereinafter, all the sensing matrices we talk about in our method are constrained by (15). A matrix described as (15) is actually a tight frame, which has been proved valid when being applied to wireless sensor networks [31], [32]. Comparing (14) with (8), this preprocessing procedure on Φ just constrains the searching space for optimal sensing matrix, and will not affect the process for calculating sparse coefficients, e.g., OMP based techniques can be employed as usual.

B. MEASUREMENT-DRIVEN FRAMEWORK

Now go to our main job on the measurement-driven framework. As seen from the discussions on sparse coefficients, no matter model (2) in sparse representation theory or (12) in CS, the sparse coefficient matrix S is always updated along with the dictionary Ψ, i.e., alternating between sparse coding for S and dictionary updating for Ψ. However in the practical CS scenario, what we receive is the low-dimension measurements Y, and our ultimate task is to reconstruct S from Y, which will be affected by both Φ and Ψ as indicated in (8) or (14). From this point of view, it is important to recover S adapted to the measurement matrix Y, rather than the high-dimension signal matrix X. We name this as measurement-driven framework to make it different from data-driven or task-driven frameworks as the latter two are popularly used in dictionary learning problem [33], and we intend to calculate the sparse coefficient matrix S from the measurements Y without considering the original data X.

In our system design strategy, mathematically, the independent update procedure of S can be formulated as

\[
\min_S \|Y − ΦΨS\|^2_F \quad \text{s.t.} \quad \|S(:, k)\|_0 \leq K, \forall k \quad (16)
\]

where Y is measured by the CS system composed of pair (Φ, Ψ), the design of Φ and Ψ will be discussed in the next subsection, and the notation S(:, k) = s_k. (16) makes the
sparse coefficients totally associate with the measurements and the measure procedure (according to sensing matrix $\Phi$ and dictionary $\Psi$).

Remark 2: It should be declared that (14) is the process concerning signal reconstruction, this is exactly the reason why we choose (16) to update sparse coefficients in the system design strategy. Make this updating procedure fully connected with the signal reconstruction, then utilize the obtained sparse coefficients to optimize sensing matrix and dictionary, so that $\Phi$ and $\Psi$ will be implicitly adapted to the reconstruction accuracy. As far as the authors know, this is the first work to separate the updates of sparse coefficients and dictionary into different cost functions. Calculate $S$ from the measurements, then design system pair $(\Phi, \Psi)$ with $S$ fixed. This is the core idea of our measurement-driven framework.

C. SIMULTANEOUS SENSING MATRIX AND DICTIONARY OPTIMIZATION

Besides the measurement-driven framework, another main contribution of this paper is to derive one cost function that can optimize sensing matrix and dictionary simultaneously.

As sparse representation is prerequisite for CS theory, when designing the dictionary, the penalty term regarding representation error

$$\min_{\Psi} \|X - \Psi S\|_F^2 \triangleq \varrho_1(\Psi)$$

should be included in the design model naturally. Once again, in our measurement-driven framework, the sparse coefficients are updated by (16), so $\varrho_1(\Psi)$ is just a cost function of $\Psi$, and $S$ is viewed as a constant matrix when minimizing $\varrho_1(\Psi)$. This is totally different from the traditional dictionary learning problem (2).

Considering the sensing matrix, a better $\Phi$ should result in less information loss when reducing the dimension of the original signal to lower one. Define $X^* \triangleq \Psi S$ as the ideal signal set without sparse representation error. The desired sensing matrix is expected to sense the key information in $X^*$ as precisely as possible. Denote the clean measurements $Y^* = \Phi X^*$, then it is obvious to derive the best linear estimate of $X^*$ with the clean measurements $Y^*$ via the following Least-Squares

$$\hat{X} = \arg\min_X \|Y^* - \Phi X\|_F^2$$

$$= \Phi^T Y^* = \Phi^T (\Phi \Phi^T)^{-1} Y^* = \Phi^T Y^*$$

as $\Phi$ is constrained by (15) in our work which leads to $\Phi \Phi^T = I_M$. $\Phi^T \triangleq (\Phi \Phi^T)^{-1} = \Phi^T$ is the pseudo-inverse of $\Phi$ [29].

With the best linear estimate $\hat{X}$ obtained above and $Y^* = \Phi X^*$, we propose the following penalty term

$$\|X^* - \hat{X}\|_F^2$$

$$= \|\Psi S - \Phi^T Y^*\|_F^2$$

$$= \|\Psi S - \Phi^T (\Phi \Phi^T)^{-1} Y^*\|_F^2$$

$$= \|Y^* - \Phi^T \Phi \Psi S\|_F^2$$

$$= \|I_M - \Phi^T \Phi \Psi^2\|_F^2$$

$$\triangleq \varrho_2(\Phi, \Psi)$$

(17)

to design the CS system for preferable performance. Regarding this penalty term, $S$ is also treated as a constant matrix which is calculated with an independent step by (16).

Based on the discussions above, the following multi-objective cost function for simultaneously optimizing sensing matrix and dictionary is proposed:

$$\min_{\Phi, \Psi} \{\varrho(\Phi, \Psi) \triangleq \varrho_1 + \omega \varrho_2\} \quad \text{s.t.} \quad (15) \quad (18)$$

where $\omega$ is the weighting factor to control the importance of each term.

Remark 3:

- The system design strategy of this paper can be summarized as: I. update sparse coefficients $S$ with (16); II. update system pair $(\Phi, \Psi)$ with (18). Iterating these two steps results in our desired CS system composed of optimal $(\Phi, \Psi)$. The first step is associated with the signal reconstruction accuracy from the measurements directly, while the second step is focused on the properties of system matrices $\Phi$ and $\Psi$. On one hand, this cost function (18) makes sensing matrix and dictionary connect to each other closely so that information is allowed to be fully exchanged between these two members. On the other hand, the dictionary is learned mainly based on the signals, and this operation makes it return to original intention of training dictionary as indicated in sparse representation theory.

- In many existing works concerning designing robust sensing matrix against sparse representation error, the penalty term $\|\Phi E\|_F^2 = \|\Phi (X - \Psi S)\|_F^2$ was taken into consideration [19]–[21], [26]. In our opinion, if we include this term in designing system pair $(\Phi, \Psi)$, the property of $\Psi$ may be destroyed from representing original signals sparsely when the tradeoff strategy is not properly made. Besides, as illustrated in [21], for cases where there are no sufficient number of data available to obtain the error matrix $E$, or where a dictionary should be trained on huge datasets with millions of training samples and in a dynamic CS system for streaming signals, it is usually prohibitive to compute and store $E$ for designing $\Phi$. Moreover, in fact, this penalty term is exactly the one for calculating sparse coefficients $S$ in our first step (16), and the effect of this term on $\Phi$ can be implicitly reflected via $\varrho_2(\Phi, \Psi)$ with the obtained $S$. This also avoids the repetitive optimization of $\|\Phi (X - \Psi S)\|_F^2$.

- Comparing with [22] and [23], besides an independent step for updating sparse coefficients, another main improvement of this paper is the cost function (18) for simultaneously optimizing sensing matrix and dictionary. In the two mentioned references, though model (12) embeds the sensing matrix in the design of dictionary, the convergence of (12) cannot be guaranteed.

\[3\] For (18), though $\varrho_1$ and $\varrho_2$ are both related with $\Psi$, a well-chosen weighting factor $\omega$ will result in a $\Psi$ with favourable sparse representation performance. This will be verified by the experiments.
strictly, as the update procedure for sensing matrix actually does not follow the same cost function. In this paper, we design the pair \((\Phi, \Psi)\) using (18) sternly, so better convergence behavior can be expected.

### IV. ITERATIVE ALGORITHM FOR CS SYSTEM DESIGN

This section is devoted to the detailed algorithm description of the CS system design. The pseudo-code is given below:

**Algorithm**

Initialization:
- \(\Phi_0\) - initial sensing matrix constrained by (15);
- \(\Psi_0\) - initial dictionary;
- \(X\) - training data;
- \(K\) - sparsity level;
- \(\text{Iter}_1\) - iteration number for system design;
- \(\text{Iter}_2\) - iteration number for sensing matrix and dictionary update.

**Begin:** For \(1 \leq i \leq \text{Iter}_1\), repeat Step I - Step III:
- **Step I:** With \(\Phi_{i-1}\) and \(\Psi_{i-1}\), compute the measurement matrix \(Y = \Phi_{i-1}X\) and solve the following
  \[
  S_i = \arg \min_{S} \|Y - \Phi_{i-1}\Psi_{i-1}S\|_F^2 \\
  \text{s.t.} \quad \|S(:, k)\|_0 \leq K, \quad \forall \ k
  \]  
  (19)
to get the sparse coefficients \(S_i\) by OMP algorithm.
  Set \(S = S_i, j = 1\) and \(\Phi_{(0)i} = \Phi_{i-1}\).
- **Step II:** While \(1 \leq j \leq \text{Iter}_2\), execute the followings:
  - For \(\Phi_{(j-1)i}\) fixed, update dictionary by
    \[
    \Psi_{(j)i} = \arg \min_{\Psi} g(\Phi_{(j-1)i}, \Psi) \\
    \]  
    (20)
  - For \(\Psi_{(j-1)i}\) fixed, update sensing matrix by
    \[
    \Phi_{(j)i} = \arg \min_{\Phi} g(\Phi, \Psi_{(j)i}) \quad \text{s.t.} \quad (15)
    \]  
    (21)

Go to Step II with \(j \to j + 1\).
- **Step III:** Set \(\Phi_i = \Phi_{\text{Iter}_2}\) and \(\Psi_i = \Psi_{\text{Iter}_2}\), then go to **Begin** with \(i \to i + 1\) to continue the iterative procedure.

**End:** End the algorithm, and the designed CS system is composed of \((\Phi_{\text{Iter}_1}, \Psi_{\text{Iter}_1})\).

As (19) is attacked by OMP method, the remaining problems in the proposed algorithm are to update the dictionary and sensing matrix, i.e., solving (20) and (21). For convenience, the superscripts and subscripts will be omitted in the following.

### A. UPDATE DICTIONARY \(\Psi\)

With some algebraic operations, \(g(\Phi, \Psi)\) indicated in (18) can be rewritten as

\[
g(\Phi, \Psi) = \|X - \Psi S\|_F^2 + \omega \|(I_N - \Phi^T \Phi) \Psi S\|_F^2 \\
= \|A - B \Psi S\|_F^2
\]  
(22)

where both

\[
A \triangleq \begin{bmatrix} X^T \ 0 \end{bmatrix}, \quad B \triangleq \begin{bmatrix} \sqrt{\omega} (I_N - \Phi^T \Phi) \end{bmatrix}
\]

are independent of \(\Psi\).

Considering

\[
\Psi = \arg \min_{\Psi} \|A - B \Psi S\|_F^2
\]  
(23)

this problem can be solved by the following theorem proposed in [23].

**Theorem 1:** Let \(A, B\) and \(S\) be known and the SVDs of \(B\) and \(S\) be given below

\[
B = U_B \begin{bmatrix} S_B & 0 \\ 0 & 0 \end{bmatrix} V_B^T, \quad S = U_S \begin{bmatrix} S_S & 0 \\ 0 & 0 \end{bmatrix} V_S^T
\]

Then the solutions to (23) are characterized by

\[
\Psi = V_B \begin{bmatrix} S_B^{-1} \tilde{A}_{11} S_S^{-1} \tilde{\Psi}_{12} \tilde{\Psi}_{21} \\ \tilde{\Psi}_{12} \tilde{\Psi}_{21} \end{bmatrix} U_S^T
\]  
(24)

where \(\tilde{A}_{11}\) is the first block of

\[
\tilde{A} \triangleq U_B^T A V_S = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}
\]

with dimension \(\text{rank}(B) \times \text{rank}(S)\), and \(\tilde{\Psi}_{12}, \tilde{\Psi}_{21}\) and \(\tilde{\Psi}_{22}\) are all arbitrary matrices of proper dimensions. In the experiment part, \(\tilde{\Psi}_{12}, \tilde{\Psi}_{21}\) and \(\tilde{\Psi}_{22}\) are all generated with normally distributed entries.

### B. UPDATE SENSING MATRIX \(\Phi\)

When considering updating sensing matrix \(\Phi\), we can rewrite \(g(\Phi, \Psi)\) as follows

\[
g(\Phi, \Psi) = \|X - \Psi S\|_F^2 + \omega \|(I_N - \Phi^T \Phi) \Psi S\|_F^2 \\
= \|X - \Psi S\|_F^2 + \omega \text{tr}[(\Psi S)^T (\Psi S)] \\
= \|X - \Psi S\|_F^2 + \omega \|\Psi S\|_F^2
\]

(25)

with \(\text{tr}[\cdot]\) denoting the trace operation and

\[
C \triangleq \|X - \Psi S\|_F^2 + \omega \|\Psi S\|_F^2
\]

independent from \(\Phi\). Note that the second equation of (25) holds because of (15) which leads to \(\Phi^T \Phi = I_M\). So the problem (21) is equivalent to

\[
\Phi = \arg \max_{\Phi} \|\Phi \Psi S\|_F^2 \quad \text{s.t.} \quad \Phi = U [I_M \ 0] V^T
\]  
(26)

Define \(Q \triangleq \Psi S (\Psi S)^T\) a positive semi-definite matrix and

\[
Q = V_Q \Sigma_Q V_Q^T
\]

be the SVD of \(Q\), where

\[
\Sigma_Q = \text{diag}(\sigma_1, \ldots, \sigma_n, \ldots, \sigma_N)
\]

with

\[
\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq \sigma_{n+1} \geq \cdots \geq \sigma_N \geq 0
\]  
(27)
Denote \( \tilde{V} \triangleq V^T_Q V \), then

\[
\| \Phi \Psi S \|_F^2 = \| U [I_M 0] V^T \Psi S \|_F^2 = \text{tr}[[I_M 0] V^T \Psi S \{[I_M 0] V^T \Psi S\}]^T
\]

\[
= \text{tr}[[I_M 0] V^T Q V [I_M 0]]
\]

\[
= \text{tr}[[I_M 0] \tilde{V}^T \Psi Q \tilde{V} [I_M 0]]
\]

\[
= \text{tr}[\Sigma Q \tilde{V} [I_M 0] V^T]
\]

\[
= \sum_{n=1}^{N} \sigma_n \tilde{V}(n,:) [I_M 0] 0 \tilde{V}(n,:)^T
\]

(28)

According to

\[
\tilde{V}(n,:) [I_M 0] 0 \tilde{V}(n,:)^T \leq \| \tilde{V}(n,:) \|_2^2 = 1
\]

and

\[
\sum_{n=1}^{N} \tilde{V}(n,:) [I_M 0] 0 \tilde{V}(n,:)^T = \text{tr} \tilde{V} [I_M 0] 0 \tilde{V}^T
\]

\[
= M
\]

(30)

one can conclude that the expression (28) satisfies

\[
\sum_{n=1}^{N} \sigma_n \tilde{V}(n,:) [I_M 0] 0 \tilde{V}(n,:)^T \leq \sum_{n=1}^{M} \sigma_n
\]

(31)

and that the equality holds if and only if

\[
\tilde{V}(n,:) [I_M 0] 0 \tilde{V}(n,:)^T = 1,
\]

\[
\forall n \leq M
\]

\[
\Leftrightarrow \tilde{V} = \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}
\]

where \( V_{11} \in \mathbb{R}^{M \times M} \) and \( V_{22} \in \mathbb{R}^{(N-M) \times (N-M)} \) are two arbitrary orthonormal matrices. Hence

\[
V = V_Q \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}
\]

with \( V_Q \) defined as the singular vector matrix of \( \Psi S(\Psi S)^T. \)

So the solution of (26), as well as (21), can be expressed as

\[
\Phi = U [I_M 0] (V_Q \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix})^T
\]

(32)

with \( U \in \mathbb{R}^{M \times M} \) an arbitrary orthonormal matrix. In the experiment part, \( U, V_{11} \) and \( V_{22} \) are all set to identity matrices of proper dimensions for convenience.

With the analytical solutions of (24) and (32), after several iterations, we can reach the optimal dictionary and sensing matrix of (18).

**Remark 4:**

- Take a close look at the convergence behavior of the new cost function for simultaneously optimizing sensing matrix and dictionary, i.e., Step II in the proposed algorithm concerning index \( j \). This is an iterative procedure, resulting in a series of \((\Phi_j^{(i)}, \Psi_j^{(i)})\). When updating dictionary

\[
\Psi_j^{(i)} = \arg \min_{\Psi} \varrho(\Phi_j^{(i-1)}, \Psi)
\]

the solution of the above problem is derive analytically as (24). So \( \varrho(\Phi_j^{(i-1)}, \Psi_j^{(i)}) \leq \varrho(\Phi_j^{(i-1)}, \Psi_j^{(i-1)}) \) holds. Then turn to consider

\[
\Phi_j^{(i)} = \arg \min_{\Phi} \varrho(\Phi, \Psi_j^{(i)}) \text{ s.t. } (15)
\]

(15) just constrains the searching space for optimal sensing matrix and does not make change to the cost function. With the closed-form solution (32), we can conclude that under the constraint (15), \( \varrho(\Phi_j^{(i)}, \Psi_j^{(i)}) \leq \varrho(\Phi_j^{(i-1)}, \Psi_j^{(i)}) \) in summary, for index \( 1 \leq j \leq \text{Iter}_2 \),

\[
\varrho(\Phi_j^{(i)}, \Psi_j^{(i)}) \leq \varrho(\Phi_j^{(i-1)}, \Psi_j^{(i)}) \leq \varrho(\Phi_j^{(i-1)}, \Psi_j^{(i-1)})
\]

i.e., \((\Phi_j^{(i)}, \Psi_j^{(i)})\) makes \( \varrho \) a monotonically nonincreasing function. So the convergence of the process for optimizing sensing matrix and dictionary can be guaranteed.

- For the iteration indexed by \( i \) in the algorithm, it is hard to say whether it converges or not, as the update of sparse coefficients is an independent procedure from the cost function \( \varrho(\Phi, \Psi) \). This is treated as a specialty of our paper. In some existing works such as [22], [23] and [26], \( S \) is included in the design with \( \Phi, \Psi \). When discussing the convergence performance, the authors always assumed that the OMP based techniques can be performed perfectly and the best sparse approximations to the signals are obtained. However, it is not guaranteed whether this assumption holds all the time for all the cases.

- Regarding the results of (24) and (32), there still exists several degrees of freedom, such as \( \Psi_{12}, \Psi_{21}, \Psi_{22} \) and orthonormal matrices \( U, V_{11}, V_{22} \). If we can find certain significative criterions to further design \( \Phi \) and \( \Psi \), these degrees of freedom may be optimized.

**V. EXPERIMENT RESULTS**

In this section, we evaluate the performance of the system design method using real image data, especially medical images. Firstly, the convergence performance of minimizing the cost function \( \varrho(\Phi, \Psi) \) is assessed with synthetic data. Then the choice of the weighting factor \( \omega \) is discussed with image data. The extensive experiments on both image data and real images are executed based on the results of the above two.
A. CONVERGENCE BEHAVIOR OF (18)

In the first place, we carry out several simulations to examine the convergence behavior of optimizing cost function \( \varrho(\Phi, \Psi) \), i.e., Step II in the proposed algorithm.

An \( N \times L \) \( \Psi \) is generated with normally distributed entries as the ground-truth dictionary. One more random \( N \times L \) dictionary \( \Psi^{(0)} \) and an \( M \times N \) matrix \( \Phi^{(0)} \) of structure (15) is respectively set as the initial condition.

The data for training and testing is generated as follows. Firstly, we produce \( 2P \) \( K \)-sparse \( L \times 1 \) vectors \( \{s_k\}_{k=1}^{2P} \), where each non-zero element of \( s_k \) is randomly positioned with an Gaussian distribution of i.i.d. that has zero-mean and unit variance. With the ground-truth dictionary \( \Psi \), the set of signal vectors \( \{x_k\}_{k=1}^{2P} \) is generated by

\[
x_k = \Psi s_k, \quad k = 1, 2, \ldots, 2P
\]

Arrange \( \{x_k\}_{k=1}^{2P} \) as an \( N \times 2P \) matrix \( X = \{x_k\}_{k=1}^{2P} \). Then divide \( X \) into two equal parts, i.e.,

\[
X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}
\]

with \( X_1 \) and \( X_2 \) both of dimension \( N \times P \). \( X_1 \) is used for training \( (\Phi, \Psi) \) and \( X_2 \) is used for testing purpose.

The measurements \( Y \triangleq \{y_k\} \) are obtained using \( Y = \Phi X_2 \) for \( \Phi \) of each iteration. The reconstruction accuracy is quantified with the mean square error (MSE) defined as [13]

\[
\sigma_{\text{mse}} \triangleq \frac{1}{N \times P} \|\hat{X}_2 - X_2\|_F^2
\]

where \( \hat{X}_2 \) is the recovered version of \( X_2 \) with its \( k \)-th atom given by \( \hat{s}_k = \Psi \hat{x}_k \) with updated \( \Psi \), and \( \hat{s}_k \) is the solution of the following minimization:

\[
\hat{s}_k = \arg \min_{s_k} \|s_k\|_0 \text{ s.t. } y_k = \Phi \Psi s_k
\]

which is solved using the OMP algorithm.

Set \( K = 4, M = 20, N = 80, L = 120, \text{Iter}_2 = 1000 \) and \( P = 1000 \).

Figure 1 depicts the evolution of the cost function \( \varrho(\Phi^{(j)}, \Psi^{(j)}) \). The corresponding evolutions of the iterates \( \Phi^{(j)} \) and \( \Psi^{(j)} \), measured using

\[
\delta \Phi(j) \triangleq \sqrt{\frac{1}{M \times N}} \|\Phi^{(j)} - \Phi^{(j-1)}\|_F
\]

\[
\delta \Psi(j) \triangleq \sqrt{\frac{1}{N \times L}} \|\Psi^{(j)} - \Psi^{(j-1)}\|_F
\]

are shown in Figures 2 and 3, respectively. The evolution of \( \sigma_{\text{mse}} \) for the CS system composed of \( (\Phi^{(j)}, \Psi^{(j)}) \) is given in Figure 4.

Remark 5: As seen from Figures 1 to 4, it is clear that the cost function \( \varrho(\Phi, \Psi) \), the iterate metrics \( \delta \Phi \) and \( \delta \Psi \), and the MSE performance all converge well within 100 iterations for three choices of \( \omega \). Especially the monotonic decrease performance of \( \varrho(\Phi, \Psi) \) verifies the statement in point one of Remark 4 very well. It is also observed that the convergence

\[
\delta \Phi \text{ seems a little fluctuant. This phenomenon can be explained by the constraint (15), which limits the searching space for sensing matrix. Broadly speaking, } \delta \Phi \text{ converges anyhow. Moreover, larger } \omega \text{ leads to slower convergence.}
\]

\[
\delta \Psi\text{ of the iterate } \delta \Phi \text{ seems a little fluctuant. This phenomenon can be explained by the constraint (15), which limits the searching space for sensing matrix. Broadly speaking, } \delta \Phi \text{ converges anyhow. Moreover, larger } \omega \text{ leads to slower convergence.}
\]
This is also caused by (15), as the penalty term \( \varrho_2(\Phi, \Psi) \) constrained by (15) is de-emphasized by taking small \( \omega \).

**B. CHOICE OF THE WEIGHTING FACTOR \( \omega \) IN (18)**

From here on, all the experiments will be carried out with image data or real images. The parameters \( K = 4, N = 64 \) and \( L = 256 \) are fixed.

The training data and testing data are real images extracted from the LabelMe database [34] as in [23]. The training and testing sets both consist of 6000 \( 64 \times 1 \) samples, which means that \( P = 6000 \). Denote the clean testing set as \( \{ x_k^* \} \) with each \( x_k^* \in \mathbb{R}^{64 \times 1}, \forall k = 1, \cdots, P \). White Gaussian noise set \( \{ e_k \} \) is added to each of the clean testing data \( x_k^* \) with certain signal-to-noise ratio (SNR) \( \sigma_{\text{snr}} \). The set of signals \( \{ x_k^* + e_k \} \) is used to evaluate the signal reconstruction accuracy of the CS systems.

For image applications, the recovery accuracy is assessed in terms of peak signal-to-noise ratio (PSNR), defined as [13]

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

with \( r = 8 \) bits per pixel and \( \sigma_{\text{mse}} \) the same as in (33).

With \( M = 20, \text{Iter}_1 = 50 \) and \( \text{Iter}_2 = 50 \) (\( \text{Iter}_1 \) and \( \text{Iter}_2 \) are fixed from here on), execute every experiment for 100 times and the average results are saved. Figure 5 shows the effects of different \( \omega \) values on the performance indicator \( \sigma_{\text{psnr}} \) with \( \sigma_{\text{snr}} = \infty \) (i.e., no additional noise is added to the image data).

**Remark 6:** Though there is no systematic way to find the best factor \( \omega \), data experiments show that \( \omega \) taking value around 1 usually leads to high reconstruction accuracy. So \( \omega = 1 \) is set in the sequel.

**C. EXPERIMENTS WITH IMAGE DATA**

Now we start to evaluate the performance of our designed system compared with several existing works. The system composed of sensing matrix and dictionary optimized by the proposed algorithm is denoted as CS\(_{\text{NEW}}\). For comparison, the system composed of K-SVD dictionary [13] with random sensing matrix, and that designed by [22], [23] and [26] will be tested. The corresponding comparisons are denoted as CS\(_{\text{KSVD}}\), CS\(_{\text{DCS}}\), CS\(_{\text{BLL}}\) and CS\(_{\text{LZWH}}\), respectively. The training data and testing data are kept the same as those used in the previous subsection.

We generate an initial sensing matrix \( \Phi_0 \in \mathbb{R}^{M \times 64} \) with all the entries i.i.d. uniformly distributed (also used for CS\(_{\text{KSVD}}\)), then constrain its structure with (15) for CS\(_{\text{NEW}}\). An overcomplete discrete cosine transformation (DCT) dictionary \( \Psi_0 \) of size \( 64 \times 256 \) is created by sampling the cosine wave at different frequencies. The iteration numbers for CS\(_{\text{KSVD}}\), CS\(_{\text{DCS}}\), CS\(_{\text{BLL}}\) and CS\(_{\text{LZWH}}\) are all kept the same as those used in the previous subsections.

When \( \sigma_{\text{snr}} = \infty \), the PSNR performance of the five compared CS systems versus measurement dimension \( M \) is depicted in Figure 6.

**Remark 7:** It can be observed from Figure 6 that the superiority of the proposed system CS\(_{\text{NEW}}\) is more notable with small \( M \) values. When \( M \) increases to more than 30,
the improvements of CS\textsubscript{BLL} and CS\textsubscript{LZWH} are prominent and the performance gets close to CS\textsubscript{NEW}. It should be emphasized that CS is a theory that focuses on both compression and sensing, so the stable performance of CS\textsubscript{NEW} on high compression ratio ($N/M$) makes it remarkable.

The above experiment is carried out with $\sigma_{\text{SNR}} = \infty$. To demonstrate the robustness of the systems, varying SNR cases are executed. For $M = 10, 20, 30$ and $40$, when the input SNR level added to the testing data is varying from $10$ dB to $28$ dB, the PSNR results are shown in Figures 7 - 10.

Remark 8: As can be observed, when considering high compression ratio scenarios, e.g., when $M = 10$, the proposed CS\textsubscript{NEW} performs much better than other methods. It achieves PSNR values almost $2$ dB higher than CS\textsubscript{LZWH} and $4$ dB higher than CS\textsubscript{BLL}. When for $M \geq 20$, the phenomena are consistent with the previous experiment. The performance of CS\textsubscript{BLL} and CS\textsubscript{LZWH} improves dramatically and gets close to CS\textsubscript{NEW}. We still can say that CS\textsubscript{NEW} is more robust than others for all these cases.

The simulations are conducted with MATLAB R2018a, running Microsoft Windows 10. Table 1 summarizes the computing time of each algorithm for four $M$ values. The computing time of CS\textsubscript{NEW} is less than others except CS\textsubscript{LZWH}. In our discussions, designing system is an off-line task and such an increasing in design complexity (compared with CS\textsubscript{LZWH}) is acceptable. As will be seen in the follow-up experiments, the superiority of CS\textsubscript{NEW} over CS\textsubscript{LZWH} is notable, especially for medical image cases.

**D. EXPERIMENTS WITH REAL IMAGES**

Besides the image patches, we carry out a great many of experiments with full images to evaluate the performance of each system.

A frequently-used image quality assessment index for assessing the perceptual recovery result of a full image is the
mean structural similarity (MSSIM) defined in [35]. As clarified by the authors of [35], the value of MSSIM denoted as $\sigma_{mssim}$ is bounded by

$$\sigma_{mssim} \leq 1$$ (35)

FIGURE 11. Twelve tested images.

FIGURE 12. Twelve tested medical images.
with equality holds if and only if the two compared images are exactly the same. The better the image recovers, the bigger of the value $\sigma_{\text{mssim}}$ will be.

Twelve popularly tested images as indicated in Figure 11 are adopted. For every $8 \times 8$ non-overlapping patch in each image, it is re-arranged to a $64 \times 1$ vector. For instance, when considering the ‘Lena’ image of size $512 \times 512$, we get $P = 4096$ vectors for testing. Tables 2 and 3 summarize the recovery PSNR and MSSIM results for these twelve images with $\sigma_{\text{snr}} = \infty$ and $M = 20$. For each test, we highlight the best result with bold.

**Remark 9:** As seen from Tables 2 and 3, though the gap is not so notable, CS$_{\text{NEW}}$ achieves all the best results. Compared with PSNR, MSSIM is a much more illustrative indicator to demonstrate the superiority on image visual effect.

Now we apply the new system design method to medical image compression and reconstruction tasks. As known, it is difficult to diagnose special diseases due to the lack of medical resources or less diagnostic experience from clinicians. Sending the case report of patient with corresponding images to the experts in center hospital by the network is an effective way to improve the diagnosis accuracy. Gastrointestinal endoscopy is an important technique to diagnose gastrointestinal diseases, which generates a huge number of images in each operation. For such a heavy transmission task in remote diagnosis, CS can be employed to improve the transmission efficiency and also improve the reconstruction quality.

The images selected for testing in this part include four upper digestive tract images (UDTI) of size $144 \times 168$ [36] and eight gastric images (GI) of size $392 \times 424$ as shown in Figure 12. Each medical image is preprocessed to a $64 \times P$ testing matrix in a similar way as above. Tables 4 and 5 summarize the recovery PSNR and MSSIM results for these medical images with $M = 10$. For each test, we also highlight the best result with bold.

**Remark 10:** For the medical image applications, the superiority of CS$_{\text{NEW}}$ is demonstrated strongly over all the other compared methods: For PSNR indicator, CS$_{\text{NEW}}$ is averagely 2.5 dB higher than CS$_{\text{LZWH}}$, 3 dB higher than CS$_{\text{BLL}}$ and 3.5 dB higher than CS$_{\text{DCS}}$; while for MSSIM indicator,
TABLE 4. Statistics of PSNR values for tested medical images compressed by different CS systems with $M = 10$.

|          | CS$_{KSY}$ | CS$_{DCS}$ | CS$_{BL}$ | CS$_{LZW}$ | CS$_{NEW}$ |
|----------|------------|------------|-----------|------------|------------|
| UDTI-1   | 33.698     | 36.4962    | 37.3731   | 38.2594    | 41.4759    |
| UDTI-2   | 31.789     | 35.3387    | 35.4651   | 36.5990    | 39.7132    |
| UDTI-3   | 31.6385    | 34.6682    | 34.8430   | 35.7375    | 38.6536    |
| UDTI-4   | 31.1663    | 34.7102    | 34.8639   | 34.9572    | 38.4389    |
| GI-1 (cardia) | 31.7027 | 35.2495    | 35.2164   | 35.3820    | 36.9557    |
| GI-2 (cardia) | 31.7564 | 34.8992    | 35.2309   | 35.7763    | 37.7076    |
| GI-3 (gastric remnant) | 27.0268 | 29.7123    | 30.3127   | 30.8011    | 34.0125    |
| GI-4 (gastric remnant) | 28.4656 | 31.2105    | 31.8506   | 32.3024    | 35.2055    |
| GI-5 (descending duodenum) | 27.6448 | 29.9642    | 30.5849   | 31.0720    | 33.9486    |
| GI-6 (mid-esophageal) | 32.3600 | 36.2767    | 36.2178   | 36.8857    | 38.1994    |
| GI-7 (gastric fundus) | 30.3483 | 32.8295    | 33.6720   | 34.1697    | 36.7674    |
| GI-8 (middle of gastric body) | 30.5095 | 33.7751    | 34.0960   | 34.8189    | 36.8049    |

TABLE 5. Statistics of MSSIM values for tested medical images compressed by different CS systems with $M = 10$.

|          | CS$_{KSY}$ | CS$_{DCS}$ | CS$_{BL}$ | CS$_{LZW}$ | CS$_{NEW}$ |
|----------|------------|------------|-----------|------------|------------|
| UDTI-1   | 0.8409     | 0.8996     | 0.9142    | 0.9298     | 0.9617     |
| UDTI-2   | 0.7976     | 0.8879     | 0.8920    | 0.9104     | 0.9477     |
| UDTI-3   | 0.7955     | 0.8810     | 0.8794    | 0.9045     | 0.9431     |
| UDTI-4   | 0.7730     | 0.8622     | 0.8682    | 0.8827     | 0.9317     |
| GI-1 (cardia) | 0.6837 | 0.8195     | 0.8154    | 0.8358     | 0.8627     |
| GI-2 (cardia) | 0.7300 | 0.8457     | 0.8472    | 0.8667     | 0.8965     |
| GI-3 (gastric remnant) | 0.6533 | 0.7968     | 0.8020    | 0.8333     | 0.8796     |
| GI-4 (gastric remnant) | 0.6684 | 0.8023     | 0.8088    | 0.8348     | 0.8839     |
| GI-5 (descending duodenum) | 0.6778 | 0.8070     | 0.8103    | 0.8350     | 0.8825     |
| GI-6 (mid-esophageal) | 0.7181 | 0.8449     | 0.8413    | 0.8634     | 0.8887     |
| GI-7 (gastric fundus) | 0.7061 | 0.8269     | 0.8236    | 0.8522     | 0.8871     |
| GI-8 (middle of gastric body) | 0.6967 | 0.8283     | 0.8294    | 0.8561     | 0.8906     |

CS$_{NEW}$ is averagely 0.04 higher than CS$_{LZW}$, 0.06 higher than CS$_{BL}$, and CS$_{DCS}$.

Figures 13 - 14 present the visual effects of two gastric images. The corresponding PSNR and MSSIM values can be found in Tables 4 and 5.

Remark 11:
- These two gastric images chosen for showing contain special information. GI-3 and GI-7 show stromal tumor in gastric remnant and small leiomyoma of gastric fundus, respectively. As seen from the reconstructed images, CS$_{NEW}$ always achieves the best visual effect, especially in the edge of interested area, for example, the tumor. This can also be verified by the PSNR and MSSIM results as indicated in Tables 4 and 5. The visual effect of a medical image is pretty important for diagnosing certain diseases.
- CS can save the storage space and speed the transmission of the medical images through the network, especially for huge amount of data. The potential of CS being applied in medical image processing is promising.

VI. CONCLUSION
In this paper, we have investigated the problem of designing optimal CS system. The contribution is three-fold. The first one is to have presented a measurement-driven framework, where the sparse coefficient matrix is calculated adapted to the low-dimension measurements. Under this novel framework, a new cost function that optimizes the sensing matrix and dictionary simultaneously has been proposed as the second contribution. The last one is to have carried out an iterative algorithm to solve for optimum sensing matrix and dictionary. In addition, the superiority of the designed CS system composed of the optimized sensing matrix and dictionary has been demonstrated with a great many of experiments based on real images, especially medical images.

REFERENCES
[1] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[3] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[4] W. Chen and I. J. Wassell, “Optimized node selection for compressive sleeping wireless sensor networks,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 2, pp. 827–836, Feb. 2016.
[5] W. Chen and I. J. Wassell, “Cost-aware activity scheduling for compressive sleeping wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2314–2323, May 2016.
[6] P. Sun, L. Wu, Z. Wang, M. Xiao, and Z. Wang, “Sparsest random sampling for cluster-based compressive data gathering in wireless sensor networks,” *IEEE Access*, vol. 6, pp. 36383–36394, 2018.
[7] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, “Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 221–224, Jan. 2013.
[8] V. Singhal, A. Majumdar, and R. K. Ward, “Semi-supervised deep blind compressed sensing for analysis and reconstruction of biomedical signals from compressive measurements,” *IEEE Access*, vol. 6, pp. 545–553, 2018.
[9] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via {} minimization,” *Proc. Nat. Acad. Sci.ences*, vol. 100, no. 5, pp. 2197–2202, Feb. 2003.

[10] Z. Zhu and M. B. Wakin, “Approximating sampled sinusoids and multiband signals using multiband modulated DPSS dictionaries.” *J. Fourier Anal. Appl.*, vol. 23, no. 6, pp. 1263–1310, Dec. 2017.

[11] Z. Zhu and M. B. Wakin, “Wall clutter mitigation and target detection using discrete prolate spheroidal sequences,” in *Proc. 3rd Int. Workshop Compressed Sens. Theory Appl. Radar, Sonar Remote Sens.* (CoSeRa), Pisa, Italy, Jun. 2015, pp. 41–45.

[12] K. Engan, S. O. Aase, and J. H. Husoy, “Method of optimal directions for frame design,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Phoenix, AZ, USA, Mar. 1999, pp. 2443–2446.

[13] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[14] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[15] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, “A survey of sparse representation: Algorithms and applications,” *IEEE Access*, vol. 3, pp. 490–530, 2015.

[16] Q. Jiang, S. Li, Z. Zhu, H. Bai, X. He, and R. C. de Lamare, “Design of compressed sensing system with probability-based prior information,” *IEEE Trans. Multimedia*, to be published.

[17] M. Elad, “Optimized projections for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[18] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, “On projection matrix optimization for compressive sensing systems,” *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887–2898, Jun. 2013.

[19] G. Li, X. Li, S. Li, H. Bai, Q. Jiang, and X. He, “Designing robust sensing matrix for image compression,” *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5389–5400, Dec. 2015.

[20] H. Bai, S. Li, and X. He, “Sensing matrix optimization based on equiangular tight frames with consideration of sparse representation error,” *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2040–2053, Oct. 2016.

[21] T. Hong and Z. Zhu, “An efficient method for robust projection matrix design,” *Signal Process.*, vol. 143, pp. 200–210, Feb. 2018.

[22] J. M. Duarte-Carvajalino and G. Sapiro, “Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization,” *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.

[23] H. Bai, G. Li, S. Li, Q. Li, Q. Jiang, and L. Chang, “Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1581–1594, Mar. 2015.

[24] X. Ding, W. Chen, and I. J. Wassell, “Joint sensing matrix and sparsifying dictionary optimization for tensor compressive sensing,” *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3632–3646, Jul. 2017.

[25] X. Li, H. Bai, and B. Hou, “A gradient-based approach to optimization of compressed sensing systems,” *Signal Process.*, vol. 139, pp. 49–61, Oct. 2017.

[26] G. Li, Z. Zhu, X. Wu, and B. Hou, “On joint optimization of sensing matrix and sparsifying dictionary for robust compressed sensing systems,” *Digit. Signal Process.*, vol. 73, pp. 62–71, Feb. 2018.

[27] W. Chen and M. R. D. Rodrigues, “Dictionary learning with optimized projection design for compressive sensing applications,” *IEEE Signal Process. Lett.*, vol. 20, no. 10, pp. 992–995, Oct. 2013.

[28] W. Chen, I. J. Wassell, and M. R. D. Rodrigues, “Dictionary design for distributed compressive sensing,” *IEEE Signal Process. Lett.*, vol. 22, no. 1, pp. 55–99, Jan. 2015.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[31] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, “On the use of unit-norm tight frames to improve the average MSE performance in compressive sensing applications,” *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 8–11, Jan. 2012.

[32] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, “Projection design for statistical compressive sensing: A tight frame based approach,” *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016–2029, Apr. 2013.

[33] J. Mairal, F. Bach, and J. Ponce, “Task-driven dictionary learning,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[34] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: A database and Web-based tool for image annotation,” *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, Oct. 2007.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[36] Healthy Upper Digestive Tract. Accessed: Mar. 13, 2019. [Online]. Available: https://www.sciencephoto.com/media/237068/view