Research Article

Optimized Projection Matrix for Compressive Sensing

Jianping Xu, Yiming Pi, and Zongjie Cao

School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

Correspondence should be addressed to Jianping Xu, xujianping1982@hotmail.com

Compressive sensing (CS) is mainly concerned with low-coherence pairs, since the number of samples needed to recover the signal is proportional to the mutual coherence between projection matrix and sparsifying matrix. Until now, papers on CS always assume the projection matrix to be a random matrix. In this paper, aiming at minimizing the mutual coherence, a method is proposed to optimize the projection matrix. This method is based on equiangular tight frame (ETF) design because an ETF has minimum coherence. It is impossible to solve the problem exactly because of the complexity. Therefore, an alternating minimization type method is used to find a feasible solution. The optimally designed projection matrix can further reduce the necessary number of samples for recovery or improve the recovery accuracy. The proposed method demonstrates better performance than conventional optimization methods, which brings benefits to both basis pursuit and orthogonal matching pursuit.

1. Introduction

Compressive sensing (CS) [1–3] has received much attention as it has shown promising results in many applications. CS is an emerging framework, stating that signals which have a sparse representation on an appropriate basis can be recovered from a number of random linear projections of dimension considerably lower than that required by the Shannon-Nyquist theorem. Moreover, compressible signals, that is, the signals' transform coefficients on appropriate basis decay rapidly, can also be sampled at a much lower rate than that required by the Shannon-Nyquist theorem and then reconstructed with little loss of information.

Consider a signal $\mathbf{X} \in R^n$ which can be sparsely represented over a fixed dictionary $\boldsymbol{\Psi} \in R^{n \times k}$ that is assumed to be redundant $(k > n)$. Accordingly, the signal can be described as

$$\mathbf{X} = \boldsymbol{\Psi}\boldsymbol{\theta}, \tag{1}$$

where $\boldsymbol{\theta} \in R^k$ is the coefficient vector which represents $\mathbf{X}$ on $\boldsymbol{\Psi}$ and $\|\boldsymbol{\theta}\|_0 \ll n$. The l^0 norm used here simply counts the number of nonzero element in $\boldsymbol{\theta}$.

CS is an innovative and revolutionary idea that offers a joint sampling and compressing process for such signal. Consider a general linear sampling process which computes m $(m < n)$ inner products between $\mathbf{X}$ and a collection of vectors $\{\boldsymbol{\varphi}_j\}_{j=1}^m$ as $y_j = \langle \mathbf{X}, \boldsymbol{\varphi}_j \rangle$. Arrange the measurements $\{y_j\}_{j=1}^m$ in an $m \times 1$ vector $\mathbf{Y}$ and the sampling vectors $\{\boldsymbol{\varphi}_j^T\}_{j=1}^m$ as rows in an $m \times n$ matrix $\boldsymbol{\Phi}$, then $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta}. \tag{2}$$

The original $\mathbf{X}$ can be reconstructed from $\mathbf{Y}$ by exploring its sparse expression, that is, among all possible $\hat{\boldsymbol{\theta}}$ that satisfies $\mathbf{Y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\hat{\boldsymbol{\theta}}$, seek the sparsest. If this representation coincides with $\boldsymbol{\theta}$, a perfect reconstruction of the signal in (1) is gotten. This reconstruction requires the solution of

$$\min \left\|\hat{\boldsymbol{\theta}}\right\|_0, \quad \text{subject to } \mathbf{Y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\hat{\boldsymbol{\theta}} = \mathbf{D}\hat{\boldsymbol{\theta}}, \tag{3}$$

where $\mathbf{D} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ is defined as the equivalent dictionary.

It is known to be NP-hard in general to solve the problem [4] and different suboptimal strategies are used in practice such as Basis Pursuit (BP) [3] and Orthogonal Matching Pursuit (OMP) [5, 6].

Until now, almost all works on CS made the assumption that $\boldsymbol{\Phi}$ is drawn at random except the one by Elad [7] and the one by Duarte-Carvajalino and Sapiro [8]. In [7], Elad proposed an iterative algorithm. The algorithm

tries to minimize the t-averaged mutual coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$, where $\mathbf{\Psi}$ is fixed. Although the reconstruction performance can be obviously improved, the method is time-consuming because it needs many iterative steps to achieve good performance. Some large mutual coherence values that are not present in the original Gram matrix are also created, which ruin completely the worst case guarantees of the reconstruction algorithms. Instead of targeting on the t-averaged mutual coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$, in [8], Duarte-Carvajalino and Sapiro addressed the problem by making any subset columns of $\mathbf{D}$ as orthogonal as possible, or equivalently, making the Gram matrix as closely as possible to identity matrix. The method is much faster than Elad's but the reconstruction performance is not very good because $\mathbf{D}$ is overcomplete and it could not be an orthogonal basis.

In this paper, a method to optimize the projection matrix is proposed from the perspective of ETF design [9]. The object is to find an equivalent dictionary whose Gram matrix is as close as possible to an ETF's because an ETF has minimum coherence [10]. It is impossible to find an exact solution so an alternating minimization type method is used to find a feasible solution. The proposed method needs few iterative steps to achieve good performance and the reconstruction performance is much better than the existed methods, with both BP and OMP.

The remainder of the paper is organized as follows. In Section 2, the basics of CS are provided along with a statement of the main results in literature relating to this paper. In Section 3, after briefly describing the methods suggested by Elad and Duarte-Carvajalino and Sapiro, a method to optimize the projection matrix is proposed from the aspect of ETF design and an alternative minimization method is proposed to solve the problem. In Section 4, experimental results are presented and the performance obtained with all the optimization methods is compared. Finally, concluding remarks and directions for future research are presented in Section 5.

2. Compressive Sensing: The Basics

CS relies on two fundamental premises: sparsity and incoherence. Sparsity means, in (1), most elements of θ are zero or they can be discarded without much loss of information. Incoherence means, in (2), the projection matrix $\mathbf{\Phi}$ and the sparsifying matrix $\mathbf{\Psi}$ should be as incoherent as possible. A possible measure of coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is given by the inner products of different columns in $\mathbf{D}$ [11, 12]:

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \max_{1 \leq i, j \leq k, \, i \neq j} \left| \left\langle d_i, d_j \right\rangle \right|. \qquad (4)$$

A different way to measure mutual coherence is considering the Gram matrix of equivalent dictionary $\mathbf{G} = \mathbf{D}^H \mathbf{D}$ which is computed after normalizing each column of $\mathbf{D}$. The off-diagonal entries of $\mathbf{G}$ are the inner products that appear in (4). Mutual coherence is the off-diagonal entry with largest magnitude.

Mutual coherence measures the maximal correlation between both matrix elements and plays an important role in the success of reconstruction algorithm. It has been demonstrated that mutual coherence should be as small as possible in CS [2].

Theorem 1. *The necessary number of samples needed to recover signal is confined by*

$$p > C \cdot \mu^2(\mathbf{\Phi}, \mathbf{\Psi}) \cdot S \cdot \log n, \qquad (5)$$

where C is a positive constant, S is the sparsity level of signal, and n is the dimension. It is obvious to see that the smaller the coherence, the fewer samples are needed [2].

Theorem 2. *If the flowing inequality holds,*

$$\left\| \hat{\theta} \right\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mu(\mathbf{\Phi}, \mathbf{\Psi})} \right), \qquad (6)$$

then, $\hat{\theta}$ is necessary the sparsest solution such that $\mathbf{X} = \mathbf{\Psi}\hat{\theta}$ and the reconstruction algorithms are guaranteed to succeed in finding the correct $\hat{\theta}$ [13–15].

From the former discussion, it is easy to see that CS deals with the case of low coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$. With the properties of $\mu(\mathbf{\Phi}, \mathbf{\Psi})$, there is a sensible reason to design the projection matrix in a way that minimizes the mutual coherence $\mu(\mathbf{\Phi}, \mathbf{\Psi})$ which may lead to better performance of reconstruction algorithms.

3. Optimizing Projection Matrix for Compressive Sensing

In this paper, only the case that $\mathbf{\Psi}$ is fixed while $\mathbf{\Phi}$ can be arbitrary is considered. Hence, the object is to optimize $\mathbf{\Phi}$ that will minimize the mutual coherence $\mu(\mathbf{\Phi}, \mathbf{\Psi})$. After reviewing the former related work, the proposed algorithm is introduced.

3.1. Elad's Method [7]. Instead of mutual coherence, Elad considered a different coherence—t-averaged mutual coherence which reflects the average behavior. The t-averaged mutual coherence of $\mathbf{D}$ is defined as the average of all absolute and normalized inner products between different columns in $\mathbf{D}$ (denoted as g_{ij}) that are above t. Formally

$$\mu_t(\mathbf{\Phi}, \mathbf{\Psi}) = \frac{\sum_{1 \leq i, j \leq k, \, i \neq j} \left(g_{ij} \geq t \right) \cdot \left| g_{ij} \right|}{\sum_{1 \leq i, j \leq k, \, i \neq j} \left(g_{ij} \geq t \right)}. \qquad (7)$$

Putting very simply, the object is to minimize $\mu_t(\mathbf{\Phi}, \mathbf{\Psi})$ with respect to $\mathbf{\Phi}$, assuming that $\mathbf{\Psi}$ and the parameter t are fixed and known. In this algorithm, the main object is the reduction of the absolute inner products $|g_{ij}|$ that are above t. The Gram matrix of the normalized equivalent dictionary is computed and the values above t are "shrinked" by multiplying with γ ($0 < \gamma < 1$). In order to preserve the order of the absolute values in the Gram matrix, entries in

G with magnitude below t but above γt are "shrinked" by a small amount using the following function:

$$g_{ij} = \begin{cases} \gamma g_{ij}, & \left| g_{ij} \right| \geq t, \\ \gamma t \ \text{sign}\left(g_{ij} \right), & t \geq \left| g_{ij} \right| \geq \gamma t, \\ g_{ij}, & \gamma t \geq \left| g_{ij} \right|. \end{cases} \tag{8}$$

The former shrinking operation causes the resulting Gram matrix to become full rank in general case. Thus, the next steps should mend this by forcing the rank to be m and find the matrix $\mathbf{\Phi}$ that could best describe the squared root of the obtained Gram matrix. The process could be realized using SVD. The details can be found in [7].

3.2. Duarte-Carvajalino and Sapiro's Method. Unlike the previous one, Duarte-Carvajalino and Sapiro's method is noniterative. Instead of targeting on t-averaged mutual coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$, this method addressed the problem by making any subset of columns in $\mathbf{D}$ as orthogonal as possible, or equivalently, making the Gram matrix as close as possible to an identity matrix. Their approach was carried out as follows.

Consider the Gram matrix of the equivalent dictionary

$$\mathbf{G} = \mathbf{D}^H \mathbf{D} = \mathbf{\Psi}^H \mathbf{\Phi}^H \mathbf{\Phi} \mathbf{\Psi}. \tag{9}$$

The object is to find such $\mathbf{\Phi}$ that makes the Gram matrix as close as possible to identity matrix

$$\mathbf{\Psi}^H \mathbf{\Phi}^H \mathbf{\Phi} \mathbf{\Psi} \approx \mathbf{I}. \tag{10}$$

Multiplying both sides of (10) with $\mathbf{\Psi}$ on the left and $\mathbf{\Psi}^H$ on the right, it becomes

$$\mathbf{\Psi}\mathbf{\Psi}^H \mathbf{\Phi}^H \mathbf{\Phi} \mathbf{\Psi}\mathbf{\Psi}^H \approx \mathbf{\Psi}\mathbf{\Psi}^H. \tag{11}$$

Now, consider the eigen-decomposition of $\mathbf{\Psi}\mathbf{\Psi}^H$ which is

$$\mathbf{\Psi}\mathbf{\Psi}^H = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H. \tag{12}$$

Then (11) becomes

$$\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \mathbf{\Phi}^H \mathbf{\Phi} \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \approx \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H, \tag{13}$$

which is equivalent to

$$\mathbf{\Lambda}\mathbf{V}^H \mathbf{\Phi}^H \mathbf{\Phi} \mathbf{V}\mathbf{\Lambda} \approx \mathbf{\Lambda}. \tag{14}$$

By denoting $\mathbf{\Gamma} = \mathbf{\Phi}\mathbf{V}$, they finally formulated the problem to minimize the following function with respect to $\mathbf{\Gamma}$:

$$\min \left\| \mathbf{\Lambda} - \mathbf{\Lambda}\mathbf{\Gamma}^H \mathbf{\Gamma}\mathbf{\Lambda} \right\|_F^2. \tag{15}$$

By solving the problem of (15), they achieved to optimize the projection matrix. The details to solve the problem can be found in [8].

3.3. The Proposed Method. Elad's method is time-consuming and the shrinkage function creates some large values that are not present in the original Gram matrix. Large off-diagonal values in the Gram matrix ruin completely the worst case guarantees of the reconstruction algorithms. Duarte-Carvajalino and Sapiro's method is noniterative and the reconstruction relative error rate is high. To overcome these drawbacks, a method based on ETF design is proposed in this paper. The object is to find an equivalent dictionary which is as close as possible to an ETF because of the minimum coherence property of ETF, and then from the equivalent dictionary, the optimized projection matrix can be constructed. It is impossible to solve the problem exactly because of the complexity, so an alternative minimization type method is used to find a feasible solution.

Firstly, model the problem as an optimization problem. Let $\mathbf{G} = \mathbf{D}^H \mathbf{D}$ be the Gram matrix of $\mathbf{D}$. The mutual coherence of $\mathbf{D}$ is the maximum absolute value of the off-diagonal entry of $\mathbf{G}$, supposing the columns of $\mathbf{D}$ are normalized. For such $\mathbf{D}$, if the magnitudes of all off-diagonal entries of $\mathbf{G}$ are equal, $\mathbf{D}$ has minimum coherence [10]. This normalized dictionary is called an ETF. Although this type of frame has many nice properties, ETF does not exist for any arbitrary selection of dimension. Therefore the optimization process aims at finding the nearest admissible solution which is as close as possible to an ETF.

For the normalized equivalent dictionary $\mathbf{D} \in R^{m \times k}$, the mutual coherence of $\mathbf{D}$ is defined as

$$\mu_{\mathbf{D}} = \max_{i,j, \ i \neq j} \left\{ \left\langle \mathbf{d}_i, \mathbf{d}_j \right\rangle \right\}. \tag{16}$$

A column normalized dictionary $\mathbf{D}$ is called ETF when there is a constant r $(0 < r < \pi/2)$ that

$$\left| \left\langle \mathbf{d}_i, \mathbf{d}_j \right\rangle \right| = \cos(r) \quad \forall \ i, j, \ i \neq j. \tag{17}$$

Strohmer and Heath Jr. in [16] showed that if there is an ETF in the set of $m \times k$ uniform frames, it is the solution of

$$\arg \min_{\mathbf{D} \in R^{m \times k}} \left\{ \mu_{\mathbf{D}} \right\}. \tag{18}$$

To study the lower bound of $\mu_{\mathbf{D}}$, the existence of an ETF and its Gram matrix, Strohmer showed that $\mu_{\mathbf{D}}$ is lower bounded by

$$\mu_{\mathbf{D}} \geq \mu_G = \sqrt{\frac{k - m}{m(k - 1)}}. \tag{19}$$

Let $\mathbf{A}_m^k$ be the set of Gram matrices of all $m \times k$ ETF. If $\mathbf{G}_G \in \mathbf{A}_m^k$, then the diagonal elements and the absolute values of the off-diagonal elements of $\mathbf{G}_G$ are one and μ_G, respectively. A nearness measure of $\mathbf{D} \in R^{m \times k}$ to the set of ETF can be defined as the minimum distance between the Gram matrix of $\mathbf{D}$ and $\mathbf{G}_G$. To minimize the distance of a dictionary to ETF, it needs to solve

$$\min_{\mathbf{G}_G \in \mathbf{A}_m^k} \left\| \mathbf{D}^H \mathbf{D} - \mathbf{G}_G \right\|_\infty. \tag{20}$$

The matrix operator $\| \cdot \|_\infty$ is defined as the maximum absolute value of the elements in the matrix. Instead, it is

better to use a different norm space which simplifies the problem. An advantage of using l_2 in the given problem is that it considers the errors of all elements. Therefore it forms the following formulation:

$$\min_{\mathbf{G}_G \in \mathbf{A}_m^k} \left\| \mathbf{D}^H \mathbf{D} - \mathbf{G}_G \right\|_F^2, \tag{21}$$

where $\| \cdot \|_F$ is Frobenius norm. This is a nonconvex optimization problem in general. It might have a set of solutions or have no solution. Extend $\mathbf{A}_m^k$ to a convex set $\mathbf{\Lambda}^k$, which is not empty for any k,

$$\mathbf{\Lambda}^k$$

$$= \left\{ \mathbf{G}_\Lambda \in R^{k \times k} : \mathbf{G}_\Lambda = \mathbf{G}_\Lambda{}^T, \ \text{diag } \mathbf{G}_\Lambda = 1, \ \max_{i \neq j} \left| g_{ij} \right| \leq \mu_G \right\}. \tag{22}$$

Relaxing (21) by replacing $\mathbf{A}_m^k$ with $\mathbf{\Lambda}^k$, it gives the following optimization problem:

$$\min_{\mathbf{G} \in \mathbf{\Lambda}^k} \left\| \mathbf{D}^H \mathbf{D} - \mathbf{G}_\Lambda \right\|_F^2. \tag{23}$$

A standard method to solve (23) is alternating projection [17]. In this work, a different method which has similarities with alternating projection is used. Although the proposed solution has similarities with alternating projection, it does not follow its steps exactly. The difference lies in the stage of updating the current solution with respect to $\mathbf{\Lambda}^k$. A point between the current solution and the projection on $\mathbf{\Lambda}^k$ is chosen, that is because after being projected onto $\mathbf{\Lambda}^k$, the structure of the Gram matrix changes significantly and the selection of a new point in the following step is very difficult. After performing alternating minimization, the optimized projection matrix can be constructed from the output Gram matrix with a rank revealing QR factorization with eigenvalue decomposition. The details can be found in [18].

The conditions under which the algorithm converges can be found in [17].

The following are the steps of the proposed algorithm for optimizing the projection matrix, supposing the sparsifying matrix is known.

(1) Initialize: the projection matrix $\mathbf{\Phi}$, sparsifying matrix $\mathbf{\Psi}$, and equivalent dictionary $\mathbf{D}$, iterative steps l

　　For $p = 1 \cdots l$

(2) Compute the Gram matrix $\mathbf{G} = \mathbf{D}^H \mathbf{D}$, denote the element of $\mathbf{G}$ as g_{ij}

(3) Project the Gram matrix onto $\mathbf{\Lambda}^k$, that is,

$$g_{ij} = \begin{cases} 1, & i \neq j, \\ g_{ij}, & \text{abs}\left(g_{ij}\right) < \mu_G, \\ \text{sign}\left(g_{ij}\right) \cdot \mu_G, & \text{else} \end{cases} \tag{24}$$

(4) Choose a point between the current solution and the projection on $\mathbf{\Lambda}^k$ to update the Gram matrix

$$\mathbf{G}_{p+1} = \alpha \mathbf{G}_p + (1 - \alpha)\mathbf{G}_{p-1}, \quad 0 < \alpha < 1 \tag{25}$$

(5) Update the projection matrix $\mathbf{\Phi}$ using QR factorization with eigenvalue decomposition

　　end

4. Experiment Results

Firstly, the distribution change of inner products between different columns of $\mathbf{D}$ before and after optimization is considered. Figure 1 presents the distribution of the off-diagonal elements of the Gram matrix in absolute value, obtained using a fixed sparsifying matrix for four different projection matrices. The four projection matrices considered are listed as follows: a Gaussian random matrix and the matrices obtained by the three optimization methods mentioned above. All the three optimization methods try to reduce the largest off-diagonal elements in the Gram matrix. However, Elad's method always presents a consistent artifact, where some off-diagonal elements in the Gram matrix actually increase their values, which ruin completely the performance of reconstruction algorithm. Both the method proposed in [8] and our proposed algorithm reduce the number of large absolute values in the Gram matrix. But in our proposed method, the absolute values concentrate around μ_G which can make the equivalent dictionary as close as possible to an ETF. This better behavior can further reduce the necessary samples for recovery and improve the recovery performance.

Then, the performance of recovery algorithm in CS is evaluated before and after projection matrix optimization. In order to compare the performance, the test proposed by Elad in [7] is used. The test includes the following steps.

Step 1. Generate data: choose a dictionary $\mathbf{\Psi} \in R^{n \times k}$ and synthesize N test signals $(\mathbf{x}_j)_{j=1}^N$ by generating N sparse vectors $(\boldsymbol{\eta}_j)_{j=1}^N$ of length k each, and computing for all j, $\mathbf{x}_j = \mathbf{\Psi}\boldsymbol{\eta}_j$. All representations are built using the same cardinality $\|\boldsymbol{\eta}\|_0 = T$.

Step 2. Initial projection: for a chosen number of measurement m, create a random projection matrix $\mathbf{\Phi}$ and apply it to the signals, obtaining $\mathbf{y}_j = \mathbf{\Phi}\mathbf{x}_j$, for all j, Compute the equivalent dictionary $\mathbf{D}$.

Step 3. Performance test: apply BP and OMP to reconstruct the signals by approximating the solution of $\hat{\boldsymbol{\eta}}_j = \arg\min_{\boldsymbol{\eta}}\|\boldsymbol{\eta}_j\|_0$ subject to $\mathbf{y}_j = \mathbf{D}\boldsymbol{\eta}_j$. Test the error $\|\mathbf{x}_j - \mathbf{\Psi}\hat{\boldsymbol{\eta}}_j\|_2$. Measure the average error rate—a reconstruction with a mean squared error above some threshold is considered as a reconstruction failure.

Step 4. Optimize the projection matrix: use the three methods mentioned above to optimize the projection matrix.
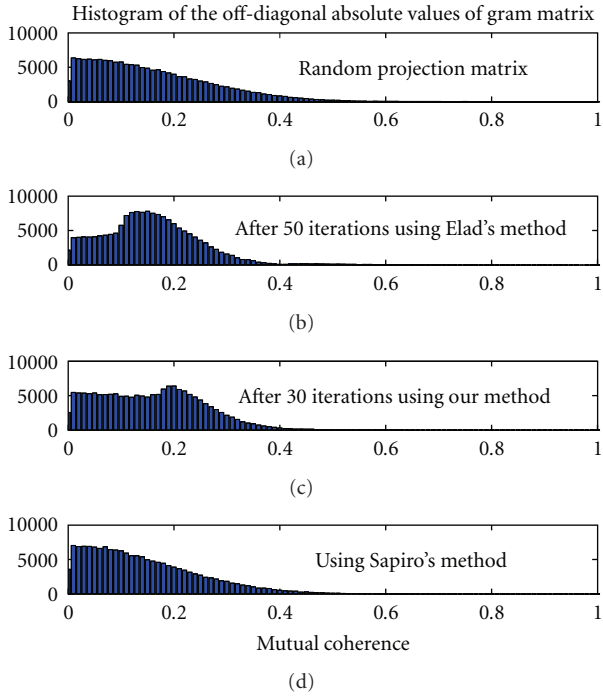
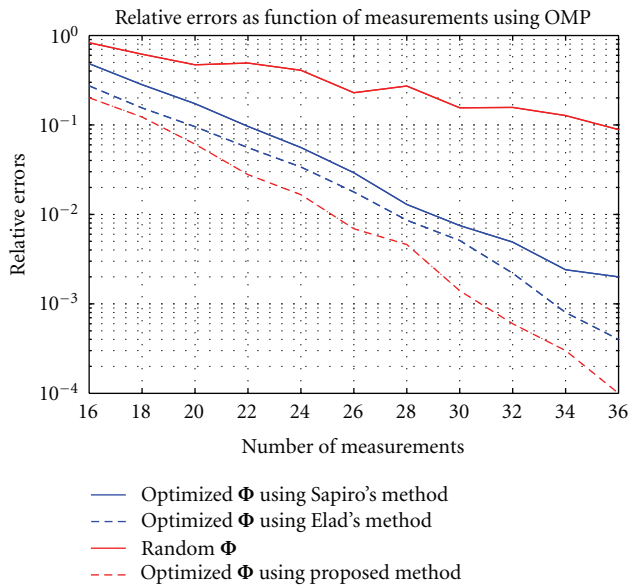FIGURE 1: Histogram of off-diagonal absolute values of *G* before and after optimization.

FIGURE 2: CS relative errors as function of the number of measurements using OMP.

Step 5. Reapply Steps 2 and 3 using the optimized projection matrix.

The following experiments followed the previously described steps. Figures 2 and 3 show the performance of CS before and after the projection matrix optimization using different optimization methods, with both OMP and BP for varying number of measurements.
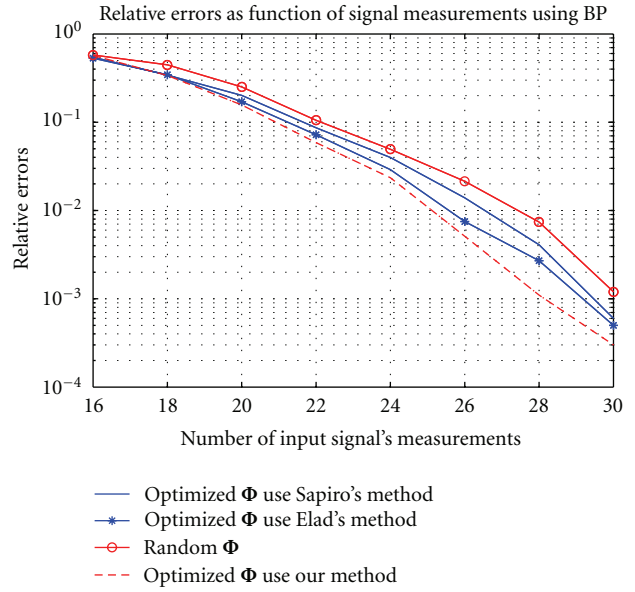

FIGURE 3: CS relative errors as function of the number of measurements using BP.

In the first experiment (Figures 2 and 3), a random dictionary of size 80×120 was used. This size was chosen since it enabled the CS performance evaluation in reasonable time. $N = 10000$ sparse vectors were generated of length $k = 120$ with cardinality $T = 4$. The nonzero's locations were chosen at random. These sparse vectors were used to create the test signals with which the CS performance was evaluated. CS performance was tested with varying values of m. The relative error rate was evaluated as a function of m for both OMP and BP before and after the optimization. Each point in the graph represents an average performance, accumulated over a possible varying number of experiments. While every point was supposed to present an average performance over N signals, in cases where more than 300 errors was accumulated, the test stopped and the average so far was used instead. This was done to reduce the overall runtime.

From Figures 2 and 3, the CS performance improves for both BP and OMP as m increases and in this experiment, BP outperforms OMP. Also, as expected, all the optimization methods lead to improved performance of CS. For some values of m, the performance is improved by 10 : 1 for BP and 100 : 1 for OMP. The optimization method proposed by us outperforms both Elad's method and Duarte-Carvajalino and Sapiro's method. For some values of m, our method improves the CS performance by 5 : 1 than Elad's method and even 10 : 1 than Duarte-Carvajalino and Sapiro's method for OMP. The improvement for BP is smaller than OMP, about 2 : 1 for some values of m.

The second experiment is almost the same as the last one, but m was fixed and T was changeable. From Figures 4 and 5, the CS performance decreases for both BP and OMP as T increases. That is because as T increases, more measurements are needed to achieve the same CS performance. Also, as
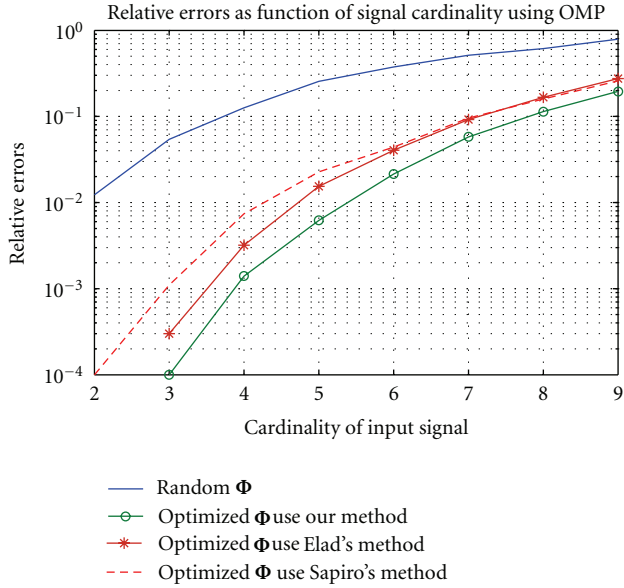
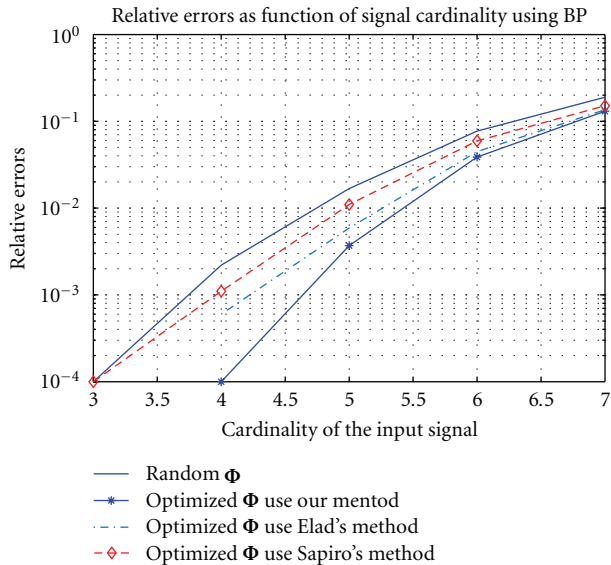FIGURE 4: CS relative errors as function of signal cardinality using OMP.

FIGURE 5: CS relative errors as function of signal cardinality using BP.

FIGURE 6: Relative errors as function of signal's dimension n using BP.
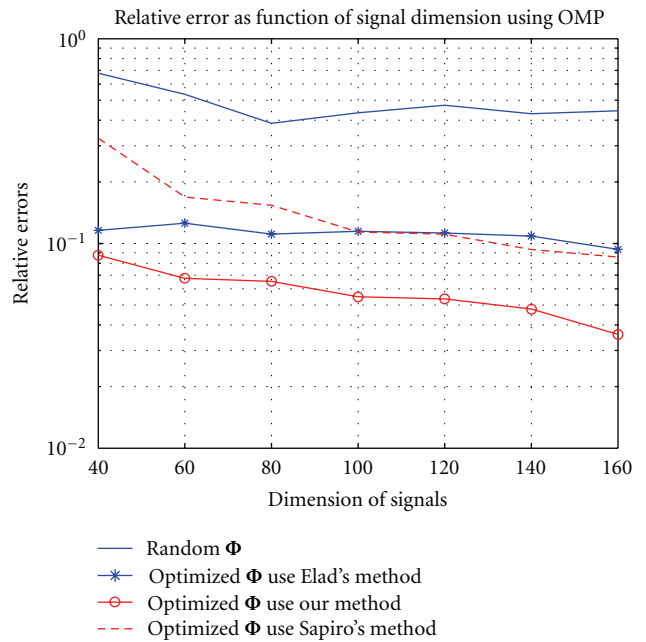
FIGURE 7: Relative errors as function of signal's dimension n using OMP.

expected, the CS performance is improved after using the optimization methods mentioned above with both BP and OMP. The improvement is much larger with OMP than that with BP. It is almost $100:1$ for some values of T with OMP. The CS performance is obviously improved using the proposed optimization method than both Elad's method and Duarte-Carvajalino and Sapiro's method. It is about $4:1$ to Elad's method with OMP. With respect to BP, when T is small, the improvement is larger and when T increases, the CS performance is almost the same for all the optimization methods.
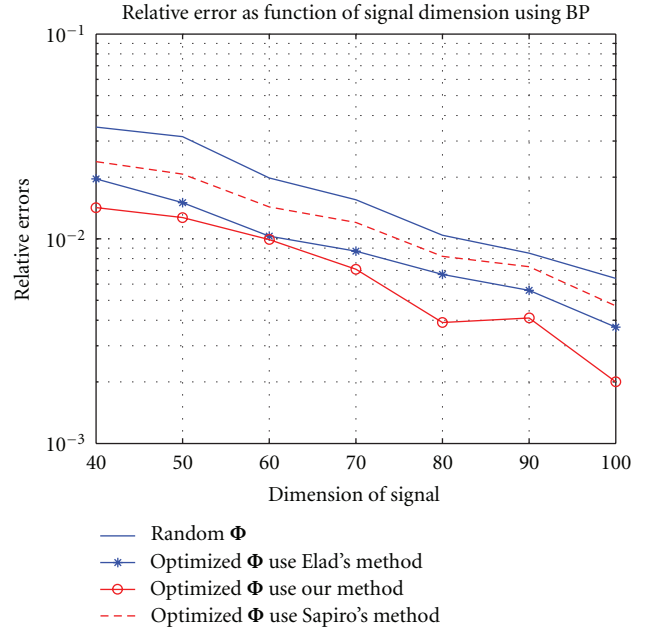
In the following experiment, the effect of signal dimension on CS performance before and after optimization of the projection matrix is considered. In this experiment, the signals' dimension n was varied while proportionally updating measurements m, dictionary dimension k and signal cardinality T with it. The object was to get a better indication to the asymptotic performance as studied in [1–3].
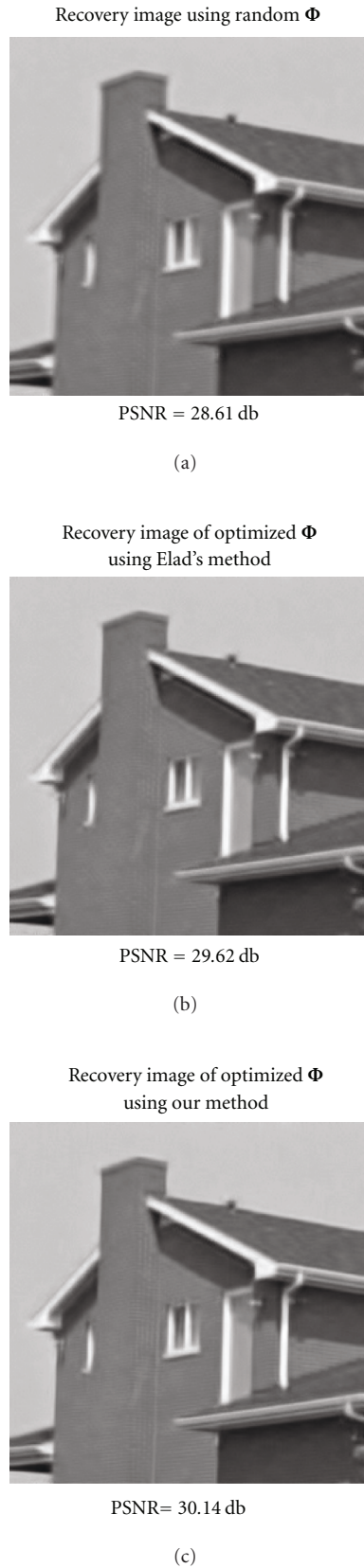
Recovery image using random Φ

PSNR = 28.61 db

(a)

Recovery image of optimized Φ
using Elad's method

PSNR = 29.62 db

(b)

Recovery image of optimized Φ
using our method

PSNR= 30.14 db

(c)

FIGURE 8: Recovered image before and after optimization. (a) Recovered using random Φ, (b) recovered using optimized Φ by Elad's method, and (c) recovered using optimized Φ by our proposed method.

From the results in Figures 6 and 7, there is an insistent improvement before and after optimization with both BP and OMP. It is also obvious that there is great performance improvement using the optimization method proposed by us.

Afterwards, a test image was used to assess the recovery algorithm performance in CS before and after optimizing the projection matrix. The testing image consists of non-overlapping 8×8 patches reconstructed from their noisy projections (5% level of noise).

In this experiment, m is set to be 15 and the sparsifying matrix to be DCT. The reconstruction algorithm used is OMP. From the pictures above, it is obvious to see that the recovery performance is improved for both the optimization methods. The Elad's method can improve PSNR about 1 db than using random Φ, and the proposed method in this paper can improve about 0.5 db than Elad's method. If larger number of measurements is used for recovery, the reconstruction performance could be better, but it is time-consuming.

5. Conclusion

A crucial ingredient in the deployment of the CS idea is the process of random linear projections that mix the signal. This operation has been traditionally chosen as a random matrix. This paper aimed to show that the optimally designed projection matrix can further improve the CS performance. The method was that constructing the projection matrix so as to minimize the coherence between Φ and Ψ based on ETF design. The experimental results demonstrated that the optimally designed projection matrix indeed lead to a better CS performance, not only improvement of the reconstruction accuracy but also reduction in the necessary number of samples for recovery. It also demonstrated that after optimizing the projection matrix using the proposed method, the CS performance was greatly improved than using random projection matrix and the already existed optimization methods.

As this is only one of the very few works to address the problem of optimization the projection matrix, there is still great work to do. Here are some advices for future research. (1) How to perform the proposed method when the signals are of high dimension. (2) Whether there is a more direct method to address the problem which may be easier.

Acknowledgment

References

[1] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[3] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Asilomar Conference on Signals, Systems & Computers*, pp. 40–44, November 1993.

[6] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[7] M. Elad, "Optimized projections for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.

[8] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, 2009.

[9] M. A. Sustik, J. A. Tropp, I. S. Dhillon, and R. W. Heath Jr., "On the existence of equiangular tight frames," *Linear Algebra and Its Applications*, vol. 426, no. 2-3, pp. 619–635, 2007.

[10] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 361–379, 2000.

[11] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[12] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[13] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, 2003.

[15] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[16] T. Strohmer and R. W. Heath Jr., "Grassmannian frames with applications to coding and communication," *Applied and Computational Harmonic Analysis*, vol. 14, no. 3, pp. 257–275, 2003.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computation*, Johns Hopkins University Press, Baltimore, Md, USA, 1996.

[18] J. A. Tropp, I. S. Dhillon, R. W. Heath Jr., and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 188–209, 2005.