Efficient Projection for Compressed Sensing

Vo Dinh Minh Nhat
Ubiquitous Computing Lab
Kyung Hee University
vdmnhat@oslab.khu.ac.kr

Duc Vo
Victorian Research Labs
University of Technology, Sydney
dvo@eng.uts.edu.au

Subhash Challa
Victorian Research Labs
The University of Melbourne
subhash.challa@nicta.com.au

SungYoung Lee
Ubiquitous Computing Lab
Kyung Hee University
sylee@oslab.khu.ac.kr

Abstract

Compressed sensing (CS), a joint compression and sensing process, is a emerging field of activity in which the signal is sampled and simultaneously compressed at a greatly reduced rate. In CS, the projection matrix is chosen at random which would lead to inefficient performance of CS. Recently, an optimized projection (OP) was chosen such that it leads to better coherence of the effective dictionary, leading substantially better CS reconstruction performance. This is the only recent work related to the optimization of projection matrix, but the major drawback of this approach is that its algorithm is an iterative and high complexity one. It is considered that an efficient and lightweight algorithm for designing projection matrix would usefully supplement and enhance the performance of CS. In this paper, we propose an algorithm to obtain such an projection matrix called efficient projection (EP) which is designed in such a way that the new projected dictionary will have the structure as much similar as the original dictionary. For this purpose, we use Multidimensional Scaling (MDS) technique which helps to find a low-dimensional new projected dictionary such that the pairwise distances between new atoms in this new projected dictionary match as well as possible the original atoms in original dictionary. This leads to a solution for EP which is very simple and can be obtained by doing singular value decomposition of original dictionary. The experiments show the novelty of our approach when it can gain a comparative performance with a very low complexity compared to that of OP.

1 Introduction

Compressed Sensing (CS) [2][3][1][8][18][16][9], a joint compression and sensing process, provides a new way to acquire and represent sparse signals that requires less sampling resources than traditional approaches.

1.1 Problem Statement

The mathematical problem statement can be summarized as follow. Given a T represent sparse signal $x \in \Re^n$ over a *dictionary* $\Psi = [\Psi_1, \Psi_2, ..., \Psi_k] \in \Re^{n \times k}$, with $n < k$ and each column of Ψ is called *atom*. Signal x can be described as $x = \Psi\alpha$ with $\alpha \in \Re^k$ and $\|\alpha\|_0 < T \ll n$. It is noted that $\|\alpha\|_0$ is the l^0-norm of α which counts the number of non-zeros in α. In CS, signal x is compressed by a projection matrix $\Phi \in \Re^{p \times n}$ with $T < p \ll n$, which yields the sensing vector $y = \Phi x \in \Re^p$ (see Fig. 1). Now the objective is to reconstruct the original signal x given *dictionary* Ψ, projection matrix Φ and $y = \Phi x = \Phi\Psi\alpha$. Let $\Omega = \Phi\Psi \in \Re^{p \times k}$, then a representation of $x \in \Re^n$ can be thought of as a vector $\alpha \in \Re^k$ satisfying

$$y = \Omega\alpha \tag{1}$$

Because $p \ll k$, the problem (1) is underdetermined (i.e. there is no unique solution to the problem), so CS reconstruct x from y by exploiting the sparsity (i.e., among all possible α satisfying $y = \Omega\alpha$ it seeks the sparsest). The reconstruction problem can be written as

$$(P_0) \quad \hat{\alpha} = \arg\min_\alpha \|\alpha\|_0 \quad s.t. \quad y = \Omega\alpha \tag{2}$$

After getting $\hat{\alpha}$, the reconstructed signal is calculated as $\hat{x} = \Psi\hat{\alpha}$.

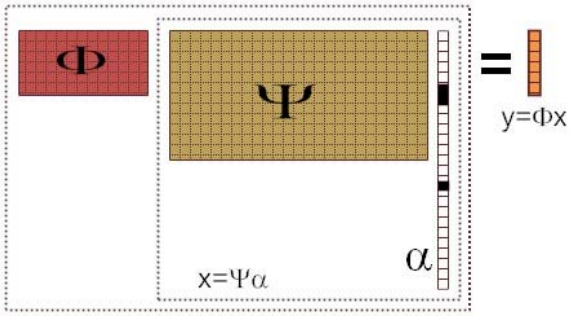


Figure 1. The joint compression and sensing process.

1.2 Current Solutions to CS

Several methods have been developed and introduced to solve the problem of finding sparse representation α to reconstruct original signal x which basically consists of Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP) and Basis Pursuit (BP). Matching pursuit (MP), a greedy algorithm, proposed in [14] is an iterative procedure of finding a sub-optimal signal's representation in a highly redundant dictionary of functions. In MP, a signal representation is iteratively built up by selecting the atom that maximally improves the representation at each iteration. While there is no guarantee that MP computes sparse representations, MP is easily implemented, converges quickly, and has good approximation properties [11]. Moreover, there is one variant of MP, Orthogonal Matching Pursuit (OMP), can be shown to compute nearly sparse representations under some conditions. In [15], they proposed an improved algorithm called Orthogonal Matching Pursuit (OMP) that maintains the backward orthogonality of the residual (error) and leads to a better convergence. OMP also can be used to solve (2) efficiently. Basically, we want to find the sparse representation via l^0-norm, but this is NP-hard problem, so we can relax the problem to l^1-norm since it typically leads to linear programs, which are convex and poly-time solvable. The l^0-norm optimization problem (2) can be relaxed to l^1-norm optimization problem as follow

$$(P_1) \quad \hat{\alpha} = \arg\min_\alpha \|\alpha\|_1 \quad s.t. \quad y = \Omega\alpha \qquad (3)$$

Basis Pursuit (BP) finds sparse representations by convex optimization. It obtains the decomposition that minimizes the l^1-norm of the coefficients occurring in the representation. Because it is based on global optimization, it can stably super-resolve in ways that Matching Pursuit can not [4]. Recent theoretical work also shows that representations computed by BP are guaranteed to be sparse under certain conditions [11].

1.3 Compressed Sensing Applications

Though CS is an emerging field of activity, with its beautiful theoretical results, in recent few years, there has been an certain amount of literature on the application of CS. In [19], they propose algorithms and hardware to support a new theory of compressive imaging which is based on a new digital image/video camera that directly acquires random projections of the signal without first collecting the pixels/voxels. It is very interesting that their camera includes the ability to obtain an image with a single detection element while measuring the image/video fewer times than the number of pixels this can significantly reduce the computation required for video acquisition/encoding. Coming along is the application of CS on medical imaging which is studied in [13]. They develop practical under sampling schemes for MR imaging based on the theory of CS. Compressed Sensing DNA Microarrays was also studied in [17]. A very comprehensive and almost complete reference of CS could be found at http://www.dsp.ece.rice.edu/cs/.

1.4 Focus and Aim

While the choice of dictionary matrix Ψ depends mostly on the model of target signal x, the projection matrix Φ is chosen at random in CS framework. The reason for simple choice of the projection matrix is to simplify its theoretical analysis and also facilitates a simple implementation. But randomly picking up Φ would lead to inefficient performance of CS which is a good motivation to design the projection matrix Φ in an efficient and optimal way. In [9], they show that by optimizing the choice of Φ such that it leads to better coherence of the effective dictionary, a substantially better CS reconstruction performance is obtained, with both BP and OMP algorithms. This is the only recent work related to the optimization of projection matrix, but the major drawback of this approach is that its algorithm is an iterative and high complexity one. It is considered that an efficient and lightweight algorithm for designing projection matrix would usefully supplement and enhance the performance of CS. In this paper, we propose an algorithm to design projection matrix Φ which is called *efficient projection* (EP) to distinguish with *optimized projection* (OP) in [9]. The EP is designed in such a way that the new projected dictionary $\Omega = \Phi\Psi$ will have the structure as much similar as the dictionary Ψ. For this purpose, we use Multidimensional Scaling (MDS) technique which helps to find a low-dimensional new projected dictionary Ω such that the pairwise distances between new atoms in this new projected dictionary Ω match as well as possible the original atoms in dictionary Ψ. This leads to a solution for EP which is very simple and can be obtained by doing singular value decomposition of dictionary Ψ. In next section, OP tech-

nique will be presented. MDS and proposed EP algorithm are presented in section 3. We demonstrate the experimental results to show the novelty of our proposed algorithm in section 4. Some conclusions and future works are drawn in section 5.

2 Optimized Projection

Objective: Minimize $\mu_t\{\Phi\Psi\}$ with respect to Φ.
Input: Use the following parameters:

1. t or $t\%$ - coherence (fixed or relative) threshold,
2. Ψ - the dictionary,
3. p - number of measurements,
4. γ - down-scaling factor, and
5. *Iter* - number of iterations.

Initialization: Set $\Phi \in \Re^{p\times n}$ to be an arbitrary random matrix.
Loop: Set k= 0 and repeat *Iter* times:

1. *Gram Matrix Construction:* Normalize the columns in the matrix $\Phi_k\Psi$ and obtain the effective dictionary Ω_k and compute Gram Matrix $G_k = \Omega_k^T\Omega_k$.
2. *Set Threshold and Shrink:* If mode of operation is fixed, use t as threshold. Otherwise, choose t such that $t\%$ of the off-diagonal entries in G_k are above it. Update the Gram matrix and obtain $\hat{G}_k$ by

$$g_{ij} = \begin{cases} \gamma g_{ij} & |g_{ij}| \geq t \\ \gamma t.sign(g_{ij}) & t > |g_{ij}| \geq \gamma t \\ g_{ij} & \gamma t > |g_{ij}| \end{cases} \tag{4}$$

3. *Reduce Rank:* Apply SVD and force the rank of $\hat{G}_k$ to be equal to p.
4. *Squared-Root:* Build the squared-root of $\hat{G}_k = S_k^T S_k$, where S_k is the size of $p \times n$.
5. *Update* Φ*:* Find Φ_{k+1} that minimizes the error $\|S_k - \Phi\Psi\|_F^2$ and set $k = k + 1$.

Output: The output of the above algorithm is Φ_{Iter}.

Table 1. Optimized Projection Algorithm.

Given a dictionary Ψ, its *mutual-coherence* is defined as the largest absolute and normalized inner product between different columns in Ψ [7][10][9] and can be written as follow:

$$\mu\{\Psi\} = \max_{i\neq j} \frac{\left|d_i^T d_j\right|}{\|d_i\|\,\|d_j\|} \tag{5}$$

where d_i is the column i^{th} in dictionary matrix Ψ. The *mutual-coherence* provides a measure of the worst similarity between the dictionary columns, a value that exposes the dictionary's vulnerability, as such two closely related columns may confuse any pursuit technique. A different way to understand the *mutual-coherence* is by considering the Gram matrix $G = \tilde{\Psi}^T\tilde{\Psi}$, computed using the dictionary after normalizing each of its columns. The *mutual-coherence* is the off-diagonal entry g_{ij} with the largest magnitude. Suppose that the signal x_0 has been constructed by $x_0 = \Psi\alpha_0$ with a sparse representation. Suppose further that the following inequality is satisfied:

$$\|\alpha_0\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu\{\Psi\}}\right) \tag{6}$$

Some theoretical works [7][10] show that if inequality (6) is satisfied, the vector α_0 can be found exactly by BP or OMP algorithms. Based on this fact, Elad in [9] proposed a more strict requirement that

$$\|\alpha_0\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu\{\Phi\Psi\}}\right) \leq \frac{1}{2}\left(1 + \frac{1}{\mu\{\Psi\}}\right) \tag{7}$$

and designed Φ such that $\mu\{\Phi\Psi\}$ is as small as possible. But he claimed that the *mutual-coherence* defined as above does not do justice to the actual behavior of sparse representations and pursuit algorithms' performance. So he considered a different *mutual-coherence*, which called *t-averaged mutual-coherence* and defined as follows:

$$\mu_t(\Psi) = \frac{\sum\limits_{i\neq j}\left(|g_{ij}| \geq t\right).|g_{ij}|}{\sum\limits_{i\neq j}\left(|g_{ij}| \geq t\right)} \tag{8}$$

Based on *t-averaged mutual-coherence*, an *optimized projection* can be obtained by the proposed algorithm in [9] which can be summarized as in Table I.

3 Efficient Projection

EP is based on the idea of Multidimensional Scaling (MDS) technique which helps to find a low-dimensional new projected dictionary Ω such that the pairwise distances between new atoms in this new projected dictionary Ω match as well as possible the original atoms in dictionary Ψ. In this section, we review MDS briefly and proposed our approach to get EP.

3.1 Multidimensional Scaling - MDS

Suppose a set of N objects is under consideration and between each pair of objects (i, j) there is a measurement δ_{ij} of the "dissimilarity" between the two objects. The typical goal of MDS is to find a low dimensional space, usually Euclidean, in which points in the space represent the objects, one point representing one object, and such that the distances between the points in the space $\{d_{ij}\}$ match as well as possible the original dissimilarities $\{\delta_{ij}\}$ [6]. Multidimensional Scaling treats dissimilarities $\{\delta_{ij}\}$ directly as Euclidean distances and then makes use of the spectral decomposition of a doubly centered matrix of dissimilarities. Formally, a symmetric $N \times N$ matrix $\Delta = [\delta_{ij}]$ is called a dissimilarity matrix if $\delta_{ij} \geq 0$ (nonnegative elements) and $\delta_{ii} = 0$ (zero diagonal elements). From a given dissimilarity matrix Δ, MDS algorithm constructs a configuration of points in a Euclidean space of specified dimension p [12]. Because the distances in the configuration space are to be Euclidean, we can have

$$\delta_{ij}^2 = d_{ij}^2(X) = \|x_i - x_j\|^2 = (x_i - x_j)^T(x_i - x_j) \quad (9)$$

By writing the squared distances as $d_{ij}^2(X) = x_i^T x_i - 2x_i x_j + x_j^T x_j$ and defining $\Xi = \left[x_1^T x_1, .., x_N^T x_N\right]^T$ [5], the squared distance matrix $D = [d_{ij}^2(X)]_{i,j=1}^N$ now can be written as

$$D = \Xi e^T - 2X^T X + e\Xi^T \quad (10)$$

where $e \in \Re^n$ is a vector of all ones, and $X = [x_1, x_2, ..., x_N]$. Let us define the inner product matrix $B = X^T X$. Matrix B can be rewritten as

$$B = -HDH \quad (11)$$

where $H = I - ee^T/N$ is the centering operator. In order to recover the coordinates from B, MDS needs to extract the eigenvectors and eigenvalues as follow

$$B = U\Lambda U^T \quad (12)$$

Let $\Lambda_p = diag(\lambda_1, .., \lambda_p)$ be the diagonal matrix of p largest non-zero eigenvalues of B, and $U_p = [u_1, .., u_p]$ is the corresponding eigenvectors. Then the new coordinate matrix $\hat{X}$ is given by $\hat{X} = \Lambda_p^{1/2} U_p^T$. Let us now derive the relationship between X and $\hat{X}$ to see what the projection matrix looks like. Based on the relation between singular value decomposition and eigenvalue decomposition, we can write

$$X^T X = U_p \Lambda_p U_p^T = \left(U_p \Lambda_p^{1/2} V^T\right)\left(V \Lambda_p^{1/2} U_p^T\right) \quad (13)$$

where $X = V\Lambda_p^{1/2} U_p^T$ shows the singular value decomposition of X. Now we can have $\hat{X} = V^T X$. It is noted that when MDS is applied on Euclidean distance, it is as same as Principal Component Analysis.

3.2 Efficient Projection

Now it is time to derive how to obtain *efficient projection*. As we stated that the EP is designed in such a way that the new projected dictionary $\Omega = \Phi\Psi$ will have the structure as much similar as the dictionary Ψ. For this purpose, we use Multidimensional Scaling (MDS) technique which helps to find a low-dimensional new projected dictionary Ω such that the pairwise distances between new atoms in this new projected dictionary Ω match as well as possible the original atoms in dictionary Ψ. Suppose that the dictionary matrix Ψ has the singular value decomposition as follow:

$$\Psi = A\Lambda B^T \quad (14)$$

By applying MDS on $\Psi \in \Re^{n\times k}$, we get the new projected matrix $\Omega \in \Re^{p\times k}$ which have the relationship with Ψ as

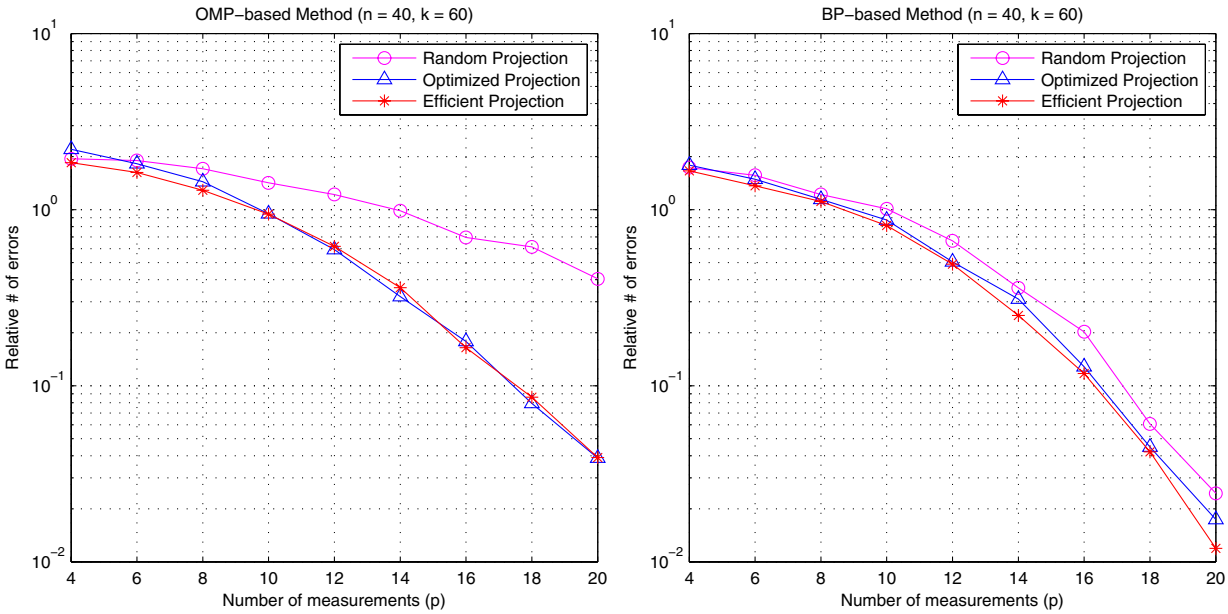
$$\Omega = A_p^T \Psi \quad (15)$$

where A_p contains the first p columns of A which are eigenvectors corresponding to p largest eigenvalues of Ψ. From (15), we can design *efficient projection* matrix $\Phi = A_p^T$.

4 Experimental Results

In this section, we conduct some experiments with generated data in order to compare the performance of random projection, *optimized projection* and *efficient projection* in CS framework. Basically, experimental protocol is carried on through following steps:

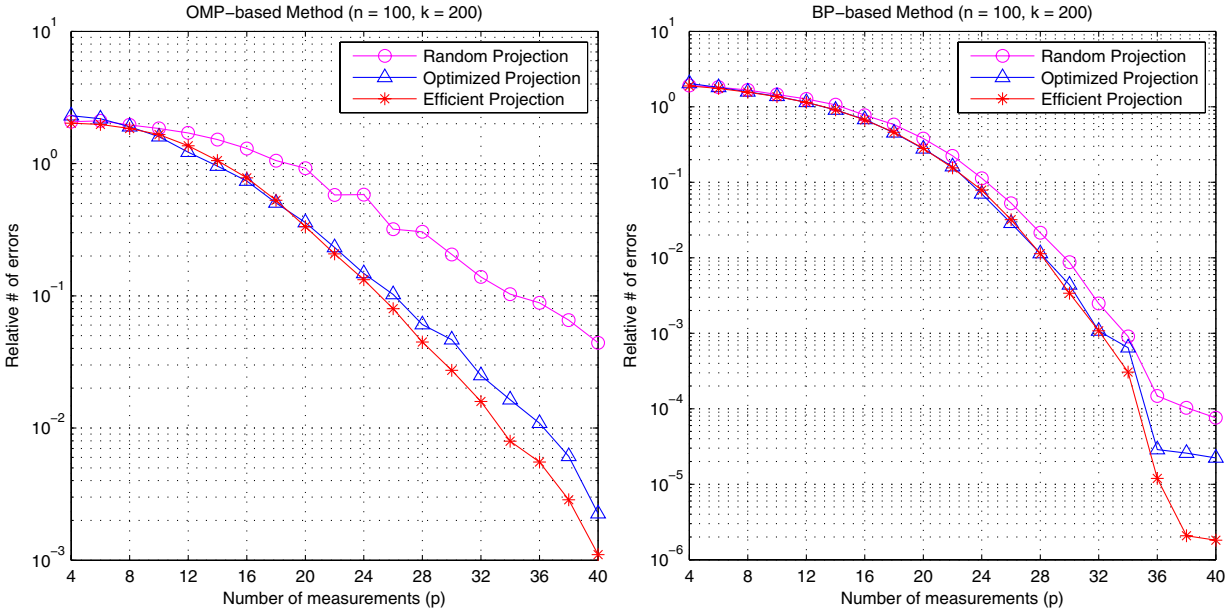
1. **Generating Data**
   - Generate a random dictionary matrix $\Psi \in \Re^{n\times k}$ with normalized columns.
   - Design projection matrix :
     (a) Generate random projection matrix $\Phi \in \Re^{p\times n}$ with normalized rows
     (b) or design OP matrix Φ by algorithm in Table 2
     (c) or design EP matrix Φ as described in previous section.
   - Create $m = 10000$ samples $X = [x_1, x_2, .., x_m] \in \Re^{n\times m}$ where $x_i = \Psi\alpha_i \in \Re^n$ and α_i is the sparse vector with T non-zero elements chosen randomly.
   - Calculate measurements $y_i = \Phi x_i \in \Re^p$, where $i = 1..m$.
2. **Decoding and Error Calculation**
   - Given Ψ, Φ, y_i with $i = 1..m$, we use OMP or BP to find sparse vectors $\hat{\alpha}_i$ subject to $\Phi\Psi\alpha_i = y_i$ and calculate estimated $\hat{x}_i = \Psi\hat{\alpha}_i$ of x_i.



(a) Decoding by OMP method with $n = 40$, $k = 60$ and $T = 4$

(b) Decoding by BP method with $n = 40$, $k = 60$ and $T = 4$

Figure 2. Case I - Comparison of CS performance based on random projection matrix, optimized projection matrix and efficient projection matrix. OMP is applied in (a) and BP is applied in (b).



(a) Decoding by OMP method with $n = 100$, $k = 200$ and $T = 4$

(b) Decoding by BP method with $n = 100$, $k = 200$ and $T = 4$

Figure 3. Case II - Comparison of CS performance based on random projection matrix, optimized projection matrix and efficient projection matrix. OMP is applied in (a) and BP is applied in (b).

- Calculate distance error between original (x_i) and estimated ($\hat{x}_i$) samples and get the averaged error over all samples.

We setup two study cases for our experiments. **Case I** is carried on with $n = 40$, $k = 60$ and $T = 4$, while $n = 100$, $k = 200$ and $T = 4$ are selected in **Case II**. We follow the above experimental protocol in all experiments. This protocol is as similar as that of [9]. It is noted that the non-zero locations of value α were chosen at random and populated with iid zero-mean and unit variance Gaussian values. These sparse vectors were used to create the sample matrix X with which to evaluate the CS performance. The purpose of these experiments is to study the performance of CS before and after the design of the optimized and efficient projections, with BP and OMP, and for varying amounts of measurements (i.e. varying the value of p). In test **Case I**, the size of dictionary matrix is 40×60 with the number of non-zeros $T = 4$, so we vary the amounts of measurements p from 4 to 20 with step size of 2. Both OMP and BP are used to evaluate the performance of CS (see Fig. 2). Experimental performance of **Case II** can be see in Fig. 3 where dictionary size is 100×200, $T = 4$ and p is varied from 4 to 40 with step size of 2. Through all experiments, some conclusions are highlighted as follow:

- In CS framework, *efficient projection* gives a comparative performance as good as *optimized projection*.
- While the algorithm of designing *optimized projection* is an iterative procedure with high time complexity, *efficient projection* can be obtained just by well-known singular value decomposition which is a non-iterative and low complexity procedure.
- In general, the performance of BP is better than that of OMP in CS framework.

5 Conclusions

The purpose of the current study was to design an efficient and novel projection for compressed sensing framework. The algorithm is based on singular value decomposition which yields a very efficient projection matrix for CS framework. The experiments show comparative results of the proposed approach compared to those of optimized projection approach. The simplicity and effectiveness of the approach make it suitable to be applied in any application of CS framework. While current work focuses on the theory of designing projection matrix, future work will be applying the *efficient projection*-based CS on some real applications for further evaluation.

References

[1] E. J. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. Comput. Math.*, 6(2):227–254, 2006.

[2] E. J. Candès, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[3] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.

[5] J. A. Costa, N. Patwari, and I. Alfred O. Hero. Distributed weighted-multidimensional scaling for node localization in sensor networks. *ACM Trans. Sen. Netw.*, 2(1):39–64, 2006.

[6] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman Hall, 1994.

[7] D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proc. Natl. Acad. Sci.*, 100:2197–2202, 2002.

[8] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[9] M. Elad. Optimized projections for compressed sensing. *IEEE Trans. on Signal Processing*, 2007.

[10] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325, 2003.

[11] P. S. Huggins and S. W. Zucker. Greedy basis pursuit. *IEEE Transactions on Signal Processing*, 55(7):3760–3772, 2007.

[12] A. Kearsley, R. Tapia, and M. Trosset. The Solution of the Metric STRESS and SSTRESS problems in Muldimensional Scaling Using Newton's Method. *Computational Statistics*, 13(3):369–396, 1998.

[13] Lustig. Sparse mri: The application of compressed sensing for rapid mr imaging. 2007.

[14] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[15] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. *Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems,*, November 1993.

[16] B. Richard. A lecture on compressive sensing. *IEEE Signal Processing Magazine*, July 2007.

[17] M. A. Sheikh, O. Milenkovic, S. Sarvotham, and R. G. Baraniuk. Technical report, Rice University, 2007.

[18] Y. Tsaig and D. L. Donoho. Extensions of compressed sensing. *Signal Processing*, 86(3):549–571, 2006.

[19] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk. An architecture for compressive imaging. In *ICIP*, pages 1273–1276, 2006.