ON THE DESIGN OF OPTIMIZED PROJECTIONS FOR SENSING SPARSE SIGNALS IN OVERCOMPLETE DICTIONARIES

Wei Chen[†] *Miguel R. D. Rodrigues*[⋆] *Ian J. Wassell*[†]

[†] Computer Laboratory, University of Cambridge
[⋆] Instituto de Telecomunicações, Departamento de Ciência de Computadores, Universidade do Porto

ABSTRACT

Sparse signals can be sensed with a reduced number of random projections and then reconstructed if compressive sensing (CS) is employed. Traditionally, the projection matrix has been chosen as a random Gaussian matrix, but improved reconstruction performance can be obtained by optimizing the projection matrix. In this paper, we are interested in projection matrix designs for sensing sparse signals in overcomplete dictionaries. In particular, we put forth a closed form design that stems from the formulation of an optimization problem, which bypasses the complexity of iterative design approaches.

1. INTRODUCTION

Consider a signal $\mathbf{f} \in \mathbb{R}^N$ which has a sparse representation $\mathbf{x} \in \mathbb{R}^K$ in a known dictionary $\mathbf{D} \in \mathbb{R}^{N\times K}(N \leq K)$. Thus, the signal can be described by

$$\mathbf{f} = \mathbf{D}\mathbf{x}. \tag{1}$$

We say the signal $\mathbf{f}$ is sparse if its representation satisfies $\|\mathbf{x}\|_0 \leq S \ll N$, where the ℓ_0 norm counts the number of nonzero elements in $\mathbf{x}$. This sparse signal can be represented by a linear combination of very few signal-atoms $\{\mathbf{d_i}\}_{i=1}^K$ which are columns of the overcomplete dictionary $\mathbf{D}$. Thus, it is wasteful to sample the signal fully and then discard most coefficients.

Compressive sensing (CS) has attracted a growing interest in the recent years, as it enables one to recover the signal with a reduced number of measurements [1, 2]. The sensing process can be described by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{f} + \mathbf{n}, \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^M (M \ll N)$ is the measurement vector, $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ is the projection matrix and $\mathbf{n} \in \mathbb{R}^M$ is the noise vector with i.i.d. random elements drawn according to $\mathcal{N}(0, \sigma^2)$. One can then reconstruct the sparse signal by solving the following problem:

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_0 \qquad \text{s.t.} \quad \|\mathbf{\Phi}\mathbf{D}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \tag{3}$$

The work of M. R. D. Rodrigues was supported by Fundação para a Ciência e a Tecnologia through the research project CMU-PT/SIA/0026/2009.

where $\epsilon > 0$ relates to an estimate of the noise level. As (3) is known to be an NP-hard problem, a number of convex optimization algorithms and greedy algorithms have been proposed in the literature for recovering very sparse signals [3].

The projection matrix $\mathbf{\Phi}$ is assumed to be randomly generated in most works owing to its convenience in theoretical analysis. However, Elad proposes to optimize the projection matrix using iterative algorithms that lead to an improved average reconstruction performance [4]. Subsequently, Duarte-Carvajalino and Sapiro assume the dictionary is not fixed and propose to optimize the projection matrix and dictionary simultaneously [5]. Xu et al. consider the equiangular tight frame (ETF) as their target design and proposed an algorithm to make the projection matrix approach that design [6]. Although improved average performances are observed, the optimized projection matrices are iteratively computed in these approaches.

In our earlier work, we focused on the study on noniterative designs and have shown that the mean squared error (MSE) performance of unit-norm tight frame based projection matrices surpasses that of the other designs mentioned previously, where we restrict the dictionary to be the orthonormal basis. In this paper, we study the design of the projection matrix for signals over some known overcomplete dictionary, which adds a lot of flexibility and significantly extends the range of applicability of the work. A noniterative design of the projection matrix is proposed and its superior reconstruction performance is demonstrated by a range of numerical results.

The following notation is used. Upper-case letters denote numbers, boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors and calligraphic upper-case letters denote sets. $(\cdot)^T$, $\text{Tr}(\cdot)$ and $\text{rank}(\cdot)$ denote the transpose, trace and rank of a matrix, respectively. $\|\mathbf{x}\|_1$ denotes the ℓ_1 norm, i.e., the summation of the absolute value of elements of the vector $\mathbf{x}$. $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$ denotes the ℓ_2 norm of the vector $\mathbf{x}$. $\|\mathbf{X}\|_F$ denotes the Frobenius norm of the matrix $\mathbf{X}$. $\mathbf{X} \succeq 0$ denotes that the matrix $\mathbf{X}$ is positive semi-definite.

2. DESIGN OF THE PROJECTION MATRIX FOR SPARSE SIGNALS

In the first instance, we consider a simplified measurement model

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}, \tag{4}$$

i.e., $\mathbf{f} = \mathbf{x}$, $\mathbf{x}$ is sparse and so $\mathbf{D}$ is the identity matrix in (1). Note that having abstained some result for this model, later on we will return to the more general model given in the introduction.

We wish to minimize the mean squared error (MSE) in estimating $\mathbf{x}$ from $\mathbf{y}$, given by

$$\texttt{MSE}(\mathbf{\Phi}) = \mathbb{E}_{\mathbf{x},\mathbf{n}}\left(\|\mathcal{F}(\mathbf{\Phi x} + \mathbf{n}) - \mathbf{x}\|_2^2\right), \tag{5}$$

where $\mathcal{F}(\cdot)$ denotes an estimator and $\mathbb{E}_{\mathbf{x},\mathbf{n}}(\cdot)$ denotes expectation with respect to the joint distribution of the random vectors $\mathbf{x}$ and $\mathbf{n}$. We assume a random signal model where the signal is exactly S-sparse and the positions of the S nonzero elements follow an equiprobable distribution.

We consider the well-known oracle estimator, which uses least square estimation based on the prior knowledge of the positions of the S non-zero elements of the sparse representation, in order to bound the MSE in (5). The reason for using the oracle estimator is based on the fact that the oracle MSE is equal to the unbiased Cramér-Rao bound (CBD) for exactly S sparse deterministic vectors [7]. The oracle MSE - for a fixed sparse vector $\mathbf{x}$ - is given by:

$$\begin{aligned}\texttt{MSE}_{\mathbf{n}}^{\text{oracle}}(\mathbf{\Phi}, \mathbf{x}) =& \mathbb{E}_{\mathbf{n}}\left(\|\mathcal{F}^{\text{oracle}}(\mathbf{\Phi x} + \mathbf{n}) - \mathbf{x}\|_2^2\right) \\ =& \sigma^2 \text{Tr}\left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{\Phi}^T \mathbf{\Phi} \mathbf{E}_{\mathcal{J}}\right)^{-1}\right),\end{aligned} \tag{6}$$

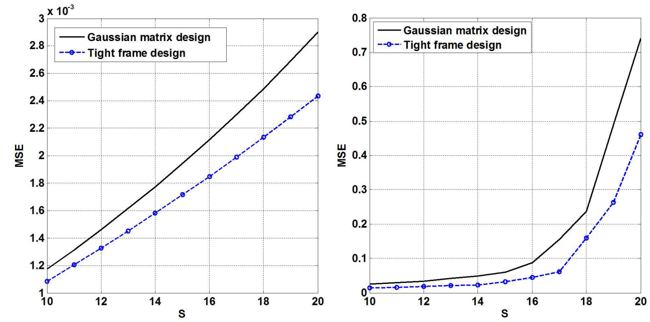
where $\mathbb{E}_{\mathbf{n}}(\cdot)$ denotes expectation with respect to the distribution of the random vector $\mathbf{n}$, $\mathcal{J}$ denotes the set containing the positions of the S nonzero elements of $\mathbf{x}$ and $\mathbf{E}_{\mathcal{J}}$ denotes the matrix that results from the identity matrix by deleting the set of columns out of the set $\mathcal{J}$.

Consequently, the average value of the oracle MSE - which acts as a lower bound to the MSE in (5) - is given by:

$$\begin{aligned}\texttt{MSE}^{\text{oracle}}(\mathbf{\Phi}) &= \mathbb{E}_{\mathbf{x}}\left(\texttt{MSE}_{\mathbf{n}}^{\text{oracle}}(\mathbf{\Phi}, \mathbf{x})\right) \\ &= \mathbb{E}_{\mathbf{x}}\left(\sigma^2 \text{Tr}\left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{\Phi}^T \mathbf{\Phi} \mathbf{E}_{\mathcal{J}}\right)^{-1}\right)\right) \\ &= \sigma^2 \mathbb{E}_{\mathcal{J}}\left(\text{Tr}\left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{\Phi}^T \mathbf{\Phi} \mathbf{E}_{\mathcal{J}}\right)^{-1}\right)\right),\end{aligned} \tag{7}$$

where $\mathbb{E}_{\mathbf{x}}(\cdot)$ and $\mathbb{E}_{\mathcal{J}}(\cdot)$ denote expectation with respect to the distribution of the random vector $\mathbf{x}$ and the random set $\mathcal{J}$, respectively. In (7), we have used the fact that the expectation with respect to the distribution of the random vector $\mathbf{x}$ is equal to the expectation with respect to the distribution of the positions of the nonzero elements of the random vector $\mathbf{x}$, owing to the use of the oracle.

We define the coherence matrix of the sensing matrix as $\mathbf{Q} = \mathbf{\Phi}^T \mathbf{\Phi}$. We now look for the coherence matrix, which up to a rotation leads to the sensing matrix, that minimizes the average value of the oracle MSE subject to an energy constraint by posing the following optimization problem:



(a) Oracle estimator (b) BPDN

Fig. 1. Comparison of the MSE of projection matrices ($M = 64$, $N = 128$ and $\sigma^2 = 10^{-4}$).

$$\begin{aligned}\min_{\mathbf{Q}} \quad & \mathbb{E}_{\mathcal{J}}\left(\text{Tr}\left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{Q} \mathbf{E}_{\mathcal{J}}\right)^{-1}\right)\right) \\ \text{s.t.} \quad & \mathbf{Q} \succeq 0, \\ & \text{Tr}(\mathbf{Q}) = N, \\ & rank(\mathbf{Q}) \leq M,\end{aligned} \tag{8}$$

However, (8) is non-convex due to the rank constraint. Therefore, we first consider a convex relaxation of (8) by ignoring the rank constraint, and then look for a feasible solution that is the closest to the solution to the relaxed problem and satisfies the rank constraint.

Proposition 1 *The solution of the optimization problem:*

$$\begin{aligned}\min_{\mathbf{Q}} \quad & \mathbb{E}_{\mathcal{J}}\left(Tr\left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{Q} \mathbf{E}_{\mathcal{J}}\right)^{-1}\right)\right) \\ s.t. \quad & \mathbf{Q} \succeq 0, \\ & Tr(\mathbf{Q}) = \mathrm{N},\end{aligned} \tag{9}$$

which represents a convex relaxation of the original optimization problem in (8), is the $\mathrm{N} \times \mathrm{N}$ *identity matrix* $\mathbf{I}_{\mathrm{N}}$.

For finite-dimentional real spaces, we define a Parseval tight frame as a matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ such that $\|\mathbf{\Phi}^T \mathbf{z}\|_2^2 = \|\mathbf{z}\|_2^2$ for any vector $\mathbf{z} \in \mathbb{R}^M$. We claim in Proposition 2 that a Parseval tight frame is the closest design - in the Frobenius norm sense - to the solution of the convex relaxation of the original optimization problem in (8).

Proposition 2 *The solution of the optimization problem*

$$\begin{aligned}\min_{\mathbf{\Phi}} \quad & \left\|\mathbf{\Phi}^T \mathbf{\Phi} - \mathbf{I}_{\mathrm{N}}\right\|_F^2 \\ s.t. \quad & Tr\left(\mathbf{\Phi}^T \mathbf{\Phi}\right) = \mathrm{N},\end{aligned} \tag{10}$$

is the $\mathrm{M} \times \mathrm{N}$ *Parseval tight frame.*

In the interest of space, the proofs are omitted, but they are similar to our proofs in [8]. Fig. 1 illustrates the average

MSE performance of the oracle estimator and the practical ℓ_1 minimization estimator using basis pursuit de-noise (BPDN). The MSE is calculated by averaging over 1000 trials, where in each trial we generate randomly a sparse vector with S randomly placed ± 1 spikes. The reconstruction performance of Parseval tight frames based sensing matrices clearly outperforms the reconstruction performance of the Gaussian matrix for both estimators.

3. DESIGN OF THE PROJECTION MATRIX FOR SENSING SPARSE SIGNALS IN AN OVERCOMPLETE DICTIONARY

We now return to the measurement model considered in the introduction,

$$\mathbf{y} = \mathbf{\Phi}\mathbf{D}\mathbf{x} + \mathbf{n}, \tag{11}$$

where $\mathbf{f} = \mathbf{D}\mathbf{x}$ is sparse in the overcomplete dictionary $\mathbf{D}$. This model does not lead to a simple formulation as the model in (4) where the dictionary can be seen to correspond to an identity matrix and the model in [8] where the dictionary corresponds to an orthonormal matrix. However, the previous analysis suggests that the projection matrix design $\mathbf{\Phi}$ ought to be such that $\mathbf{\Phi}\mathbf{D}$ is close to a Parseval tight frame.

Let $\mathbf{U} \in \mathbb{R}^{M \times K}$ be a Parseval tight frame. Therefore, we define the following optimization problem:

$$\min_{\hat{\mathbf{\Phi}}} \quad \left\| \hat{\mathbf{\Phi}}\mathbf{D} - \mathbf{U} \right\|_F^2 + \alpha \left\| \hat{\mathbf{\Phi}} \right\|_F^2, \tag{12}$$

where $\alpha \geq 0$ is a given scalar. The term, $\left\| \hat{\mathbf{\Phi}} \right\|_F^2$, in (12) denotes the penalty for the amplified energy of the projection matrix. In other words, the solution of (12) is the design that makes $\mathbf{\Phi}\mathbf{D}$ close to the Parseval tight frame and consumes low energy in the projection. The solution of the optimization problem gives

$$\hat{\mathbf{\Phi}} = \mathbf{U}\mathbf{D}^T \left(\mathbf{D}\mathbf{D}^T + \alpha \mathbf{I}_N \right)^{-1}, \tag{13}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix.

To enable fair comparison to other projection matrix designs, we pose an energy constrain, i.e., $\mathrm{Tr}\left(\mathbf{\Phi}^T \mathbf{\Phi} \right) = N$. Thus, we derive our projection matrix design for the overcomplete dictionary as follows,

$$\mathbf{\Phi} = \frac{\sqrt{N}\hat{\mathbf{\Phi}}}{\|\hat{\mathbf{\Phi}}\|_F} = \frac{\sqrt{N}\mathbf{U}\mathbf{D}^T \left(\mathbf{D}\mathbf{D}^T + \alpha \mathbf{I}_N \right)^{-1}}{\|\mathbf{U}\mathbf{D}^T \left(\mathbf{D}\mathbf{D}^T + \alpha \mathbf{I}_N \right)^{-1}\|_F}. \tag{14}$$

The scalar α controls the weight for the energy penalty of the projection. If the penalty is not considered, i.e. $\alpha = 0$, we have the projection matrix design $\mathbf{\Phi} = \frac{\mathbf{U}\mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}}{\|\mathbf{U}\mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}\|_F}$. In contrast, for a very high penalty, i.e., $\alpha = +\infty$, we have the design $\mathbf{\Phi} = \frac{\mathbf{U}\mathbf{D}^T}{\|\mathbf{U}\mathbf{D}^T\|_F}$.

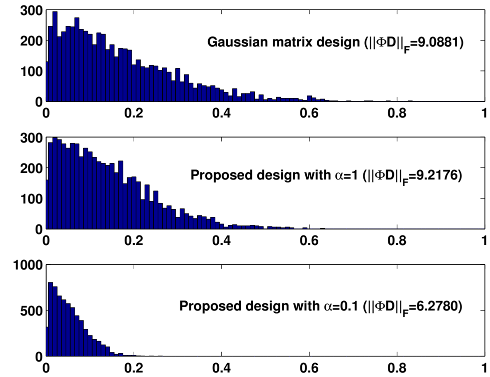


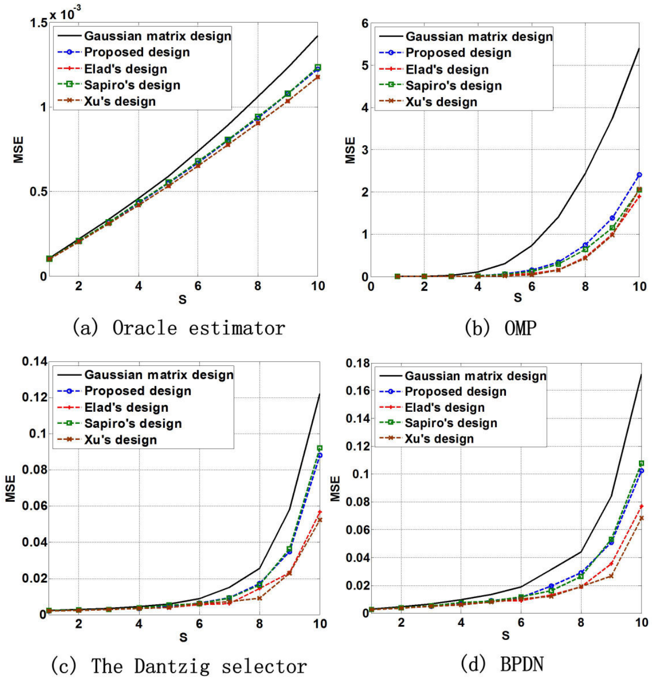
Fig. 2. Histogram of the absolute off-diagonal entries of $\mathbf{\Psi}$.

To illustrate the behavior of the proposed design, we present the histogram of the absolute values of the off-diagonal entries of the coherence matrix $\mathbf{\Psi} = \mathbf{D}^T \mathbf{\Phi}^T \mathbf{\Phi}\mathbf{D}$ in Fig. 2. The overcomplete dictionary $\mathbf{D}$ of size 64×80 is generated with entries drawn from the i.i.d. zero mean and unit variance Gaussian distribution. We derive two projection matrices containing 40 projections using (14) with $\alpha = 1$ and $\alpha = 0.1$ and compare the two designs with Gaussian matrix design. Generally, it is known that a projection matrix with small off-diagonal entries has a high reconstruction performance according to the mutual coherence reconstruction condition [9]. Fig. 2 shows that the distributions of off-diagonal entries in both designs are better than the Gaussian matrix design. We note the design with $\alpha = 0.1$ has smaller off-diagonal entries in absolute value than the design with $\alpha = 1$. However, the sensed energy $\|\mathbf{\Psi}\mathbf{D}\|_F = 6.2780$ for the $\alpha = 0.1$ design is lower than the others owing to the low penalty $\alpha = 0.1$ for the amplified energy of the projection matrix in (14), which results in a noise amplification effect compared to the Gaussian matrix design.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed projection matrix design in the CS setting with an overcomplete dictionary. Note that the proposed matrix $\mathbf{\Phi}$ is generated in a noniterative manner.

In the first experiment, we use an overcomplete discrete cosine transform (DCT) dictionary of size 64×80, i.e., $N = 64$ and $K = 80$. The four estimators used to reconstruct the signal are listed as follows: the oracle estimator, orthogonal matching pursuit (OMP), the dantzig selector and BPDN. We compare the reconstruction performances of five projection matrix designs including a random Gaussian matrix, the proposed design (with $\alpha = 1$) and three other optimized iterative designs, i.e., Elad's design [4], Duarte-Carvajalino and Sapiro's design [5], and Xu's design [6]. The reconstruc-



(a) Oracle estimator (b) OMP

(c) The Dantzig selector (d) BPDN

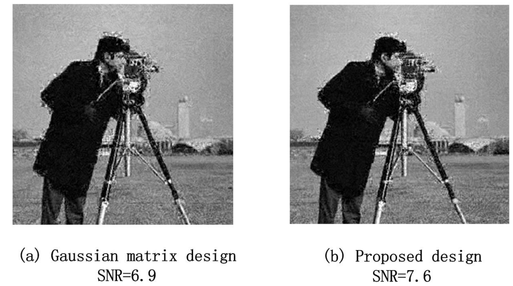
Fig. 3. Comparison of the MSE of different projection matrix designs ($M = 40$, $N = 64$, $K = 80$ and $\sigma^2 = 10^{-4}$).

tion performance in terms of average MSE is calculated by averaging over 1000 trials, where in each trial we generate randomly a sparse vector with S randomly placed ± 1 spikes. As shown in Fig. 3, the proposed design improves the reconstruction performance for all the four estimators, compared to the Gaussian matrix design. The iterative algorithms including Elad's design, Duarte-Carvajalino and Sapiro's design, and Xu's design, slightly outperform the proposed design, but their computation complexity associated with the generation of the projection matrix is much higher than our method.

The cameraman image of size 256×256 pixels is used to demonstrate the advantage of the proposed projection matrix design in Fig. 4. The image is partitioned into 1024 nonoverlapping patches of size 8×8 pixels, i.e., $N = 64$. We use the overcomplete DCT dictionary of size 64×80 as the sparsifying dictionary and set $M = 40$. OMP is used to reconstruct the image from its noisy projections owing to its fast execution. Let $\mathbf{P}$ be original image and $\tilde{\mathbf{P}}$ be the reconstruction result. We define the reconstructed signal to noise ratio by

$$\text{SNR} = \frac{\|\tilde{\mathbf{P}}\|_F}{\|\mathbf{P} - \tilde{\mathbf{P}}\|_F}. \quad (15)$$

Two reconstructed images, i.e., one using a Gaussian matrix and the other using our design, and their SNRs are shown in Fig. 4. As can be seen, a higher reconstruction quality and a higher SNR is obtained by using the proposed projection matrix design.



(a) Gaussian matrix design SNR=6.9 (b) Proposed design SNR=7.6

Fig. 4. Reconstructed images using Gaussian projection matrix and the proposed projection matrix design.

5. CONCLUSIONS

In this paper, we investigate the problem of projection matrix design for sensing signals which are sparse in overcomplete dictionaries. We derive a close-form expression for generating the optimized projection matrix, which has been shown to lead to MSE performance gains for standard CS reconstruction algorithms. The proposed approach leads to a closed-form design that bypasses the complexity of previous approaches.

6. REFERENCES

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

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

[3] M. Zibulevsky and M. Elad, "L1-l2 optimization in signal and image processing," *Signal Processing Magazine, IEEE*, vol. 27, no. 3, pp. 76 –88, May 2010.

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

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

[6] Jianping Xu, Yiming Pi, and Zongjie Cao, "Optimized projection matrix for compressive sensing," *EURASIP J. Adv. Signal Process*, vol. 43, pp. 1–8, Feb. 2010.

[7] Z. Ben-Haim and Y.C. Eldar, "The cramér-rao bound for estimating a sparse parameter vector," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3384 –3389, June 2010.

[8] Wei 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," *Signal Processing Letters, IEEE*, vol. 19, no. 1, pp. 8 –11, Jan. 2012.

[9] D.L. Donoho, M. Elad, and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6 – 18, 2006.