ON OPTIMIZATION OF THE MEASUREMENT MATRIX FOR COMPRESSIVE SENSING

Vahid Abolghasemi[1], *Saideh Ferdowsi*[1], *Bahador Makkiabadi*[1,2], *and Saeid Sanei*[1]

[1]Centre of Digital Signal Processing, School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK
[2]Electrical Engineering Department, Islamic Azad University, Ashtian Branch, Ashtian, IRAN
{abolghasemiv, ferdowsis, makkiabadib, saneis}@cf.ac.uk

ABSTRACT

In this paper the problem of Compressive Sensing (CS) is addressed. The focus is on estimating a proper measurement matrix for compressive sampling of signals. The fact that a small *mutual coherence* between the measurement matrix and the representing matrix is a requirement for achieving a successful CS is now well known. Therefore, designing measurement matrices with smaller coherence is desired. In this paper a gradient descent method is proposed to optimize the measurement matrix. The proposed algorithm is designed to minimize the mutual coherence which is described as absolute off-diagonal elements of the corresponding Gram matrix. The optimization is mainly applied to random Gaussian matrices which is common in CS. An extended approach is also presented for *sparse* signals with respect to redundant dictionaries. Our experiments yield promising results and show higher reconstruction quality of the proposed method compared to those of both unoptimized case and previous methods.

1. INTRODUCTION

Compressive Sensing (CS) [1] [2] is one of the recent interesting fields in signal and image processing communities. It has been utilized in many different applications from biomedical signal and image processing [3] to communication [4] and astronomy [5]. The core idea in CS is a novel sampling technique which under certain conditions can lead to a smaller rate compared to conventional Shannon's sampling rate. The key requirement for achieving a successful CS is *compressibility* or more precisely *sparsity* of the input signal. A sparse signal has a small number of active (nonzero) components compared to its total length. This property can either exist in the sampling domain of the signal or with respect to other basis such as Fourier, wavelet, curvelet or any other basis. A CS scenario mainly consists of two crucial parts; encoding (sampling) and decoding (recovery). We formally explain both parts in this section, but the focus in this work is on the first part where we try to improve the sensing process which consequently affects the reconstruction performance.

Let $\mathbf{x} \in \mathbb{R}^n$ be a sparse signal, meaning to have at most $s \ll n$ nonzero elements. We now want to find p linear measurements termed as $\mathbf{y} = \Phi\mathbf{x}$, where $s < p < n$, and $\Phi \in \mathbb{R}^{p\times n}$ is the measurement matrix. Obviously, $p = s$ is the best possible choice. However, since we are not aware of the locations and the values of the nonzero elements when reconstructing $\mathbf{x}$ from $\mathbf{y}$, a larger p is needed to guarantee the recovery of $\mathbf{x}$ [6]. In addition, the structure of Φ is not arbitrary, and should be adopted based upon some specific rules. For instance, it has been proved [6] that for a Φ with i.i.d Gaussian entries with zero mean and variance $1/p$ the bound $p \geq C \cdot s \cdot \log(n/p)$ is achievable and can guarantee the exact reconstruction of $\mathbf{x}$ with overwhelming probability[1]. The same condition can be applied to binary matrices with independent entries taking values $\pm 1/\sqrt{p}$. As another example, we imply *Fourier ensembles* which are obtained by selecting p renormalized rows from the $n \times n$ discrete Fourier transform, with the bound $p \geq C \cdot s \cdot (\log n)^6$.

Now, assume that $\mathbf{x}$ is not sparse in the present form (e.g. time domain), but it can be sparsely represented in another basis (e.g. Fourier, wavelet, etc.). Mathematically speaking, we name α the sparse version of $\mathbf{x}$ satisfying $\mathbf{x} = \Psi\alpha$, where $\Psi \in \mathbb{R}^{n\times m}$ is called the representing (or sparsifying) dictionary (or basis). Ψ is called a *complete* dictionary if $n = m$, otherwise $n < m$ and we call it *overcomplete* (or *redundant*). For consistency in notations we always consider complete dictionaries thorough the paper, unless it is stated. The extension to overcomplete dictionaries is easy though. In order to achieve a successful CS, we must choose a Φ having least possible coherence with Ψ. In other words, a $\mathbf{D} := \Phi\Psi$ is desired to have columns with small correlations [7]. This property is termed as *mutual coherence*, and usually denoted by μ. More details about μ will be discussed later in this paper. The mutual coherence also affects the bound for p such that $p \geq C \cdot \mu^2 \cdot s \cdot (\log n)^6$ [6]. Interestingly, it is shown that random ensembles are largely incoherent with any fixed basis [7]. This is a useful property which allows us to non-adaptively choose a random Φ for any type of signal. More details and proofs in this context can be found in [6] and the references therein.

The second crucial job in CS is reconstruction, which can be described as solving the well known underdetermind problem with sparsity constraint,

$$\min \|\alpha\|_0 \quad s.t. \quad \mathbf{y} = \Phi\Psi\alpha \equiv \mathbf{D}\alpha \tag{1}$$

where $\|\cdot\|_k$ denotes the ℓ^k-norm in general, and here, ℓ^0-norm gives the support of α, which is actually the level of sparsity. Although this problem is sometimes described in other forms, the main issue is to find the sparsest possible α satisfying $\mathbf{y} = \mathbf{D}\alpha$. Unfortunately, the problem (1) is nonconvex in general, and the solution needs a combinatorial search among all possible sparse α, which is not feasible. However, there are some greedy methods trying to iteratively solve (1). The family of Matching Pursuit (MP) [8] methods such as Orthogonal Matching Pursuit (OMP), stagewise OMP [9], and Iterative Hard Thresholding (IHT) [10] mainly attempt

[1] C is a computable constant.

 ISSN 2076-1465

to solve (1) by selecting the vectors of $\mathbf{D}$ which are mostly correlated with $\mathbf{y}$. Thanks to the greedy nature of these algorithms; they are fast. In contrast, there have been proposed some optimization-based methods, mainly achievable by linear programming, called Basis Pursuit (PB) [11] which attempt to recover α by converting (1) into a convex problem which relaxes the ℓ^0-norm to an ℓ^1-norm problem:

$$\min \|\alpha\|_1 \quad s.t. \quad \mathbf{y} = \Phi\Psi\alpha \equiv \mathbf{D}\alpha \tag{2}$$

Although random matrices are suitable choices for Φ, it has been recently shown that optimizing Φ with the hope of reducing the mutual coherence can improve the performance [12–15]. Elad [12] attempts to iteratively decrease the average mutual coherence using a shrinkage operation followed by a Singular Value Decomposition (SVD) step. Duarte-Carvajalino et al. [13] take the advantage of an eigenvalue decomposition process followed by a KSVD-based algorithm [13] (see also [16]) to optimize Φ and train Ψ, respectively. Overally, the results of the current methods show enhancement in terms of both reconstruction error and compression rate. That motivates us to work more on optimizing the measurement matrix to obtain better results and also attempt to make the optimization process more efficient, specially for large-scale signals.

In this paper we propose a gradient-based optimization approach to decrease the mutual coherence between the measurement matrix and the representing matrix. This can be achieved by minimizing the absolute off-diagonal elements of the corresponding Gram matrix $\mathbf{G} = \tilde{\mathbf{D}}^T\tilde{\mathbf{D}}$, where $\tilde{\mathbf{D}}$ is the column-normalized version of $\mathbf{D}$, and $(.)^T$ denotes the transposition. Our idea is to approximate $\mathbf{G}$ with an identity matrix using a gradient descent method.

In the next section we formally express the sensing problem and the required mathematics related to optimization of the measurement matrix. Then, in section 3, the gradient-based optimization method is described in details. In section 4, the simulations are given to examine the proposed method from practical perspectives. Finally, the paper is concluded in section 5.

2. PROBLEM FORMULATION

Similar to the notations in the previous section, we consider the signal $\mathbf{x}$ to be sparse with cardinality s over the dictionary $\Psi \in \mathbb{R}^{n\times m}$. Consider the noiseless case, we are going to take $p < n$ linear measurements which based on CS rules can be done by multiplying a Φ with random i.i.d Gaussian entries such that $\mathbf{y} = \Phi\Psi\alpha$[2]. However, p cannot exceed the bounds mentioned in the previous section, which highly depends on the coherence between Φ and Ψ and the level of sparsity s. One of the suitable ways to measure the coherence between Φ and Ψ is to look at the columns of $\mathbf{D}$, instead. As $\mathbf{D} = \Phi\Psi$, the mutual coherence (which is desired to be minimized) can be defined as the maximum absolute value and normalized inner product between all columns in $\mathbf{D}$ which can be described as follows [11],

$$\mu(\mathbf{D}) = \max_{i\neq j, 1\leq i,j\leq m} \left\{ \frac{|\mathbf{d}_i^T \mathbf{d}_j|}{\|\mathbf{d}_i\| \cdot \|\mathbf{d}_j\|} \right\} \tag{3}$$

[2]Note that in cases where $\mathbf{x}$ is sparse in the current domain, we ignore Ψ and consider $\mathbf{x} = \alpha$ and $\Phi = \mathbf{D}$ without loss of generality.

Another suitable way to describe μ, especially for the purpose of this paper, is to compute the Gram matrix $\mathbf{G} = \tilde{\mathbf{D}}^T\tilde{\mathbf{D}}$, where $\tilde{\mathbf{D}}$ is column-normalized version of $\mathbf{D}$. We then define μ_{mx} (which is equal to $\mu(\mathbf{D})$), as the maximum absolute off-diagonal elements in $\mathbf{G}$,

$$\mu_{mx} = \max_{i\neq j, 1\leq i,j\leq m} |g_{ij}| . \tag{4}$$

Moreover, the average absolute off-diagonal elements in $\mathbf{G}$ is another useful measure defined as

$$\mu_{av} = \frac{\sum_{i\neq j} |g_{ij}|}{m(m-1)}. \tag{5}$$

Mainly, there are two important reasons why we are interested in matrices with small coherence and these motivate us to optimize Φ with the aim of decreasing the coherence. Suppose that the following inequality holds:

$$\|\alpha\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right) \tag{6}$$

then, α is necessarily the sparsest reconstructed signal when $\mathbf{y}$ and $\mathbf{D}$ are known, and both BP and OMP are guaranteed to succeed [12] [17] [18]. This implies that as long as μ is very small, a successful reconstruction is achievable for a wider range of sparsity degree. Another key notion here is considering the *Restricted Isometry Property* (RIP) [6] [11], which implies that for a proper isometry constant, RIP ensures any subset of columns in $\mathbf{D}$ with cardinality less than sparsity level s, is nearly orthogonal. This results in a better CS behavior and guarantees the identifiability of the original signal by both OMP and BP [13].

3. PROPOSED APPROACH

So far, it has been realized that a low coherence between columns of $\mathbf{D}$ is a desired property in CS framework. In section 2, the mathematical expression of coherence led to computing the Gram matrix and from that point we concluded that small absolute off-diagonal elements in $\mathbf{D}$ is desired. Let us now consider the ideal case, where minimum possible coherence ($\mu_{mx} = \mu_{av} = 0$) occurs. This situation will give us $\mathbf{G} = \mathbf{I}_{mm}$, where $\mathbf{I}_{mm}$ is identity matrix with indicated dimensions (note that $\tilde{\mathbf{D}}$ is column-normalized). Unfortunately, this may only occur when $p \geq m$, which is meaningless in CS. However, we might be able to introduce a measurement matrix leading to a $\mathbf{G}$ as close as possible to identity, even if $p < m$. One way to provide such a matrix is to solve the following problem,

$$\hat{\mathbf{G}} = \arg\min_{\mathbf{G}} \| \mathbf{G} - \mathbf{I} \|_\infty^2 \tag{7}$$

where $\| \, . \, \|_\infty$ is defined as the maximum absolute off-diagonal elements of $\mathbf{G}$. However, we prefer to use the Frobenius norm denoted as $\| \, . \, \|_F$, which has the advantages of simplifying the minimization problem, and also participating not only the maximum absolute off-diagonal, but all off-diagonal elements of $\mathbf{G}$ in the minimization process. Therefor, minimizing (7) with Frobenius norm will effectively, but not directly, minimize μ_{mx} and μ_{av}.

In order to set up a practical cost function for our problem, assume either the case where $\mathbf{x}$ is sparse in current domain (i.e. $\mathbf{D} = \Phi$), or the dictionary Ψ is square ($n = m$).

Note, however, the extension to the case of redundant dictionaries ($n < m$) is easy and will be explained later in the sequel. If we then express $\mathbf{G}$ in terms of $\tilde{\mathbf{D}}$ and replace it in (7) we obtain the following unconstrained minimization problem,

$$\hat{\mathbf{D}} = \arg\min_{\tilde{\mathbf{D}}} \| \tilde{\mathbf{D}}^T \tilde{\mathbf{D}} - \mathbf{I} \|_F^2 \tag{8}$$

To minimize (8), we adopt a gradient-descent method. To do this, we first define the corresponding error as,

$$E = \| \tilde{\mathbf{D}}^T \tilde{\mathbf{D}} - \mathbf{I} \|_F^2 = \mathbf{Tr}\{(\tilde{\mathbf{D}}^T \tilde{\mathbf{D}} - \mathbf{I})(\tilde{\mathbf{D}}^T \tilde{\mathbf{D}} - \mathbf{I})^T\} \tag{9}$$

where $\mathbf{Tr}\{\cdot\}$ denotes the trace operation.

Then, we need to compute the gradient of E with respect to elements of $\tilde{\mathbf{D}}$, denoted by $\tilde{d}_{ij}$,

$$\nabla E \equiv \frac{\partial E}{\partial \tilde{d}_{ij}} = 4\tilde{\mathbf{D}}(\tilde{\mathbf{D}}^T \tilde{\mathbf{D}} - \mathbf{I}) \tag{10}$$

Once the gradient of E is obtained, the solution for (8) can be described as an iterative process to update $\tilde{\mathbf{D}} \leftarrow \tilde{\mathbf{D}} - \eta \nabla E$, where $\eta > 0$ is the stepsize which can be fixed or variable. Consequently, the full description of update equation in each iteration is expressed as,

$$\tilde{\mathbf{D}}_{(i+1)} = \tilde{\mathbf{D}}_{(i)} - \hat{\eta} \tilde{\mathbf{D}}_{(i)} (\tilde{\mathbf{D}}_{(i)}^T \tilde{\mathbf{D}}_{(i)} - \mathbf{I}) \tag{11}$$

where i is the iteration index and $\hat{\eta}$ is the new stepsize after merging the scaler 4 in (10) with η. The proposed algorithm starts with an initial random $\tilde{\mathbf{D}}_{(0)}$ and then iteratively updating $\tilde{\mathbf{D}}$. In addition, a normalization step for columns of $\tilde{\mathbf{D}}$, denoted by $\tilde{\mathbf{d}}^j$ with $j = 1, 2 \ldots n$, is required at each iteration:

$$\tilde{\mathbf{d}}_{(i+1)}^j \leftarrow \tilde{\mathbf{d}}_{(i)}^j / \|\tilde{\mathbf{d}}_{(i)}^j\|_2 \tag{12}$$

After K iterations, when the convergence condition(s) is(are) met, $\hat{\mathbf{D}} = \tilde{\mathbf{D}}_{(K)}$ is given as the solution for (8). Finally, if $\mathbf{x}$ is sparse in its current domain, $\hat{\Phi} = \hat{\mathbf{D}}$ would actually be the required measurement matrix and the algorithm is terminated. However, if $\mathbf{x}$ is sparse over Ψ_{nn}, then an inverse or pseudoinverse is required to obtain the measurement matrix $\hat{\Phi} = \hat{\mathbf{D}}\Psi^{-1}$. Note that the gradient descent methods, mainly do not guarantee a global minimum, but can normally provide a local minimum. The pseudocode of the proposed algorithm is given in Algorithm 1.

In order to extend the above algorithm for the case of sparse signals over a dictionary Ψ_{nm}, with $n < m$, we only need to consider Ψ in converting (7) to (8). Since $\mathbf{G} = \tilde{\mathbf{D}}^T \tilde{\mathbf{D}} = \Psi^T \Phi^T \Phi \Psi$, we modify the minimization problem and obtain,

$$\hat{\Phi} = \arg\min_{\tilde{\Phi}} \| \Psi^T \Phi^T \Phi \Psi - \mathbf{I} \|_F^2 \tag{13}$$

The gradient of the error is then computed,

$$\frac{\partial E}{\partial \phi_{ij}} = 4\Phi\Psi(\Psi^T \Phi^T \Phi \Psi - \mathbf{I})\Psi^T \tag{14}$$

and the update equation which directly updates Φ is expressed as:

$$\Phi_{(i+1)} = \Phi_{(i)} - \eta \Phi_{(i)} \Psi (\Psi^T \Phi_{(i)}^T \Phi_{(i)} \Psi - \mathbf{I}) \Psi^T. \tag{15}$$

Algorithm 1: Gradient-descent optimization.

Input: Sparse representation basis Ψ_{nn} (if necessary), Stepsize η, Maximum number of iterations K.
Output: Measurement matrix Φ_{pn}.

```
begin
    Initialize D to a random matrix.
    for k=1 to K do
        for j=1 to n do
            d^j <--- d^j/||d^j||_2
        end
        D <--- D - ηD(D^T D - I)
    end
    if Ψ_nn has been given as input then
        Φ̂ <--- DΨ^{-1}
    else
        Φ̂ <--- D
    end
end
```

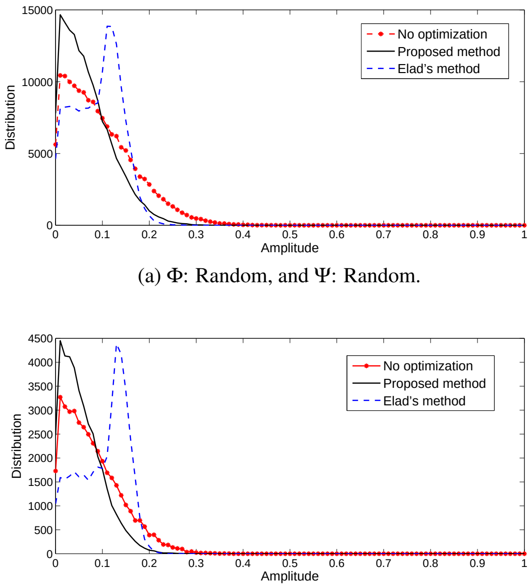
4. SIMULATION RESULTS

In this section we illustrate some results of our experiments to show the ability of the proposed method in optimizing the measurement matrix and consequently its effect in the reconstruction process. The results are encouraging and show that the idea of optimizing the measurement matrix can increase the performance in CS framework.

In the first experiment we built up a random dictionary $\Psi_{(200\times400)}$ and a random $\Phi_{(100\times200)}$ both with i.i.d Gaussian elements. We then ran the proposed algorithm with $\eta = 0.02$ and 60 number of iterations to optimize Φ. We also applied Elad's algorithm [12] to the same Φ and Ψ. The parameters we used for Elad's method were: the shrinkage parameter $\gamma = 0.5$, a *fixed* threshold $t = 0.25$ and 60 number of iterations. Figure 1(a) shows the distribution of absolute off-diagonal elements of $\mathbf{G}$ for this experiment. As seen from the figure, after applying the proposed method, the distribution becomes denser with more coefficients close to zero. This change verifies a smaller coherence between Φ and Ψ, which is also confirmed by having $\mu_{av} = 0.0075$ and $\mu_{mx} = 0.3887$ in this experiment, compared to $\mu_{av} = 0.0148$ and $\mu_{mx} = 0.4995$, for unoptimized Φ. Elad's method improves the coherence, almost similarly; $\mu_{av} = 0.0187$ and $\mu_{mx} = 0.4255$. However, large μ_{av} in Elad's method, also reported in [13], is due to an undesired peak around 0.15 in Figure 1(a), which may weaken the RIP conditions [13]. This drawback has been well mitigated in the proposed method as can be noticed from Figure 1(a). The same results are also seen for random i.i.d Gaussian $\Phi_{(100\times200)}$ and Discrete cosine transform (DCT) $\Psi_{(200\times200)}$, shown in Figure 1(b).

In the next experiment we generated a set of 10000 sparse signals with the length of $n = 120$, at random locations and random amplitudes, for nonzero samples. We chose DCT bases for Ψ, with the size of 120×120, i.e. $n = m = 120$. In this experiment we used $\eta = 0.01$ and 150 number of iterations. For Elad's method we used $\gamma = 0.95$, *relative* threshold $t = 20\%$ and 1000 number of iterations:

- First, we fixed the sparsity level to $s = 8$ and varied p from 20 to 80 in taking the measurements $\mathbf{y} = \Phi_{pn}\Psi_{nn}\alpha$. We then applied three reconstruction methods: IHT, OMP, and BP on all 10000 signals. This experiment was repeated for each $p \in [20\ 80]$, and then the relative error



(a) Φ: Random, and Ψ: Random.

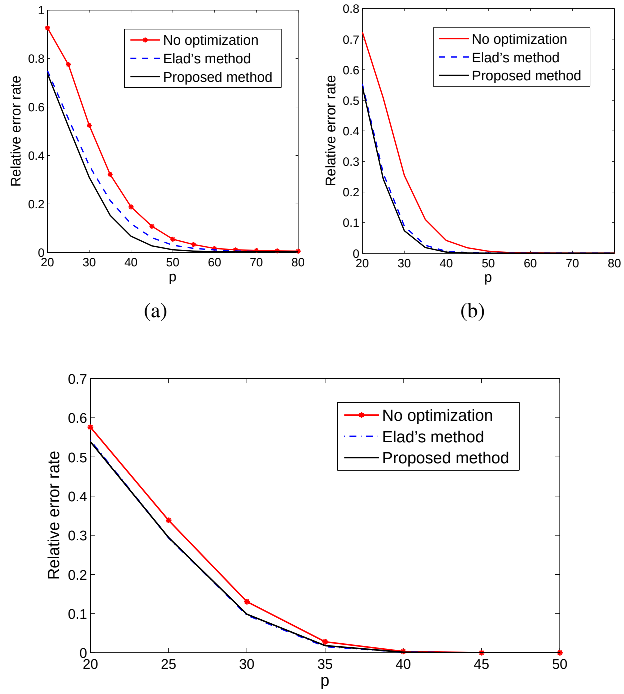
(b) Φ: Random, and Ψ: DCT.

Figure 1: Distribution of off-diagonal elements of $\mathbf{G}$.

rate between the reconstructed signals and the original signals were computed. The result of this experiment is shown in Figure 2. It is seen that optimization of Φ has a considerable influence in reducing the reconstruction error. In addition, it is observed that the proposed method works better compared to the work in [12]. It is also seen that the error rate is almost similar for both proposed and Elad's, when using BP for recustruction (Figure 2 (c)).

- Second, in order to evaluate the performance of the proposed method against changes in cardinality, we fixed $p = 30$ and varied the sparsity level from 1 to 20, and then reconstructed $\mathbf{x}$ by applying IHT and OMP for all 10000 sparse signals. The computed relative error rate for this experiment is shown in Figure 3. Again we see improvement compared to unoptimized case.

In the last experiment, we studied the effects of the proposed optimization on sampling and reconstruction of real images. Due to huge number of pixels, it is mainly impossible to process the whole image, directly. Hence, we applied the multi-scale strategy proposed in [19] and [20], considering wavelet transform as the representing dictionary to sparsify the input image. We used *symmlet8* wavelet with coarsest scale at 4 and the finest scale at 5, and followed the same procedure as in [19]. As an illustrative example, the *"Mondrian"* image of size 256×256 was compressed using unoptimized and optimized measurement matrices. Following the same procedure in [19] to keep the approximation coefficients, the whole image was then compressed to 870 samples. Then, the image was reconstructed using HALS_CS method [19, 21]. As can be seen from Figure 4, the reconstruction error computed as $\varepsilon = \|\hat{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$, where $\hat{\mathbf{x}}$ is the reconstructed image, is less when we compress the image using the proposed optimized measurement matrix. We also recorded the running time of this experiment where a desktop computer with a Core 2 Duo CPU of 3.00 GHz, and 2 G Bytes of RAM was used. It is seen from Figure 4 that adding the optimization step increases the running time, as expected. However, less computation time of the proposed method compared to Elad's is noticeable.

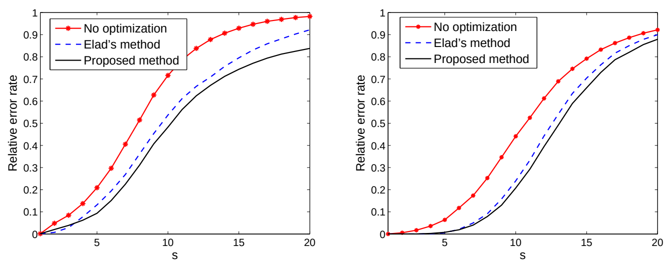


(a)

(b)

(c)

Figure 2: Relative error rate vs. the number of measurements p, using (a) IHT, (b) OMP, and (c) BP for reconstruction.



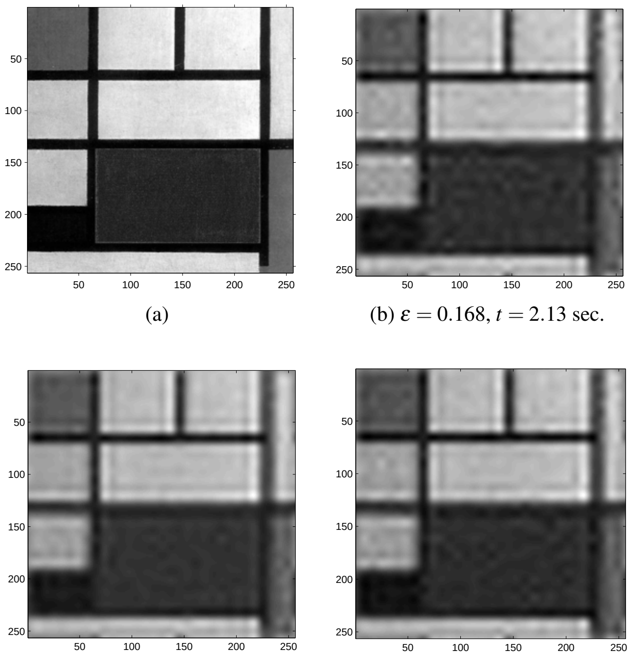
(a)

(b)

Figure 3: Relative error rate vs. the sparsity level s, using (a) IHT and (b) OMP for reconstruction.

5. DISCUSSION AND CONCLUSION

In this paper the problem of compressive sensing is investigated. A new gradient-based approach is proposed in which the aim is to optimize the measurement matrix in order to decrease the mutual coherence. An extension to the proposed algorithm is also presented which is suitable for sparse signals with respect to overcomplete dictionaries. The results of our simulations on both real and synthetic signals are promising and confirm that optimization of the measurement matrix increases the performance, which is introduced in terms of reconstruction error and the number of measurements taken. In practice, however, the proposed method is still not applicable to very large-scale problems. Lower complexity of the proposed method, compared to the previous methods, indi-



(a) (b) $\varepsilon = 0.168$, $t = 2.13$ sec.

(c) $\varepsilon = 0.157$, $t = 19.37$ sec. (d) $\varepsilon = 0.163$, $t = 444.3$ sec.

Figure 4: Reconstruction of Mondrian image using HALS_CS method. (a) Original image. Reconstruction with (b) no optimization, (c) proposed optimization, and (d) Elad's optimization of the measurement matrix.

cates the possibility of appropriateness of such methods for very large-scale problems. This fact has not been reported in the literature yet and needs to be challenged.

REFERENCES

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

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

[3] M. Lustig, D. Donoho, J.M. Santos, and J.M. Pauly. Compressed sensing MRI. *Signal Processing Magazine, IEEE*, 25(2):72–82, March 2008.

[4] S.F. Cotter and B.D. Rao. Sparse channel estimation via matching pursuit with application to equalization. *Communications, IEEE Transactions on*, 50(3):374–377, Mar 2002.

[5] J. Bobin, J.-L. Starck, and R. Ottensamer. Compressed sensing in astronomy. *Selected Topics in Signal Processing, IEEE Journal of*, 2(5):718–726, Oct. 2008.

[6] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[7] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[8] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, Dec. 2007.

[9] D. Donoho, Y. Tsaig, I. Drori, and J-l. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Technical report, 2006.

[10] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 2004.

[11] D. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49(3):906–931, 1989.

[12] M. Elad. Optimized projections for compressed sensing. *Signal Processing, IEEE Transactions on*, 55(12):5695–5702, Dec. 2007.

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

[14] P. Vinuelas-Peris and A. Artes-Rodriguez. Sensing matrix optimization in distributed compressed sensing. In *Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on*, pages 638 –641, Aug. 2009.

[15] V. Abolghasemi, S. Sanei, S. Ferdowsi, F. Ghaderi, and A. Belcher. Segmented compressive sensing. In *Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on*, pages 630 –633, Aug. 2009.

[16] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.

[17] J.A. Tropp. Greed is good: algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, 50(10):2231 – 2242, Oct. 2004.

[18] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l-minimization. *PNAS*, 100(5):2197–2202, March 2003.

[19] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley, 2009.

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

[21] S. Sanei, A.H. Phan, Jen-Lung Lo, V. Abolghasemi, and A. Cichocki. A compressive sensing approach for progressive transmission of images. In *Digital Signal Processing, 2009 16th International Conference on*, pages 1 –5, Jul. 5-7 2009.