Projection Matrix Optimization for Block-sparse Compressive Sensing

Shuang Li, Zhihui Zhu, Gang Li, Liping Chang and Qiuwei Li
Zhejiang Provincial Key Lab. for Signal Processing
College of Information Engineering
Zhejiang University of Technology
310023 Hangzhou, Zhejiang, P.R. China
Email: lishuang0809@gmail.com

***Abstract*—Traditionally, the projection matrix in compressive sensing (CS) is chosen as a random matrix. In recent years, we have seen that the performance of CS systems can be improved by using a carefully designed projection matrix rather than a random one. In particular, we can reduce the coherence between the columns of the equivalent dictionary thanks to a well-designed projection matrix. Then, we can get a lower reconstruction error and a higher successful reconstruction rate. In some applications, the signals of interest have nonzero entries occurring in clusters — i.e., block-sparse signals. In this paper, we use the equiangular tight frame (ETF) to approach the Gram matrix of equivalent dictionary rather than the identity matrix used in [1]. Then, we minimize a weighted sum of the subblock coherence and the interblock coherence of the equivalent dictionary. The simulation results show that our novel method for projection matrix optimization significantly improves the ability of block-sparse approximation techniques to reconstruct and classify signals than the method proposed by Lihi Zelnik-Manor (LZM) [1].**

***Index Terms*—Compressive sensing, projection matrix optimization, block-sparsity, coherence, ETF.**

I. Introduction

CS declares that sparse signals can be exactly recovered from a number of linear projections of dimension lower than the number of samples required by Shannon-Nyquist Theorem [2]. Recent work has demonstrated that the performance of CS can be improved by using a carefully designed projection matrix rather than a random one. The goal of projection matrix optimization is to construct a projection matrix which can improve the recovery ability for a given sparsifying dictionary. In other words, the projection matrix is optimized to improve the ability of sparse apporximation algorithms such as BP and OMP to recover the sparsest representation from an underdetermined system.

In some applications, there are some sparse vectors which have nonzero entries appearing in blocks rather than arbitrarily spreading throughout the vectors. They are the so-called block-sparse signals. In this paper, we are interested in these block-sparse signals which arise naturally when dealing with the multi-band signals [3] or in the measurements of gene expression levels [4].

We define two separate notions of coherence to analyze the block-sparse model: one is the coherence within a block, referred to as the subblock coherence which captures the local properties of a dictionary, and the other is the coherence between the different blocks, referred to as the interblock coherence which describes the global dictionary properties.

The main objective and contribution of this paper are:

- **Objective**: in this paper, we consider the problem of optimizing a projection matrix Φ for a CS system in which the dictionary Ψ is given. We use the ETF rather than the identity matrix to approach the Gram matrix of equivalent dictionary. The projection matrix optimization is formulated in terms of finding those Φ such that the *Frobenius* norm of the difference between the Gram matrix of the equivalent dictionary and the ETF Gram is minimized.
- **Contribution**: we solve the problem of projection matrix optimization for block-sparse signals which have nonzero entries occuring in clusters. At the same time, we propose an improved WCM algorithm based on ETF. Experiments are given to show that the projection matrix obtained by using our novel method outperforms others both in the normalized reconstruction error and the successful reconstruction rate.

The outline of this paper is arranged as follows. In Section II, we briefly review the basic work on projection matrix optimization and the basic knowledge of block-sparsity. In Section III, we present our objective for projection matrix optimization and solve it with an improved WCM algorithm based on ETF. Simulations are presented in Section IV to show the outperformance of our proposed method in improving signal reconstruction accuracy. Some concluding remarks are given in Section V to end this paper.

II. Preliminaries

A. Basic Work on Projection Matrix Optimization

In this subsection we briefly review the basic work on projection matrix optimization. The process of CS system can be formulated as follows:

$$y = \Phi x = \Phi\Psi\theta \overset{\triangle}{=} A\theta \tag{1}$$

978-1-4799-1027-4/13/$31.00 ©2013 IEEE

where $y \in \Re^M$ is an observation vector and $\Phi \in \Re^{M\times N}$ with $M < N$ is the sensing matrix or projection matrix. $x \in \Re^N$ is an unknown vector which can be represented as $x = \Psi\theta$, where $\Psi \in \Re^{N\times L}$ with $N \leq L$ is an overcomplete dictionary and θ is sufficiently sparse. $A = \Phi\Psi \in \Re^{M\times L}$ is referred to as the equivalent dictionary of the CS system.

The work of Tropp has shown that BP and OMP can succeed in recovering θ from the measurement $y = A\theta$ when the following condition holds [5]:

$$||\theta||_0 \leq \frac{1}{2}(1+\frac{1}{\mu(A)}) \tag{2}$$

where $\mu(A)$ is the mutual coherence of the equivalent dictionary A, which is defined as

$$\mu(A) \triangleq \max_{i\neq j} \frac{|A_i^{\mathcal{T}} A_j|}{\|A_i\|_2 \|A_j\|_2}$$

with A_j is the jth column vector of A.

The mutual coherence $\mu(A)$ characterizes the maximum linear dependency possibly achieved by any two columns of A and only reflects the most extreme correlation in A. Condition (2) is a worst-case bound and does not reflect the average recovery ability of sparse approximation methods. The smaller $\mu(A)$, the higher successful reconstruction rate. That is to say, the recovery can be improved when A is as orthogonal as possible.

Let $A = \begin{bmatrix} A_1 & A_2 & \cdots & A_L \end{bmatrix} \in \Re^{M\times L}$. Denote $g_{ij} \triangleq A_i^{\mathcal{T}} A_j$ as the (i,j)th element of the Gram matrix of A and

$$S_c \triangleq diag(g_{11}^{-1/2} \quad \cdots \quad g_{jj}^{-1/2} \quad \cdots \; g_{LL}^{-1/2})$$

The Gram matrix of $\bar{A} \triangleq AS_c$, denoted as $\bar{G} = \{\bar{g}_{ij}\}$, is normalized such that $\bar{g}_{jj} = 1,\ \forall\ j$. Obviously, $\mu(A) = \max_{i\neq j} |\bar{g}_{ij}|$.

Motivated by the above observations, LZM proposed a method for projection matrix design by minimizing $||G-I||_F^2$ [1]. In other words, their goal is to solve the following function

$$\begin{aligned} &\min_{\Phi} ||A^{\mathcal{T}}A - I_L||_F^2 = \min_{\Phi} ||\Psi^{\mathcal{T}}\Phi^{\mathcal{T}}\Phi\Psi - I_L||_F^2 \\ &= \min_{\Phi} \sum_{i\neq j} |g_{ij}|^2 + \sum_{j=1}^{L} |g_{jj}-1|^2 \end{aligned} \tag{3}$$

where I_L is the identity matrix of dimension L. As can be seen, this objective has a very clear physical meaning. The first term is the averaged coherence factor while the second term can be interpreted as a set of constraints on the norms of the equivalent atoms A_j to be one.

B. Block-Sparsity

In this subsection, we consider the problem of representing a vector $x \in \Re^N$ in a given dictionary $\Psi \in \Re^{N\times L}$ with $N \leq L$, i.e., $x = \Psi\theta$ with $\theta \in \Re^L$. As the system of equation $x = \Psi\theta$ is underdetermined, there are infinitely many solutions. In this paper, we consider the case of sparse vector θ with nonzero entries appearing in blocks rather than arbitrarily spread throughout the vector. Specific examples include signals which lie in unions of subspaces please refer to [6].

We view θ as a concatenation of blocks in order to define the block-sparsity. s is the length of the blocks and $\theta[j]$ denotes the jth block [7], i.e.,

$$\theta = [\underbrace{\theta_1 \cdots \theta_s}_{\theta^{\mathcal{T}}[1]} \underbrace{\theta_{s+1} \cdots \theta_{2s}}_{\theta^{\mathcal{T}}[2]} \cdots \underbrace{\theta_{L-s+1} \cdots \theta_L}_{\theta^{\mathcal{T}}[B]}]^{\mathcal{T}} \tag{4}$$

where $L = Bs$.

In a similar way, we can represent the dictionary Ψ as follows

$$\Psi = [\underbrace{\Psi_1 \cdots \Psi_s}_{\Psi[1]} \underbrace{\Psi_{s+1} \cdots \Psi_{2s}}_{\Psi[2]} \cdots \underbrace{\Psi_{L-s+1} \cdots \Psi_L}_{\Psi[B]}] \tag{5}$$

The k-sparsity of a vector θ is defined as

$$||\theta||_{2,0} = \sum_{j=1}^{B} I(||\theta[j]||_2 > 0) \leq k$$

which means that $x[j]$ has nonzero *Euclidean* norm for at most k indices j.

As $A = \Phi[\Psi[1]\ \Psi[2]\ \cdots\ \Psi[B]] = [A[1]\ A[2]\ \cdots\ A[B]]$, then the (i,j)th block of the Gram matrix $G \in \Re^{L\times L}$, $A[i]^{\mathcal{T}}A[j]$, is denoted as $G[i,j]$.

III. Sensing Matrix Optimization Based on ETF

It can be shown [8] that for a matrix A with $M\times L$, $\mu(A)$ is bounded with $\underline{\mu} \leq \mu(A) \leq 1$ where the low bound is given by

$$\underline{\mu} \triangleq \sqrt{\frac{L-M}{M(L-1)}} \tag{6}$$

We can note that when $M << L$, the low bound is approximately equal to $\frac{1}{\sqrt{M}}$. It shows that we can only make $\mu(A)$ infinitely close to $\underline{\mu}$ rather than 0.

A unit-norm frame $\{A_i\}$ (i.e., $||A_i||_2 = 1,\ \forall\ i$) is said to be *equiangular* if $|A_i^{\mathcal{T}} A_j| = c,\ \forall\ i \neq j$, where c is some positive constant. It can be shown [8] that the matrix A with $||A_i||_2 = 1,\ \forall\ i$ achieves $\mu(A) = \underline{\mu}$ *if and only if* $\{A_i\}$ is ETF, and that $\mu(A) = \underline{\mu}$ can only hold if $L \leq M(M+1)/2$. For convenience, such an Gram is called ETF Gram, denoted with G_{etf}.

ETF has already been used in optimal dictionary design[9]. As it is also a very nice averaged mutual coherence behavior, we formulate the optimal sensing matrix design problem as

$$\min_{\Phi, G_t \in H_\mu} ||G - G_t||_F^2 \tag{7}$$

where the dictionary Ψ is assumed to be given, $G = \Psi^{\mathcal{T}}\Phi^{\mathcal{T}}\Phi\Psi$ is the Gram matrix of the equivalent dictionary $A = \Phi\Psi$, and the space H_μ is defined as

$$\begin{aligned} H_\mu \triangleq \{G_t \in \Re^{L\times L}:\ G_t = G_t^{\mathcal{T}},\ G_t(j,j) = 1,\ \forall\ k, \\ \max_{i\neq j} |G_t(i,j)| \leq \mu\} \end{aligned} \tag{8}$$

The objective (7) can be solved by using the alternating minimization strategy.

Based on (8), we define the shrinking Gram G_t as

$$G_t[i,j]_n^m = \begin{cases} 1, & i=j, m=n \\ G[i,j]_n^m, & |G[i,j]_n^m| \le \underline{\mu} \\ sign(G[i,j]_n^m)\underline{\mu}, & |G[i,j]_n^m| > \underline{\mu} \end{cases} \quad (9)$$

with $G[i,j]_n^m$ denoting the (m,n)th entry of $G[i,j]$. Then

$$||G-G_t||_F^2 = \sum_{j=1}^{B}\sum_{i\neq j} ||G[i,j]-G_t[i,j]||_F^2$$

$$+(\sum_{j=1}^{B}||G[j,j]-G_t[j,j]||_F^2 - \sum_{m=1}^{K}(G_m^m-1)^2) + \sum_{m=1}^{K}(G_m^m-1)^2$$

As we define the total interblock coherence μ_B^t, the total subblock coherence ν^t and η as follows

$$\mu_B^t \triangleq \sum_{j=1}^{B}\sum_{i\neq j}||G[i,j]-G_t[i,j]||_F^2$$

$$\nu^t \triangleq \sum_{j=1}^{B}||G[j,j]-G_t[j,j]||_F^2 - \sum_{m=1}^{K}(G_m^m-1)^2$$

$$\eta \triangleq \sum_{m=1}^{K}(G_m^m-1)^2$$

which lead to

$$||G-G_t||_F^2 = \mu_B^t + \nu^t + \eta \quad (10)$$

Instead of minimize μ_B^t and ν^t directly, we propose a weighting factor α. Then, our goal is

$$\Phi = \arg\min_{\Phi} \ \frac{1}{2}\eta + (1-\alpha)\mu_B^t + \alpha\nu^t \quad (11)$$

where $0 < \alpha < 1$.

To obtain a surrogate function we define $f(G,G_t)$ as

$$f(G,G_t) \triangleq \frac{1}{2}\eta(G) + (1-\alpha)\mu_B^t(G,G_t) + \alpha\nu^t(G,G_t)$$

$$= \frac{1}{2}||u_\eta(G)||_F^2 + (1-\alpha)||u_\mu(G,G_t)||_F^2 + \alpha||u_\nu(G,G_t)||_F^2$$

where the matrix operators u_η, u_μ and u_ν are defined as

$$u_\eta(G)[i,j]_n^m = \begin{cases} G[i,j]_n^m - 1, & i=j, m=n \\ 0, & else \end{cases}$$

$$u_\mu(G,G_t)[i,j]_n^m = \begin{cases} G[i,j]_n^m - G_t[i,j]_n^m, & i\neq j \\ 0, & else \end{cases}$$

$$u_\nu(G,G_t)[i,j]_n^m = \begin{cases} G[i,j]_n^m - G_t[i,j]_n^m, & i=j, m\neq n \\ 0, & else \end{cases}$$

Now, we can write

$$\begin{aligned} f(G,G_t) = &\frac{1}{2}||G-h_\eta(G)||_F^2 + (1-\alpha)||G-h_\mu(G,G_t)||_F^2 \\ &+\alpha||G-h_\nu(G,G_t)||_F^2 \end{aligned} \quad (12)$$

where the matrix operators h_η, h_μ and h_ν are defined as

$$h_\eta(G)[i,j]_n^m = \begin{cases} 1, & i=j, m=n \\ G[i,j]_n^m, & else \end{cases}$$

$$h_\mu(G,G_t)[i,j]_n^m = \begin{cases} G_t[i,j]_n^m, & i\neq j \\ G[i,j]_n^m, & else \end{cases}$$

$$h_\nu(G,G_t)[i,j]_n^m = \begin{cases} G_t[i,j]_n^m, & i=j, m\neq n \\ G[i,j]_n^m, & else \end{cases}$$

Based on (12), we define a surrogate objective $g(G,G^{(n)},G_t^{(n)})$ at the nth iteration as

$$\begin{aligned} g(G,G^{(n)},G_t^{(n)}) = &\frac{1}{2}||G-h_\eta(G^{(n)})||_F^2 \\ &+(1-\alpha)||G-h_\mu(G^{(n)},G_t^{(n)})||_F^2 \\ &+\alpha||G-h_\nu(G^{(n)},G_t^{(n)})||_F^2 \end{aligned} \quad (13)$$

It can be shown that the surrogate function (13) satisfies the conditions of a surrogate objective for the bound-potimization method and we no longer prove here.

Then, our goal is formulated as

$$\begin{aligned} &\min_{\Phi} g(G,G^{(n)},G_t^{(n)}) \\ &= \min_{\Phi} tr(A^{\mathcal{T}}AA^{\mathcal{T}}A - 2A^{\mathcal{T}}Ah_t(G,G^{(n)},G_t^{(n)}) \\ &= \min_{\Phi} ||G-h_t(G^{(n)},G_t^{(n)})||_F^2 \end{aligned} \quad (14)$$

where

$$\begin{aligned} h_t(G^{(n)},G_t^{(n)}) \triangleq &\frac{2}{3}(\frac{1}{2}h_\eta(G^{(n)}) + (1-\alpha)h_\mu(G^{(n)},G_t^{(n)}) \\ &+\alpha h_\nu(G^{(n)},G_t^{(n)})) \end{aligned}$$

Then, find the SVD of $h_t(G^{(n)},G_t^{(n)})$:

$$h_t(G^{(n)},G_t^{(n)}) = V_M\Delta_M V_M^{\mathcal{T}}.$$

then we can get

$$\Phi = \Delta_M^{\frac{1}{2}} V_M^{\mathcal{T}} \Psi^{\dagger}$$

where $\Psi^{\dagger}$ is the pseudo-inverse of Ψ.

A summary of the improved Weighted Coherence Minimization (WCM) algorithm is given below.

Algorithm : *WCM Based on ETF*

Objective: Sensing matrix optimization with a given block-sparsifying dictionary $\Psi \in \Re^{N\times L}$

$$\Phi = \arg\min_{\Phi} \ \frac{1}{2}\eta + (1-\alpha)\mu_B^t + \alpha\nu^t.$$

Initialization: Calculate the SVD of $\Psi^{\mathcal{T}}\Psi$. Set $\Phi^{(0)}$as the outcome of (3), i.e., $\Phi^{(0)} = [I_M \ 0]\Lambda^{-\frac{1}{2}}U^{\mathcal{T}}$.

Loop: Set $n = 1$.

- **Step I**: While $1 \le n \le iter$, compute
  $$G^{(n)} = (\Phi^{(n-1)}\Psi)^{\mathcal{T}}(\Phi^{(n-1)}\Psi).$$
- **Step II**: Obtain the shrunken Gram $G_t^{(n)}$ by applying the *shrinking* operation (9)
- **Step III**:Calculate $h_t(G^{(n)},G_t^{(n)})$.
- **Step IV**: Find the SVD of $h_t(G^{(n)},G_t^{(n)})$:
  $$h_t(G^{(n)},G_t^{(n)}) = V_M\Delta_M V_M^{\mathcal{T}}.$$
- **Step V**:Set $\Phi^{(n)} = \Delta_M^{\frac{1}{2}}V_M^{\mathcal{T}}\Psi^{\dagger}$.
- **Step VI**: End while.

End Algorithm

IV. Simulations Results

In this section, we compare the recovery and classification abilities of BOMP [7] when using projection matrix designed by our algorithm to the original WCM algorithm [1].

We generate a random dictionary $\Psi_{N\times L}$ with normally distributed entries and normalized atoms. We divide the dictionary into $B=\frac{L}{s}$ blocks of size s. We then generate $Num=1000$ test signals Θ of dimension L which is the k block-sparse representations of X with respect to Ψ. We compare four options for designing $\Phi_{M\times N}$: (1) random; (2) the outcome of DS[10]; (3) the WCM proposed by LZM; (4) the improved WCM based on ETF denoted as WCMetf. The last two methods are initialized as the outcome of DS.

We use two measures to illustrate the success of the simulations based on their outputs Φ and $\widehat{X}$. (1) The successful reconstruction rate of X: $r=\frac{||\widehat{\Theta}\odot\Theta||_0}{Lks}$ with $\odot$ denotes element-wise multiplication. (2) The normalized reconstruction error: $e=\frac{||X-\Psi\widehat{\Theta}||_F}{||X||_F}$.

In this experiment, we set $N=60, k=2, s=3, B=40, M=14$, and $L=Bs=120$ to illustrate the performance of the improved WCM algorithm based on ETF as a function of the weighting factor α. The result is shown in Fig. 1 and Fig. 2. It indicates that our method for projection matrix design significantly improves the ability of block-sparse approximation techniques to reconstruct and classify signals than the other three methods. We can conclude that when designing sensing matrices for block sparse decoding, the best results are obtained by choosing α close enough to 1 in our method as well as in the original WCM. That is to say, in the improved WCM algorithm based on ETF, the best recovery results are obtained when the total subblock coherence of equivalent dictionary is minimized. Besides, from the Fig. 1 and Fig. 2, we can see that as α increases, the reconstruction error and the reconstruction rate change very slowly except for $\alpha=0.5$.

V. Conclusions

In this paper, we propose a novel method for the projection matrix optimization. We use the ETF to approach the Gram matrix of the equivalent dictionary rather than the identity matrix used before. Then, we minimize a weighted sum of the subblock coherence and the interblock coherence of the equivalent dictionary. We can conclude from the simulations that the best recovery results are obtained when the total subblock coherence of equivalent dictionary is minimized. Besides, the experiments also demonstrate that the projection matrix obtained by using our method significantly outperforms others in signal reconstruction accuracy.

Acknowledgment

This work was supported by the NSFC-Grants 61273195, CPSF-Grant 2012M511386 and ZJNSF-Grant LY13F010009.

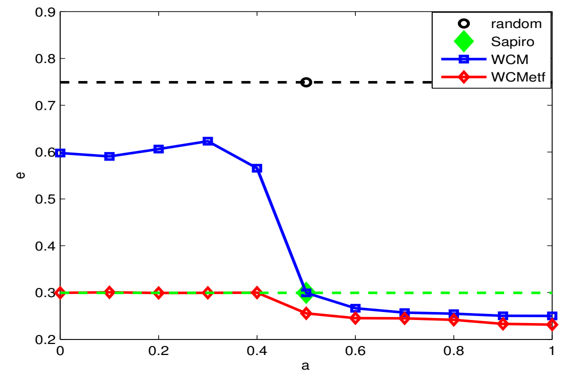

Fig. 1. The normalized reconstruction error corresponding to different weighting factor α.

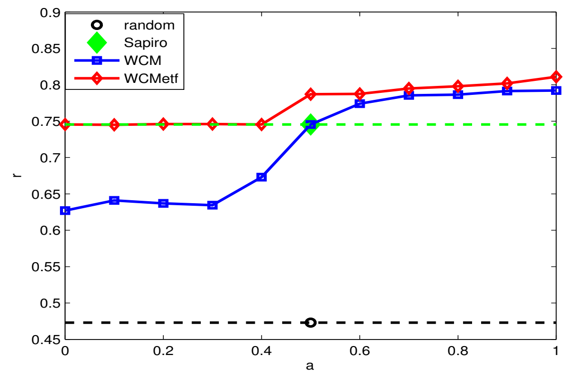

Fig. 2. The successful reconstruction rate corresponding to different weighting factor α.

References

[1] L. Zelnik-Manor, K. Rosenblum and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300-4312, Sep., 2011.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Sept., 2006.

[3] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993-1009, Mar., 2009.

[4] F. Parvaresh, H. Vikalo, S. Misra and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275-285, Jun., 2008.

[5] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231-2242, Oct., 2004.

[6] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302-5316, Nov., 2009.

[7] Y. C. Eldar, P. Kuppinger and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042-3054, Apr., 2010.

[8] T. Strohmer and R. W. Heath, "Grassmannian frames with applications to coding and communication," Appl. Comp. Harmonic Anal., vol. 14, no. 3, pp. 257-275, May 2003.

[9] M. Yaghoobi, L. Daudet, and M. E. Davies, "Parametric disctionary design for sparse coding," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4800-4810, 2009.

[10] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IMA Preprint Ser.*, no. 2211, May 2008.