USE OF TIGHT FRAMES FOR OPTIMIZED COMPRESSED SENSING

Evaggelia Tsiligianni, Lisimachos P. Kondi

University of Ioannina
Dept. of Computer Science
Ioannina, Greece

Aggelos K. Katsaggelos

Northwestern University
Dept. of EECS
Evanston, IL, USA

ABSTRACT

Compressed sensing (CS) theory relies on sparse representations in order to recover signals from an undersampled set of measurements. The sensing mechanism is described by the projection matrix, which should possess certain properties to guarantee high quality signal recovery, using efficient algorithms. Although the major breakthrough in compressed sensing results is obtained for random matrices, recent efforts have shown that CS performance could be improved with optimized non-random projections. Designing matrices that satisfy CS theoretical requirements is closely related to the construction of equiangular tight frames, a problem that has applications in various scientific fields like sparse approximations, coding, and communications. In this paper, we employ frame theory and propose an algorithm for the optimization of the projection matrix that improves sparse signal recovery.

***Index Terms*—** Compressed sensing, tight frames, Grassmannian frames

1. INTRODUCTION

The conventional approach to sampling signals or images follows the Shannon sampling theorem, that is, the sampling rate must be at least twice the maximum frequency present in the signal (the so-called Nyquist rate). Compressed sensing (CS) theory [1, 2], asserts that one can recover certain signals from far fewer samples.

Let $x \in \mathbb{R}^K$ be a vector representing a signal of length K. We want the number of available measurements, m, to be much smaller than the dimension, K, of the signal. We need a linear sensing mechanism described by the following equation

$$y = Px, \tag{1}$$

with $y \in \mathbb{R}^m$ and $P \in \mathbb{R}^{m\times K}$, a proper *sensing or projection matrix*.

The reconstruction of the original signal x from $m \ll K$ measurements leads to an underdetermined linear system of m equations and K unknowns. The question that arises is under what conditions would such a system have a unique solution and how it could be obtained. An important result is that uniqueness is guaranteed for the sparsest possible solution [3]. Obviously, CS can be applied to signals with sparse representations. We expand x in a basis or dictionary $D \in \mathbb{R}^{KXN}$, $K \leq N$, such that the number T of the most significant coefficients, carrying the useful information of the signal, is much smaller than the signal length. Therefore, we obtain a T-*sparse representation* of the original signal x

$$x = D\alpha, \tag{2}$$

with $\alpha \in \mathbb{R}^N$ and $\|\alpha\|_0 = T$. We denote by $\|\cdot\|_0$ the so-called ℓ_0-norm (which is actually not a norm) counting the nonzero coefficients of the respective signal.

According to equations (1) and (2), the sensing process of a signal $x \in \mathbb{R}^K$, with a sparse representation $\alpha \in \mathbb{R}^N$, can be described by the following equation:

$$y = PD\alpha, \quad \text{or} \quad y = F\alpha, \tag{3}$$

with $F = PD$, $F \in \mathbb{R}^{m\times N}$.

Considering the recovery process, seeking the sparsest solution leads to the ℓ_0-minimization problem,

$$\min_{\alpha\in\mathbb{R}^N} \|\alpha\|_0 \quad \text{subject to} \quad y = F\alpha, \tag{4}$$

which is known to be NP-hard. Approximate methods of solving (4) have been proposed and appropriate algorithms such as BP, OMP [4, 5] exist. The efficiency of these algorithms depends on the properties of the matrix F; incoherence, which expresses a low correlation between the projection matrix P and the representation matrix D, is an important one. A matrix drawn randomly from a suitable distribution is incoherent with any arbitrary orthonormal basis [6].

The idea that non-random matrices could be more effective than random projections has been expressed only recently with a few publications [7, 8, 9, 10, 11, 12]. An optimally designed projection matrix could improve the reconstruction accuracy or further reduce the necessary number of samples. Thus, designing a projection matrix is a challenge.

In this paper, we borrow some concepts from frame theory and propose a method for optimizing the projection matrix based on the work of [7]. The rest of the paper is organized as follows: In Section 2 we discuss the impact of

This work was supported by a Marie Curie International Reintegration Grant within the 7th European Community Framework Programme and in part by a grant from the Department of Energy (DE-NA0000457).

incoherent matrices in CS and survey previous work on the optimization of projection matrices. Section 3 reviews some basic concepts of frame theory. In Section 4 we summarize the basic steps of the algorithm proposed in [7] and explain the proposed modifications for its improvement. Finally, in Section 5 experimental results are presented, and conclusions are drawn in Section 6.

2. INCOHERENCE

In order to acquire maximal signal information with an undersampled set of measurements, CS theory requires that the projection matrix P and the representation matrix D must be incoherent. Let $P = [p_1, p_2, \ldots, p_m]^T, p_i \in \mathbb{R}^K, 1 \le i \le m$ and $D = [d_1, d_2, \ldots, d_N], d_j \in \mathbb{R}^K, 1 \le j \le N$. *Coherence* measures the maximal correlation between the two matrices, that is

$$\mu(P, D) = \max_{\substack{1 \le i \le m \\ 1 \le j \le N}} \frac{\left|p_i^T d_j\right|}{\|p_i\|_2 \, \|d_j\|_2}, \tag{5}$$

and must be as small as possible. The necessary number of measurements m depends on this property. Considering an orthonormal basis D, a T-sparse signal of length N can be reconstructed exactly with overwhelming probability from m measurements, if

$$m \ge C \cdot \mu^2(P, D) \cdot T \cdot \log N, \tag{6}$$

where C is some positive constant [6]. Similar results for redundant dictionaries can be found in [13].

An equivalent analysis that takes into consideration noisy measurements utilizes the correlation between the columns of the matrix $F = PD$. Considering the underdetermined linear system (3), it would be desirable that F had properties similar to an orthonormal basis. This concept is expressed by the Restricted Isometry Property (RIP) [6]. We will loosely say that when a matrix F obeys RIP of order s, then all subsets of s columns of F are nearly orthogonal.

A more convenient way of analyzing the recovery abilities of the matrix F is the *mutual coherence* that measures the maximal correlation between different columns of F, that is

$$\mu\{F\} = \max_{\substack{1 \le i,j \le N \\ i \ne j}} \frac{\left|f_i^T f_j\right|}{\|f_i\|_2 \, \|f_j\|_2}. \tag{7}$$

It is known [3] that if the representation satisfies the following condition,

$$\|\alpha\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu\{F\}}\right), \tag{8}$$

then α is the sparsest solution to (4) and, therefore, it is unique. The above relation implies that smaller values of mutual coherence enable CS reconstruction of signals with denser representations.

It is evident from (6) that optimizing the choice of the projection and the representation matrix, in a way to minimize the coherence between the two matrices, would offer higher quality signal recovery with reduced number of measurements. However, direct optimization of these matrices would be prohibitive as it involves combinatorial search. Recently, some algorithms for indirect design of incoherent matrices have been proposed. In [7] Elad presents an algorithm to optimize the choice of the projection matrix P, which decreases the average coherence of the matrix $F = PD$, keeping fixed the dictionary D. In [9] the authors aim at producing a matrix F approximate to the identity matrix by introducing a training based framework for the joint design and optimization of the representation dictionary and the projection matrix. Similar treatment can be found in [10, 11, 12].

The minimization of the maximal correlation between the columns of an $m \times N$ matrix, with $m < N$, is a problem that arises in numerous other contexts, besides CS. In the present work, we address this problem as a finite-dimensional frame design problem and borrow ideas from numerical linear algebra for its solution. Next, we will discuss some basic concepts of frame theory and explain our contribution to the optimization of projection matrices.

3. FRAMES

Viewing the problem of incoherent matrices from a theoretical perspective, we are led to Grassmannian frames [14]. Frames are a generalization of the idea of bases to sets that may be linearly dependent. A finite frame F_m^N in the complex Hilbert space $\mathbb{C}^m$ is a sequence of $N \ge m$ vectors $\{f_k\}_{k=1}^N$, $f_k \in \mathbb{C}^m$ satisfying the following condition

$$\alpha \|f\|_2^2 \le \sum_{k=1}^{N} |\langle f, f_k \rangle|^2 \le \beta \|f\|_2^2 \quad \forall f \in \mathbb{C}^m, \tag{9}$$

where α, β are positive constants, called the lower and upper frame bounds, respectively. The redundancy of the frame is defined by $\rho = N/m$.

When constructing a frame, our goal is to combine the advantages of orthonormal bases with the advantages provided by the frame redundancy. Compared to the expansion of a signal in an orthonormal basis, a representation in an overcomplete or redundant frame can reveal certain signal characteristics such as its sparsity. However, the representation is not unique as the frame elements may be linearly dependent. On the other hand, the zero correlation between the basis vectors is the main advantage of an orthonormal basis. Let $\{e_k\}_{k=1}^m$ be an orthonormal basis in $\mathbb{C}^m$. Then $\langle e_k, e_\ell \rangle = 0$, for $k \ne \ell$, which expresses the linear independence between the basis vectors, responsible for the unique representation of every element $f \in \mathbb{C}^m$ as a linear combination of the e_k's. Searching for frames "as close as possible" to orthonormal bases, we focus on unit norm frames with minimal cross-correlation.

3.1. Unit norm tight frames

Let $F_m^N = \{f_k\}_{k=1}^N$ be a finite redundant frame in $\mathbb{C}^m$. Then it is possible to take $\alpha = \beta$ such that

$$f = \frac{1}{\alpha} \sum_{k=1}^{N} \langle f, f_k \rangle f_k, \quad \forall f \in \mathbb{C}^m. \tag{10}$$

We call these frames *α-tight frames*. If, moreover, $\|f_k\|_2 = 1$ for all k, then we obtain a *unit norm tight frame*. Considering the matrix formed by the vectors $\{f_k\}_{k=1}^N$ of an α-tight frame as its columns, the rows of this matrix have norms equal to $\sqrt{\alpha}$ and form an orthogonal family. It is obvious that we cannot design a unit norm tight frame with an arbitrary tightness parameter. For α-tight frames the following relation holds

$$\sum_{k=1}^{N} \|f_k\|_2^2 = \alpha m. \tag{11}$$

Thus, a unit norm tight frame F_m^N exists only for $\alpha = N/m$.

3.2. Equiangular frames

Orthonormal bases exhibit equal correlation between column vectors. Considering a redundant frame F_m^N, we want

$$|\langle f_k, f_\ell \rangle| = c, \quad \text{for } k \neq \ell. \tag{12}$$

For a unit norm frame, the absolute value of the inner product between two frame vectors equals the cosine of the acute angle between the two vectors. For this reason, frames satisfying (12) are called *equiangular*.

The maximal correlation between columns depends on the frame dimensions m, N. The lower bound on the minimal achievable correlation for equiangular frames is known [14] to be

$$\mu\{F_m^N\} \geq \sqrt{\frac{N-m}{m(N-1)}}. \tag{13}$$

Equiangular frames are important in applications like CS because they minimize the correlation between columns, a property that plays significant role in the performance of the reconstruction algorithms (OMP, BP) used in signal recovery.

3.3. Grassmannian frames

Among all unit norm frames with the same redundancy, the ones characterized by the property of minimal cross-correlation between their elements are called *Grassmannian* frames [14]. If minimal cross-correlation is the lowest achievable, as (13) implies, then we obtain an *optimal Grassmannian* frame. According to [14], an equiangular unit norm tight frame is an optimal Grassmannian frame.

The construction of optimal Grassmannian frames is not trivial. Actually, optimal Grassmannian frames exist only for a few frame dimensions. Therefore, our research focuses on frames that have properties similar to optimal Grassmannian.

3.4. Nearest tight frames

As unit norm tight frames with dimensions m, N exist for specific tightness parameters ($\alpha = N/m$), constructing an optimal Grassmannian frame equals to constructing an equiangular N/m-tight frame [14]. However, there is no explicit way of constructing such frames, even if we know that they exist. In this paper, we propose a construction that combines the minimization of frame correlation with the computation of an N/m-tight frame. The following theorem of numerical linear algebra [15] will help us to obtain an N/m-tight frame that is closest to a matrix with desirable properties.

Theorem 1: Given a matrix $F \in \mathbb{R}^{m \times N}$, $N \geq m$, suppose F has singular value decomposition $U\Sigma V^*$. With respect to the Frobenius norm, a nearest α-tight frame F' to F is given by αUV^*. Assume in addition that F has full low-rank. Then αUV^* is the *unique* α-tight frame closest to F. Moreover, one may compute UV^* using the formula $F' = \alpha(FF^*)^{-1/2}$.

4. CONSTRUCTING A TIGHT FRAME NEAREST TO OPTIMIZED PROJECTIONS

It is common practice in designing matrices with low column correlation to work with the Gram matrix [16, 7]. Given a matrix $F \in \mathbb{R}^{m \times N}$, formed by the frame vectors $\{f_k\}_{k=1}^N$ as its columns, the Gram matrix is the Hermitian matrix of the column inner products, that is $G = F^T F$. Having computed the Gram matrix of the normalized F, the maximal correlation is obtained as the off-diagonal entry of G with the largest absolute value.

In [7] Elad proposed an algorithm for the optimization of the Gram matrix leading to *optimized projections*. The algorithm uses a "shrinkage" process to optimize the values of the off-diagonal elements of the Gram matrix. Entries in G with magnitude above a threshold t are "shrunk" by a factor γ. Entries with magnitude below t but above γt are "shrunk" by a smaller amount. The new Gram matrix elements, $\widehat{g}_{ij}$, are obtained according to

$$\widehat{g}_{ij} = \begin{cases} \gamma g_{ij}, & |g_{ij}| \geq t, \\ \gamma t \cdot sign(g_{ij}), & t > |g_{ij}| \geq \gamma t, \\ g_{ij}, & \gamma t > |g_{ij}|. \end{cases} \tag{14}$$

We propose to combine the above "shrinkage" process with Theorem 1 to improve the optimization results. Starting from a projection matrix $P \in \mathbb{R}^{m \times K}$ and a fixed dictionary $D \in \mathbb{R}^{K \times N}$, the "shrinkage" aims at derceasing the mutual coherence of the effective dictionary $F = PD$, $F \in \mathbb{R}^{m \times N}$. Based on the observation that optimal Grassmannian frames not only exhibit minimal mutual coherence but N/m-tightness as well, we propose to apply Theorem 1 to the effective dictionary produced by "shrinkage" to obtain an N/m-tight frame.

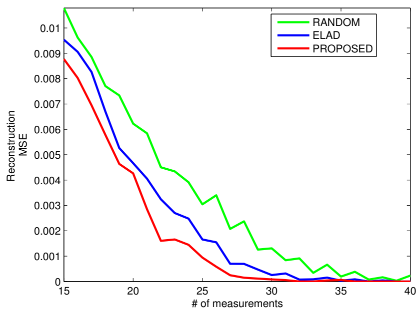

Fig. 1. Average reconstruction MSE as a function of the number of measurements.

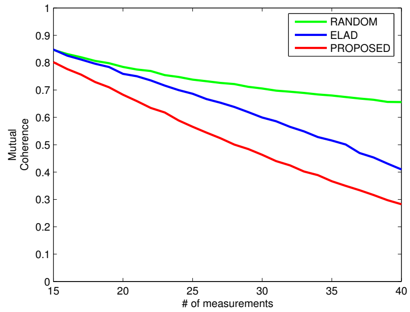

Fig. 2. Average mutual coherence for $m \times N$ matrices, with $m \in [15, 40]$ and $N = 120$.

Let S be the square root of the "shrunk" Gram matrix $\hat{G}$, i.e., $S^T S = \hat{G}$. We expect S to have lower mutual coherence compared to the initial matrix. We obtain an N/m-tight frame S' that is nearest to S according to $S' = (N/m) \cdot (SS^*)^{-1/2}$. The new matrix S' preserves low correlation between its columns as it is close to S, while it also exhibits N/m-tightness.

The optimization algorithm we propose is iterative. We choose the initial P to be a random Gaussian matrix and D an arbitrary fixed dictionary. First, we apply Theorem 1 to the initial $F_0 = PD$, obtaining an N/m-tight frame F_0' as input frame. After column normalization we obtain $\hat{F}_0$. The q-th iteration involves "shrinkage" and tightness including the following steps:

1. Obtain the matrix $\hat{F}_q$, after column normalization of F_q'.
2. Apply (14) to the off-diagonal elements of the Gram $G_q = \hat{F}_q^T \hat{F}_q$, to produce a "shrunk" Gram matrix $\hat{G}_q$.
3. Apply SVD to $\hat{G}_q$ to force the matrix rank to be m.
4. A matrix $S_q \in \mathbb{R}^{m \times N}$ is obtained as the square root of $\hat{G}_q$, i.e., $S_q^T S_q = \hat{G}_q$.
5. Obtain the nearest N/m-tight frame S_q' according to $S_q' = (N/m) \cdot (S_q S_q^*)^{-1/2}$. Set $F_q' = S_q'$.

The projection matrix P_{q+1} is obtained as a solution to $\min_P \left\|F_q' - PD\right\|$.

Regarding convergence, there is no theoretical guarantee that the above algorithm will converge as the problem is not convex. In practice, the algorithm behaves well, especially when the parameter γ is chosen close to 1 (see also [7]).

5. EXPERIMENTAL RESULTS

In order to test the proposed method in CS, we deploy the proposed optimized tight frames to acquire sparse synthetic signals. The reconstruction process involves solving a convex optimization problem, with a pursuit algorithm like OMP.

In every experiment, we generate a T-sparse vector of length N, $\alpha \in \mathbb{R}^N$, which constitutes a sparse representation of the K-length synthetic signal $x = D\alpha$, $x \in \mathbb{R}^K$, $K \leq N$. We choose the dictionary $D \in \mathbb{R}^{K \times N}$ to be a random Gaussian matrix. The nonzero coefficients' locations of the sparse vector are chosen at random. Besides the effectiveness of the projection matrix P, the reconstruction results also depend on the number of measurements m and the sparsity level of the representation T. Thus, our experiments include varying values of these parameters. For a specified number of measurements $m \ll K$, we create a random projection matrix $P \in \mathbb{R}^{m \times K}$. We obtain m projections of the original signal according to (3). The approximation $\widehat{\alpha}$ of the original sparse representation is obtained by solving (4).

In all experiments presented here, the synthetic signals were of length $K = 80$ and the respective sparse representations, under the dictionary D, of length $N = 120$. The execution of the optimization algorithm included up to 50 iterations, with parameters $\gamma = 0.95$ and $t = 0.2$, which are also used in [7]. Two sets of experiments have been considered; the first one includes varying values of the number of measurements and the second one includes varying values of the number of sparsity level of the signals under testing. For every value of the aforementioned parameters we made 300 experiments.

Figure 1 presents the reconstruction results as a function of the number of measurements m, for a fixed sparsity level ($T = 4$) of the treated signal, by means of average reconstruction MSE. In Fig. 2 we can see the average mutual coherence of the matrix F used in these experiments. The average mutual coherence and MSE were calculated for every measurement value, m, over 300 experiments. Figures 3 and 4 present the results for reconstruction and mutual coherence, respectively, for a fixed number of measurements ($m = 25$) and varying values of the sparsity level of the signal.

It is clear that regarding the reconstruction results of the acquired signal, the matrix produced by the proposed algorithm leads to a smaller reconstruction error compared to random matrices and to matrices produced by [7]. We see that trying to make a construction that is closer to Grassmannian frames yields an implicit substantial improvement of the mutual coherence of the effective dictionary F, which is respon-

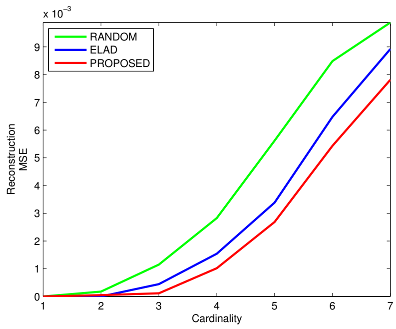

Fig. 3. Average reconstruction MSE as a function of the signal sparsity level.

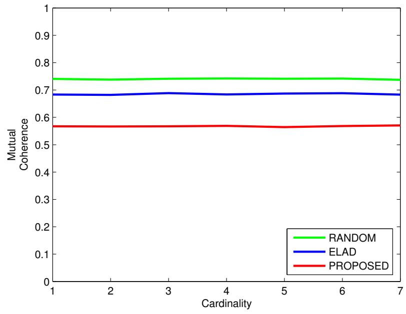

Fig. 4. Average mutual coherence for 25×120 matrices used in the experiments with varying signal sparsity level.

sible for the better reconstruction results we observe.

6. CONCLUSIONS

According to CS theory, it is possible to recover a sparse signal from incomplete measurements as far as the sensing matrix columns exhibit minimal correlation, a property that can be measured by mutual coherence. Based on an existing algorithm for minimizing the mutual coherence of a given matrix, we introduce into the minimization process the concept of tightness, towards a construction that is close to Grassmannian frames. The produced matrix yields better results regarding mutual coherence and reconstruction quality as well, advocating that any progress in the construction of Grassmannian frames would be an important contribution to CS.

7. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Th.*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Th.*, vol. 52, no. 2, pp. 489–509, 2006.

[3] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1-minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[5] G. Davis, S. Mallat, and M. Avellaneda, "Greedy adaptive approximation," *Constr. Approx.*, vol. 13, pp. 57–58, 1997.

[6] E. J. Candès and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

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

[8] Y. Weiss, F. S. Chang, and W. Freeman, "Learning compressed sensing," in *Allerton Conf. on Commun. Control and Comput.*, 2007.

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

[10] V. Abolghasemi, D. Jarchi, and S. Sanei, "A robust approach for optimization of the measurement matrix in compressed sensing," in *2nd Int. Wor. on Cogn. Inf. Proc. (CIP)*, June 2010, pp. 388 –392.

[11] J. Xu, Y. Pi, and Z. Cao, "Optimized projection matrix for compressive sensing," *EURASIP J. Adv. Signal Process.*, 2010.

[12] H. Huang and A. Makur, "Optimized measurement matrix for compressive sensing," in *Intl. Conf. Sampling Th. App.*, May 2011.

[13] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed Sensing and Redundant Dictionaries," *IEEE Trans. Inf. Th.*, vol. 54, no. 5, pp. 2210–2219, 2008.

[14] T. Strohmer and R. W. Heath, "Grassmannian frames with applications to coding and communication," *App. Comp. Harm. Anal.*, vol. 14, no. 3, pp. 257–275, 2003.

[15] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[16] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Th.*, vol. 51, no. 1, pp. 188 – 209, 2005.