Preconditioning for Underdetermined Linear Systems with Sparse Solutions

Evaggelia Tsiligianni, *Student Member, IEEE*, Lisimachos P. Kondi, *Senior Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—Performance guarantees for the algorithms deployed to solve underdetermined linear systems with sparse solutions are based on the assumption that the involved system matrix has the form of an incoherent unit norm tight frame. Learned dictionaries, which are popular in sparse representations, often do not meet the necessary conditions for signal recovery. In compressed sensing (CS), recovery rates have been improved substantially with optimized projections; however, these techniques do not produce binary matrices, which are more suitable for hardware implementation. In this paper, we consider an underdetermined linear system with sparse solutions and propose a preconditioning technique that yields a system matrix having the properties of an incoherent unit norm tight frame. While existing work in preconditioning concerns greedy algorithms, the proposed technique is based on recent theoretical results for standard numerical solvers such as BP and OMP. Our simulations show that the proposed preconditioning improves the recovery rates both in sparse representations and CS; the results for CS are comparable to optimized projections.**

***Index Terms*—Compressed sensing, incoherent unit norm tight frames, preconditioning, sparse representations.**

I. Introduction

SPARSE signal recovery was introduced in signal processing in the context of sparse and redundant representations as the problem of finding a signal representation with a few nonzero coefficients. Sparsity has improved the performance of many signal processing applications such as compression, feature extraction, pattern classification, and noise reduction [1].

A recent branch of sparse representations that has become a center of interest of its own, is compressed sensing (CS) [2], [3]. Exploiting sparsity, CS acquires signals at a drastically smaller rate than the Shannon/Nyquist theorem imposes, performing a number of measurements that is much smaller than the signal length.

At the heart of sparse representations and CS lies an underdetermined linear system defined by

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

with $A \in \mathbb{R}^{m\times N}$, $m < N$. Having more unknowns than equations, system (1) either has no solutions or infinitely many solutions. Assuming that the system matrix A is full rank to avoid the anomaly of having no solution, one way to guarantee a single solution is to enforce sparsity [4]. Therefore, we assume that $\|x\|_0 = T$, $T \ll m$, where $\|\cdot\|_0$ is the so-called ℓ_0-norm (which is actually not a norm) counting the non-vanishing components of the respective vector. Seeking a sparse solution, we are led to the following minimization problem

$$\hat{x} = \arg\min_x \|x\|_0 \quad \text{s.t.} \quad y = Ax. \tag{2}$$

Problem (2), known as *sparse recovery*, is NP-hard and can be solved with numerical methods including greedy algorithms and convex relaxation. Performance guarantees for standard numerical solvers such as OMP and BP underline that A must be an incoherent unit norm tight frame [5]. Incoherence is a property that characterizes the similarity between the columns of A, denoted by a_i. The worst similarity, defined by

$$\mu(A) = \max_{\substack{1\le i,j\le N\\ i\ne j}} \frac{|\langle a_i, a_j\rangle|}{\|a_i\|_2\|a_j\|_2}, \tag{3}$$

is known as *mutual coherence*. A matrix with small mutual coherence is referred to as *incoherent*.

In sparse representations, CS and many other problems assuming the above formulation, the involved system matrix does not always satisfy the necessary conditions for sparse recovery. In this paper, we transform (1) into a form that is more suitable for finding numerically a sparse solution, a process referred to as *preconditioning*. Using frame theory and following recent theoretical results for sparse recovery, the proposed preconditioning yields a system matrix having the properties of an incoherent unit norm tight frame (UNTF). The proposed technique is based on an algorithm for building incoherent UNTFs presented in [6], [7]. Besides sparse representation problems, for the first time to the best of our knowledge, we apply preconditioning in CS. Using binary random matrices for sensing, we show that preconditioning can improve the efficiency of the acquisition hardware substantially; according to our simulations, the performance of the deployed sparse recovery algorithms is similar to the one observed for optimized projections in [7].

The rest of the paper is organized as follows. In Section II we review the sparse recovery problem in the context of sparse representations and CS. As our construction employs frame theory, Section II also provides the necessary background regarding

Manuscript received September 19, 2014; revised December 25, 2014; accepted January 08, 2015. Date of publication January 13, 2015; date of current version January 28, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gitta Kutyniok.

E. Tsiligianni and L. P. Kondi are with the Department of Computer Science & Engineering, University of Ioannina, Ioannina, Greece (e-mail: etsiligia@cs.uoi.gr; lkon@cs.uoi.gr).

A. K. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University, IL 60208 USA (e-mail: aggk@eecs.northwestern.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2392000

frames. The proposed preconditioning and the algorithm for building incoherent UNTFs are presented in Section III. Our experiments in Section IV include signal recovery both in sparse representation problems and CS, deploying OMP and BP. Finally, conclusions are drawn in Section V.

II. Background

A. Sparse and Redundant Representations

The weakness of orthogonal transforms to provide highly sparse representations has promoted the development of overcomplete dictionaries. An overcomplete or redundant dictionary is an $m \times N$ matrix, $m < N$, with unit norm columns known as atoms, spanning the m-dimensional signal space. When we expand a signal $y \in \mathbb{R}^m$ in an overcomplete dictionary $D \in \mathbb{R}^{m \times N}$, we obtain the underdetermined linear system

$$y = Dx, \tag{4}$$

where $x \in \mathbb{R}^N$ is the vector of the unknown coefficients. Seeking a sparse vector x satisfying (4) leads to a minimization problem of the form (2). Considering the necessary conditions for signal recovery, D must be an incoherent UNTF.

Although constructions of incoherent tight dictionaries appear often in signal processing applications, such dictionaries have a limited ability of sparsifying signals or are suitable only for certain signal types. Learning based dictionaries that have been proposed as an alternative, contain atoms generated from instances belonging to a particular signal family. Every signal in the family can then be represented as a linear combination of a few atoms from the dictionary. The weakness of such dictionaries to satisfy incoherence properties, motivated the authors in [8] to propose a modification of thresholding and OMP, such that in the estimation of the unknown support, a matrix other than the original representation dictionary is employed. Regarding thresholding, an explicit formula for calculating the optimal matrix for support estimation is given in [9].

B. Compressed Sensing

Compressed sensing offers simultaneous acquisition and compression of sparse signals, using a sensing mechanism described by

$$y = PDx, \tag{5}$$

where $P \in \mathbb{R}^{d \times m}$, $d \ll m$, is the *sensing* or *projection* matrix and D is the $m \times N$ representation dictionary. System (5) is underdetermined with d equations and N unknowns. Seeking a sparse solution, we are led to a problem of the form (2), which can be solved numerically as long as the *effective dictionary* $\Phi = PD$ is close to an incoherent UNTF.

Successful signal reconstruction in CS is based on the choice of the projection matrix. While random matrices are considered a universal solution, the demand to increase reconstruction accuracy and reduce the necessary number of measurements has led to new theoretical and practical results. Towards this direction, a substantial improvement has been achieved with optimized projections [10], [6], [7]. Nevertheless, when talking about projection matrices, a significant issue is the design of acquisition hardware. Binary random matrices are considered the best option for practical implementation. However, the recovery rates they yield are similar to the ones achieved with random Gaussian matrices at best [11], [12], while certain types of binary projections work well only when combined with specific representation dictionaries [13].

C. Frames

Frames [14] have been popular in sparse and redundant representations as they are a natural extension of orthogonal bases. A finite frame F_m^N in a real or complex m-dimensional Hilbert space $\mathbb{H}^m$ is a sequence of $N \geq m$ vectors $\{f_k\}_{k=1}^N$, $f_k \in \mathbb{H}^m$, satisfying the following condition

$$\alpha \|f\|_2^2 \leq \sum_{k=1}^{N} |\langle f, f_k \rangle|^2 \leq \beta \|f\|_2^2 \quad \forall f \in \mathbb{H}^m, \tag{6}$$

with positive constants α and β. When $\alpha = \beta$, we obtain an α-tight frame, that is

$$f = \frac{1}{\alpha} \sum_{k=1}^{N} \langle f, f_k \rangle f_k \quad \forall f \in \mathbb{H}^m. \tag{7}$$

In this case, the rows of $\alpha^{-1/2} F_m^N$ form an orthonormal family. A tight frame with unit norm columns, referred to as unit norm tight frame (UNTF), exists only for $\alpha = N/m$ [14].

Frame theory has become important in sparse reconstruction problems due to equiangular tight frames (ETFs), which exhibit equal correlation between frame elements [15]. Equiangular UNTFs, also known as optimal Grassmannian frames, are ideal candidates for sparse recovery algorithms as they meet the minimum possible bound regarding mutual coherence [15], that is

$$\mu(F) \geq \sqrt{\frac{N-m}{m(N-1)}}. \tag{8}$$

Frames satisfying (8) with equality do not exist for arbitrary frame dimensions m, N, while their construction has been proven extremely difficult.

Next, we present a design methodology for building incoherent UNTFs that are close to optimal Grassmannian frames.

III. Incoherent Unit Norm Tight Frames for Sparse Recovery

A. Preconditioning for Underdetermined Linear Systems with Sparse Solutions

In linear algebra and numerical analysis, preconditioning is a process that conditions a given problem into a form that is more suitable for numerical solution [16]. Given a linear system $y = Ax$, a preconditioner C^{-1} of the matrix A is a matrix such that CA has a smaller condition number than A. Considering an underdetermined linear system with sparse solutions, recent results have shown that problem (2) can be efficiently solved with greedy algorithms and convex relaxation, if the system matrix is an incoherent UNTF [5]. Therefore, a preconditioner for a minimization problem of the form (2) should yield a system matrix as close to an incoherent UNTF as possible.

Let A be an arbitrary $m \times N$ matrix, not satisfying the necessary conditions for sparse recovery. Suppose there exists an $m \times m$ matrix C such that the product CA be an incoherent UNTF. Multiplying both sides of (1) by C, we obtain

$$Cy = CAx \quad \text{or} \quad z = CAx, \tag{9}$$

where $z = Cy$. Requiring C to be invertible, implies that system (1) is equivalent to (9). Therefore, solving the following minimization problem

$$\hat{x} = \arg\min_{x} \|x\|_0 \quad \text{s.t.} \quad z = CAx, \tag{10}$$

we obtain a solution that satisfies also (2).

Problem (10) involves the *effective system matrix* $F = CA$; thus, the efficiency of the numerical algorithms deployed to solve it depends on the properties of F. The question that naturally arises is how can we construct an invertible $m \times m$ matrix C such that the effective matrix F is an incoherent UNTF?

B. Construction of Preconditioner

Incoherent UNTFs are frames close to optimal Grassmannian frames. Optimal Grassmannian frames not only exhibit minimal mutual coherence, but N/m-tightness as well. Thus, we propose the following design methodology: First, we compute a matrix with small mutual coherence. Then, we find a UNTF that is nearest to the computed incoherent matrix in the Frobenius norm.

Regarding the first step, we work with the Gram matrix. 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 $G = F^*F$. For unit norm frame vectors, the maximal correlation is obtained as the largest absolute value of the off-diagonal entries of G. We propose to bound the off-diagonal entries according to

$$\tilde{g}_{ij} = \begin{cases} \operatorname{sgn}(g_{ij}) \cdot \frac{1}{\sqrt{m}}, & \text{if} \quad 1/\sqrt{m} < |g_{ij}| < 1, \\ \quad g_{ij}, & \text{otherwise}, \end{cases} \tag{11}$$

where g_{ij} is the (i,j) entry of the Gram matrix. The selected bound $1/\sqrt{m}$ is approximately equal to the lowest bound (see eq. (8)) for large values of N. Other choices of the bound might be considered depending on the frame dimensions.

Regarding the second step, we must solve a matrix nearness problem. We can solve this problem algebraically by employing the following theorem [17].

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

The algorithm we propose is iterative. We select the initial matrix C_{init} to be an $m \times m$ random Gaussian matrix; a square random matrix will almost never be singular [18]. Setting $F_0 = C_{\text{init}}A$, the q-th iteration of the proposed algorithm involves the following steps:

1. Obtain the matrix $\hat{F}_q$, after column normalization of F_q.
2. Calculate the Gram matrix $\hat{G}_q = \hat{F}_q^* \hat{F}_q$ and apply (11) to bound the absolute values of the off-diagonal entries, producing $\tilde{G}_q$. (Our experience indicates that this step preserves positive semidefiniteness of $\tilde{G}_q$, which is necessary to proceed.)
3. Obtain $\check{G}_q$ of rank m by computing the truncated SVD of $\tilde{G}_q$, i. e. set the $N - m$ smallest singular values to zero.
4. Let USU^* the SVD of $\check{G}_q$. The matrix $S_q = (S)^{1/2}U^*$ satisfies $S_q^* S_q = \check{G}_q$.
5. Apply Theorem 1 to obtain $S'_q = \sqrt{N/m} \cdot (S_q S_q^*)^{-1/2} S_q$.
6. Find the $m \times m$ matrix C_q by solving the minimization problem $\min_C \|CA - S'_q\|_2$.
7. Set $F_{q+1} = C_q A$.

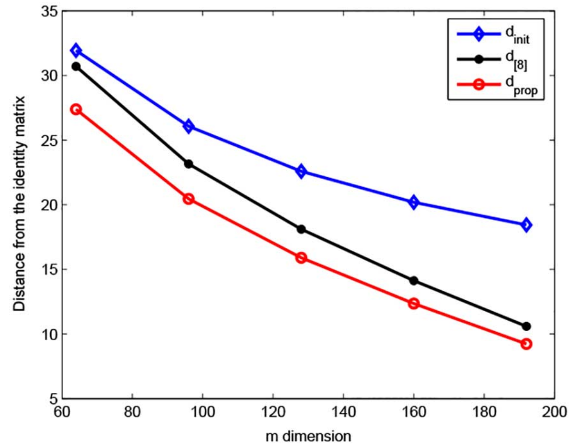

Fig. 1. Discrepancy between the Gram or pseudo-Gram matrices involved in support estimation and the identity matrix of same dimensions. The experiments involve $m \times N$ matrices with $m \in [64:32:192]$ and $N = 256$.

We cannot guarantee that the above algorithm yields an invertible matrix C. However, according to our analysis in [7], there is strong evidence that the algorithm converges locally, meaning that the output matrix C is close to the initial matrix C_{init}. Having selected an invertible initial matrix, the probability that the obtained matrix is singular is very small.

IV. Experimental Results

In this section we present simulations showing the effect of the proposed preconditioning approach in the numerical solution of sparse recovery problems that we encounter in sparse representations and CS. In our experiments we deploy OMP, a typical greedy approach, and BP, a numerical solver for convex relaxation programs [1].

A. Sparse Representations

Considering a sparse representation given by (4), when deploying greedy algorithms, the recovery of the unknown support depends on the inner products $D^*y = D^*Dx$. If D was an orthonormal basis, then $D^*D = I_N$, where I_N is the $N \times N$ identity matrix, and the product D^*y would recover the unknown support. Similarly, when employing overcomplete dictionaries, successful recovery is achieved if the Gram matrix has small off-diagonal entries. For dictionaries with high coherence, the solution proposed in [8] involves support estimation with a dictionary Ψ other than the original dictionary D. With Ψ being incoherent to D, the product $\Psi^*y = \Psi^*Dx$ yields higher recovery rates.

One way to estimate the appropriateness of the dictionaries involved in support recovery is to compute the discrepancy between the corresponding Gram or pseudo-Gram matrix and the identity matrix, that is, $d_{\text{init}} = \|D^*D - I_N\|_{\mathcal{F}}$ for the initial dictionary, $d_{\text{prop}} = \|F^*F - I_N\|_{\mathcal{F}}$ for the proposed preconditioning, and $d_{[8]} = \|\Psi^*D - I_N\|_{\mathcal{F}}$ for [8], where $\mathcal{F}$ denotes the Frobenius norm. Results using random Gaussian dictionaries averaged over 500 experiments are presented in Fig. 1, involving varying matrix dimensions. The results are best with the proposed construction, indicating improved performance in numerical recovery.

To test the performance of the deployed algorithms in sparse representations, we use random Gaussian dictionaries 128×256 and produce sparse synthetic signals with varying support size.

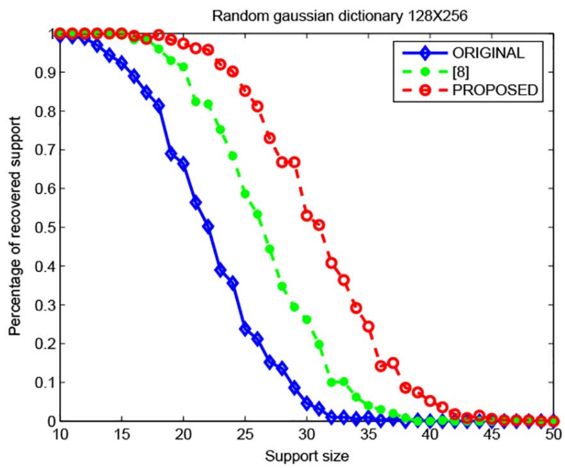

Fig. 2. Support recovery rates for sparse representations using OMP for signals with varying support size.

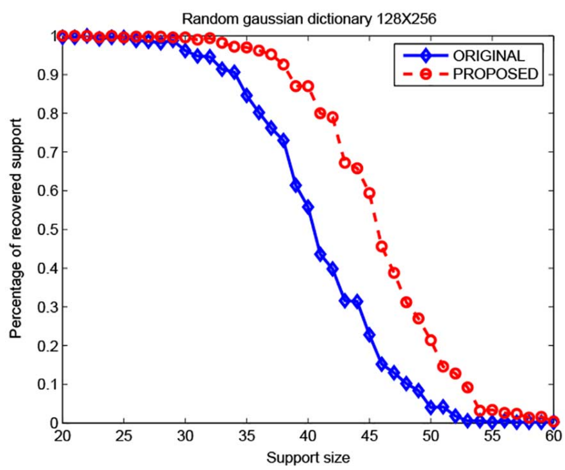

Fig. 3. Support recovery rates for sparse representations using BP for signals with varying support size.

The percentage of fully recovered support, referred to as recovery rate, is used to quantify the algorithms' performance. Averaged over 500 experiments, the recovery rates for OMP in Fig. 2 show that the proposed technique improves algorithm's performance and surpasses the results in [8]. Similar results obtained for BP ([8] is not applicable here) in Fig. 3 confirm that the proposed preconditioning is appropriate for finding sparse representations efficiently.

B. Compressed Sensing

In our previous work [7], starting from Gaussian random projections, we employ the algorithm for building incoherent UNTFs to construct an optimized projection matrix P such that the effective dictionary $\Phi = PD$ becomes an incoherent UNTF. In this paper, we consider a more practical problem, assuming that the sensing mechanism is implemented by a binary random matrix and improve signal recovery using preconditioning.

In our experiments, we use a 64×128 Bernoulli random projection matrix with entries 0,1. We perform signal acquisition according to (5). Preconditioning leads to the underdetermined linear system $Cy = CPDx$ or $z = Fx$, where $C \in \mathbb{R}^{64\times 64}$ is the preconditioner and $F = CPD$ is the new system matrix having the properties of an incoherent UNTF. The first group of experiments involves 128×255 Haar-DCT dictionaries. Recovery rates obtained for OMP and BP are presented in Fig. 4. Averaged over 500 realizations, the results show that preconditioning yields significant improvement in the performance of OMP, and particularly of BP, implying that the proposed technique can be applied successfully in CS.

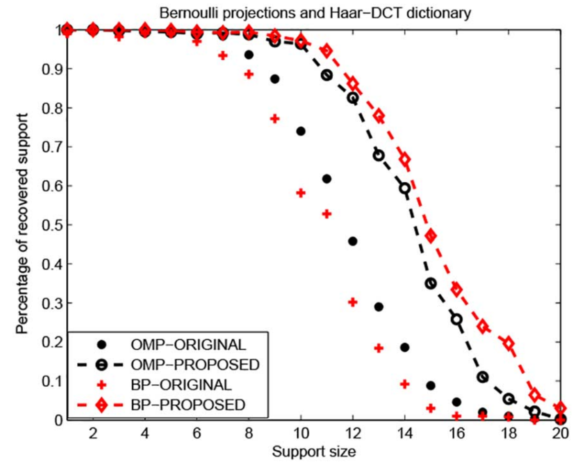

Fig. 4. Support recovery rates for OMP and BP, for signals with varying support size acquired with Bernoulli random projections.

TABLE I
RECOVERY RATE COMPARISON OF THE PROPOSED METHOD WITH OPTIMIZED PROJECTIONS [7]

Support Size	OMP		BP	
	Proposed	[7]	Proposed	[7]
4	1.000	1.000	1.000	1.000
8	0.996	1.000	0.998	1.000
12	0.860	0.870	0.924	0.928
16	0.250	0.248	0.398	0.380
20	0.004	0.006	0.042	0.054

Next, we compare the proposed method to optimized projections, for signals that are sparse under 128×256 random Gaussian dictionaries. Table I demonstrates recovery rates achieved with preconditioning and optimized projections [7]. The results are similar for both methods, showing that the performance of the deployed algorithms when used with Bernoulli projections and preconditioning is comparable to optimized projections. Considering that Bernoulli matrices are more convenient for hardware implementation, this is an important result for practical compressed signal acquisition.

V. CONCLUSIONS

In this paper, we propose a preconditioning technique for underdetermined linear systems that are encountered in sparse representations and CS. When the involved system matrix does not satisfy the necessary conditions for numerical solution, the proposed technique improves the performance of the deployed algorithms. Preconditioning is shown to increase the recovery rate of binary matrices used for sensing, matching that of optimized projections, and, therefore, is useful in practical CS systems. Future work involves the application of the proposed technique in other problems assuming the same formulation.

ACKNOWLEDGMENT

The authors would like to thank Dr. Karin Schnass for providing us the code implementing the method in [8].

REFERENCES

[1] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[2] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

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

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

[5] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harmno. Anal.*, vol. 25, pp. 1–24, 2008.

[6] E. Tsiligianni, L. P. Kondi, and A. K. Katsaggelos, "Use of tight frames for optimized compressed sensing," in *Eur. Signal Process. Conf. (EUSIPCO)*, Bucharest, Aug. 2012.

[7] E. Tsiligianni, L. P. Kondi, and A. K. Katsaggelos, "Construction of incoherent unit norm tight frames with application to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2319–2330, 2014.

[8] K. Schnass and P. Vandergheynst, "Dictionary preconditioning for greedy algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1994–2002, May 2008.

[9] K. Schnass and P. Vandergheynst, "Average performance analysis for thresholding," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 828–831, Nov 2007.

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

[11] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[12] W. Lua, W. Li, K. Kpalma, and J. Ronsin, *Near-optimal binary compressed sensing matrix*, 2013, arXiv:1304.4071.

[13] M. F. Duarte and Y. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[14] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA, USA: Birkhäuser, 2003.

[15] T. Strohmer and R. Heath, "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, 2003.

[16] O. Axelsson, *Iterative Solution Methods*. New York: Cambridge University Press, 1994.

[17] R. Horn and C. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1985.

[18] M. Rudelson and R. Vershynin, "The Littlewood-Offord Problem and invertibility of random matrices," Jan. 2008 [Online]. Available: http://arxiv.org/abs/math/0703503v2