Design of incoherent frames via convex optimization

Cristian Rusu

***Abstract*—This paper describes a new procedure for the design of incoherent frames used in the field of sparse representations. We present an efficient algorithm for the design of incoherent frames that works well even when applied to the construction of relatively large frames. The main advantage of the proposed method is that it uses a convex optimization formulation that operates directly on the frame, and not on its Gram matrix. Solving a sequence of convex optimization problems allows for the introduction of constraints on the frame that were previously considered impossible or very hard to include, such as non-negativity. Numerous experimental results validate the approach.**

***Index Terms*—Grassmannian frames, sparse representations.**

I. INTRODUCTION

Recent developments in the field of sparse representations [1] have received a lot of attention both in the theoretical and practical applications settings. At the core of the field lies the sparse approximation problem that can be loosely stated as: given a signal $\boldsymbol{y} \in \mathbb{R}^n$, provide a good approximation by using an underdetermined linear model such that $\boldsymbol{y} \approx \boldsymbol{D}\boldsymbol{x}$ where the matrix $\boldsymbol{D} \in \mathbb{R}^{n\times m}$, with $n \leq m$, is called dictionary, its columns $\boldsymbol{d}_j, j = 1, \ldots, m$, are called atoms and the sparse vector $\boldsymbol{x} \in \mathbb{R}^m$ is called the representation vector. The precise problem definition is given by

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \|\boldsymbol{x}\|_0 \\ \text{subject to} \quad & \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}, \end{aligned} \tag{1}$$

where $\|\boldsymbol{x}\|_0$ is the ℓ_0 pseudo-norm, which is defined as the number of non-zero elements in $\boldsymbol{x}$. Since in this context we operate with underdetermined systems, we call the dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times kn}$ to be k times overcomplete.

A theoretical framework has been developed where the performance of these algorithms can be studied using concepts such as the mutual coherence [2] and the restricted isometry property (RIP) [3]. The mutual coherence of a matrix is defined as the maximum absolute value of the normalized dot products between different columns and it is thus very easy to compute, while the RIP characterizes how close the current matrix is to an orthogonal one when operating on sparse vectors. In terms of the mutual coherence, the main research direction is to find structures as incoherent as possible (e.g. Equiangular Tight Frames [4] or Grassmannian Frames [5]) and for the RIP, classes of matrices that meet desirable bounds have been found (we mention here Gaussian matrices) and used with great success in Compressive Sensing (CS) [6]. Furthermore, even though highly incoherent frames have not been extensively used in signal processing, they are of great importance in coding theory and communications [7] and for the development of spherical codes [8].

In terms of design algorithms, probably the most influential work consists of a design procedure based on the alternating projection method [9] with an implementation presented in [10]. Recent work in this direction, that improves upon these results, includes [11] where the frame factors into a constant projection matrix and a variable representation matrix that is updated, $\boldsymbol{F} = \boldsymbol{P}\boldsymbol{D}$. The procedures can be summarized by the following iterative procedure: decrease the values of a fixed number of the largest entries in the Gram matrix $\boldsymbol{G} = \boldsymbol{F}^T\boldsymbol{F}$ by a given amount, use the singular value decomposition (SVD) to reduce the rank of $\boldsymbol{G}$ back to n and then factor out the new dictionary. This class of methods works well but there are some internal parameters that need to be tuned: the number of entries of the Gram matrix to be updated and how big this update is. Previous to these methods, in the field of communications, algorithms were developed to construct frames, by minimizing the total squared correlation, that are the optimal sequences for DS-CDMA systems [12]. Recently, in the context of dictionary learning, a formulation has been proposed [13] that contains an additional penalization term (in the Frobenius norm) for the mutual coherence. This way, the method controls a trade-off between the representation error and the coherence value.

This paper outlines a completely new approach to the problem by considering directly the frame itself as the optimization variable, instead of the Gram matrix. The iterative optimization algorithm is based on a sequence of convex optimization problems and, as we will see, it provides an effective and versatile design method.

The paper is structured as follows: Section II presents basic properties of frames and the incoherent frame design problem, Section III proposes a new solution called Iterative Decorrelation by Convex Optimization (IDCO), Section IV presents various experimental results including the design of very large frames.

II. FRAMES

A sequence sequence of m vectors $\boldsymbol{f}_j, j = 1, \ldots, m$, is called a frame for $\mathbb{R}^n$ if:

$$\alpha\|\boldsymbol{v}\|_2^2 \leq \sum_{j=1}^{m} |\boldsymbol{v}^T\boldsymbol{f}_j|^2 \leq \beta\|\boldsymbol{v}\|_2^2,\ \forall \boldsymbol{v} \in \mathbb{R}^n, \tag{2}$$

This work was supported by the Romanian National Authority for Scientific Research, CNCS - UEFISCDI, project number PN-II-ID-PCE-2011-3-0400.

The author is with the Department of Automatic Control and Computers, University Politehnica of Bucharest, Spl. Independenţei 313, Bucharest 060042, Romania (e-mail: cristian.rusu@schur.pub.ro).

and we introduce its associated matrix $\boldsymbol{F} \in \mathbb{R}^{n \times m}$ consisting of the concatenated frame vectors

$$\boldsymbol{F} = [\boldsymbol{f}_1 \ \boldsymbol{f}_2 \ \dots \ \boldsymbol{f}_m]. \tag{3}$$

Constants α, $\beta \in \mathbb{R}$, with $0 < \alpha \leq \beta < \infty$, are called the lower and upper bounds of the frame. When $\alpha = \beta$ the frame is called α-tight and when $\alpha = \beta = 1$ the frame is called a Parseval frame (orthonormal if vectors have unit norm) and it exists only if $m = n$. In this paper, we deal with the construction of unit ℓ_2 norm frames.

We recall that the Gram matrix is symmetric and positive semidefinite with rank n, it has a unit diagonal and the off-diagonal entries are equal to the dot products between any two distinct frame vectors. We define the mutual coherence

$$\mu(\boldsymbol{F}) = \max_{1 \leq i < j \leq m} |g_{ij}|, \tag{4}$$

which is useful in the context of sparse representations since it provides a bound on the performance of sparse approximation algorithms [14]: considering the system $\boldsymbol{F}\boldsymbol{x} = \boldsymbol{b}$, the solution $\boldsymbol{x}$ is the sparsest if

$$\|\boldsymbol{x}\|_0 \leq \frac{1}{2}\left(\frac{1}{\mu(\boldsymbol{F})} + 1\right) = \bar{s}. \tag{5}$$

The mutual coherence μ takes values between 0 and 1, with the lower bound reached for orthogonal frames. Frames with low mutual coherence values are called incoherent.

Since the properties of these frames revolve around the Gram matrix, most design procedures consider the $\boldsymbol{G}$ matrix the central object, while their constructions are mostly based on the following theorem presented in [15] and the numerical procedure of polar decomposition to recover the frame.

Theorem 1. [15] Given a frame $\boldsymbol{F} \in \mathbb{R}^{n \times m}$ and its singular value decomposition $\boldsymbol{U\Sigma V}^T$, with $\boldsymbol{\Sigma}$ square. Then the matrix, called the polar factor, $\alpha \boldsymbol{U}\boldsymbol{V}^T$ is the closest α^2-tight frame in Frobenius norm to $\boldsymbol{F}$ and coincides with $\alpha(\boldsymbol{F}\boldsymbol{F}^{\boldsymbol{T}})^{-1/2}\boldsymbol{F}$.

We next define the Equiangular Tight Frames (ETF) [4] as the sequence of m vectors $\boldsymbol{f}_j$ for which all the dot products between them are the same and equal, in absolute value, to

$$\mu = \sqrt{\frac{m-n}{n(m-1)}}. \tag{6}$$

It is trivial to observe that orthonormal frames are ETFs and actually reach the lowest possible coherence, zero. Sadly, ETFs are not frequent, they only exist for a few pairs (n, m) (for example, these do not exist if $m > n(n+1)/2$) [4]. When they exist, ETFs provide the lowest possible mutual coherence and they are considered the generalization of orthonormal basis because they preserve (a generalization of) Parseval's identity.

A generalization of ETFs is embodied by Grassmannian Frames (GFs) [5]. In this setting, the largest absolute value dot product is minimal for the given dimensions n and m. Notice that this relaxed condition means that GFs exist for every pair (n, m), but still this does not help in decreasing the computational complexity of finding incoherent frames. Considerable effort was allocated to the creation of explicit frames using tools developed by abstract algebra and we mention here: the constructions by conference matrices [16] and Kerdock codes [17].

Actually, relation (6) describes a general lower bound on μ that is also known as the Welch bound (WB). For a detailed discussion on frames the reader is advised to consult [18].

From this brief outline it should be clear that creating maximally incoherent frames (either ETF or Grassmanian) is an extremely difficult task in the general case and there are no guarantees that the lower bound of coherence can be reached for a given pair (n, m).

III. Incoherent frame design

This subsection is dedicated to the description of a novel numerical procedure for the design of incoherent frames. Unlike the previous approaches, we develop an optimization procedure that directly designs the frame $\boldsymbol{F}$ with a mutual coherence as low as possible.

The proposed method, called Iterative Decorrelation by Convex Optimization (IDCO), tries to gradually decrease the coherence of an initial frame by finding a new frame that is close to the previous one, but that is able to achieve higher incoherence. The algorithmic description of IDCO is next and it is followed by a detailed discussion on how it operates.

Iterative Decorrelation by Convex Optimization (IDCO). Given dimensions n and m, construct frame $\boldsymbol{F} \in \mathbb{R}^{n \times m}$ as incoherent as possible until the maximum number of iterations K is reached.

1) Initialization:
   a) Create the frame $\boldsymbol{F}_{-1}$ with random entries extracted from a Gaussian distribution and normalize its columns.
   b) With $\boldsymbol{F}_{-1} = \boldsymbol{U\Sigma V}^T$ update the frame by the unit polar decomposition (by Theorem 1): $\boldsymbol{F}_0 = \boldsymbol{U}\boldsymbol{V}^T$.
   c) To finalize, normalize the columns of $\boldsymbol{F}_0$.
2) Iterations $k = 1, \dots, K$
   a) Solve the following optimization problem where $\boldsymbol{F}_{k-1}$ is fixed and create the working frame $\boldsymbol{H} \in \mathbb{R}^{n \times m}$

$$\begin{aligned} \underset{\boldsymbol{H}}{\text{minimize}} \quad & \max_{i \neq j} |\boldsymbol{h}_i^T \boldsymbol{f}_j| \\ \text{subject to} \quad & \|\boldsymbol{h}_j - \boldsymbol{f}_j\|_2 \leq T, \ 1 \leq j \leq m, \end{aligned} \tag{7}$$

   where $T \in \mathbb{R}, \ 0 < T \ll 1$, is a measure of how far we are willing to leave the reference model in search for the new frame, $\boldsymbol{h}_j$ and $\boldsymbol{f}_j$ are the columns of $\boldsymbol{H}$ and $\boldsymbol{F}_{k-1}$ respectively.
   b) Normalize the columns of $\boldsymbol{H}$ to get the new frame $\boldsymbol{F}_k$.
   c) Compute the new mutual coherence $\mu(\boldsymbol{F}_k)$. If the result is worse than in previous step, update $T = T/2$.

The IDCO approach is somewhat similar to the sequential convex programming framework that approximates solutions to non-convex optimization problems by solving a series of locally convex optimization problems inside a defined trust region. Next follows a discussion on each step of IDCO.

The initialization. A random frame is created since this type of initialization generically produces a relatively low coherence frame. In fact, it has been shown that random frames drawn uniformly from the unit sphere of $\mathbb{R}^n$ have very good μ values and actually, as m tends to infinity, they converge to

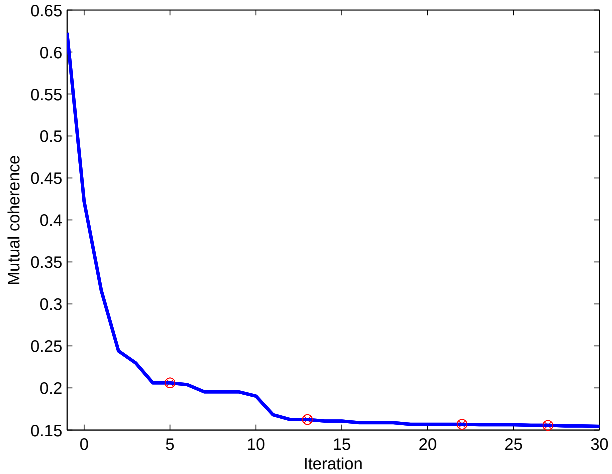

Fig. 1. Incoherence reduction evolution produced by 30 iterations of IDCO for the frame $\boldsymbol{F} \in \mathbb{R}^{64\times 256}$. Plot also contains the initialization results and marks for the iterations when the value of T changed.

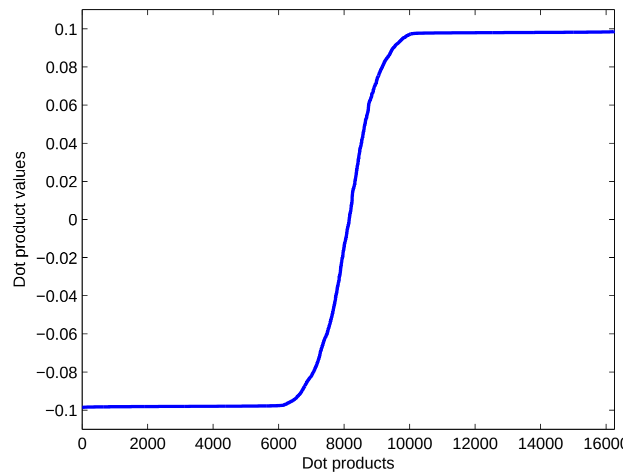

Fig. 2. Off-diagonal entries of $\boldsymbol{G} = \boldsymbol{F}^T\boldsymbol{F}$ for frame $\boldsymbol{F} \in \mathbb{R}^{64\times 128}$ designed by IDCO after 100 iterations.

tight frames [19]. For further decrease μ we apply the polar decomposition (see Theorem 1).

The iterations. The first step solves the convex optimization problem that lies at the heart of IDCO. We consider the previous given frame $\boldsymbol{F}_{k-1}$ as a reference for a new frame. We are searching for this new frame $\boldsymbol{H}$, such that each new frame vector is T-close in the ℓ_2 norm to the reference, that achieves a lower mutual coherence value. Observe that minimizing directly the coherence of $\boldsymbol{H}$ would lead to a non-convex optimization problem. In order to make the optimization problem tractable, the objective in (7) is an approximation of the mutual coherence, ie. we consider $\boldsymbol{H}^T\boldsymbol{H} \approx \boldsymbol{H}^T\boldsymbol{F}_{k-1}$. As the algorithm progresses and the value of T lowers this approximation becomes more precise.

Still, if no further progress is made in $\mu(\boldsymbol{F}_k)$ then we diminish the maximum allowed change from the reference model to the new frame, thus providing a better approximation for the mutual coherence. Experimentally we have observed that this approach further decreases μ but at the expense of slower progress. Typically we start with $T = 0.1, \ldots, 0.3$. In Figure 1 we see the evolution of the mutual coherence at each iteration with marks when the value of T updates.

Also, notice that since (7) is actually an ℓ_∞ optimization problem we expect many of the dot products to be equal in magnitude. To illustrate this, Figure 2 shows a typical result of IDCO, all the dot products for the design with $n = 64$ and $m = 128$. Notice that most values are close in absolute value to the mutual coherence. Ideally, for an ETF, this graph should be as close as possible to the graph of the signum function.

IV. Results

Now that IDCO was fully described, we apply it to design incoherent frames of various sizes.

In a concrete design setting, IDCO is used to design frames $\boldsymbol{F} \in \mathbb{R}^{n\times 120}$ where we vary the dimension $n \in \{15, \ldots, 40\}$. To show the progress of IDCO we also show $\mu(\boldsymbol{F}_0)$ and

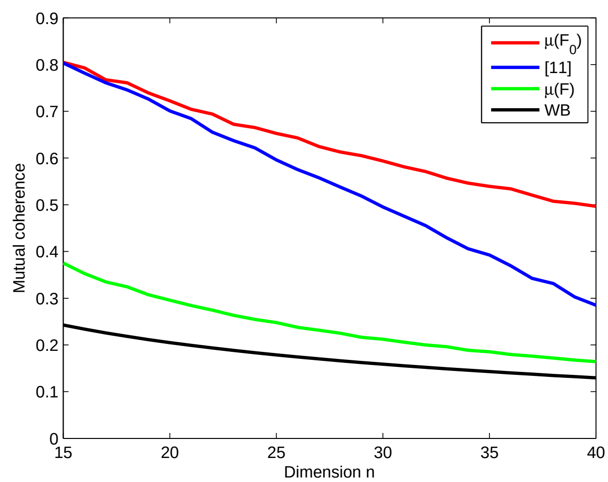

Fig. 3. Coherence of frames $\boldsymbol{F} \in \mathbb{R}^{n\times 120}$ designed using IDCO for $n \in \{15, \ldots, 40\}$. The plot also includes results from [11].

the Welch lower bound (6). The values of n and m were chosen like in [11] to allow for an easy comparison with the best results presented there (results better than previous works). The results are depicted in Figure 3 and it is clear that the results produced by IDCO are the best. Notice that the distance to the Welch bound decreases with the increase in the dimension n.

In the second design setting, IDCO is used to design frames $\boldsymbol{F} \in \mathbb{R}^{40\times m}$ where we vary the dimension $m \in \{80, \ldots, 240\}$ with step 10. To show the progress of IDCO we also show $\mu(\boldsymbol{F}_{-1})$, $\mu(\boldsymbol{F}_0)$, to highlight of effect of the polar decomposition, and the Welch lower bound (6), for perspective. The experimental tests are meant to show how IDCO behaves in a fixed dimension with an increasing number of vector frames, culminating with an extreme case where a six times overcomplete frame is designed. The results are depicted in Figure 4. Notice that the distance to the Welch bound increases with the increase in the dimension m.

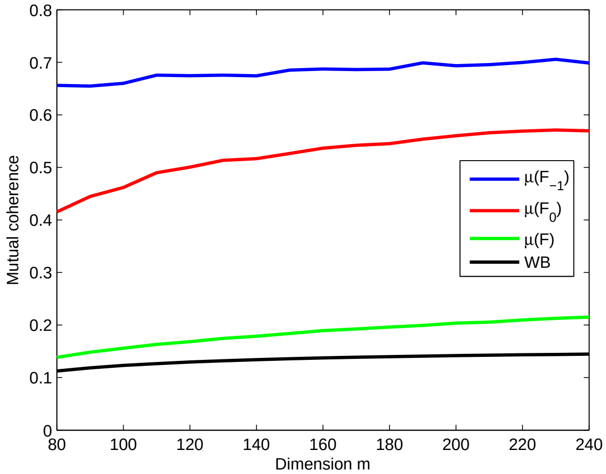

Fig. 4. Coherence of frames $\boldsymbol{F} \in \mathbb{R}^{40\times m}$ designed using IDCO for $m \in \{80, \ldots, 240\}$ with step 10.

TABLE I
INCOHERENCE RESULTS REACHED BY IDCO FOR FINAL FRAMES $\boldsymbol{F} \in \mathbb{R}^{64\times m}$ WITH INCREASING VALUES OF m. THE INITIAL FRAMES ARE ALSO SHOWN TOGETHER WITH THE WELCH BOUND FOR PERSPECTIVE.

m	$\mu(\boldsymbol{F}_{-1})$	$\mu(\boldsymbol{F}_0)$	$\mu(\boldsymbol{F})$	WB	$\lfloor \bar{s} \rfloor$
70	0.5508	0.1724	0.0425	0.0369	12
80	0.5833	0.2764	0.0662	0.0563	8
90	0.5617	0.2495	0.0786	0.0676	6
100	0.5719	0.2291	0.0812	0.0754	6
128	0.5738	0.3606	0.0987	0.0887	5
192	0.6140	0.4596	0.1206	0.1023	4
256	0.6184	0.4176	0.1371	0.1085	4
320	0.6263	0.4755	0.1451	0.1120	3
384	0.6466	0.5395	0.1536	0.1143	3
448	0.6026	0.4993	0.1599	0.1159	3
512	0.6305	0.5040	0.1634	0.1170	3
576	0.6276	0.5009	0.1720	0.1180	3
640	0.6265	0.5469	0.1764	0.1187	3
768	0.6581	0.5250	0.1884	0.1198	3
960	0.6877	0.6023	0.2136	0.1208	2
1280	0.7131	0.6386	0.2257	0.1219	2

We describe in Table I a set of results obtained for frames $\boldsymbol{F} \in \mathbb{R}^{64\times m}$. Here, we are also interested in the actual sparse recovery guarantee from (5). Also, to show the versatility of IDCO and a possible application to non-negative matrix factorization (NMF) problems, Table II describes results obtained for the design of frames $\boldsymbol{F} \in \mathbb{R}^{64\times m}$ with $f_{ij} >= 0$. In this situation, the unit polar decomposition is avoided in order to satisfy the positivity constraints at all times.

In each context, IDCO starts with $T = 0.1$ and stops after $K = 150$ iterations. The running times vary from a few minutes ($m = 70$) to an over-week run ($m = 1280$) on a modern system using the generic CVX solver (http://cvxr.com). Of course, results are dependent on initialization.

V. CONCLUSIONS

This article describes an efficient new method for the design of incoherent frames. The procedure constructs highly incoherent frames in an iterative manner by solving a sequence of convex optimization problems. Unlike previous methods in the literature, the central object of these optimization problems is the actual frame, not its Gram matrix. Numerous experimental tests show that IDCO produces the best results obtained so far.

TABLE II
INCOHERENCE RESULTS REACHED BY IDCO FOR FINAL NON-NEGATIVE FRAMES $\boldsymbol{F} \in \mathbb{R}^{64\times m}$ WITH INCREASING VALUES OF m.

m	$\mu(\boldsymbol{F}_0)$	$\mu(\boldsymbol{F})$
100	0.5970	0.2403
128	0.5835	0.2635
192	0.6085	0.2873
256	0.6124	0.3093
320	0.6231	0.3292

ACKNOWLEDGMENTS

The author would like to thank B. Dumitrescu and S. A. Tsaftaris for comments that improved the quality of this paper.

REFERENCES

[1] A. M. Bruckstein, D. L. Donoho and M. Elad, From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images, SIAM Review, vol. 51, pp. 34–81, 2009.
[2] J. A. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inf. Theory, vol. 50, pp. 2231-2242, 2004.
[3] E. Candes and T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory, vol. 51, pp. 4203–4215, 2005.
[4] M. Sustik, J. Tropp, I. Dhillon and R. Heath, On the existence of equiangular tight frames, Linear Algebra and Its Applications, vol. 426, no. 2–3, pp. 619–635, 2007.
[5] J. H. Conway, R. H. Hardin and N. J. A. Sloane, Packing lines, planes, etc.: Packings in Grassmannian spaces, Exper. Math., vol. 5, no. 2, pp. 139–159, 1996.
[6] E. Candes, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory, vol. 52, pp. 489–509, 2006.
[7] T. Strohmer and R. Heath, Grassmannian frames with applications to coding and communication, Applied and Computational Harmonic Analysis, vol. 14, no. 3, pp. 257–275, 2003.
[8] V. Tarokh and I.-M. Kim, Existence and construction of noncoherent unitary space-time codes, IEEE Trans. Inf. Theory, vol. 48, pp. 3112–3117, 2002.
[9] J. A. Tropp, I. S. Dhillon, R. W. Heath and T. Strohmer, Designing structured tight frames via an alternating projection method, IEEE Trans. Inf. Theory, vol. 51, no.1, pp. 188–209, 2005.
[10] M. Elad, Optimized projections for compressed sensing, IEEE Trans. Sig. Process., vol. 55, no. 12, pp. 5695–5702, 2007.
[11] E. Tsiligianni, L. Kondi and A. Katsaggelos, Use of Tight Frames for Optimized Compressed Sensing, EUSIPCO, pp. 1439–1443, 2012.
[12] S. Ulukus and R. D. Yates, Iterative construction of optimum signature sequence sets in synchronous CDMA systems, IEEE Trans. Inf. Theory, vol. 47, no. 5, pp. 1989–1998, 2001.
[13] C. D. Sigg, T. Dikk and J. M. Buhmann, Learning Dictionaries with Bounded Self-Coherence, IEEE Sig. Process. Letters., vol. 19, no. 12, pp. 861–864, 2012.
[14] D. L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, Proc. of the National Academy of Sciences, vol. 100, no. 5, pp. 2197–2202, 2003.
[15] R. A. Horn and C. R. Johnson, Matrix Analysis. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[16] T. Strohmer, A note on equiangular tight frames, Linear Alg. Appl., vol. 429, pp. 326–330, 2008.
[17] A. R. Calderbank, P. J. Cameron, W. M. Kantor and J. J. Seidel, $\mathbb{Z}4$-Kerdock codes, orthogonal spreads and extremal Euclidean line sets, Proc. London Math. Soc., vol. 75, no. 2, pp. 436–480, 1997.
[18] O. Christensen, Frames and Bases, Birkhauser Boston, 2008.
[19] V. K. Goyal, M. Vetterli and N. T. Thao, Quantized overcomplete expansions in $\mathbb{R}^n$: Analysis, synthesis and algorithms, IEEE Trans. Inf. Theory, vol. 44, no. 1, pp. 16–31, 1998.