



On block coherence of frames



Robert Calderbank^a, Andrew Thompson^{a,*}, Yao Xie^b

^a Department of Mathematics, Duke University, Durham, NC, United States

^b H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, GA, United States

ARTICLE INFO

Article history:

Received 12 August 2013

Received in revised form 18 January 2014

Accepted 12 March 2014

Available online 21 March 2014

Communicated by Joel A. Tropp

Keywords:

Block coherence

Compressed sensing

Grassmann packing

Frame constructions

Random subspaces

ABSTRACT

Block coherence of matrices plays an important role in analyzing the performance of block compressed sensing recovery algorithms (Bajwa and Mixon, 2012). In this paper, we characterize two block coherence metrics: worst-case and average block coherence. First, we present lower bounds on worst-case block coherence, in both the general case and also when the matrix is constrained to be a union of orthobases. We then present deterministic matrix constructions based upon Kronecker products which obtain these lower bounds. We also characterize the worst-case block coherence of random subspaces. Finally, we present a flipping algorithm that can improve the average block coherence of a matrix, while maintaining the worst-case block coherence of the original matrix. We provide numerical examples which demonstrate that our proposed deterministic matrix construction performs well in block compressed sensing.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

In compressed sensing [1,2], a new paradigm in signal processing, we are interested in recovering a sparse signal x from a reduced number of measurements $y = Ax$. Rather than sparsity, where x only has a few nonzero entries, a more general model is to consider *block sparsity*, where the signal has a block structure in that the entries in each block are nonzero simultaneously, and there are only a few blocks with nonzero entries. Recovery of block sparse signal arises from a wide range of applications (as described in Section 1.1).

Various algorithms have been proposed to recover block sparse signals, including *group LASSO* [3] and the lower complexity *one-step group thresholding* [4], see Section 1.1 for a brief survey. It has been shown that the performance of one-step group thresholding depends crucially on two block coherence properties of the sensing matrix A : *worst-case block coherence* and *average block coherence*, which we next define. Suppose the matrix $A \in \mathbb{C}^{n \times mr}$ consists of the concatenation of m equally sized blocks, namely

* Corresponding author.

E-mail addresses: robert.calderbank@duke.edu (R. Calderbank), thompson@math.duke.edu (A. Thompson), yao.c.xie@gmail.com (Y. Xie).

$$A = [A_1 \ A_2 \ \dots \ A_m],$$

$$A_i \in \mathbb{C}^{n \times r}, \quad i \in \{1, 2, \dots, m\},$$

where we assume that the columns of A are unit-norm, that the columns within each block are orthonormal, and also that $mr > n$ and $r < n$. Thus, provided the columns of A span \mathbb{C}^n , A is a unit-norm frame in \mathbb{C}^n . At times, we will also consider the special case in which A is real. Following [3,5], we define $\mu(A)$, the worst-case block coherence of A , to be

$$\mu(A) := \max_{i \neq j} \|A_i^* A_j\|_2, \quad (1)$$

and following [4], we define $\nu(A)$, the average block coherence of A , to be

$$\nu(A) := \frac{1}{m-1} \max_i \left\| \sum_{j \neq i} A_i^* A_j \right\|_2, \quad (2)$$

where X^* denotes the conjugate transpose of a matrix X , and $\|X\|_2$ denotes the spectral norm of X , which is also equal to the maximum absolute singular value of X . Note that $\nu(A)$ is referred to as average *group* coherence in [4].

Intuitively, a matrix A with good coherence properties should have subspaces formed by submatrices of A as orthogonal to each other as possible. Hence, the problem of designing a matrix with good coherence properties is related to the problem of packing subspaces, also known as *Grassmann packing* [6,7]. Let $G(n, r)$ be the Grassmann manifold of r -dimensional subspaces in \mathbb{C}^n . Now suppose $\{A_1, A_2, \dots, A_m\}$ are orthonormal bases for the subspaces $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_m\}$, so that $\mathcal{S}_i \in G(n, r)$ for each $i = 1, 2, \dots, m$. A Grassmann packing of m subspaces in $G(n, r)$ with respect to some distance metric is said to be optimal if it maximizes the minimum distance between the subspaces. Various distance metrics have been considered, including the chordal distance, spectral distance and geodesic distance [6,7].

In a seminal paper on Grassmann packings, Conway et al. [6] proved an upper bound on the chordal distance given (m, n, r) , from which an upper bound on the spectral distance was deduced in [7]. Meanwhile, a fundamental lower bound on worst-case block coherence was essentially (though not explicitly) determined in [8]. It can be shown that optimal Grassmann packings with respect to the spectral distance yield matrices with minimum worst-case block coherence. In fact, it is straightforward to deduce the lower bound on worst-case block coherence in [8] from the upper bound on spectral distance in [7] (see Section 2.1 for further elaboration).

Concerning achievability of the above-mentioned bounds, most explicit constructions currently existing in the literature are for the case of $r = 1$ (worst-case and average column coherence respectively [9]). In this case, the lower bound on worst-case coherence, known as the Welch bound [10], is obtained if and only if the matrix A is an equiangular tight frame (ETF) [11], see Section 2.1.2. Several infinite families of ETFs have been found [12–15], as well as other infinite families which nearly meet the Welch bound [16]. For the case $r \geq 2$ which is our interest in the present paper, various optimal Grassmann packings (thereby yielding matrices with optimal worst-case block coherence) were constructed for small dimensions in [6]. Infinite families of nearly-optimal Grassmann packings have also been found [17,18], both of which have an underlying group-theoretic structure. Numerical methods for empirically constructing Grassmann packings have also been proposed, based upon nonlinear optimization [19] and alternating projection [7].

It was shown in [8] that optimal Grassmann packings can also be constructed as the Kronecker product of an ETF with a unitary matrix, from which it immediately follows that every infinite family of ETFs gives rise to infinite families of Grassmann packings. In this paper, we will extend this technique of employing Kronecker products.

The contributions of this paper are as follows

1. We establish a connection between worst-case coherence and optimal Grassmann packings with respect to both the chordal distance and spectral distance. Using this connection, we review and unite existing results concerning optimal Grassmann packings [6,7] and a fundamental lower bound on worst-case block coherence [8]. We then review an optimal construction, originally given in [8]: these matrices are constructed as the Kronecker product of an equiangular tight frame with a unitary matrix. The result implies the existence of several families of matrices with optimal worst-case block coherence, and furthermore the existence of several infinite families of optimal Grassmann packings with respect to both the chordal and spectral distances. We also translate bounds on the maximum number of subspaces in optimal Grassmann packings into the world of frame coherence, which gives upper bounds on the number of blocks in the matrix A .
2. We also prove a tighter lower bound on the worst-case block coherence of matrices which are constrained to be unions of orthonormal bases. By extending the Kronecker product approach, we explicitly construct matrices which nearly meet the fundamental lower bound which allows for an increased number of blocks. This result implies the existence of several infinite families of matrices with nearly optimal worst-case block coherence, which in turn gives infinite families of near optimal Grassmann packings. We also obtain upper bounds on the number of blocks in the matrix A for optimal packings in this relaxed sense.
3. We analyze the worst-case block coherence of matrices formed from random orthonormal bases. In a proportional–dimensional asymptotic framework, we prove an upper bound on worst-case block coherence which is a small multiple above optimal. Our result can also be used to quantify the asymptotic spectral distance of Grassmann packings formed from the union of random subspaces.
4. We present a flipping algorithm for subspaces, inspired by a similar flipping algorithm for columns [20], which can reduce the average block coherence of a frame while maintain the worst-case block coherence. We also prove that, starting with a frame with low worst-case block coherence, the flipping algorithm can produce a frame with average block coherence that meets the requirement of the one-step thresholding algorithm [4].

An outline of the paper is as follows: We first extend our introduction in Section 1.1 by exploring some of the applications of block sparsity in compressed sensing, and of optimal Grassmann packings. We focus on worst-case block coherence in Section 2: we review the existing results on fundamental bounds and optimal constructions in Section 2.1, before presenting our new bounds and constructions for unions of orthonormal bases in Section 2.2, and analyzing random designs in Section 2.3. We then turn to average block coherence in Section 3: we prove results for our deterministic constructions in Section 3.1, and we present our flipping algorithm for reducing average block coherence in Section 3.2. Numerical illustrations are provided in Section 4.

1.1. Applications

There exist several applications of block sparsity in compressed sensing. A far from exhaustive list includes the blind sensing of multi-band signals [3], DNA microarrays [21], and also in medical imaging including ECG imaging [22] and source localization in EEG/MEG brain imaging [23]. The same problem has also received much attention in statistical regression, where one predictor may often imply the presence of several others [4]. Another special case is the multiple measurement vector (MMV) model, also referred to as joint or simultaneous sparsity, where the matrix is a suitable rearrangement of the rows of a matrix of the form $I \otimes A$, and which has found applications in sensor networks and MIMO channels [24].

We recall that block coherence is closely related to optimal Grassmann packing, which has found applications in the design of codes for multiple-antenna communication systems, such as MIMO. In this application, complex Grassmann packings consisting of a small number of low dimensional subspaces in a much larger ambient dimension are required, see [7] for references therein.

1.2. Algorithms for block compressed sensing: further background

Block coherence is crucial for the analysis of the one-step group thresholding algorithm for block sparse compressed sensing. To give additional context, several algorithms have been proposed for this problem, including the group LASSO convex relaxation (also known as ℓ_2/ℓ_1 mixed optimization) [3], the greedy block OMP algorithm [3] and iterative thresholding [25]. It is celebrated that compressed sensing allows the recovery of a signal whose sparsity is proportional to its dimension, and several optimal-order guarantees of this form have been obtained in the case of random matrices for algorithms for block sparsity by means of the block-based Restricted Isometry Property and the analysis of message passing algorithms [3,25]. The only current optimal-order recovery guarantee for deterministic matrices is for one-step group thresholding [4], and relies crucially upon both the worst-case and average block coherence. Another important consideration is that one-step group thresholding is an extremely simple algorithm and has the lowest computational complexity among all the algorithms mentioned.

2. Worst-case block coherence

2.1. Fundamental bounds and constructions for worst-case block coherence

In this section, we present a review of existing results concerning optimal Grassmann packings [6,7] and optimal worst-case block coherence [8], emphasizing the connection between the two. We also present optimal constructions which generalize the one given in [8].

2.1.1. Bounds

We begin by making the observation that $\mu(A) = 1$ whenever $n < 2r$, since in this case the two subspaces corresponding to any pair of blocks have a non-trivial intersection. We will therefore assume throughout Section 2 that $n \geq 2r$.

Two choices of distance metric for Grassmann packings which have been previously studied [6,7] are the chordal distance $d_C(\mathcal{S}_i, \mathcal{S}_j)$ and the spectral distance $d_S(\mathcal{S}_i, \mathcal{S}_j)$, defined as follows [7]

$$d_C(\mathcal{S}_i, \mathcal{S}_j) := \sqrt{r - \|A_i^* A_j\|_F^2}, \quad (3)$$

$$d_S(\mathcal{S}_i, \mathcal{S}_j) := \sqrt{1 - \|A_i^* A_j\|_2^2}, \quad (4)$$

where $\|X\|_F$ denotes the Frobenius norm of X , defined as $\|X\|_F = \sum_i \sum_j |X_{ij}|^2$.

Both metrics may also be expressed in terms of the principal angles between \mathcal{S}_i and \mathcal{S}_j [7]: the spectral norm is the sine of the smallest principal angle and the square of the chordal distance is the sum of the squared sines of all the principal angles.

The following upper bound on the minimum chordal distance was obtained in a seminal paper by Conway et al. [6]. It was named the Rankin bound since it was obtained by projecting Grassmannian space onto a sphere and applying the Rankin bound for spherical codes.

Lemma 2.1 (Rankin bound for chordal distance). (See [6, Corollary 5.2].)

$$\min_{i \neq j} [d_C(\mathcal{S}_i, \mathcal{S}_j)]^2 \leq \frac{r(n-r)}{n} \cdot \frac{m}{m-1}.$$

If the bound is met, all subspaces are equidistant.

The subspaces $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_m\}$ are said to be *equi-isoclinic* if the singular values among all $A_i^* A_j$ for $i \neq j$ are equal [8], which is also equivalent to saying that the principal angles are all equal for each pair of subspaces. Dhillon et al. [7] deduced from Lemma 2.1 an upper bound on the minimum spectral distance.

Lemma 2.2 (Spectral distance bound). (See [7, Corollary 4.2].)

$$\min_{i \neq j} [d_S(\mathcal{S}_i, \mathcal{S}_j)]^2 \leq \frac{n-r}{n} \cdot \frac{m}{m-1}. \quad (5)$$

If the bound is met, the subspaces are equi-isoclinic.

A lower bound on worst-case block coherence may be derived from Lemma 2.2.

Theorem 2.3 (Universal lower bound on worst-case coherence). (See [8, Theorem 3.6].)

$$\mu(A) \geq \sqrt{\frac{mr-n}{n(m-1)}}. \quad (6)$$

If the bound is met, all singular values among all matrices $A_i^* A_j$ for $i \neq j$ are equal in modulus.

Proof. Combining (4) and (5), we have

$$1 - \max_{i \neq j} \|A_i^* A_j\|_2^2 \leq \frac{n-r}{n} \cdot \frac{m}{m-1}, \quad (7)$$

which rearranges to give

$$\max_{i \neq j} \|A_i^* A_j\|_2^2 \geq 1 - \frac{n-r}{n} \cdot \frac{m}{m-1} = \frac{n(m-1) - m(n-r)}{n(m-1)} = \frac{mr-n}{n(m-1)}, \quad (8)$$

and (6) follows. Finally, if (5) holds with equality, it follows from Lemma 2.2 that the subspaces are equi-isoclinic, which implies that all singular values among all matrices $A_i^* A_j$ for $i \neq j$ are equal in modulus. \square

Theorem 2.3 was essentially proved in [8], which gives an upper bound on the number of subspaces m for a given principal angle. Theorem 2.3 extends the Welch bound for coherence of columns of a matrix [10], and reduces to the Welch bound for $r = 1$. Similarly, the condition that the subspaces are equi-isoclinic is an extension of the condition for equality in the Welch bound, namely that the columns of the matrix are required to be an equiangular tight frame [11].

Frames with optimal block coherence cannot have arbitrarily many blocks. We next give an upper bound on m , the number of blocks in the matrix, or equivalently the number of subspaces in an optimal Grassmann packing. Since we require the subspaces to be equi-isoclinic, we may use the following bound on the maximum number of equi-isoclinic subspaces, along with a tighter bound in the case where A is real.

Theorem 2.4 (Subspace bound for exact optimality). (See [8, Theorem 3.5], [7, Theorem 8].) The number of r -dimensional equi-isoclinic subspaces in \mathbb{C}^n cannot exceed $n^2 - r^2 + 1$. The number of r -dimensional equi-isoclinic subspaces in \mathbb{R}^n cannot exceed $\frac{1}{2}n(n+1) - \frac{1}{2}r(r+1) + 1$.

2.1.2. Constructions

We turn now to constructions which achieve the lower bound (6). Both our constructions in Section 2 will take the form of a Kronecker product between a matrix with good column coherence properties (block coherence with $r = 1$) and a unitary matrix. We begin with a definition. First a definition: a matrix A is said to be an *equiangular tight frame* (ETF) if it satisfies the following conditions:

- The columns of A are unit norm.
- The inner products between pairs of different columns are equal in modulus.
- The columns form a tight frame, that is $AA^* = (mr/n)I$.

ETFs have the important property that they minimize the worst-case column coherence (block coherence with $r = 1$), achieving the Welch bound.

Lemma 2.5 (Welch bound equality). (See [11].) Let $P \in \mathbb{C}^{n \times m}$ be an ETF and let $r = 1$. Then

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

The next result gives a construction with minimum worst-case block coherence. This result appear in various forms throughout [8], but we include a proof of it here for completeness.

Theorem 2.6 (Kronecker product construction 1). (See [8].) Let $A = P \otimes Q$ where $P \in \mathbb{C}^{(n/r) \times m}$ is an ETF and $Q \in \mathbb{C}^{r \times r}$ is a unitary matrix. Then the columns in each block are orthonormal, and

$$\mu(A) = \sqrt{\frac{mr-n}{n(m-1)}}. \tag{9}$$

Proof. Let A be defined as in the theorem. Let us write $P = [p_1 \ p_2 \ \dots \ p_m]$ for the columns of P , and

$$p_i = \begin{bmatrix} P_{1i} \\ P_{2i} \\ \vdots \end{bmatrix}$$

for the entries in each column. Then we have

$$\|A_i^* A_j\|_2 = \left\| \sum_{s=1}^{n/r} (P_{si} Q)^* (P_{sj} Q) \right\|_2 = \|\langle p_i, p_j \rangle Q^* Q\|_2 = \langle p_i, p_j \rangle, \tag{10}$$

where the last equality holds since Q is unitary. Since P is an ETF, it follows from Lemma 2.5 that, for $i \neq j$,

$$\max_{i \neq j} |\langle p_i, p_j \rangle|^2 = \frac{m - \frac{n}{r}}{\frac{n}{r}(m-1)} = \frac{mr-n}{n(m-1)},$$

which combines with (10) to yield

$$\mu(A) = \max_{i \neq j} \|A_i^* A_j\|_2 = \sqrt{\frac{mr-n}{n(m-1)}},$$

which proves (9). It remains to show that the columns in each block are orthonormal. Writing $Q = [q_1 \ q_2 \ \dots \ q_r]$ for the columns of Q , then the columns of A are $p_i \otimes q_j$ for $1 \leq i \leq (n/r)$ and $1 \leq j \leq r$, and a standard Kronecker product identity gives

$$(p_i \otimes q_j)^*(p_i \otimes q_j) = (p_i^* \otimes p_i)(q_j^* \otimes q_j) = 1,$$

which proves that the columns are unit norm. Finally,

$$A_i^* A_i = \sum_{s=1}^{n/r} (P_{si} Q)^*(P_{si} Q) = \langle p_i, p_i \rangle Q^* Q = I,$$

where the last equality follows since P has unit norm columns and Q is unitary, which implies that the columns within each block are orthonormal, and the proof is complete. \square

Kronecker product constructions have also previously been considered in [3].

To construct a matrix for a given triple (m, n, r) in this way, there must exist an ETF of size $(n/r) \times m$, and therefore r must necessarily divide n . ETFs can only exist, however, for certain sizes, see [13,14] for insightful studies. Various ETFs with small dimensions have been explicitly found [13,14], and several infinite families of ETFs are given in [12–15]. Alternating projection algorithms were also used in [26] to construct approximations to ETFs.

In our construction, the ETF P must obey $m \leq (n/r)^2$ [27], and division by r^2 means that the number of blocks is sub-optimal compared with the bound established in Theorem 2.4. This sub-optimality is clearly more marked as r increases. If the Kronecker product construction is used in conjunction with any of the infinite families of ETFs mentioned in the previous paragraph, further sub-optimality results from the fact that, for these families, the number of columns in the ETF is far short of the optimal number [27].

2.2. Bounds and constructions for nearly optimal worst-case block coherence

We now present new results concerning bounds and constructions for optimal worst-case block coherence for a more restrictive family of matrices, namely where the matrix A consists of a union of orthonormal bases. Such constructions have proved important since they can possess almost optimal coherence properties while allowing for more frame elements than would be possible with equiangular tight frames [28]. We will see that unions of orthonormal bases will also provide useful constructions with almost optimal block coherence.

2.2.1. Bounds

Let us suppose then that A consists of orthonormal bases of size $n \times n$ which are split into n/r blocks of size $n \times r$, for which we necessarily require that r divides n . We have the following result.

Theorem 2.7 (Unions of orthonormal bases). *Let A be a union of orthonormal bases. Then*

$$\mu(A) \geq \sqrt{\frac{r}{n}}. \quad (11)$$

If the bound is met, $\|A_i^ A_j\|_2$ is equal for all pairs of blocks A_i and A_j from different orthonormal bases.*

Proof. Since $mr > n$, we have at least two orthonormal bases. Suppose A includes the bases $P \in \mathbb{C}^{n \times n}$ and $Q \in \mathbb{C}^{n \times n}$, which are divided into the blocks $\{P_1, P_2, \dots, P_{n/r}\}$ and $\{Q_1, Q_2, \dots, Q_{n/r}\}$ respectively. Then we have

$$\sum_{i=1}^{n/r} \sum_{j=1}^{n/r} \|P_i^* Q_j\|_F^2 = \|P^* Q\|_F^2 = n, \quad (12)$$

where the last equality follows since $P^* Q$ is unitary. It follows from (12) by a standard norm inequality that

$$\sum_{i=1}^{n/r} \sum_{j=1}^{n/r} \|P_i^* Q_j\|_2^2 \geq \frac{n}{r},$$

from which it follows that

$$\left(\frac{n}{r}\right)^2 \max_{i \neq j} \|P_i^* Q_j\|_2^2 \geq \frac{n}{r}, \tag{13}$$

and dividing by $(n/r)^2$ gives $\max_{i \neq j} \|P_i^* Q_j\|_2^2 \geq \frac{r}{n}$. The same argument applies to all pairs of orthonormal bases, which yields (11). Finally note that the bound only holds if (13) holds, which requires $\|P_i^* P_j\|_2$ to be equal for all i and j , and since this holds for all pairs of orthonormal bases, the theorem is proved. \square

Note that the bound (11) is tighter than (6), since

$$\sqrt{\frac{r}{n}} > \sqrt{\frac{mr - n}{n(m - 1)}}$$

whenever $r < n$. However, they are asymptotically equivalent for large m , that is to say, for fixed n and r ,

$$\lim_{m \rightarrow \infty} \sqrt{\frac{mr - n}{n(m - 1)}} = \sqrt{\frac{r}{n}}.$$

Analogous to Section 2.1, a bound on the number of subspaces for unions of orthonormal bases can also be derived, by making another link with the world of Grassmannian packings. As well as the Rankin bound on the chordal distance of Lemma 2.1, the authors of [6] also proved a tighter bound which allows for a larger number of subspaces, which we state next.

Lemma 2.8 (Tighter Rankin bound for chordal distance). (See [6, Corollary 5.3], [27, Corollary 1].) For $m > n^2$,

$$\min_{i \neq j} [d_C(\mathcal{S}_i, \mathcal{S}_j)]^2 \leq \frac{r(n - r)}{n}. \tag{14}$$

If the bound is met, $m \leq 2(n + 1)(n - 1)$. Furthermore, if A is real and $m > \frac{1}{2}n(n + 1)$, the bound (14) also holds, and if the bound is met, $m \leq (n - 1)(n + 2)$.

We can deduce from Lemma 2.8 a relaxed bound on the maximum number of blocks.

Theorem 2.9 (Subspace bound for unions of orthonormal bases). Let A be a union of orthonormal bases, and suppose $\mu(A) = \sqrt{\frac{r}{n}}$. Then $m \leq 2(n + 1)(n - 1)$, and furthermore if A is real, $m \leq (n - 1)(n + 2)$.

Proof of Theorem 2.9. Suppose $\mu(A) = \sqrt{\frac{r}{n}}$. Then, by (1),

$$r \left(1 - \max_{i \neq j} \|A_i^* A_j\|_2^2\right) = r \left(1 - \frac{r}{n}\right) = \frac{r(n - r)}{n},$$

which combines with (3) and a standard norm inequality to give

$$\min_{i \neq j} [d_C(\mathcal{S}_i, \mathcal{S}_j)]^2 = r - \max_{i \neq j} \|A_i^* A_j\|_F^2 \geq r \left(1 - \max_{i \neq j} \|A_i^* A_j\|_2^2\right) = \frac{r(n - r)}{n}.$$

It follows that the bound in Lemma 2.8 is met, and so we may apply the lemma to deduce both results. \square

2.2.2. Constructions

In order to obtain matrices with a larger number of blocks, we now consider constructions of the form $P \otimes Q$ in which P is a union of orthobases. We have the following result.

Theorem 2.10 (Kronecker product construction 2). *Let $A = P \otimes Q$ where $P \in \mathbb{C}^{(n/r) \times m}$ is a union of orthobases such that the any pair of columns (p_i, p_j) of P from different orthobases satisfies*

$$p_i^* p_j = \sqrt{\frac{r}{n}}, \quad (15)$$

and where Q is a unitary matrix. Then A itself is a union of orthobases, the columns in each block are orthonormal, and

$$\mu(A) = \sqrt{\frac{r}{n}}. \quad (16)$$

Proof. Adopting the notation in the proof of [Theorem 2.6](#), we may follow the argument in the proof of [Theorem 2.6](#) to deduce [\(10\)](#). It follows from [\(15\)](#) that

$$\max_{i \neq j} |\langle p_i, p_j \rangle| = \sqrt{\frac{r}{n}},$$

which combines with [\(10\)](#) to yield

$$\mu(A) = \max_{i \neq j} \|A_i^* A_j\|_2 = \sqrt{\frac{r}{n}},$$

which yields [\(16\)](#). Since P has unit norm columns, we may rehearse the argument of [Theorem 2.6](#) to deduce that the blocks of A are orthonormal. It remains to show that A itself is a union of orthobases. In this regard, let us write P as

$$P = [P_1 \ P_2 \ \dots \ P_{mr/n}],$$

where P_i is an orthonormal basis for all $i = 1, 2, \dots, mr/n$. Then, for all $i \in \{1, 2, \dots, mr/n\}$, $A_i = P_i \otimes Q$, which is unitary, and the theorem is proved. \square

We have therefore constructed a matrix which satisfies the worst-case block coherence lower bound [\(11\)](#). To construct a matrix for a given triple (m, n, r) in this way, there must exist a union of orthobases, P , of size $(n/r) \times m$, and therefore r must necessarily divide n . In addition, we require the columns of P to satisfy [\(15\)](#). Several families of matrices have been proposed which satisfy these conditions, including Alltop Gabor frames, discrete chirp frames and Kerdock frames, see [\[16\]](#). Each of these frames allow P to have size $(n/r) \times (n/r)^2$, leading to a maximum of $(n/r)^2$ blocks in A .

2.3. Random designs with low worst-case block coherence

We now analyze the block coherence of matrices consisting of random blocks, restricting our attention to the case of A being real.¹ By “random”, we mean that each block in A is an orthonormal basis for a subspace

¹ The analysis becomes more involved in the complex case.

drawn from the uniform distribution on the Grassmann manifold $G(n, r)$ defined in Section 1, which is also the distribution with probability measure invariant under orthogonal transformations of $G(n, r)$ [29].

We first introduce some definitions of some special functions and distributions. Given $p > 0$, define the gamma function [30, Sections 6.1] to be

$$\Gamma(p) := \int_0^\infty t^{p-1} e^{-t} dt, \tag{17}$$

and, given $p > 0$ and r a positive integer, define the multivariate gamma function [29] to be

$$\Gamma_r(p) := \pi^{\frac{1}{4}r(r-1)} \prod_{j=1}^r \Gamma\left(p + \frac{1-j}{2}\right). \tag{18}$$

Let $A_i \in \mathbb{R}^{n \times r}$ and $A_j \in \mathbb{R}^{n \times r}$ be orthonormal bases for two random subspaces, and let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ be the squared singular values of $A_i^* A_j$. Then, if $n \geq 2r$, $(\lambda_1, \lambda_2, \dots, \lambda_r)$ are shown in [29] to follow the multivariate beta distribution.

Lemma 2.11 (Distribution of the squared singular values of $A_i^* A_j$). (See [29].) Let $A_i \in \mathbb{R}^{n \times r}$ and $A_j \in \mathbb{R}^{n \times r}$ be orthonormal bases for two real random subspaces, where $n \geq 2r$, and let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ be the squared singular values of $A_i^* A_j$. Then $(\lambda_1, \lambda_2, \dots, \lambda_r) \sim \text{Beta}_r(\frac{r}{2}, \frac{n-r}{2})$, with pdf

$$f(\lambda_1, \lambda_2, \dots, \lambda_r) = c_{n,r} \prod_{i < j} (\lambda_i - \lambda_j) \cdot \prod_{i=1}^r \lambda_i^{-\frac{1}{2}} (1 - \lambda_i)^{\frac{1}{2}(n-2r-1)}; \quad \lambda_i \geq 0 \ \forall i, \tag{19}$$

where

$$c_{n,r} := \frac{\pi^{\frac{1}{2}r^2} \Gamma_r(\frac{n}{2})}{[\Gamma_r(\frac{r}{2})]^2 \Gamma_r(\frac{n-r}{2})},$$

and where $\Gamma_r(\cdot)$ is defined in (18).

We are interested in λ_1 , the largest squared singular value. Note that the restriction $n \geq 2r$ is quite natural, since we have $\lambda_1 = 1$ if $n < 2r$ since the subspaces have nontrivial intersection.

While (19) is somewhat opaque, it is possible to derive from it a quantitative *asymptotic* bound on λ_1 in a framework in which the dimensions (n, r) grow proportionally. More precisely, we consider a sequence of matrices A in which the dimensions (n, r) tend to infinity in proportion to each other. We define $\beta \in (0, 1]$ to be the limiting value of the ratio r/n , that is

$$\beta = \lim_{n \rightarrow \infty} \frac{r}{n}.$$

Given $\beta \in (0, 1/2)$, we determine a small constant $\hat{a}(\beta)$ (see later in the section for a quantification) such that the probability that $[\mu(A)]^2$ exceeds $\hat{a}(\beta) \cdot \beta$ vanishes exponentially as $n \rightarrow \infty$. Practically speaking, our asymptotic result will accurately describe the behavior of random subspaces in large dimensions.

The significance of this result is the way it relates the worst-case block coherence of random subspaces to the fundamental limits derived in Section 2.1. There it was shown that a lower bound on $[\mu(A)]^2$ is given by $(mr - n)/[n(m - 1)]$, which tends to β as $(m, n, r) \rightarrow \infty$ proportionally. It follows that, in this asymptotic sense, random subspaces gives worst-case block coherence with optimal order, though the multiple involved is slightly sub-optimal.

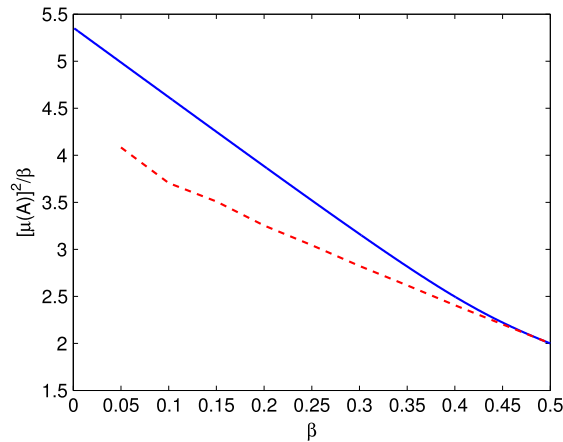


Fig. 1. Worst-case block coherence for random subspaces: theoretical upper bound (unbroken) and the empirical average from 1000 trials with $n = 1000$ and $m = (n/r)^2$ (dashed).

We next define our threshold $\hat{a}(\beta)$. Let $\beta \in (0, 1/2)$, and let $\hat{a}(\beta)$ be the unique solution in $2 \leq a < 1/\beta$ to the equation

$$\beta \ln a + \left(\frac{1 - 2\beta}{2} \right) \ln(1 - a\beta) - (1 - \beta) \ln(1 - \beta) = 0. \quad (20)$$

Eq. (20) has a unique solution in $a > 2$, since the left-hand side is positive for $a = 2$, tends to $-\infty$ as $a \rightarrow 1/\beta$, and is monotonically decreasing on $[2, 1/\beta)$. We proceed to our main result for random subspaces, the proof of which can be found in [Appendix A](#). Note that the restriction $\beta \in (0, 1/2)$ corresponds to the restriction $n > 2r$. The case $n = 2r$ is not covered by our analysis, since in that case the proportional–dimensional analysis does not apply (see [Appendix A](#)). However it can be shown that $\hat{a}(\beta) \rightarrow 2$ as $\beta \rightarrow 1/2$, and therefore one expects the result $a(1/2) = 2$, which would imply that the subspaces have non-trivial intersection with high probability.

Theorem 2.12. *Let $r/n \rightarrow \beta \in (0, 1/2)$ as $n \rightarrow \infty$ and choose $\epsilon > 0$. Let A be a real matrix consisting of random blocks, as defined earlier in this section, where the number of blocks m is polynomial in n . Then*

$$\mathbb{P}\{[\mu(A)]^2 \geq \hat{a}(\beta) \cdot \beta + \epsilon\} \rightarrow 0,$$

where $\hat{a}(\beta)$ is defined in (20).

Fig. 1 plots the square of the asymptotic upper bound on worst-case block coherence of random subspaces, $\hat{a}(\beta)$, for $\beta \in (0, 1/2)$. We see that worst-case block coherence is never more than a small multiple² of $\sqrt{\beta} \approx \sqrt{r/n}$. Also plotted is the empirical behavior of worst-case coherence for random subspaces: fixing $n = 1000$, for varying r , 1000 trials of $(n/r)^2$ random subspaces were generated (mimicking the Kronecker product constructions of [Section 2.2](#)), and the worst-case coherence calculated as an average over all trials. We observe that the theoretical upper bound is somewhat tight, and that the expected asymptotic behavior is in evidence even for this modest problem size of $n = 1000$. Note that empirical tests take $\beta \geq 0.05$, since large-scale testing is required to obtain high-dimensional limiting behavior for smaller values of β .

² As $\beta \rightarrow 0$, $\hat{a}(\beta)$ tends to ≈ 5.357 , which is the solution in $2 \leq a < 1/\beta$ to $a = 2(1 + \ln a)$, and this upper bounds $\hat{a}(\beta)$ on $\beta \in (0, 1/2)$.

Table 1

Worst-case and average block coherence for $A = P \otimes Q$, for Q unitary and P belonging to various families of frames.

Frame family	$\mu(A)$	$\nu(A)$
Alltop Gabor	$\frac{1}{\sqrt{n}}$	$\frac{1}{n+1}$
Discrete chirp	$\frac{1}{\sqrt{n}}$	$\frac{m-n}{n(m-1)}$
Kerdock	$\frac{1}{\sqrt{n}}$	$\frac{1}{m-1}$
Dual BCH	$\sqrt{\frac{2}{n}}$	$\frac{m-n}{n(m-1)}$

3. Average coherence

3.1. Theoretical results

The constructions which consist of unions of orthobases are also amenable to the analysis of their average block coherence, defined in (2). Given a construction $A = P \otimes Q$, where Q is unitary, we relate the average block coherence of A to the average column coherence of P [16], defined to be

$$\nu_1(P) := \frac{1}{m-1} \max_i \left\| \sum_{j \neq i} p_i^* p_j \right\|_2, \tag{21}$$

where $\{p_i\}$ are the columns of P .

Lemma 3.1. *Let $A = P \otimes Q$ where $P \in \mathbb{C}^{(n/r) \times m}$, and where $Q \in \mathbb{C}^{r \times r}$ is a unitary matrix. Then*

$$\nu(A) = \nu_1(P). \tag{22}$$

Proof. Adopting the notation in the proof of Theorem 2.6, we may follow the argument in the proof of Theorem 2.6 to deduce (10). Then, using (2) and (21), we have

$$\nu(A) = \frac{1}{m-1} \max_i \left\| \sum_{j \neq i} A_i^* A_j \right\|_2 = \frac{1}{m-1} \max_i \left| \sum_{j \neq i} p_i^* p_j \right| = \nu_1(P),$$

which proves (22). \square

Using results from [16] in combination with Lemma 3.1, we obtain specific results for P belonging to several families of frames, which are displayed in Table 1. See [16] for more details on how these frames are constructed.

3.2. Flipping algorithm

In the following we will present an algorithm that can improve the average coherence by random rotations, while preserving the worst-case coherence, since $\|X\|_2 = \|-X\|_2$ for a matrix X . This algorithm is inspired by a similar flipping algorithm for columns [20].

Theorem 3.2. *Let $\{A_i\}_{i=1}^m$ be a set of subspaces with $A_i \in \mathbb{R}^{n \times r}$, where $a_i = \{1, -1\}$ are equally likely and mutually independent. If $c > 16\sqrt{e} \sqrt{\frac{n}{r(m-1)}}$, then there exists a set of flipped subspaces $\{a_i A_i\}$ with the same worst-case coherence and satisfies $\nu \leq c\mu \sqrt{\frac{r \log m}{n}}$.*

Proof. Note that the average coherence of the wiggled subspaces is given by

$$\nu = \frac{1}{m-1} \max_i \left\| \sum_{j \neq i} a_i a_j A_i^* A_j \right\|_2.$$

Now consider for fixed $i = 1$,

$$\left\| \sum_{j \neq i} a_i a_j A_i^* A_j \right\|_2 = \left\| \sum_{j=2}^m a_i a_j A_i^* A_j \right\|_2$$

We can construct a sequence $M_k = \sum_{j=1}^{k+1} a_i a_j A_i^* A_j$, and let $M_0 = 0$. The sequence M_k , $k = 0, 1, \dots, m-1$ forms a martingale, since for $k = 1, \dots, m-1$

$$\begin{aligned} \mathbb{E}\{M_k - M_{k-1}\} &= \mathbb{E}\{a_1 a_{k+1} A_1^* A_{k+1}\} \\ &= \mathbb{E}\{a_1\} \mathbb{E}\{a_{k+1}\} A_1^* A_{k+1} = 0, \end{aligned}$$

where we have used the fact that a_i 's are mutually independent. The martingale difference is bounded

$$\begin{aligned} \|M_k - M_{k-1}\|_2 &= \|a_1 a_{k+1} A_1^* A_{k+1}\|_2 \\ &= \|A_1^* A_{k+1}\|_2 \leq \mu, \quad \forall k = 0, \dots, m-1. \end{aligned}$$

Using the Banach-space-valued Azuma's inequality, we obtain the following

$$\mathbb{P}\{\|M_k - M_0\|_2 \geq \delta\} \leq e^2 \exp\left\{-\frac{c_0 \delta^2}{(m-1)\mu^2}\right\}$$

where $c_0 \triangleq \frac{e^{-1}}{256}$. Note that

$$\frac{1}{m-1} \|M_k - M_0\|_2 = \frac{1}{m-1} \left\| \sum_{j=2}^m A_1^* A_j \right\|_2.$$

Hence taking a union bound with respect to i :

$$\begin{aligned} &\mathbb{P}\left\{\frac{1}{m-1} \max_i \left\| \sum_{j \neq i} a_i a_j A_i^* A_j \right\|_2 \geq \nu\right\} \\ &= \mathbb{P}\left\{\max_i \left\| \sum_{j \neq i} A_i^* A_j \right\|_2 \geq (m-1)\nu\right\} \\ &\leq e^2 m \exp\left\{-\frac{c_0 \nu^2 (m-1)}{\mu^2}\right\} \end{aligned}$$

Set

$$\nu \leq c\mu \sqrt{\frac{r \log m}{n}},$$

then

Algorithm 1 Flipping algorithm

```

1: Input:  $A = [A_1 \dots A_m]$ 
2: Output:  $B = [B_1 \dots B_m]$ 
3:  $B_1 = A_1, F_1 = B_1$ 
4: for  $k = 1 \rightarrow m - 1$  do
5:   if  $\|F_k + A_{k+1}\|_2 \leq \|F_k - A_{k+1}\|_2$  then
6:      $B_{k+1} = A_{k+1}$ 
7:   else
8:      $B_{k+1} = -A_{k+1}$ 
9:   end if
10:   $F_{k+1} = F_k + B_{k+1}$ 
11: end for
    
```

$$\begin{aligned} & \mathbb{P} \left\{ \frac{1}{m-1} \max_i \left\| \sum_{j \neq i} a_i a_j A_i^* A_j \right\|_2 \geq \nu \right\} \\ & \leq \exp \left\{ 2 + \left(1 - c_0 c^2 (m-1) \frac{r}{n} \right) \log m \right\} \end{aligned} \tag{23}$$

where $c_0 = e^{-1}/256$. Hence when $c > 16\sqrt{e}\sqrt{\frac{n}{r(m-1)}}$ (note that $n \ll r(m-1)$), $1 - c_0 c^2 (m-1) \frac{r}{n} < 0$. When $\frac{mr}{n}$ is sufficiently large and m is sufficiently large, the right hand side of (23) is less than 1. Under such conditions,

$$\mathbb{P} \left\{ \frac{1}{m-1} \max_i \left\| \sum_{j \neq i} a_i a_j A_i^* A_j \right\|_2 \leq \nu \right\} > 0,$$

which proves the theorem. \square

In [4], it is established that for the one-step group thresholding algorithm to perform well, the average block coherence ν has to be sufficiently small: $\nu \leq c\mu\sqrt{\frac{r \log m}{n}}$. In the following theorem, we demonstrate that the flipping algorithm can produce a set of subspaces $\{A_i\}_{i=1}^m$ that satisfies the condition on block coherence.

Theorem 3.3. *Let $\{A_i\}_{i=1}^m, A_i \in \mathbb{R}^{n \times r}$ be a set of subspaces. Suppose*

$$\frac{m-1}{m-n/r} \cdot \frac{1}{\log m} \cdot \frac{\sqrt{m}+1}{m-1} \leq c^2 \left(\frac{r}{n} \right)^2$$

for some constant $c > 0$, then the flipping algorithm in Algorithm 1 outputs a set of subspaces $\{B_i\}_{i=1}^m$ with the same worst-case block coherence μ and the average block coherence satisfies $\nu \leq c\mu\sqrt{\frac{r \log m}{n}}$.

Proof. Lemma 2.3 sets a lower-bound on the worse-case block coherence. Combine this with Lemma 3.4, which says that the flipping algorithm produces a set of subspaces with $\nu \leq \frac{\sqrt{m}+1}{m-1}$. Hence, to have $\nu \leq c\mu\sqrt{\frac{r \log m}{n}}$ for $\{B_i\}$, it suffices to have

$$\frac{\sqrt{m}+1}{m-1} \leq c\sqrt{\frac{mr-n}{n(m-1)}} \cdot \sqrt{\frac{r \log m}{n}},$$

or equivalently

$$\frac{m-1}{m-n/r} \cdot \frac{1}{\log m} \cdot \frac{\sqrt{m}+1}{m-1} \leq c^2 \left(\frac{r}{n} \right)^2,$$

for the constant c . \square

The following lemma finds an upper bound on the average block coherence for the set of subspaces produced by the flipping algorithm:

Lemma 3.4. *The flipping algorithm produces $\{B_k\}$ whose average coherence satisfies $\nu \leq \frac{\sqrt{m+1}}{m-1}$.*

Proof. Note that

$$\begin{aligned} \left\| \sum_{j \neq i} A_i^* A_j \right\|_2 &= \left\| \sum_{j=1}^m A_i^* A_j - A_i^* A_i \right\|_2 \\ &\leq \left\| \sum_{j=1}^m A_i^* A_j \right\|_2 + \|A_i^* A_i\|_2 = \left\| \sum_{j=1}^m A_i^* A_j \right\|_2 + 1 \\ &\leq \|A_i\|_2 \left\| \sum_{j=1}^m A_j \right\|_2 + 1 = \left\| \sum_{j=1}^m A_j \right\|_2 + 1, \end{aligned}$$

where we have used the triangle inequality of the matrix norm $\|A + B\|_2 \leq \|A\|_2 + \|B\|_2$, the inequality $\|AB\|_2 \leq \|A\|_2 \|B\|_2$, $A_i^* A_j = I_r$, and $\|A_i\|_2 = 1$.

Next, we prove the result by induction. First, $\|F_1\|_2 = \|A_1\|_2 = 1$. Next assume that $\|F_k\|_2 = k$. Use the parallelogram law $2\|X\|_2^2 + 2\|Y\|_2^2 = \|X + Y\|_2^2 + \|X - Y\|_2^2$ for matrices X and Y , we have that

$$\begin{aligned} \|F_k\|_2^2 &= -\|A_{k+1}\|_2^2 + \frac{1}{2} (\|F_k + A_{k+1}\|_2^2 + \|F_k - A_{k+1}\|_2^2) \\ &\geq -\|A_{k+1}\|_2^2 + \frac{1}{2} \cdot 2\|F_k + A_{k+1}\|_2^2 \\ &= -1 + \|F_{k+1}\|_2^2. \end{aligned}$$

Hence $\|F_{k+1}\|_2^2 \leq 1 + \|F_k\|_2^2 \leq k + 1$, by induction. Hence $\|F_m\|_2^2 \leq m$. By definition of F_m , this means that $\frac{1}{m-1} \left\| \sum_{i=1}^m B_i \right\|_2 \leq \frac{\sqrt{m+1}}{m-1}$. \square

4. Numerical examples

In this section, we present numerical examples, to compare random frames formed by random subspaces, with the deterministic frames formed as a Kronecker product of a matrix with an orthogonal matrix.

In Fig. 2, we compare the performance of the one-step block thresholding algorithm for support recovery, comparing deterministic frames with those formed from the union of random subspaces. The performance metric is the non-discovery proportion (NDP), which is the fraction of entries in the original support of the signal that are not present in the recovered support. In Fig. 2a, the deterministic frame is constructed as $A = P \otimes Q$, where P is an equiangular tight frame [14], and Q is a Hadamard matrix. The size of the frame is $n = 12$, the number of blocks $m = 16$, and the size of the blocks is $r = 2$. The random subspaces are formed by first generating a Gaussian matrix $A \in \mathbb{R}^{n \times mr}$, performing eigendecomposition for each block $A_i \in \mathbb{R}^{n \times 2}$ to find the left singular vectors U_i , and finally forming the block matrix $[U_1 \ \dots \ U_m]$. The different lines in the same block correspond to different dynamic ranges for amplitudes of non-zero elements. Note that the deterministic construction has much better performance than that of the union of random subspaces. Also note the NDP does not depend on dynamic ranges, which is consistent with the result in [4].

In Fig. 2b, we consider larger frames with $n = 128$, $m = 2048$, and $r = 2$. We compare a deterministic construction with $A = P \otimes Q$, where P is a real Kerdock matrix, and Q is a Hadamard matrix. Using a Kerdock matrix for P allows us to pack more blocks in forming a frame, while still obtaining a low coherence property. The random subspaces are constructed from random Gaussian matrices as above. Similarly, the

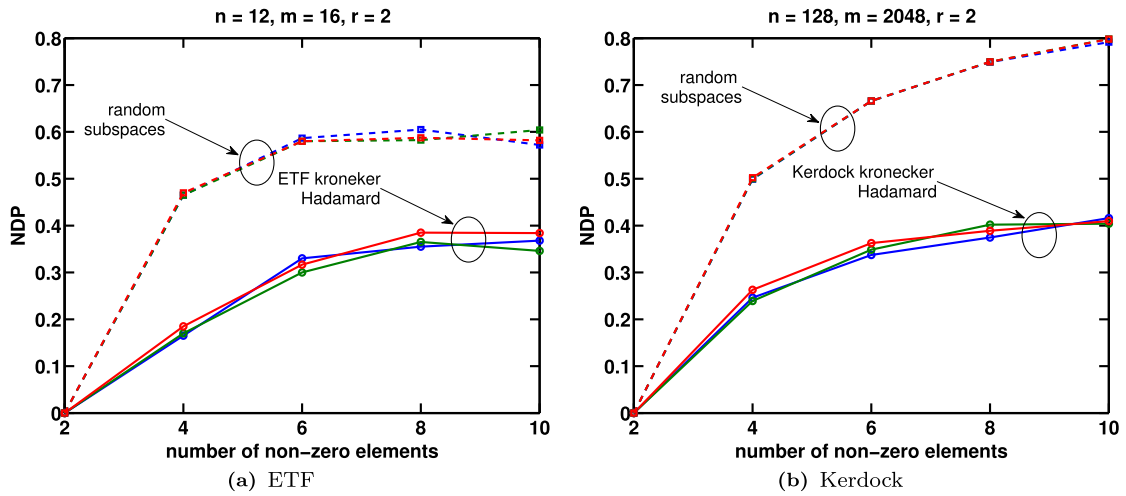


Fig. 2. Non-discovery proportion (NDP) for the random subspaces and (a) when A is formed as equiangular tight frame (ETF) with a Hadamard matrix; (b) when A is formed as the Kronecker product of a Kerdock matrix with a Hadamard matrix.

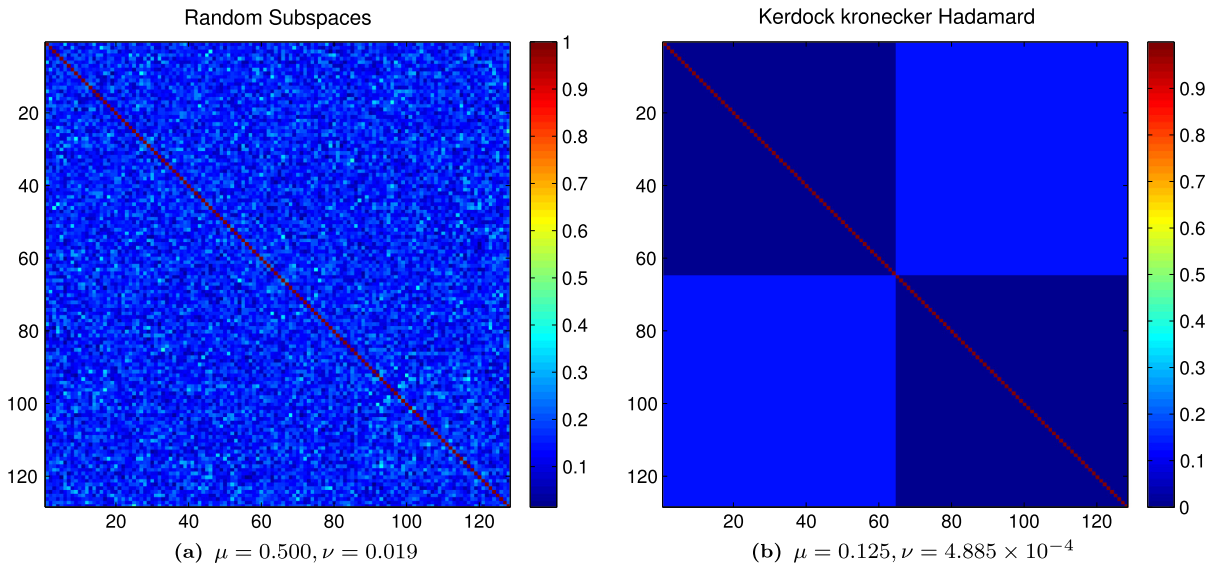


Fig. 3. Maps $\|A_i^* A_j\|_2$ for random subspaces, and A formed by Kerdock matrix Kronecker with Hadamard matrix.

deterministic construction has better performance than the random subspaces, and the performance does not depend on the dynamic range of the non-zero elements. Fig. 3 demonstrates $\|A_i^* A_j\|_2$ for $i, j = 1, 2, \dots, 128$ for the deterministic frame and the random subspaces, as well as their worst-case block coherence μ and average block coherence ν . Note that the deterministic frame has much lower coherence properties and the magnitudes of $\|A_i^* A_j\|_2$ have a structure: the magnitude of $\|A_i^* A_j\|_2, i \neq j$ takes only 2 values $\{0, \sqrt{r/n}\}$.

In the next example, we demonstrate the effectiveness of the flipping algorithm in reducing the average coherence of a frame consisting of random subspaces. We apply the flipping algorithm on 10 realization of random frames with size $n = 128, m = 2048$, and $r = 1, 2, 3$, respectively (see Table 2, Fig. 4). The flipping algorithm reduces the average coherence of the random frame by a significant percentage.

Table 2
 ν before and after flipping algorithm for union of random subspaces.

	$r = 1$	$r = 2$	$r = 3$
Before	0.0096	0.0181	0.0111
After	0.0015	0.0027	0.0037
% of improvement	84.6%	85.2%	66.6%

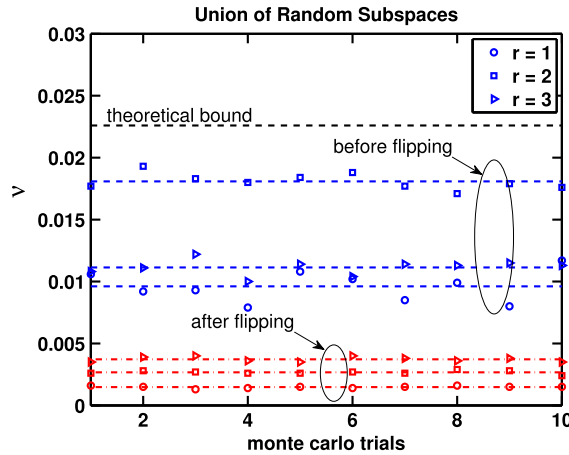


Fig. 4. Average block coherence of a random subspaces, for $r = 1, 2, 3$, respectively, before (dashed) and after (dash-dot) passing through the flipping algorithm. The dashed lines represent mean of random realization of the average block coherence, before and after the flipping algorithm is applied. The theoretical bound corresponds to [Lemma 3.4](#).

5. Conclusion

In this paper, we have studied worst-case and average block coherence of frames: from the perspective of compressed sensing, we are interested in matrices for which both of these quantities are low/optimal. We have shown that worst-case block coherence is subject to a fundamental lower bound, and we have presented deterministic constructions which achieve it. On the other hand, our analysis of random subspaces shows that they have only slightly sub-optimal worst-case block coherence. Meanwhile, low average block coherence can either be achieved by construction, or by means of a “flipping” algorithm designed to reduce average block coherence while preserving worst-case block coherence. A byproduct of our work is numerous constructions of families of optimal and almost optimal Grassmann packings.

Some questions remain: does there exist a fundamental lower bound on average block coherence, and if so which frames achieve it? What is the average block coherence in the case of random subspaces? Our optimal Grassmann packings give a smaller number of subspaces than the present theoretical limits: what is the true limit for the number of subspaces in an optimal Grassmann packing, and are there constructions which achieve it?

Appendix A. Proof of [Theorem 2.12](#)

We first introduce some further definitions.

Given $p, q > 0$ and $x \in [0, 1]$, define the beta function [[30, Sections 6.2](#)] to be

$$B(p, q) := \int_0^1 t^{p-1} (1-t)^{q-1} dt, \quad (\text{A.1})$$

and define the regularized incomplete beta function [[30, Section 6.6](#)] to be

$$I_x(p, q) := \frac{1}{B(p, q)} \int_0^x t^{p-1}(1-t)^{q-1} dt. \tag{A.2}$$

The gamma and beta functions defined in (17) and (A.1) are related as follows.

Lemma A.1 (Beta–gamma identity). (See [31, Theorem 7].) For any $p, q > 0$,

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \tag{A.3}$$

We define $H(\rho)$, the Shannon entropy with base e logarithms, in the usual way as

$$H(\rho) := -\rho \ln \rho - (1 - \rho) \ln(1 - \rho), \quad \text{for } \rho \in (0, 1). \tag{A.4}$$

The next result describes the asymptotic behavior of the gamma function.

Lemma A.2 (Log-gamma asymptotic). (See [31, Theorem 12 and (4)].) For $p > 0$,

$$\ln \Gamma(p) = \left(p - \frac{1}{2}\right) \ln p - p + \frac{1}{2} \ln(2\pi) + o(1).$$

We may use Lemma A.2 to deduce the following limiting result for the beta function when the parameters grow proportionally.

Lemma A.3 (Beta function limit). Let $(p, q) \rightarrow \infty$ such that $p/(p+q) \rightarrow \rho$ as $(p, q) \rightarrow \infty$. Then

$$\lim_{(p,q) \rightarrow \infty} \frac{1}{p+q} \ln B(p, q) = -H(\rho). \tag{A.5}$$

Proof. By (A.3) and (A.5), we have

$$\begin{aligned} & \lim_{(p,q) \rightarrow \infty} \frac{1}{p+q} \ln B(p, q) \\ &= \lim_{(p,q) \rightarrow \infty} \frac{1}{p+q} \ln \left[\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \right] \\ &= \lim_{(p,q) \rightarrow \infty} \frac{1}{p+q} [p \ln p - p + q \ln q - q - (p+q) \ln(p+q) + (p+q)] \\ &= \lim_{(p,q) \rightarrow \infty} \frac{1}{p+q} \ln \left[\frac{p^p q^q}{(p+q)^{p+q}} \right] \\ &= \lim_{(p,q) \rightarrow \infty} \frac{1}{p+q} \ln \left[\frac{\left(\frac{p}{p+q}\right)^p \left(\frac{q}{p+q}\right)^q}{\left(\frac{p+q}{p+q}\right)^{p+q}} \right] \\ &= \lim_{(p,q) \rightarrow \infty} \left[\frac{p}{p+q} \ln \left(\frac{p}{p+q} \right) + \frac{q}{p+q} \ln \left(\frac{q}{p+q} \right) \right] \\ &= \rho \ln \rho + (1 - \rho) \ln(1 - \rho), \end{aligned}$$

and the result follows by (A.4). \square

The next result gives the limiting behavior of the incomplete regularized beta function in the case where the parameters grow proportionally.

Lemma A.4 (Regularized incomplete beta limit). (See [32, Theorem 4.18].) Fix $0 < x < 1$ and let $p, q > 0$ satisfy

$$\frac{p}{p+q} > x.$$

Let $(p, q) \rightarrow \infty$ such that

$$\lim_{p \rightarrow \infty} \frac{p}{p+q} = \rho,$$

where $\rho > x$. Then

$$\lim_{p \rightarrow \infty} \frac{1}{p+q} \ln I_x(p, q) = - \left[\rho \ln \left(\frac{\rho}{x} \right) + (1 - \rho) \ln \left(\frac{1 - \rho}{1 - x} \right) \right]. \quad (\text{A.6})$$

A random variable X follows the (univariate) beta distribution $\text{Beta}(a, b)$ for $a, b > 0$ if it has probability density function (pdf)

$$f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}; \quad 0 \leq x \leq 1,$$

and we have the following result concerning its distribution function.

Lemma A.5. Let $X \sim \text{Beta}(a, b)$. Then its distribution function is

$$F(x) = I_x(a, b),$$

where $I_x(a, b)$ is the regularized incomplete beta function.

We may deduce from (19) the following bound on the pdf of the largest squared singular value λ_1 .

Lemma A.6. Let $(\lambda_1, \lambda_2, \dots, \lambda_r)$ be distributed as in Lemma 2.11. Then the following bound holds for the pdf of λ_1 .

$$f(\lambda_1) \leq \frac{\sqrt{\pi}}{[B(\frac{r}{2}, \frac{n-r}{2})]^2} \cdot \lambda_1^{\frac{1}{2}(2r-1)-1} (1 - \lambda_1)^{\frac{1}{2}(n-2r+1)-1}, \quad (\text{A.7})$$

where $B(\cdot, \cdot)$ is defined in (A.1).

Proof. Let R be the region of $(r-1)$ -dimensional space $\lambda_i \geq 0$; $i = 2, 3, \dots, r$, and given $\lambda_1 > 0$, let R_{λ_1} be the sub-region of R consisting of all $(\lambda_2, \dots, \lambda_r)$ such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$. Then

$$\begin{aligned} f(\lambda_1) &= c_{n,r} \lambda_1^{-\frac{1}{2}} (1 - \lambda_1)^{\frac{1}{2}(n-2r-1)} \\ &\quad \cdot \int_{R_{\lambda_1}} \prod_{i < j} (\lambda_1 - \lambda_j) \cdot \prod_{i=2}^r \lambda_i^{-\frac{1}{2}} (1 - \lambda_i)^{\frac{1}{2}(n-2r-1)} d\lambda_i \\ &\leq c_{n,r} \lambda_1^{r-\frac{3}{2}} (1 - \lambda_1)^{\frac{1}{2}(n-2r-1)} \\ &\quad \cdot \int_{R_{\lambda_1}} \prod_{2 \leq i < j} (\lambda_1 - \lambda_j) \cdot \prod_{i=2}^r \lambda_i^{-\frac{1}{2}} (1 - \lambda_i)^{\frac{1}{2}(n-2r-1)} d\lambda_i \end{aligned}$$

$$\begin{aligned} &\leq c_{n,r} \lambda_1^{r-\frac{3}{2}} (1-\lambda_1)^{\frac{1}{2}(n-2r-1)} \\ &\quad \cdot \int_R \prod_{2 \leq i < j} (\lambda_i - \lambda_j) \cdot \prod_{i=2}^r \lambda_i^{-\frac{1}{2}} (1-\lambda_i)^{\frac{1}{2}(n-2r-1)} d\lambda_i \\ &= \frac{c_{n,r}}{c_{n-2,r-1}} \lambda_1^{\frac{1}{2}(2r-1)-1} (1-\lambda_1)^{\frac{1}{2}(n-2r+1)-1}, \end{aligned} \tag{A.8}$$

where the last line follows by observing that we are integrating a multivariate beta distribution with $(n, r) \rightarrow (n-2, r-1)$. It remains to calculate $c_{n,r}/c_{n-2,r-1}$. We have

$$\begin{aligned} \frac{c_{n,r}}{c_{n-2,r-1}} &= \frac{\pi^{\frac{1}{2}r^2} \Gamma_r(\frac{n}{2})}{[\Gamma_r(\frac{r}{2})]^2 \Gamma_r(\frac{n-r}{2})} \cdot \frac{[\Gamma_{r-1}(\frac{r-1}{2})]^2 \Gamma_{r-1}(\frac{n-r-1}{2})}{\pi^{\frac{1}{2}(r-1)^2} \Gamma_{r-1}(\frac{n-2}{2})} \\ &= \sqrt{\pi} \cdot \prod_{j=1}^r \frac{\Gamma(\frac{n+1-j}{2})}{[\Gamma(\frac{r+1-j}{2})]^2 \Gamma(\frac{n-r+1-j}{2})} \cdot \prod_{j=1}^{r-1} \frac{[\Gamma(\frac{r-j}{2})]^2 \Gamma(\frac{n-r-j}{2})}{\Gamma(\frac{n-1-j}{2})} \\ &= \sqrt{\pi} \cdot \frac{\Gamma(\frac{n}{2}) \Gamma(\frac{n-1}{2})}{[\Gamma(\frac{r}{2})]^2 [\Gamma(\frac{n-r}{2})]^2} \\ &\leq \sqrt{\pi} \cdot \frac{[\Gamma(\frac{n}{2})]^2}{[\Gamma(\frac{r}{2})]^2 [\Gamma(\frac{n-r}{2})]^2} \\ &= \frac{\sqrt{\pi}}{[B(\frac{r}{2}, \frac{n-r}{2})]^2}, \end{aligned}$$

which combines with (A.8) to prove (A.7). \square

Now let us write $\bar{F}(\lambda_1)$ for complementary distribution function of λ_i , which we next bound.

Lemma A.7. *Let $(\lambda_1, \lambda_2, \dots, \lambda_r)$ be distributed as in Lemma 2.11. Then the following bound holds for $\bar{F}(\lambda_1)$.*

$$\begin{aligned} \bar{F}(\lambda_1) &\leq G(\lambda_1) := \sqrt{\pi} \cdot \frac{B[\frac{1}{2}(2r-1), \frac{1}{2}(n-2r+1)]}{[B(\frac{r}{2}, \frac{n-r}{2})]^2} \\ &\quad \cdot I_{1-\lambda_1} \left[\frac{1}{2}(n-2r+1), \frac{1}{2}(2r-1) \right], \end{aligned} \tag{A.9}$$

where $B(\cdot, \cdot)$ is defined in (A.1) and where $I_x(a, b)$ is defined in (A.2).

Proof. We may rewrite (A.7) as

$$f(\lambda_1) \leq \sqrt{\pi} \cdot \frac{B[\frac{1}{2}(2r-1), \frac{1}{2}(n-2r+1)]}{[B(\frac{r}{2}, \frac{n-r}{2})]^2} \cdot \frac{\lambda_1^{\frac{1}{2}(2r-1)-1} (1-\lambda_1)^{\frac{1}{2}(n-2r+1)-1}}{B[\frac{1}{2}(2r-1), \frac{1}{2}(n-2r+1)]}, \tag{A.10}$$

and we observe using Lemma A.7 that the right-hand expression is the pdf of the $B[\frac{1}{2}(2r-1), \frac{1}{2}(n-2r+1)]$ distribution. We therefore have, by Lemma A.7,

$$\begin{aligned} \bar{F}(\lambda_1) &\leq \sqrt{\pi} \cdot \frac{B[\frac{1}{2}(2r-1), \frac{1}{2}(n-2r+1)]}{[B(\frac{r}{2}, \frac{n-r}{2})]^2} \cdot \left\{ 1 - I_{\lambda_1} \left[\frac{1}{2}(2r-1), \frac{1}{2}(n-2r+1) \right] \right\} \\ &= \sqrt{\pi} \cdot \frac{B[\frac{1}{2}(2r-1), \frac{1}{2}(n-2r+1)]}{[B(\frac{r}{2}, \frac{n-r}{2})]^2} \cdot I_{1-\lambda_1} \left[\frac{1}{2}(n-2r+1), \frac{1}{2}(2r-1) \right], \end{aligned}$$

as required. \square

The next lemma analyzes the exponent of the upper tail of the distribution.

Lemma A.8. *Let $r/n \rightarrow \beta \in (0, 1/2)$ as $n \rightarrow \infty$. Then, for any $2 \leq a < 1/\beta$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln G(a\beta) = \beta \ln a + \left(\frac{1-2\beta}{2} \right) \ln(1-a\beta) - (1-\beta) \ln(1-\beta), \quad (\text{A.11})$$

where $H(\cdot)$ is defined in (A.4).

Proof. Applying (A.6), we obtain, for a sufficiently large,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \ln I_{1-a\beta} \left[\frac{1}{2}(n-2r+1), \frac{1}{2}(2r-1) \right] \\ &= - \left[\left(\frac{1-2\beta}{2} \right) \ln \left(\frac{1-2\beta}{1-a\beta} \right) + \beta \ln \left(\frac{2\beta}{a\beta} \right) \right], \end{aligned}$$

which may be combined with (A.5) to give

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln G(a\beta) = -\frac{1}{2}H(2\beta) + H(\beta) + \frac{1}{2}H(2\beta) + \beta \ln(a\beta) + \left(\frac{1-2\beta}{2} \right) \ln(1-a\beta),$$

which simplifies to give (A.11). \square

This threshold will be used to upper bound the square of the spectral norm of two random subspaces, and hence the worst-case coherence. Note that the analysis presented here does not apply to the case where $\beta = 1/2$, since in this case one of the parameters in the regularized incomplete beta function term is fixed and does not tend to infinity. We proceed to the proof of the main result.

Proof of Theorem 2.12. Consider a matrix $A_i^* A_j$ for some pair of blocks in A . It follows from (A.9) and (A.11) that, for any $\eta > 0$,

$$\frac{1}{n} \ln \mathbb{P} \{ \|A_i^* A_j\|_2^2 \geq a\beta \} \leq \beta \ln a + \left(\frac{1-2\beta}{2} \right) \ln(1-a\beta) - (1-\beta) \ln(1-\beta) + \eta, \quad (\text{A.12})$$

for all n sufficiently large. By the definition of $\hat{a}(\beta)$ in (20), and since the left-hand side of (20) is strictly decreasing in a for $\beta \in (0, 1/2)$, for any $\epsilon > 0$, setting $a := \hat{a}(\beta) + \epsilon$, and choosing η sufficiently small in (A.12) ensures

$$\frac{1}{n} \ln \mathbb{P} \{ \|A_i^* A_j\|_2^2 \geq \hat{a}(\beta) \cdot \beta + \epsilon \} \leq -C \quad \text{for all } n \text{ sufficiently large,}$$

where C is some positive constant, from which it follows that

$$\mathbb{P} \{ \|A_i^* A_j\|_2^2 \geq \hat{a}(\beta) \cdot \beta + \epsilon \} \leq e^{-Cn} \quad \text{for all } n \text{ sufficiently large.} \quad (\text{A.13})$$

Now we use (A.13) to union bound over all pairs of blocks, obtaining

$$\begin{aligned} \mathbb{P} \{ [\mu(A)]^2 \geq \hat{a}(\beta) \cdot \beta + \epsilon \} &= \mathbb{P} \left\{ \bigcup_{i \neq j} (\|A_i^* A_j\|_2^2 \geq \hat{a}(\beta) \cdot \beta + \epsilon) \right\} \\ &\leq \sum_{i \neq j} \mathbb{P} \{ \|A_i^* A_j\|_2^2 \geq \hat{a}(\beta) \cdot \beta + \epsilon \} \end{aligned}$$

$$\leq \binom{m}{2} e^{-Cn} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

since m is polynomial in n . \square

References

- [1] D. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (4) (2006) 1289–1306.
- [2] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory* 52 (2) (2006) 489–509.
- [3] Y. Eldar, P. Kuppinger, H. Bölcskei, Compressed sensing of block-sparse signals: uncertainty relations and efficient recovery, *IEEE Trans. Signal Process.* 58 (6) (2010) 3042–3054.
- [4] W. Bajwa, D. Mixon, Group model selection using marginal correlations: the good, the bad and the ugly, in: 50th Annual Allerton Conference on Communication, Control and Computing, 2012, pp. 494–501.
- [5] W. Bajwa, A. Calderbank, M. Duarte, On the conditioning of random block subdictionaries, Tech. Rep. TR-2010-06, Duke University, 2010.
- [6] J. Conway, R. Hardin, N. Sloane, Packing lines, planes, etc.: packing in Grassmannian spaces, *Experiment. Math.* 5 (2) (1996) 139–159.
- [7] I. Dhillon, R. Heath Jr., T. Strohmer, J. Tropp, Constructing packings in Grassmannian manifolds via alternating projection, *Experiment. Math.* 17 (1) (2008) 9–35.
- [8] P. Lemmens, J. Seidel, Equi-isoclinic subspaces of Euclidean space, *Indag. Math. (N.S.)* A76 (1973) 98–107.
- [9] W. Bajwa, R. Calderbank, D. Mixon, Why Gabor frames? Two fundamental measures of coherence and their role in model selection, *J. Commun. Netw.* 12 (4) (2010) 289–307.
- [10] L. Welch, Lower bounds on the maximum cross correlation of signals, *IEEE Trans. Inform. Theory* 20 (3) (1974) 397–399.
- [11] T. Strohmer, R. Heath Jr., Grassmannian frames with applications to coding and communication, *Appl. Comput. Harmon. Anal.* 14 (3) (2003) 257–275.
- [12] M. Fickus, D. Mixon, J. Treiman, Steiner equiangular tight frames, *Linear Algebra Appl.* 436 (5) (2012) 1014–1027.
- [13] M. Sustik, J. Tropp, I. Dhillon, R. Heath Jr., On the existence of equiangular tight frames, *Linear Algebra Appl.* 426 (2–3) (2007) 619–635.
- [14] J. Tropp, Complex equiangular tight frames, in: *Wavelets XI*, in: Proc. SPIE, vol. 5914, 2005.
- [15] J. Jasper, D. Mixon, M. Fickus, Kirkman equiangular tight frames and codes, preprint, 2013.
- [16] W. Bajwa, R. Calderbank, S. Jafarpour, Revisiting model selection and recovery of sparse signals using one-step thresholding, in: 48th Annual Allerton Conference on Communication, Control and Computing, 2010.
- [17] P. Shor, N. Sloane, A family of optimal packings in Grassmannian spaces, *J. Algebraic Combin.* 7 (1998) 157–163.
- [18] R. Calderbank, R. Hardin, E. Rains, P. Shor, N. Sloane, A group-theoretic framework for the construction of packings in Grassmannian spaces, *J. Algebraic Combin.* 9 (2) (1999) 129–140.
- [19] M. Trosset, Approximate maximin distance designs, in: *Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, 2001, pp. 223–227.
- [20] W. Bajwa, R. Calderbank, D. Mixon, Two are better than one: fundamental parameters of frame coherence, *Appl. Comput. Harmon. Anal.* 33 (1) (2012) 58–78.
- [21] S. Erickson, C. Sabatti, Empirical Bayes estimation of a sparse vector of gene expression, in: *Statistical Applications in Genetics and Molecular Biology*, 2005.
- [22] Z. Zhang, T. Jung, S. Makeig, B. Rao, Compressed sensing for energy-efficient telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning, *IEEE Trans. Biomed. Eng.* 60 (2) (2013) 300–309.
- [23] Z. Zhang, T. Jung, S. Makeig, B. Rao, Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware, arXiv:1206.3493, 2012.
- [24] M. Rossi, A. Haimovich, Y. Eldar, Global methods for compressive sensing in MIMO radar with distributed sensors, in: *Asilomar Conference on Signals, Systems and Computers*, 2011.
- [25] D. Donoho, I. Johnstone, A. Montanari, Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising, Tech. Rep. 2011-10, Stanford University, 2011.
- [26] J. Tropp, I. Dhillon, R. Heath Jr., T. Strohmer, Designing structured tight frames via an alternating projection method, *IEEE Trans. Inform. Theory* 51 (1) (2005) 188–209.
- [27] R.-A. Pitaval, O. Tirkkonen, S. Blostein, Low complexity MIMO precoding codebooks from orthoplex packings, in: *IEEE International Conference on Communications*, Kyoto, Japan, 2011, pp. 1–5.
- [28] A. Calderbank, P. Cameron, W. Kantor, J. Seidel, Z4-kerdock codes, orthogonal spreads and Euclidean line-sets, *Proc. Lond. Math. Soc.* 75 (3) (1997) 436–480.
- [29] P.-S. Absil, A. Edelman, P. Koev, On the largest principal angle between random subspaces, *Linear Algebra Appl.* 414 (1) (2006) 288–294.
- [30] M. Abramowitz, I. Stegun, *A Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables*, 9th edition, Dover, 1964.
- [31] E. Rainville, *Special Functions*, Macmillan, 1960.
- [32] A. Thompson, Quantitative analysis of algorithms for compressed signal recovery, Ph.D. thesis, School of Mathematics, University of Edinburgh, 2012.