

An analysis of blocks sampling strategies in compressed sensing

J eremie Bigot⁽¹⁾, Claire Boyer⁽²⁾ and Pierre Weiss^(3,2)

⁽¹⁾ DMIA, Institut Sup erieur de l'A eronautique et de l'Espace, Toulouse, France
jeremie.bigot@isae.fr

⁽²⁾ Institut de Math ematiques de Toulouse, Universit e de Toulouse, France
claire.boyer@math.univ-toulouse.fr

⁽³⁾ Institut des Technologies Avanc ees du Vivant, Toulouse, France
pierre.armand.weiss@gmail.com

July 21, 2014

Abstract

Compressed sensing is a theory which guarantees the exact recovery of sparse signals from a small number of linear projections. The sampling schemes suggested by current compressed sensing theories are often of little practical relevance since they cannot be implemented on real acquisition systems. In this paper, we study a new random sampling approach that consists in projecting the signal over blocks of sensing vectors. A typical example is the case of blocks made of horizontal lines in the 2D Fourier plane. We provide theoretical results on the number of blocks that are required for exact sparse signal reconstruction. This number depends on two properties named intra and inter-support block coherence. We then show through a series of examples including Gaussian measurements, isolated measurements or blocks in time-frequency bases, that the main result is sharp in the sense that the minimum amount of blocks necessary to reconstruct sparse signals cannot be improved up to a multiplicative logarithmic factor. The proposed results provide a good insight on the possibilities and limits of block compressed sensing in imaging devices such as magnetic resonance imaging, radio-interferometry or ultra-sound imaging.

Key-words: Compressed Sensing, blocks of measurements, MRI, exact recovery, ℓ_1 minimization.

1 Introduction

Compressive Sensing is a new sampling theory that guarantees accurate recovery of signals from a small number of linear projections using three ingredients listed below:

- **Sparsity:** the signals to reconstruct should be sparse, meaning that they can be represented as a linear combination of a small number of atoms in a well-chosen basis. A vector $\mathbf{x} \in \mathbb{C}^n$ is said to be s -sparse if its number of non-zero entries is equal to s .
- **Nonlinear reconstruction:** a key feature ensuring recovery is the use of non linear reconstruction algorithms. For instance, in the seminal papers [Don06, CRT06a], it is suggested to reconstruct \mathbf{x} via the following ℓ_1 -minimization problem:

$$\min_{\mathbf{z} \in \mathbb{C}^n} \|\mathbf{z}\|_1 \quad \text{such that} \quad \mathbf{Az} = \mathbf{y}, \quad (1)$$

where $\mathbf{A} \in \mathbb{C}^{q \times n}$ ($q \leq n$) is a sensing matrix, $\mathbf{y} = \mathbf{Ax} \in \mathbb{C}^q$ represents the measurements vector, and $\|\mathbf{z}\|_1 = \sum_{i=1}^n |z_i|$ for all $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{C}^n$.

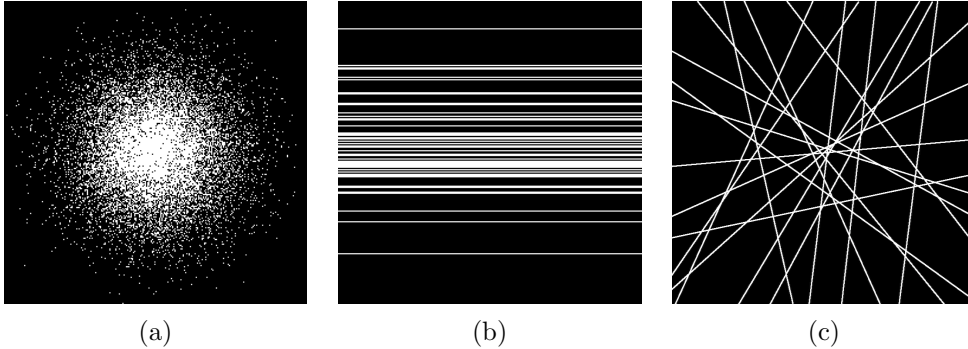


Figure 1: **An example of MRI sampling schemes in the k -space (the 2D Fourier plane where low frequencies are centered)** (a): Isolated measurements drawn from a probability distribution with radial distribution. (b): Sampling scheme in the case of non-overlapping blocks of measurements that correspond to horizontal lines in the 2D Fourier domain. (c): Sampling scheme in the case of overlapping blocks of measurements that correspond to straight lines.

- **Incoherence of the sensing matrix:** the matrix \mathbf{A} should satisfy an incoherence property described later. If \mathbf{A} is perfectly incoherent (e.g. random Gaussian measurements or Fourier coefficients drawn uniformly at random) then it can be shown that only $q = O(s \ln(n))$ measurements are sufficient to perfectly reconstruct the s -sparse vector \mathbf{x} .

The construction of good sensing matrices \mathbf{A} is a keystone for the successful application of compressed sensing. The use of matrices with independent random entries has been popularized in the early papers [CRT06b, Can08]. Such sensing matrices have limited practical interest since they can hardly be stored on computers or implemented on practical systems. More recently, it has been shown that partial random circulant matrices [FR13, PVGW12, RRT12] may be used in the compressed sensing context. With this structure a matrix-vector product can be efficiently implemented on a computer by convolving the signal \mathbf{x} with a random pulse and by subsampling the result. This technique can also be implemented on real systems such as magnetic resonance imaging (MRI) or radio-interferometry [PMG⁺12]. However this demands to modify the acquisition device physics, which is often uneasy and costly. Another way to proceed consists in drawing q sampling locations among n possible ones, see [CRT06a, RV08]. This setting, which is the most widespread in applications, is a promising avenue to implement compressed sensing strategies on nearly all existing devices. Its efficiency depends on the *incoherence* between the acquisition and sparsity bases [DH01, CR07]. It is successfully used in radio interferometry [WJP⁺09], digital holography [MAAOM10] or MRI [LDP07] where the measurements are Fourier coefficients.

To the best of our knowledge, all current compressed sensing theories suggest that the measurements should be drawn independently at random. This is impossible for most acquisition devices which have specific acquisition constraints. A typical example is MRI, where the samples should lie along continuous curves in the Fourier domain (see e.g. [Wri97, LKP08]). As a result, most current implementations of compressed sensing do not comply with theory. For instance, in the seminal work on MRI [LDP07], the authors propose to sample parallel lines of the Fourier domain (see Figure 1 and 2).

Contributions In this paper, we aim at bridging the gap between theory and practice. We consider a sensing matrix \mathbf{A} that is constructed by stacking blocks of measurements and not just isolated measurements. In the proposed formalism, the blocks can be nearly arbitrary random matrices. For instance, our main result covers the case of blocks made of groups of rows of a deterministic sensing matrix (e.g. lines in the Fourier domain) or blocks with random entries (e.g. Gaussian blocks). We study the problem of exact non-uniform sparse recovery in a noise-

free setting. This sampling strategy raises various questions. How many blocks of measurements are needed to ensure exact reconstruction? Is the required number of blocks compatible with faster acquisition?

Our first contribution is to extend the standard compressed sensing theorems to the case of blocks of measurements. We then show that our result is sharp in a few practical examples and extends the best currently known results in compressed sensing. We also prove that in many cases, imposing a block structure has a dramatic effect on the recovery guarantees since it strongly impoverishes the variety of admissible sampling patterns. Overall, we believe that the presented results give a good theoretical basis to the use of block compressed sensing and show the limits to this setting. This work also provides some insight on many currently used sampling patterns in MRI, echography, computed tomography scanners,...

Related work After submitting the first version of this paper, the authors of [PDG14] attracted our attention to the fact that their work dealt with a very similar setting. We therefore make a comparison between the results in Section 4.3.3.

Outline of the paper The remaining of the paper is organized as follows. In Section 2, we first describe the notation and the main assumptions necessary to derive a general theory on blocks of measurements. We present the main result of this paper about blocks-sampling acquisition in Section 3. In Section 4, we discuss the sharpness of our results. First, we show that our approach provides the same guarantees that existing results when using isolated measurements (either Gaussian or randomly extracted from deterministic transforms). We conclude on a pathological example to show sharpness in the case of blocks sampled from separable transforms.

2 Preliminaries

2.1 Notation

Let $S = (S_1, \dots, S_s)$ be a subset of $\{1, \dots, n\}$ of cardinality s . We denote by $P_S \in \mathbb{C}^{n \times s}$ the matrix with columns $(\mathbf{e}_i)_{i \in S}$ where \mathbf{e}_i denotes the i -th vector of the canonical basis of \mathbb{C}^n . For given $\mathbf{M} \in \mathbb{C}^{n \times n}$ and $\mathbf{v} \in \mathbb{C}^n$, we also define $\mathbf{M}_S = \mathbf{M}P_S$, and $\mathbf{v}_S = P_S^* \mathbf{v}$.

2.2 Main assumptions

Recall that we consider the following ℓ_1 -minimization problem:

$$\min_{\mathbf{z} \in \mathbb{C}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{z}, \quad (2)$$

where \mathbf{A} is the sensing matrix, $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{C}^q$ is the measurements vector, $\mathbf{x} \in \mathbb{C}^n$ is the unknown vector to be recovered. In this paper, we assume that the sensing matrix \mathbf{A} can be written as

$$\mathbf{A} = \frac{1}{\sqrt{m}} \begin{pmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_m \end{pmatrix}, \quad (3)$$

where $\mathbf{B}_1, \dots, \mathbf{B}_m$ are i.i.d. copies of a random matrix \mathbf{B} , satisfying

$$\mathbb{E}(\mathbf{B}^* \mathbf{B}) = \text{Id}, \quad (4)$$

where Id is the $n \times n$ identity matrix. This condition is the extension of the isotropy property described in [CP11] in a blocks-constrained acquisition setting.

In most cases studied in this paper, the random matrix \mathbf{B} is assumed to be of fixed size $p \times n$ with $p \in \mathbb{N}^*$. This assumption is however not necessary. The number of blocks of measurements

is denoted m , while the overall number of measurements is denoted q . When \mathbf{B} has a fixed size $p \times n$, $q = mp$.

The following quantities will be shown to play a key role to ensure sparse recovery in the sequel.

Definition 2.1. *Let $S \subset \{1, \dots, n\}$ be a set of cardinality s . We denote by $(\mu_i(S))_{1 \leq i \leq 3}$ the smallest positive reals such that the following bounds hold either deterministically or stochastically (in a sense discussed later)*

$$\begin{aligned} \|\mathbf{B}_S^* \mathbf{B}_S\|_{2 \rightarrow 2} &\leq \mu_1(S), & \sqrt{s} \max_{i \in S^c} \|\mathbf{B}_S^* \mathbf{B} \mathbf{e}_i\|_2 &\leq \mu_2(S), \\ s \max_{i \in S^c} \|\mathbb{E}[\mathbf{B}_S^* (\mathbf{B} \mathbf{e}_i) (\mathbf{B} \mathbf{e}_i)^* \mathbf{B}_S]\|_{2 \rightarrow 2} &\leq \mu_3(S). \end{aligned} \quad (5)$$

Define

$$\gamma(S) := \max_{1 \leq i \leq 3} \mu_i(S).$$

The quantities introduced in Definition 2.1 can be interpreted as follows. The number $\mu_1(S)$ can be seen as an *intra-support* block coherence, whereas $\mu_2(S)$ and $\mu_3(S)$ are related to the *inter-support* block coherence, that is the coherence between blocks restricted to the support of the signal and blocks restricted to the complementary of this support. Note that the factors \sqrt{s} and s involved in the definition of $\mu_2(S)$ and $\mu_3(S)$ ensure homogeneity between all of these quantities.

2.3 Application examples

The number of applications of the proposed setting is very large. For instance, it encompasses those proposed in [CP11]. Let us provide a few examples of new applications below.

2.3.1 Partition of orthogonal transforms

Let $\mathbf{A}_0 \in \mathbb{C}^{n \times n}$ denote an orthogonal transform. Blocks can be constructed by partitioning the rows $(\mathbf{a}_i^*)_{1 \leq i \leq n}$ from \mathbf{A}_0 :

$$\mathbf{B}_j = (\mathbf{a}_i^*)_{i \in I_j} \quad \text{for } I_j \subset \{1, \dots, n\} \quad \text{s.t.} \quad \bigsqcup_{j=1}^M I_j = \{1, \dots, n\},$$

where \bigsqcup stands for the disjoint union. This case is the one studied in [PDG14].

Let $\Pi = (\pi_1, \dots, \pi_M)$ be a discrete probability distribution on the set of integers $\{1, \dots, M\}$. A random sensing matrix \mathbf{A} can be constructed by stacking m i.i.d. copies of the random matrix \mathbf{B} defined by $\mathbb{P}(\mathbf{B} = \mathbf{B}_k / \sqrt{\pi_k}) = \pi_k$ for all $k \in \{1, \dots, M\}$. Note that the normalization by $1/\sqrt{\pi_k}$ ensures that the isotropy condition $\mathbb{E}[\mathbf{B}^* \mathbf{B}] = \text{Id}_n$ is verified.

2.3.2 Overlapping blocks issued from orthogonal transforms

In the last example, we concentrated on partitions, i.e. non-overlapping blocks of measurements. The case of overlapping blocks can also be handled. To do so, define the blocks $(\mathbf{B}_j)_{1 \leq j \leq M}$ as

follows: $\mathbf{B}_j = \left(\frac{1}{\sqrt{\alpha_i}} \mathbf{a}_i^* \right)_{i \in I_j}$, where $\bigcup_{j=1}^M I_j = \{1, \dots, n\}$, and α_i denotes the multiplicity of the

row \mathbf{a}_i^* , i.e. the number of appearances $\alpha_i = |\{j, i \in I_j\}|$ of this row in different blocks. This renormalization is sufficient to ensure $\mathbb{E}[\mathbf{B}^* \mathbf{B}] = \text{Id}_n$ where \mathbf{B}_k is defined similarly to the previous example. See Appendix D for an illustration of this setting in the case of 2D Fourier measurements.

2.3.3 Blocks issued from tight or continuous frames

Until now, we have concentrated on projections over a fixed set of n vectors. This is not necessary and the projection set can be redundant and even infinite. A typical example is the Fourier transform with a continuous frequency spectrum. This example is discussed in more details in [FR13, CP11].

2.3.4 Random blocks

In the previous examples, the blocks were predefined and extracted from deterministic matrices or systems. The proposed theory also applies to random blocks. For instance, one could consider blocks with i.i.d. Gaussian entries since these blocks satisfy the isotropy condition (4). This example is of little practical relevance since stacking random Gaussian matrices produces a random Gaussian matrix that can be analyzed with standard compressed sensing approaches. It however presents a theoretical interest in order to show the sharpness of our main result. Another example with potential interest is that of blocks generated randomly using random walks over the acquisition space [CCKW14].

3 Main result

Our main result reads as follows.

Theorem 3.1. *Let $S \subset \{1, \dots, n\}$ be a set of indices of cardinality s and suppose that $\mathbf{x} \in \mathbb{C}^n$ is an s -sparse vector supported on S . Fix $\varepsilon \in (0, 1)$. Suppose that the sampling matrix \mathbf{A} is constructed as in (3), and that the isotropy condition (4) holds. Suppose that the bounds (5) hold deterministically. If the number of blocks m satisfies the following inequality*

$$m \geq c\gamma(S) \log(4n) \log(12\varepsilon^{-1}),$$

then \mathbf{x} is the unique solution of (2) with probability at least $1 - \varepsilon$. The constant c can be taken equal to 3×534 .

The proof of Theorem 3.1 is detailed in Section C.1. It is based on the so-called golfing scheme introduced in [Gro11] for matrix completion, and adapted by [CP11] for compressed sensing from isolated measurements. Note that Theorem 3.1 is a non uniform result in the sense that reconstruction holds for a given support S and not for all s -sparse signals. It is likely that uniform results could be derived by using the so-called Restricted Isometry Property. However, this strong property is usually harder to prove and leads to narrower classes of admissible matrices and to larger number of required measurements.

Remark 3.2 (The case of stochastic bounds). *In Definition 2.1, we say that the bounds deterministically hold if the inequalities (5) are satisfied almost surely. This assumption is convenient to simplify the proof of Theorem 3.1. Obviously, it is not satisfied in the setting where the entries of \mathbf{B} are i.i.d. Gaussian variables. To encompass such cases, the bounds in Definition 2.1 could stochastically hold, meaning that the inequalities (5) are satisfied with large probability. The proof of the main result can be modified by conditioning the deviation inequalities in the Lemmas of Appendix C.1 to the event that the bounds in Definition 2.1 hold. Therefore, even though we do not provide a detailed proof, the lower bound on the required number of blocks in Theorem 3.1 remains accurate. Hence, we will propose in Section 4.2 some estimates of the quantities (5) in the case of Gaussian measurements.*

The lower bound on the number m of blocks of measurements in Theorem 3.1 depends on $\gamma(S)$ and thus on the support S of the vector \mathbf{x} to reconstruct. In the usual compressed sensing framework, the matrix \mathbf{A} is constructed by stacking realizations of a random vector \mathbf{a} . The best known results state that $O(s\mu \log(n))$ isolated measurements are sufficient to reconstruct \mathbf{x} with

high probability. The coherence μ is the smallest number such that $\|\mathbf{a}\|_\infty^2 \leq \mu$. The quantity $\gamma(S)$ in Theorem 3.1 therefore replaces the standard factor $s\mu$. The coherence μ is usually much simpler to evaluate than $\gamma(S)$ which depends on three properties of the random matrix \mathbf{B} : the intra-support coherence μ_1 and the inter-support coherences μ_2 and μ_3 . As will be seen in Section 4, it is important to keep all those quantities in order to obtain tight reconstruction results. Nevertheless, a rough upper bound of $\gamma(S)$, reminiscent of the coherence, can be used as shown in Proposition 3.3.

Proposition 3.3. *Let S be a subset of $\{1, \dots, n\}$ of cardinality s . Assume that the following inequality holds either deterministically or stochastically*

$$\|\mathbf{B}^* \mathbf{B}\|_{1 \rightarrow \infty} \leq \mu_4$$

with $\|\mathbf{B}^* \mathbf{B}\|_{1 \rightarrow \infty} = \sup_{\|\mathbf{v}\|_1 \leq 1} \|\mathbf{B}^* \mathbf{B} \mathbf{v}\|_\infty$. Then

$$\gamma(S) \leq s\mu_4. \quad (6)$$

The proof of Proposition 3.3 is given in Appendix C.2. The bound given in Proposition 3.3 is an upper bound on $\gamma(S)$ that should not be considered as optimal. For instance, for Gaussian measurements, it is important to precisely evaluate the three quantities $(\mu_i(S))_{1 \leq i \leq 3}$.

Remark 3.4 (Noisy setting). *In this paper, we concentrate on a noiseless setting. It is likely that noise can be accounted for mimicking the proofs in [CP11] for instance.*

4 Sharpness of the main result

In this section, we discuss the sharpness of the lower bound given by Theorem 3.1 by comparing it to the best known results in compressed sensing.

4.1 The case of isolated measurements

First, let us show that our result matches the standard setting where the blocks are made of only one row, that is $p = 1$. This is the standard compressed sensing framework considered e.g. by [CRT06a, FR13, CP11]. Consider that $\mathbf{A}_0 = (\mathbf{a}_i^*)_{1 \leq i \leq n}$ is a deterministic matrix, and that the sensing matrix \mathbf{A} is constructed by drawing m rows of \mathbf{A}_0 according to some probability distribution $\mathcal{P} = (p_1, \dots, p_n)$, i.e. one can write \mathbf{A} as follows:

$$\mathbf{A} = \begin{pmatrix} \frac{\mathbf{a}_{J_1}^*}{\sqrt{p_{J_1}}} \\ \vdots \\ \frac{\mathbf{a}_{J_m}^*}{\sqrt{p_{J_m}}} \end{pmatrix},$$

where the $(J_j)_{1 \leq j \leq m}$'s are i.i.d. random variables taking their value in $\{1, \dots, n\}$ with probability \mathcal{P} . According to Proposition 3.3, for a support S of cardinality s the following upper bound holds:

$$\gamma(S) \leq s \max_{1 \leq j \leq M} \frac{\|\mathbf{a}_j \mathbf{a}_j^*\|_{1 \rightarrow \infty}}{p_j}.$$

Therefore, according to Theorem 3.1, it is sufficient that

$$q \geq cs \max_{1 \leq j \leq M} \frac{\|\mathbf{a}_j \mathbf{a}_j^*\|_{1 \rightarrow \infty}}{p_j} \log(4n) \log(12\varepsilon^{-1}). \quad (7)$$

to obtain perfect reconstruction with probability $1 - \varepsilon$. Noting that $\|\mathbf{a}_j\|_\infty^2 = \|\mathbf{a}_j \mathbf{a}_j^*\|_{1 \rightarrow \infty}$, for all $j \in \{1, \dots, n\}$, it follows that Condition (7) is the same (up to a multiplicative constant) to that of [CP11].

In addition, choosing \mathcal{P}^* in order to minimize the right-hand side of (7) leads to

$$p_j^* = \frac{\|\mathbf{a}_j \mathbf{a}_j^*\|_{1 \rightarrow \infty}}{\sum_{k=1}^n \|\mathbf{a}_k \mathbf{a}_k^*\|_{1 \rightarrow \infty}}, \quad \forall k \in \{1, \dots, n\},$$

which in turn leads to the following required number of measurements:

$$q \geq cs \sum_{k=1}^n \|\mathbf{a}_k^*\|_{\infty}^2 \log(4n) \log(12\varepsilon^{-1}). \quad (8)$$

Contrarily to common belief, the probability distribution minimizing the required number of measurements is not the uniform one, but the one depending on the ℓ_{∞} -norm of the considered row. Let us highlight this fact. Consider that $\mathbf{A}_0 = \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{F}_{n-1} \end{pmatrix}$, where \mathcal{F}_{n-1} denotes the 1D Fourier matrix of size $(n-1) \times (n-1)$. If a uniform drawing distribution is chosen, the right hand side of (7) is $O(sn \ln^2(n))$. This shows that uniform random sampling is not interesting for this sensing matrix. Note that the coherence $\|\mathbf{A}_0\|_{1 \rightarrow \infty}^2$ of \mathbf{A}_0 is equal to 1, which is the worst possible case for orthogonal matrices. Nevertheless, if the optimal drawing distribution is chosen, i.e.

$$p_j^* = \begin{cases} \frac{1}{2} & \text{if } j = 1 \\ \frac{1}{2(n-1)} & \text{otherwise} \end{cases}$$

then, the right hand side of (7) becomes $O(2s \ln^2(n))$. Using this sampling strategy, compressed sensing therefore remains relevant. Furthermore, note that the latter bound could be easily reduced by a factor 2 by systematically sampling the location associated to the first row of \mathbf{A}_0 , and uniformly picking the $q-1$ remaining isolated measurements. Similar remarks were formulated in [KW14] which promote non-uniform sampling strategies in compressed sensing.

4.2 The case of Gaussian measurements

We suppose that the entries of $\mathbf{B} \in \mathbb{R}^{p \times n}$ are i.i.d. Gaussian random variables with zero-mean and variance $1/p$. This assumption on the variance ensures that the isotropy condition (4) is satisfied. The matrix \mathbf{A} constructed by concatenating such blocks is also a Gaussian random matrix with i.i.d. entries and does not differ from an acquisition setting based on isolated measurements. Therefore, if Theorem 3.1 is sharp, one can expect that $q = O(s \log(n))$ measurements are enough to perfectly reconstruct \mathbf{x} . In what follows, we show that this is indeed the case.

Proposition 4.1. *Assume that the entries of $\mathbf{B} \in \mathbb{R}^{p \times n}$ are i.i.d. Gaussian random variables with zero-mean and variance $1/p$. Then, $\gamma(S) = O\left(\frac{s \log(s)}{p}\right)$. Therefore, $O\left(\frac{s \log(s) \log(n)}{p}\right)$ Gaussian blocks are sufficient to ensure perfect reconstruction with high probability.*

This is similar to an acquisition based on isolated Gaussian measurements and this is optimal up to a logarithmic factor, see [Don06].

4.3 The case of separable transforms

In this section, we consider d -dimensional deterministic transforms obtained as Kronecker products of orthogonal one-dimensional transforms. This setting is widespread in applications. Indeed, separable transforms include d -dimensional Fourier transforms met in astronomy [BSO08] or products of Fourier and wavelet transforms met in MRI [LDSP08] or radio-interferometry [WJP⁺09]. A specific scenario encountered in many settings is that of blocks made of lines in the acquisition space. For instance, parallel lines in the 3D Fourier space are used in [LDP07]. The authors propose to undersample the 2D k_x - k_y plane and sample continuously along the orthogonal direction k_z (see Figure 2).

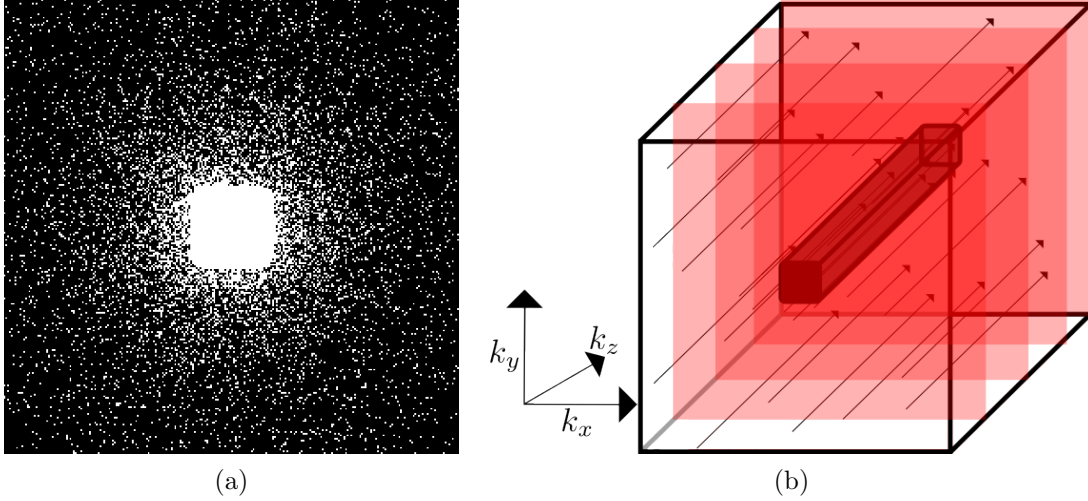


Figure 2: Example of sampling pattern used in MRI [LDP07]. (a) Visualization in the k_x - k_y plane. (b) Visualization in 3D.

The remaining of this Section is as follows. We first introduce the notation. We then provide theoretical results about the minimal amount of blocks necessary to reconstruct all s -sparse vectors. Next, we show that Theorem 3.1 is sharp in this setting since the amount of blocks required to reconstruct s -sparse vectors coincides with the minimal amount. Finally, we perform a comparison with the results in [PDG14].

4.3.1 Preliminaries

Let $\Psi \in \mathbb{C}^{\sqrt{n} \times \sqrt{n}}$ denote an arbitrary orthogonal transform, with $\sqrt{n} \in \mathbb{N}$. Let

$$\mathbf{A}_0 = \Psi \otimes \Psi = \begin{bmatrix} \Psi_{1,1}\Psi & \cdots & \Psi_{1,\sqrt{n}}\Psi \\ \Psi_{2,1}\Psi & \cdots & \Psi_{2,\sqrt{n}}\Psi \\ \vdots & \ddots & \vdots \\ \Psi_{\sqrt{n},1}\Psi & \cdots & \Psi_{\sqrt{n},\sqrt{n}}\Psi \end{bmatrix} \in \mathbb{C}^{n \times n},$$

where \otimes denote the Kronecker product. Note that \mathbf{A}_0 is also orthogonal. We define groups of measurements from \mathbf{A}_0 as follows:

$$\mathbf{B}_k = \Psi_{k,:} \otimes \Psi \quad (9)$$

$$= [\Psi_{k,1}\Psi, \dots, \Psi_{k,\sqrt{n}}\Psi] \in \mathbb{C}^{\sqrt{n} \times n}. \quad (10)$$

For instance, if Ψ is the 1D discrete Fourier transform, this strategy consists in constructing \sqrt{n} blocks as horizontal discrete lines of the discrete Fourier plane. This is similar to the blocks used in [LDP07]. Similarly to paragraph 2.3.1, a sensing matrix \mathbf{A} can be constructed by drawing m i.i.d. blocks with distribution Π . Letting $K = (k_1, \dots, k_m) \in \{1, \dots, \sqrt{n}\}^m$ denote the drawn blocks indexes, \mathbf{A} reads:

$$\mathbf{A} = \frac{1}{\sqrt{m}} \begin{bmatrix} \frac{\mathbf{B}_{k_1}}{\sqrt{\pi_{k_1}}} \\ \vdots \\ \frac{\mathbf{B}_{k_m}}{\sqrt{\pi_{k_m}}} \end{bmatrix} \quad (11)$$

$$= \left(\frac{\mathbf{D}(\pi)^{-1/2}}{\sqrt{m}} \cdot \Psi_{K,:} \right) \otimes \Psi$$

$$= \tilde{\Psi}_{K,:} \otimes \Psi \quad (12)$$

where $\mathbf{D}(\pi) := \text{diag}(\pi_{k_1}, \dots, \pi_{k_m})$ and $\tilde{\Psi}_{K,:} := \frac{\mathbf{D}(\pi)^{-1/2}}{\sqrt{m}} \cdot \Psi_{K,:}$. By combining the results in Theorem 3.1 and Proposition 3.3, we easily get the following reconstruction guarantees.

Proposition 4.2. *Let $S \subset \{1, \dots, n\}$ be the support of cardinality s of the signal $\mathbf{x} \in \mathbb{C}^n$ to reconstruct. Under the above hypotheses, if*

$$m \geq cs \max_{1 \leq j \leq M} \frac{\|\mathbf{B}_j^* \mathbf{B}_j\|_{1 \rightarrow \infty}}{\pi_j} \log(4n) \log(12\varepsilon^{-1}), \quad (13)$$

then the vector \mathbf{x} is the unique solution of (2) with probability at least $1 - \varepsilon$.

Using the above result we also obtain the following Corollary.

Corollary 4.3. *The drawing probability distribution Π^* minimizing the right hand side of Inequality (13) on the required number of measurements is defined by*

$$\pi_j^* = \frac{\|\mathbf{B}_j^* \mathbf{B}_j\|_{1 \rightarrow \infty}}{\sum_{k=1}^M \|\mathbf{B}_k^* \mathbf{B}_k\|_{1 \rightarrow \infty}}, \quad \forall j \in \{1, \dots, M\}. \quad (14)$$

For this particular choice of Π^* , the right hand side of Inequality (13) can be written as follows

$$m \geq cs \sum_{j=1}^M \|\mathbf{B}_j^* \mathbf{B}_j\|_{1 \rightarrow \infty} \log(4n) \log(12\varepsilon^{-1}). \quad (15)$$

The sharpness of the bounds on the required number of measurements in Proposition 4.3 will be discussed in the following paragraph.

4.3.2 The limits of separable transforms

Considering a 2D discrete Fourier transform and a dictionary of blocks made of horizontal lines in the discrete Fourier domain, one could hope to only require $m = O(s/p \log(n))$ blocks of measurements to perfectly recover all s -sparse vectors. Indeed, it is known since [CRT06a] that $O(s \log(n))$ isolated measurements uniformly drawn at random are sufficient to achieve this. In this paragraph, we show that this expectation cannot be satisfied since at least $2s$ blocks are necessary to reconstruct all s -sparse vectors. It means that this specific block structure is inadequate to obtain strong reconstruction guarantees. This result also shows that Theorem 4.3 is nearly optimal.

In order to prove those results, we first recall the following useful lemma. We define a decoder as any mapping $\Delta : \mathbb{C}^q \rightarrow \mathbb{C}^n$. Note that Δ is not necessarily a linear mapping.

Lemma 4.4. [CDD09, Lemma 3.1] *Set Σ_s to be the set of s -sparse vectors in \mathbb{C}^n . If \mathbf{A} is any $m \times n$ matrix, then the following propositions are equivalent:*

- (i) *There is a decoder Δ such that $\Delta(\mathbf{A}\mathbf{x}) = \mathbf{x}$, for all s -sparse \mathbf{x} in \mathbb{C}^n .*
- (ii) $\Sigma_{2s} \cap \text{Ker } \mathbf{A} = \{0\}$.
- (iii) *For any set $T \subset \{1, \dots, n\}$ of cardinality $2s$, the matrix \mathbf{A}_T has rank $2s$.*

Looking at (iii) of Lemma 4.4, since the rank of \mathbf{A}_T is smaller than $\min(2s, m)$, we deduce that $m \geq 2s$ is a necessary condition to have a decoder. Therefore, if the number of isolated measurements is less than $2s$ with s the degree of sparsity of \mathbf{x} , we cannot reconstruct \mathbf{x} . This property is an important step to prove Proposition 4.5.

Proposition 4.5. *Assume that the sensing matrix \mathbf{A} has the special block structure described in (11). If $m < \min(2s, \sqrt{n})$, then there exists no decoder Δ such that $\Delta(\mathbf{A}\mathbf{x}) = \mathbf{x}$ for all s -sparse vector $\mathbf{x} \in \mathbb{C}^n$. In other words, the minimal number m of distinct blocks required to identify every s -sparse vectors is necessarily larger than $\min(2s, \sqrt{n})$.*

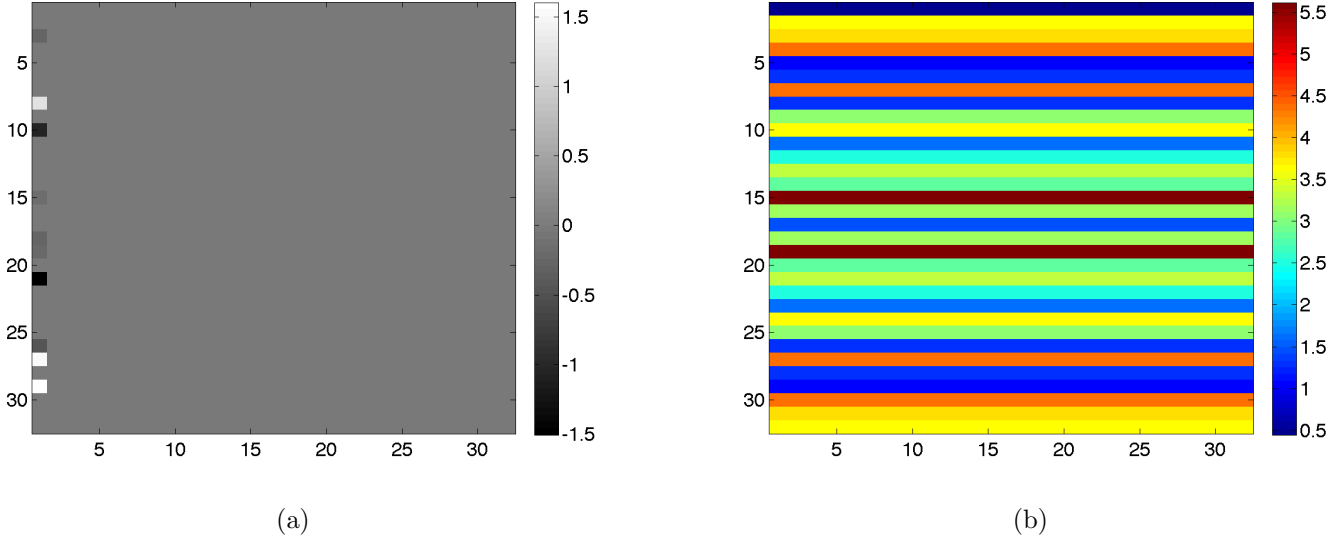


Figure 3: A pathological case where $n = 32 \times 32$ (a): The signal is s -sparse for $s = 10$ and its support is concentrated on its first column. (b) Its 2D Fourier transform is constant along horizontal lines in the Fourier plane.

Proposition 4.5 shows that there is no hope to reconstruct all s -sparse vectors with less than $m = O(s)$ blocks of measurements, using sensing matrices \mathbf{A} made of blocks such as (9). Moreover, since the blocks are of length $p = \sqrt{n}$, it follows that whenever $s \geq \frac{\sqrt{n}}{2}$, the full matrix \mathbf{A}_0 should be used to identify every s -sparse \mathbf{x} . Let us illustrate this result on a practical example. Set \mathbf{A}_0 to be the 2D Fourier matrix, i.e. the Kronecker product of two 1D Fourier matrices. Consider that the dictionary of blocks is made of horizontal lines. Now consider a vector $\mathbf{x} \in \mathbb{R}^{32 \times 32}$ to be 10-sparse in the spatial domain and only supported on the first column as illustrated in Figure 3(a). Due to this specific signal structure, the Fourier coefficients of \mathbf{x} are constant along horizontal lines, see Figure 3(b). Therefore, for this type of signal, the information captured by a block of measurements (i.e. a horizontal line) is as informative as one isolated measurement. Clearly, at least $O(s)$ blocks are therefore required to reconstruct all s -sparse vectors supported on a vertical line of the 2D Fourier plane. Using Corollary 4.3, one can derive the following result.

Proposition 4.6. *Let $\mathbf{A}_0 \in \mathbb{C}^{n \times n}$ denote the 2D discrete Fourier matrix and consider a partition in $M = \sqrt{n}$ blocks that consist of lines in the 2D Fourier domain. Assume that $\mathbf{x} \in \mathbb{C}^n$ is s -sparse. The drawing probability minimizing the right hand side of (13) is given by*

$$\pi_j^* = \frac{1}{\sqrt{n}}, \quad \forall j \in \{1, \dots, \sqrt{n}\}$$

and for this particular choice, the number m of blocks of measurements sufficient to reconstruct \mathbf{x} with probability $1 - \varepsilon$ is

$$m \geq cs \log(4n) \log(12\varepsilon^{-1}).$$

This result is disappointing but optimal up to a logarithmic factor, due to Proposition 4.5. We refer to Appendix C.5 for the proof. This Proposition indicates that $O(s \log(n))$ blocks are sufficient to reconstruct \mathbf{x} which is similar to the minimal number given in Proposition 4.5 up to a logarithmic factor.

4.3.3 Relation to previous work

To the best of our knowledge, the only existing compressed sensing results based on blocks of measurements appeared in [PDG14]. In this paragraph, we outline the differences between both approaches.

First, in our work, no assumption on the sign pattern of the non-zero signal entries is required. Furthermore, while the result in [PDG14] only covers the case described in Paragraph 2.3.1 (i.e. partitions of orthogonal transforms), our work covers the case of overlapping blocks of measurements (see Paragraph 2.3.2), subsampled tight or continuous frames (see Paragraph 2.3.3), and it can also be extended to the case of randomly generated blocks (see Paragraph 2.3.4). Last but not least, the work [PDG14] only deals with *uniform* sampling densities which is well known to be of little interest when dealing with partially coherent matrices (see e.g. end of Paragraph 4.1 for an edifying example).

Apart from those contextual differences, the comparison between the results in [PDG14] and the ones in this paper is not straightforward. The criterion in [PDG14] that controls the overall number of measurements q depends on the following quantity:

$$\Upsilon(\mathbf{A}_0, S, \mathbf{B}) := \|\overline{\mathbf{B}}_S\|_{2 \rightarrow 1},$$

where $\overline{\mathbf{B}}_S$ stands for the block restricted to the columns in S with renormalized rows. The total number of measurements required in the approach [PDG14] is

$$q_{PDG} \geq C\Upsilon(\mathbf{A}_0, S, \mathbf{B}) \max_{i,j} |\mathbf{A}_0(i, j)|^3 n^{3/2} \log(n) \quad (16)$$

which should be compared to our result

$$q \geq cp\gamma(S) \log(4n) \log(12\varepsilon^{-1}). \quad (17)$$

As shown in the previous paragraphs, the number (17) is sharp in various settings of interest, while (16) is usually hard to explicitly compute or too large in the case of partially incoherent transforms. It therefore seems that our results should be preferred over those of [PDG14].

5 Outlook

We have introduced new sensing matrices that are constructed by stacking random blocks of measurements. Such matrices play an important role in applications since they can be implemented easily on many imaging devices. We have derived theorems that guarantee exact reconstruction using these matrices via ℓ_1 -minimization algorithms and outlined the crucial role of two properties: the *extra* and *intra* support block-coherences introduced in Definition 2.1. We have showed that our main result (Theorem 3.1) is sharp in a few settings of practical interest, suggesting that it cannot be improved in the general case up to logarithmic factors.

Apart from those positive results, this work also reveals some limits of block sampling approaches. First, it seems hard to evaluate the *extra* and *intra* support block-coherences - except in a few particular cases - both analytically and numerically. This evaluation is however central to derive optimal sampling approaches. More importantly, we have showed in Paragraph 4.3.2 that not much could be expected from this approach in the specific setting where separable transforms and blocks consisting of lines of the acquisition space are used. Despite the peculiarity of such a dictionary, we believe that this result might be an indicator of a more general weakness of block sampling approaches. Since the best known compressed sensing strategies heavily rely on randomness (e.g. Gaussian measurements or uniform drawings of Fourier atoms), one may wonder whether the more rigid sampling patterns generated by block sampling approaches have a chance to provide decent results. It is therefore legitimate to ask the following question: is it reasonable to use variable density sampling with pre-defined blocks of measurements in compressed sensing?

Numerical experiments indicate that the answer to this question is positive. For instance, it is readily seen in Figure 4 (a,b,c) and (j,k,l), that block sampling strategies can produce comparable results to acquisitions based on isolated measurements. The first potential explanation to this phenomenon is that $\gamma(S)$ is low for the dictionaries chosen in those experiments. However, even acquisitions based on horizontal lines in the Fourier domain (see Figure 4 (d,e,f)) produce rather good reconstruction results while Proposition 4.6 seems to indicate that this strategy is doomed.

This last observation suggests that a key feature is missing in our study to fully understand the potential of block sampling in applications. Recent papers [AHP13, AHR14] highlight the central role of *structured sparsity* to explain the practical success of compressed sensing. A very promising perspective is therefore to couple the ideas of structured sparsity in [AHP13, AHR14] and the ideas of block sampling proposed in this paper to finely understand the results in Figure 4 and perhaps design new optimal and applicable sampling strategies.

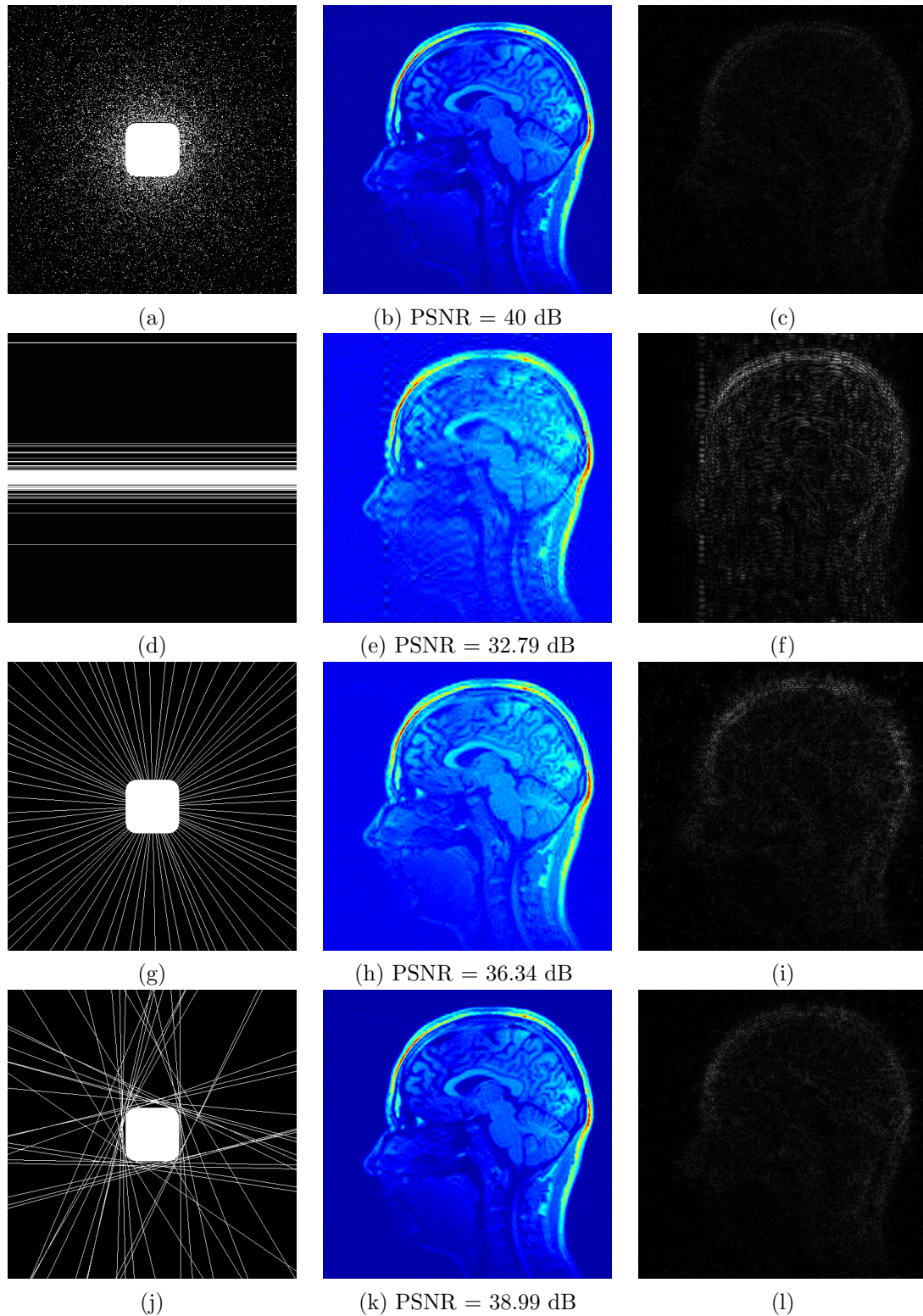


Figure 4: Reconstruction results using different sampling strategies. Each sampling pattern contains 10% of the total number of possible measurements. From top to bottom: measurements drawn independently at random with a radial distribution - horizontal lines in the Fourier domain - deterministic radial sampling - heuristic method proposed in [BWB14]. From left to right: sampling scheme - corresponding reconstruction - difference with the reference (the same colormap is used in every experiment).

A Bernstein's inequalities

Theorem A.1 (Scalar Bernstein Inequality). *Let x_1, \dots, x_m be independent random variables such that $|x_\ell| \leq K$ almost surely for every $\ell \in \{1, \dots, m\}$. Assume that $\mathbb{E}|x_\ell|^2 \leq \sigma_\ell^2$ for $\ell \in \{1, \dots, m\}$. Then for all $t > 0$,*

$$\mathbb{P}\left(\left|\sum_{\ell=1}^m x_\ell\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right),$$

with $\sigma^2 \geq \sum_{\ell=1}^m \sigma_\ell^2$.

Theorem A.2 (Rectangular Matrix Bernstein Inequality). *[Tro12, Theorem 1.6]*

Let $(\mathbf{Z}_k)_{1 \leq k \leq m}$ be a finite sequence of rectangular independent random matrices of dimension $d_1 \times d_2$. Suppose that \mathbf{Z}_k is such that $\mathbb{E}\mathbf{Z}_k = 0$ and $\|\mathbf{Z}_k\|_{2 \rightarrow 2} \leq K$ a.s. for some constant $K > 0$ that is independent of k . Define

$$\sigma^2 \geq \max\left(\left\|\sum_{k=1}^m \mathbb{E}\mathbf{Z}_k \mathbf{Z}_k^*\right\|_{2 \rightarrow 2}, \left\|\sum_{k=1}^m \mathbb{E}\mathbf{Z}_k^* \mathbf{Z}_k\right\|_{2 \rightarrow 2}\right).$$

Then, for any $t > 0$, we have that

$$\mathbb{P}\left(\left\|\sum_{k=1}^m \mathbf{Z}_k\right\|_{2 \rightarrow 2} \geq t\right) \leq (d_1 + d_2) \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right)$$

Theorem A.3 (Vector Bernstein Inequality (V1)). *[CP11, Theorem 2.6] Let $(\mathbf{y}_k)_{1 \leq k \leq m}$ be a finite sequence of independent and identically distributed random vectors of dimension n . Suppose that $\mathbb{E}\mathbf{y}_1 = 0$ and $\|\mathbf{y}_1\|_2 \leq K$ a.s. for some constant $K > 0$ and set $\sigma^2 \geq \sum_k \mathbb{E}\|\mathbf{y}_k\|_2^2$. Let $Z = \|\sum_{k=1}^m \mathbf{y}_k\|_2$. Then, for any $0 < t \leq \sigma^2/K$, we have that*

$$\mathbb{P}(Z \geq t) \leq \exp\left(-\frac{(t/\sigma - 1)^2}{4}\right) \leq \exp\left(-\frac{t^2}{8\sigma^2} + \frac{1}{4}\right),$$

where $\mathbb{E}Z^2 = \sum_{k=1}^m \mathbb{E}\|\mathbf{y}_k\|_2^2 = m\mathbb{E}\|\mathbf{y}_1\|_2^2$.

Theorem A.4 (Vector Bernstein Inequality (V2)). *[FR13, Corollary 8.44] Let $(\mathbf{y}_k)_{1 \leq k \leq m}$ be a finite sequence of independent and identically distributed random vectors of dimension n . Suppose that $\mathbb{E}\mathbf{y}_1 = 0$ and $\|\mathbf{y}_1\|_2 \leq K$ a.s. for some constant $K > 0$. Let $Z = \|\sum_{k=1}^m \mathbf{y}_k\|_2$. Then, for any $t > 0$, we have that*

$$\mathbb{P}\left(Z \geq \sqrt{\mathbb{E}Z^2} + t\right) \leq \exp\left(-\frac{t^2/2}{\mathbb{E}Z^2 + 2K\sqrt{\mathbb{E}Z^2} + Kt/3}\right),$$

where $\mathbb{E}Z^2 = \sum_{k=1}^m \mathbb{E}\|\mathbf{y}_k\|_2^2 = m\mathbb{E}\|\mathbf{y}_1\|_2^2$. Note that the previous inequality still holds by replacing $\mathbb{E}Z^2$ by σ^2 where $\sigma^2 \geq \mathbb{E}Z^2$.

B Estimates: auxiliary results

Let S be the support of the signal to be reconstructed such that $|S| = s$. Note that the isotropy condition (4) ensures that the following properties hold

- (i) $\mathbb{E}(\mathbf{B}^* \mathbf{B}) = \text{Id}_n$ and $\mathbb{E}(\mathbf{B}_S^* \mathbf{B}_S) = \text{Id}_s$.
- (ii) for any vector $\mathbf{w} \in \mathbb{C}^s$, $\mathbb{E}[\mathbf{B}_S \mathbf{w}]^2 = \|\mathbf{w}\|_2^2$.

(iii) for any $i \in S^c$, $\mathbb{E}(\mathbf{B}_S^* \mathbf{B} e_i) = 0$.

The above properties will be repeatedly used in the proof of the following lemmas.

Lemma B.1. *Let $S \subset \{1, \dots, n\}$ be of cardinality of s . Then, for any $\delta > 0$, one has that*

$$\mathbb{P}(\|\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s\|_{2 \rightarrow 2} \geq \delta) \leq 2s \exp\left(-\frac{m\delta^2/2}{\mu_1(S) + \max(\mu_1(S) - 1, 1)\delta/3}\right). \quad (\text{E1})$$

Proof. We decompose the matrix $\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s$ as

$$\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s = \frac{1}{m} \sum_{k=1}^m (\mathbf{B}_{k,S}^* \mathbf{B}_{k,S} - \text{Id}_s) = \frac{1}{m} \sum_{k=1}^m X_k,$$

where $X_k := (\mathbf{B}_{k,S}^* \mathbf{B}_{k,S} - \text{Id}_s)$. It is clear that $\mathbb{E}X_k = 0$, and since $\|\mathbf{B}_{k,S}^* \mathbf{B}_{k,S}\|_{2 \rightarrow 2} \leq \mu_1(S)$, we have that

$$\|X_k\|_{2 \rightarrow 2} = \max\left(\|\mathbf{B}_{k,S}^* \mathbf{B}_{k,S}\|_{2 \rightarrow 2} - 1, 1\right) \leq \max(\mu_1(S) - 1, 1).$$

Lastly, we remark that

$$0 \preceq \mathbb{E}X_k^2 = \mathbb{E}[\mathbf{B}_{k,S}^* \mathbf{B}_{k,S}]^2 - \text{Id}_s \preceq \mathbb{E}\|\mathbf{B}_{k,S}^* \mathbf{B}_{k,S}\|_2 \mathbf{B}_{k,S}^* \mathbf{B}_{k,S} \preceq \mu_1(S) \text{Id}_s.$$

Therefore, $\sum_{k=1}^m \mathbb{E}X_k^2 \preceq m\mu_1(S) \text{Id}_s$ which implies that $\|\sum_{k=1}^m \mathbb{E}X_k^2\|_2 \leq m\mu_1(S)$. Hence, inequality (E1) follows immediately from Bernstein's inequality for random matrices (see Theorem A.2). \blacksquare

Lemma B.2. *Let $S \subset \{1, \dots, n\}$, such that $|S| = s$. Let \mathbf{w} be a vector in \mathbb{C}^s . Then, for any $t > 0$, one has that*

$$\begin{aligned} \mathbb{P}\left(\|(\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s) \mathbf{w}\|_2 \geq \left(\sqrt{\frac{\mu_1(S) - 1}{m}} + t\right) \|\mathbf{w}\|_2\right) \\ \leq \exp\left(-\frac{mt^2/2}{(\mu_1(S) - 1) + 2\sqrt{\frac{\mu_1(S) - 1}{m}} \mu_1(S) + \mu_1(S)t/3}\right). \end{aligned} \quad (\text{E2})$$

Proof. Without loss of generality we may assume that $\|\mathbf{w}\|_2 = 1$. We remark that

$$(\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s) \mathbf{w} = \frac{1}{m} \sum_{k=1}^m (\mathbf{B}_{k,S}^* \mathbf{B}_{k,S} - \text{Id}_s) \mathbf{w} = \frac{1}{m} \sum_{k=1}^m \mathbf{y}_k,$$

where $\mathbf{y}_k = (\mathbf{B}_{k,S}^* \mathbf{B}_{k,S} - \text{Id}_s) \mathbf{w}$ is a random vector with zero mean. Simple calculations yield that

$$\begin{aligned} \left\|\frac{1}{m} \mathbf{y}_k\right\|_2^2 &= \frac{1}{m^2} \left(\mathbf{w}^* (\mathbf{B}_{k,S}^* \mathbf{B}_{k,S})^2 \mathbf{w} - 2\mathbf{w}^* \mathbf{B}_{k,S}^* \mathbf{B}_{k,S} \mathbf{w} + \mathbf{w}^* \mathbf{w}\right) \\ &\leq \frac{1}{m^2} (\mu_1(S) \mathbf{w}^* \mathbf{B}_{k,S}^* \mathbf{B}_{k,S} \mathbf{w} - 2\mathbf{w}^* \mathbf{B}_{k,S}^* \mathbf{B}_{k,S} \mathbf{w} + 1) \\ &= \frac{1}{m^2} ((\mu_1(S) - 2) \mathbf{w}^* \mathbf{B}_{k,S}^* \mathbf{B}_{k,S} \mathbf{w} + 1) \\ &\leq \frac{1}{m^2} ((\mu_1(S) - 2) \mu_1(S) \|\mathbf{w}\|_2^2 + 1) = \frac{1}{m^2} ((\mu_1(S) - 2) \mu_1(S) + 1) \\ &\leq \frac{1}{m^2} (\mu_1(S) - 1)^2 \leq \frac{1}{m^2} \mu_1^2(S). \end{aligned}$$

Now, let us define $Z = \left\| \frac{1}{m} \sum_{k=1}^m \mathbf{y}_k \right\|_2$. By independence of the random vectors \mathbf{y}_k , it follows that

$$\begin{aligned} \mathbb{E}[Z^2] &= \frac{1}{m} \mathbb{E} \|\mathbf{y}_1\|_2^2 = \frac{1}{m} \mathbb{E} [\langle \mathbf{B}_S^* \mathbf{B}_S \mathbf{w}, \mathbf{B}_S^* \mathbf{B}_S \mathbf{w} \rangle - 2 \langle \mathbf{B}_S^* \mathbf{B}_S \mathbf{w}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle] \\ &= \frac{1}{m} \mathbb{E} \left[\left\langle (\mathbf{B}_S^* \mathbf{B}_S)^2 \mathbf{w}, \mathbf{w} \right\rangle - 2 \|\mathbf{B}_S \mathbf{w}\|_2^2 + 1 \right]. \end{aligned}$$

To bound the first term in the above equality, one can write

$$\begin{aligned} \mathbb{E} \left[\left\langle (\mathbf{B}_S^* \mathbf{B}_S)^2 \mathbf{w}, \mathbf{w} \right\rangle \right] &= \left\langle \mathbb{E} \left[(\mathbf{B}_S^* \mathbf{B}_S)^2 \right] \mathbf{w}, \mathbf{w} \right\rangle \\ &\leq \mu_1(S) \langle \mathbb{E} [(\mathbf{B}_S^* \mathbf{B}_S)] \mathbf{w}, \mathbf{w} \rangle \leq \mu_1(S) \|\mathbf{w}\|_2^2 = \mu_1(S). \end{aligned}$$

One immediately has that $\mathbb{E} \langle \mathbf{B}_S \mathbf{w}, \mathbf{B}_S \mathbf{w} \rangle = \|\mathbf{w}\|_2^2 = 1$. Therefore, one finally obtains that

$$\mathbb{E}[Z^2] \leq \frac{\mu_1(S) - 1}{m}.$$

Using the above upper bounds, namely $\left\| \frac{1}{m} \mathbf{y}_k \right\|_2 \leq \frac{\mu_1(S)}{m}$ and $\mathbb{E}[Z^2] \leq \frac{\mu_1(S) - 1}{m}$, the result of the lemma is thus a consequence of the Bernstein's inequality for random vectors (see Theorem A.4), which completes the proof. \blacksquare

Lemma B.3. *Let $S \subset \{1, \dots, n\}$, such that $|S| = s$. Let \mathbf{v} be a vector of \mathbb{C}^s . Then we have*

$$\mathbb{P}(\|\mathbf{A}_{S^c}^* \mathbf{A}_S \mathbf{v}\|_\infty \geq t \|\mathbf{v}\|_2) \leq 4n \exp \left(-\frac{mt^2/4}{\frac{\mu_3(S)}{s} + \frac{\mu_2(S)}{\sqrt{s}} t/3} \right). \quad (\text{E3})$$

Proof. Suppose without loss of generality that $\|\mathbf{v}\|_2 = 1$. Then,

$$\|\mathbf{A}_{S^c}^* \mathbf{A}_S \mathbf{v}\|_\infty = \max_{i \in S^c} \langle \mathbf{e}_i, \mathbf{A}^* \mathbf{A}_S \mathbf{v} \rangle = \max_{i \in S^c} \frac{1}{m} \sum_{k=1}^m \langle \mathbf{e}_i, \mathbf{B}_k^* \mathbf{B}_{k,S} \mathbf{v} \rangle.$$

Let us define $Z_k = \frac{1}{m} \langle \mathbf{e}_i, \mathbf{B}_k^* \mathbf{B}_{k,S} \mathbf{v} \rangle$. Note that $\mathbb{E} Z_k = 0$. From the Cauchy-Schwarz inequality, we get

$$|Z_k| = \left| \frac{1}{m} \langle \mathbf{e}_i, \mathbf{B}_k^* \mathbf{B}_{k,S} \mathbf{v} \rangle \right| = \left| \frac{1}{m} \mathbf{v}^* \mathbf{B}_{k,S}^* (\mathbf{B}_k \mathbf{e}_i) \right| \leq \frac{1}{m} \|\mathbf{v}\|_2 \|\mathbf{B}_{k,S}^* (\mathbf{B}_k \mathbf{e}_i)\|_2 \leq \frac{1}{m} \frac{\mu_2(S)}{\sqrt{s}}.$$

Furthermore,

$$\begin{aligned} \mathbb{E}|Z_k|^2 &= \frac{1}{m^2} \mathbb{E} \langle (\mathbf{B}_k \mathbf{e}_i), \mathbf{B}_{k,S} \mathbf{v} \rangle^2 \\ &\leq \frac{1}{m^2} \mathbf{v}^* \mathbb{E} [\mathbf{B}_S^* (\mathbf{B} \mathbf{e}_i) (\mathbf{B} \mathbf{e}_i)^* \mathbf{B}_S] \mathbf{v} \\ &\leq \frac{1}{m^2} \max_{i \in S^c} \|\mathbb{E} [\mathbf{B}_S^* (\mathbf{B} \mathbf{e}_i) (\mathbf{B} \mathbf{e}_i)^* \mathbf{B}_S]\|_{2 \rightarrow 2} = \frac{1}{m^2} \frac{\mu_3(S)}{s}. \end{aligned}$$

Using Bernstein's inequality A.1 for complex random variables, we end to

$$\begin{aligned} &\mathbb{P} \left(\frac{1}{m} \left| \sum_{k=1}^m \langle \mathbf{e}_i, \mathbf{B}_k^* \mathbf{B}_k \mathbf{v} \rangle \right| \geq t \right) \\ &\leq \mathbb{P} \left(\frac{1}{m} \left| \sum_{k=1}^m \operatorname{Re} \langle \mathbf{e}_i, \mathbf{B}_k^* \mathbf{B}_k \mathbf{v} \rangle \right| \geq t/\sqrt{2} \right) + \mathbb{P} \left(\frac{1}{m} \left| \sum_{k=1}^m \operatorname{Im} \langle \mathbf{e}_i, \mathbf{B}_k^* \mathbf{B}_k \mathbf{v} \rangle \right| \geq t/\sqrt{2} \right) \\ &\leq 4 \exp \left(-\frac{mt^2/4}{\frac{\mu_3(S)}{s} + \frac{\mu_2(S)}{\sqrt{s}} t/3} \right). \end{aligned}$$

Taking the union bound over $i \in S^c$ completes the proof. \blacksquare

Lemma B.4. *Let S be a subset of $\{1, \dots, n\}$. Then, for any $0 < t < \frac{\mu_1(S)}{\mu_2(S)}$, one has that*

$$\mathbb{P}\left(\max_{i \in S^c} \|\mathbf{A}_S^* \mathbf{A} \mathbf{e}_i\|_2 \geq t\right) \leq n \exp\left(-\frac{\left(\sqrt{m/\mu_1(S)}t - 1\right)^2}{4}\right). \quad (\text{E4})$$

Proof. Let us fix some $i \in S^c$. For $k = 1, \dots, m$, we define the random matrix

$$\mathbf{x}_k := \frac{1}{m} \mathbf{B}_{k,S}^* \mathbf{B}_k \mathbf{e}_i.$$

One has that $\mathbb{E} \mathbf{x}_k = 0$. Then, we remark that

$$\|\mathbf{A}_S^* \mathbf{A} \mathbf{e}_i\|_2 = \left\| \frac{1}{m} \sum_{k=1}^m \mathbf{B}_{k,S}^* \mathbf{B}_k \mathbf{e}_i \right\|_2 = \left\| \sum_{k=1}^m \mathbf{x}_k \right\|_2.$$

It follows that

$$\|\mathbf{x}_k\|_2 = \frac{1}{m} \|\mathbf{B}_{k,S}^* \mathbf{B}_k \mathbf{e}_i\|_2 \leq \frac{1}{m} \frac{\mu_2(S)}{\sqrt{s}}.$$

Furthermore, using Cauchy-Schwarz inequality, one has that

$$\begin{aligned} \mathbb{E} \|\mathbf{x}_1\|_2^2 &= \frac{1}{m^2} \mathbb{E} \|\mathbf{B}_{1,S}^* \mathbf{B}_1 \mathbf{e}_i\|_2^2 \leq \frac{1}{m^2} \mathbb{E} \|\mathbf{B}_{1,S}^*\|_{2 \rightarrow 2}^2 \|\mathbf{B}_1 \mathbf{e}_i\|_2^2 \leq \frac{1}{m^2} \mu_1(S) \mathbb{E} \|\mathbf{B}_1 \mathbf{e}_i\|_2^2 = \frac{1}{m^2} \mu_1(S) \|\mathbf{e}_i\|_2^2 \\ &\leq \frac{1}{m^2} \mu_1(S). \end{aligned}$$

Hence, using the above upper bounds, it follows from Bernstein's inequality for random vectors (see Theorem A.3) that

$$\mathbb{P}(\|\mathbf{A}_S^* \mathbf{A} \mathbf{e}_i\|_2 \geq t) \leq \exp\left(-\frac{\left(\sqrt{m/\mu_1(S)}t - 1\right)^2}{4}\right),$$

Finally, Inequality (E4) follows from a union bound over $i \in S^c$, which completes the proof. \blacksquare

C Proofs of the main results

C.1 Proof of Theorem 3.1

In this section, we recall an inexact duality formulation of the minimization problem (2) in the form of sufficient conditions to guarantee that the vector \mathbf{x} is the unique minimizer of (2), see [CP11]. These conditions give the properties that an inexact dual vector must satisfy to ensure the uniqueness of the solution of (2). In what follows, the notation $\mathbf{M}|_R$ denotes the restriction of a square matrix \mathbf{M} to its range R , and we define

$$\|\mathbf{M}|_R^{-1}\|_{2 \rightarrow 2} = \sup_{\mathbf{x} \in R; \|\mathbf{x}\|_2=1} \|\mathbf{M}|_R^{-1} \mathbf{x}\|_2$$

as the operator norm of the inverse of $\mathbf{M}|_R$ restricted to its range.

Lemma C.1 (Inexact duality [CP11]). *Suppose that $\mathbf{x} \in \mathbb{R}^n$ is supported on $S \subset \{1, \dots, n\}$. Then, assume that*

$$\|(\mathbf{A}_S^* \mathbf{A}_S)|_S^{-1}\|_{2 \rightarrow 2} \leq 2 \quad \text{and} \quad \max_{i \in S^c} \|\mathbf{A}_S^* \mathbf{A} \mathbf{e}_i\|_2 \leq 1. \quad (18)$$

Moreover, suppose that there exists $\mathbf{v} \in \mathbb{R}^n$ in the row space of \mathbf{A} obeying

$$\|\mathbf{v}_S - \text{sign}(\mathbf{x}_S)\|_2 \leq 1/4 \quad \text{and} \quad \|\mathbf{v}_{S^c}\|_\infty \leq 1/4, \quad (19)$$

Then, the vector \mathbf{x} is the unique solution of the minimization problem (2)

First, let us focus on Conditions (18). We can remark that

$$\|(\mathbf{A}_S^* \mathbf{A}_S)_{|S}^{-1}\|_{2 \rightarrow 2} = \left\| \sum_{k=1}^{\infty} (\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s)^k \right\|_{2 \rightarrow 2} \leq \sum_{k=1}^{\infty} \|\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s\|_{2 \rightarrow 2}^k.$$

Therefore, if the condition $\|\mathbf{A}_S^* \mathbf{A}_S - \text{Id}_s\|_{2 \rightarrow 2} \leq 1/2$ is satisfied, then $\|(\mathbf{A}_S^* \mathbf{A}_S)_{|S}^{-1}\|_{2 \rightarrow 2} \leq 2$. Hence, by Lemma B.1, it is clear that $\|(\mathbf{A}_S^* \mathbf{A}_S)_{|S}^{-1}\|_{2 \rightarrow 2} \leq 2$ with probability at least $1 - \varepsilon$, provided that

$$m \geq 8 \left(\mu_1(S) + \frac{1}{6} \max(\mu_1(S) - 1, 1) \right) \log \left(\frac{2s}{\varepsilon} \right).$$

By definition of $\gamma(S)$, the first inequality of Conditions (18) is ensured with probability larger than $1 - \varepsilon$ if

$$m \geq 8 \left(\gamma(S) + \frac{1}{6} \max(\gamma(S) - 1, 1) \right) \log \left(\frac{2s}{\varepsilon} \right). \quad (20)$$

Furthermore, using Lemma B.4, we obtain that

$$\max_{i \in S^c} \|\mathbf{A}_S^* \mathbf{A} \mathbf{e}_i\|_2 \leq 1$$

with probability larger than $1 - \varepsilon$ if

$$m \geq \mu_1(S) \left(1 + 4\sqrt{\log \left(\frac{n}{\varepsilon} \right)} + 4\log \left(\frac{n}{\varepsilon} \right) \right).$$

Again by definition of $\gamma(S)$, the second part of Conditions (19) is ensured if

$$m \geq 9\gamma(S) \log \left(\frac{n}{\varepsilon} \right). \quad (21)$$

Conditions (19) remain to be verified. The rest proof of Theorem 3.1 relies on the construction of a vector \mathbf{v} satisfying the conditions described in Lemma C.1 with high probability. To do so, we adapt the so-called golfing scheme introduced by Gross [Gro11] to our setting. More precisely, we will iteratively construct a vector that converges to a vector \mathbf{v} satisfying (19) with high probability.

Let us first partition the sensing matrix \mathbf{A} into blocks of blocks so that, from now on, we denote by $\mathbf{A}^{(1)}$ the first m_1 blocks of \mathbf{A} , $\mathbf{A}^{(2)}$ the next m_2 blocks, and so on. The L random matrices $\{\mathbf{A}^{(\ell)}\}_{\ell=1, \dots, L}$ are independently distributed, and we have that $m = m_1 + m_2 + \dots + m_L$. As explained before, $\mathbf{A}_S^{(\ell)}$ denotes the matrix $\mathbf{A}^{(\ell)} P_S$. The golfing scheme starts by defining $\mathbf{v}^{(0)} = 0$, and then it inductively defines

$$\mathbf{v}^{(\ell)} = \frac{m}{m_\ell} \mathbf{A}^{(\ell)*} \mathbf{A}_S^{(\ell)} \left(\mathbf{e} - \mathbf{v}_S^{(\ell-1)} \right) + \mathbf{v}^{(\ell-1)}, \quad (22)$$

for $\ell = 1, \dots, L$. In the rest of the proof, we set $\mathbf{v} = \mathbf{v}^{(L)}$. By construction, \mathbf{v} is in the row space of \mathbf{A} . The main idea of the golfing scheme is then to combine the results from the various Lemmas in Section B with an appropriate choice of L and the number m of measurements, to show that the random vector \mathbf{v} will satisfy the assumptions of Lemma C.1 with large probability. Using the shorthand notation $\mathbf{v}_S^{(\ell)} = P_S^* \mathbf{v}^{(\ell)}$, let us define

$$\mathbf{w}^{(\ell)} = \mathbf{e} - \mathbf{v}_S^{(\ell)}, \quad \ell = 1, \dots, L,$$

where $\mathbf{e} = \text{sign}(\mathbf{x}_S)$, and $\mathbf{x} \in \mathbb{R}^n$ is an s -sparse vector supported on S .

From the definition of $\mathbf{v}_S^{(\ell)}$, it follows that, for any $1 \leq \ell \leq L$,

$$\mathbf{w}^{(\ell)} = \left(\text{Id}_s - \frac{m}{m_\ell} \mathbf{A}_S^{(\ell)*} \mathbf{A}_S^{(\ell)} \right) \mathbf{w}^{(\ell-1)} = \prod_{j=1}^{\ell} \left(\text{Id}_s - \frac{m}{m_j} \mathbf{A}_S^{(j)*} \mathbf{A}_S^{(j)} \right) \mathbf{e}, \quad (23)$$

and

$$\mathbf{v} = \sum_{\ell=1}^L \frac{m}{m_\ell} \mathbf{A}^{(\ell)*} \mathbf{A}_S^{(\ell)} \mathbf{w}^{(\ell-1)}. \quad (24)$$

Note that in particular, $\mathbf{w}^{(0)} = \mathbf{e}$ and $\mathbf{w}^{(L)} = \mathbf{e} - \mathbf{v}_S$. In what follows, it will be shown that the matrices $\text{Id}_s - \frac{m}{m_\ell} \mathbf{A}_S^{(\ell)*} \mathbf{A}_S^{(\ell)}$ are contractions, and that the norm of the vector $\mathbf{w}^{(\ell)}$ decreases geometrically fast as ℓ increases. Therefore, $\mathbf{v}_S^{(\ell)}$ becomes close to \mathbf{e} as ℓ tends to L . In particular, we will prove that $\|\mathbf{w}^{(L)}\|_2 \leq 1/4$ for a suitable choice of L . In addition, we also show that \mathbf{v} satisfies the condition $\|\mathbf{v}_{S^c}\|_\infty \leq 1/4$. All these conditions will be shown to be satisfied with a large probability (depending on ε).

For all $1 \leq \ell \leq L$, we assume that with high probability

$$\|\mathbf{w}^{(\ell)}\|_2 \leq \underbrace{\left(\sqrt{\frac{\mu_1(S) - 1}{m_\ell}} + r_\ell \right)}_{r'_\ell} \|\mathbf{w}^{(\ell-1)}\|_2 \quad (25)$$

$$\left\| \frac{m}{m_\ell} \left(\mathbf{A}_{S^c}^{(\ell)} \right)^* \mathbf{A}_S^{(\ell)} \mathbf{w}^{(\ell-1)} \right\|_\infty \leq t_\ell \|\mathbf{w}^{(\ell-1)}\|_2. \quad (26)$$

The values of the quantities t_ℓ and r_ℓ , introduced in the above equations, will be specified later in the proof. Note that using (25), we can write that

$$\|\mathbf{e} - \mathbf{v}_S\|_2 = \|\mathbf{w}^{(L)}\|_2 \leq \|\mathbf{e}\|_2 \prod_{\ell=1}^L r'_\ell \leq \sqrt{s} \prod_{\ell=1}^L r'_\ell. \quad (27)$$

Furthermore, Equation (26) implies that

$$\begin{aligned} \|\mathbf{v}_{S^c}\|_\infty &= \left\| \sum_{\ell=1}^L \frac{m}{m_\ell} \left(\mathbf{A}_{S^c}^{(\ell)} \right)^* \mathbf{A}_S^{(\ell)} \mathbf{w}^{(\ell-1)} \right\|_\infty \\ &\leq \sum_{\ell=1}^L \left\| \frac{m}{m_\ell} \left(\mathbf{A}_{S^c}^{(\ell)} \right)^* \mathbf{A}_S^{(\ell)} \mathbf{w}^{(\ell-1)} \right\|_\infty \\ &\leq \sum_{\ell=1}^L t_\ell \|\mathbf{w}^{(\ell-1)}\|_\infty \\ &\leq \sqrt{s} \sum_{\ell=1}^L t_\ell \prod_{j=1}^{\ell-1} r'_j. \end{aligned} \quad (28)$$

We denote by $p_1(\ell)$ and $p_2(\ell)$ the probability that the upper bound (25), (26) do not hold. Now, let us set the number of blocks of blocks L , the number of blocks m_ℓ in each $\mathbf{A}^{(\ell)}$ and the values of the parameters t_ℓ and r_ℓ that have been introduced above. We propose to make the following choices :

- (i) $L = 2 + \left\lceil \frac{\log(s)}{2 \log 2} \right\rceil$,
- (ii) $m_1, m_2 \geq c\gamma(S) \log(4n) \log(2\varepsilon^{-1})$
 $m_\ell \geq c\gamma(S) \log(2L\varepsilon^{-1})$, for $\ell = 3, \dots, L$, for some sufficiently large $c \geq 1$,
- (iii) $r_1, r_2 = \frac{1}{4\sqrt{\log 4n}}$,
 $r_\ell = \frac{1}{4}$, for $\ell = 3, \dots, L$,

$$(iv) \quad t_1, t_2 = \frac{1}{8\sqrt{s}},$$

$$t_\ell = \frac{\log(4n)}{8\sqrt{s}}, \text{ for } \ell = 3, \dots, L.$$

With such choices, we obtain that

$$r'_1, r'_2 = \sqrt{\frac{\mu_1(S) - 1}{m_\ell}} + \frac{1}{4\sqrt{\log n}} \leq \frac{1}{2\sqrt{\log n}} \leq \frac{1}{2},$$

and

$$r'_\ell = \sqrt{\frac{\mu_1(S) - 1}{m_\ell}} + \frac{1}{4} \leq \frac{1}{2}$$

Furthermore, using (27), we obtain that

$$\|\mathbf{e} - \mathbf{v}_S\|_2 \leq \sqrt{s} \prod_{\ell=1}^L r'_\ell \leq \frac{\sqrt{s}}{2^L} \leq \frac{1}{4}, \quad (29)$$

where the last inequality follows from the previously specified choice on L . Moreover, using (28), we have that

$$\begin{aligned} \|\mathbf{v}_{S^c}\|_\infty &\leq \sqrt{s} \sum_{\ell=1}^L t_\ell \prod_{j=1}^{\ell-1} r'_j = \sqrt{s} (t_1 + t_2 r'_1 + t_3 r'_1 r'_2 + \dots) \\ &\leq \left(\frac{1}{8} + \frac{1}{16\sqrt{\log n}} + \frac{1}{32} + \dots \right) \\ &\leq \frac{1}{4}. \end{aligned} \quad (30)$$

For such a choice of parameters, and by Lemmas B.2 and B.3, if we fix $\varepsilon \in (0, 1/6)$, the bound $c \geq 534$ ensures $p_1(1), p_1(2), p_2(1), p_2(2) \leq \varepsilon/2$ and $p_1(\ell), p_2(\ell) \leq \varepsilon/2L$ for $\ell = 3, \dots, L$. Therefore, $\sum_{\ell=1}^L p_1(\ell) \leq 2\varepsilon$ and $\sum_{\ell=1}^L p_2(\ell) \leq 2\varepsilon$. From the above calculation, and by Lemmas B.2 and B.3 we finally obtain that if the overall number m of blocks samples obeys the condition

$$m = \sum_{\ell=1}^L m_\ell \geq c\gamma(S) (2 \log(4n) \log(2\varepsilon^{-1}) + (L-2) \log(2L\varepsilon^{-1})),$$

which can be simplified into

$$m \geq c\gamma(S) (2 \log(4n) \log(2\varepsilon^{-1}) + \log s \log(2e \log(s)\varepsilon^{-1})), \quad (31)$$

then the random vector \mathbf{v} , defined by (24), satisfies Assumptions 19 of Lemma C.1 with probability larger than $1 - 4\varepsilon$.

Hence, we have thus shown that if m satisfies the conditions (20), (21) and (31), then the Assumptions 18 and 19 of Lemma C.1 simultaneously hold with probability larger than $1 - 6\varepsilon$. Note that the bound (31) is stronger than (20) and (21). We complete the proof of Theorem 3.1 by replacing ε by $\varepsilon/6$. The final result on the required number of blocks measurements reads as follows

$$m \geq c\gamma(S) (2 \log(4n) \log(12\varepsilon^{-1}) + \log s \log(12e \log(s)\varepsilon^{-1})),$$

for $c = 534$, but in the statement we simplify the expression to improve the readability. Moreover, note that in our proof, for the sake of concision, there is no attempt to strengthen the previous result. Yet, we could have used the clever trick used in [CP11], and reused in [FR13] which consists of oversampling blocks in the golfing scheme.

C.2 Proof of Proposition 3.3

Since $\gamma(S) = \max(\mu_1, \mu_2, \mu_3)$, it suffices to show that setting $\mu_i = s\mu_4$ for $i \in \{1, 2, 3\}$ is sufficient to ensure the inequalities (5).

The first inequality in (5) can be shown as follows:

$$\|\mathbf{B}_S^* \mathbf{B}_S\|_{2 \rightarrow 2} \leq \|\mathbf{B}_S^* \mathbf{B}_S\|_{\infty \rightarrow \infty} \leq s \|\mathbf{B}^* \mathbf{B}\|_{1 \rightarrow \infty} \leq s\mu_4.$$

The second inequality in (5) can be shown as follows:

$$\sqrt{s} \max_{i \in S^c} \|\mathbf{B}_S^* \mathbf{B}e_i\|_2 \leq \sqrt{s} \sqrt{s} \|\mathbf{B}^* \mathbf{B}\|_{1 \rightarrow \infty} \leq s\mu_4.$$

Finally, fix $i \in S^c$. One can write

$$\begin{aligned} s \mathbb{E} \mathbf{B}_S^* (\mathbf{B}e_i) (\mathbf{B}e_i)^* \mathbf{B}_S &\preceq s \|(\mathbf{B}e_i) (\mathbf{B}e_i)^*\|_{2 \rightarrow 2} \mathbb{E} \mathbf{B}_S^* \mathbf{B}_S \\ &\preceq s \max_i \|\mathbf{B}e_i\|_2^2 \text{Id} \\ &\preceq s \|\mathbf{B}^* \mathbf{B}\|_{1 \rightarrow \infty} \text{Id} \\ &\preceq s\mu_4 \text{Id}. \end{aligned}$$

C.3 Proof of Proposition 4.1

Let us evaluate the quantities $(\mu_i(S))_{1 \leq i \leq 3}$ introduced in Definition 2.1 to upper bound $\gamma(S)$ with high probability. For this purpose, using Theorem 2 in [LR10], we get that for any $0 < t < 1$

$$\mathbb{P} \left(\|\mathbf{B}_S^* \mathbf{B}_S\|_{2 \rightarrow 2} \geq \left(1 + \sqrt{\frac{s}{p}}\right)^2 (1+t) \right) \leq C \exp \left(-\sqrt{pst}^{3/2} \left(\frac{1}{\sqrt{t}} \wedge \left(\frac{s}{p}\right)^{1/4} \right) / C \right), \quad (32)$$

for C a universal constant, under the assumption that $s > p$. We could also treat the case where $p > s$ by inverting the role of s and p in the above deviation inequality. We restrict our study to the case $s > p$ for simplicity.

By Inequality (32), we can consider that $\mu_1(S) \lesssim \frac{s}{p}$ with large probability (provided that s is sufficiently large). For evaluating $\mu_2(S)$, we use the following upper bound,

$$\max_{i \in S^c} \|\mathbf{B}_S^* \mathbf{B}e_i\|_2 \leq \max_{i \in S^c} \|\mathbf{B}_S^*\|_{2 \rightarrow 2} \|\mathbf{B}e_i\|_2 \leq \sqrt{\|\mathbf{B}_S^* \mathbf{B}_S\|_{2 \rightarrow 2}} \max_{i \in S^c} \sqrt{\|\mathbf{B}e_i\|_2^2}.$$

We already know that the first term $\sqrt{\|\mathbf{B}_S^* \mathbf{B}_S\|_{2 \rightarrow 2}}$ in the above inequality is bounded by $\sqrt{\frac{s}{p}}$ (up to a constant) with high probability, thanks to the previous discussion on $\mu_1(S)$. As for the second term, we use a union bound and the sub-gamma property of the chi-squared distribution, see [BLM13, p.29], to derive that

$$\mathbb{P} \left(\max_{i \in S^c} \|\mathbf{B}e_i\|_2^2 \geq 2 \left(\sqrt{\frac{t}{p}} + \frac{t}{p} \right) \right) \leq (n-s) \exp(-t) \leq n \exp(-t).$$

Let $\delta > 1$. Using the above deviation inequality, we get that

$$\max_{i \in S^c} \sqrt{\|\mathbf{B}e_i\|_2^2} \lesssim \sqrt{\frac{\delta \log(s)}{p}},$$

with probability larger than $1 - ns^{-\delta}$. Thus, we get the following upper bound for $\mu_2(S)$:

$$\mu_2(S) \lesssim \frac{s \sqrt{\delta \log(s)}}{p},$$

that holds with high probability provided that s is sufficiently large. Finally, by conditioning with respect to \mathbf{B}_S and using the independence of \mathbf{B}_S and $\mathbf{B}\mathbf{e}_i$ for $i \in S^c$, we have that

$$\begin{aligned} s \max_{i \in S^c} \|\mathbb{E}(\mathbf{B}_S^* (\mathbf{B}\mathbf{e}_i) (\mathbf{B}\mathbf{e}_i)^* \mathbf{B}_S)\|_{2 \rightarrow 2} &= s \max_{i \in S^c} \|\mathbb{E}[\mathbb{E}(\mathbf{B}_S^* (\mathbf{B}\mathbf{e}_i) (\mathbf{B}\mathbf{e}_i)^* \mathbf{B}_S | \mathbf{B}_S)]\|_{2 \rightarrow 2}, \\ &= s \max_{i \in S^c} \|\mathbb{E}[\mathbf{B}_S^* \mathbb{E}((\mathbf{B}\mathbf{e}_i) (\mathbf{B}\mathbf{e}_i)^*) \mathbf{B}_S]\|_{2 \rightarrow 2} = s \max_{i \in S^c} \left\| \mathbb{E} \left[\mathbf{B}_S^* \frac{1}{p} \text{Id} \mathbf{B}_S \right] \right\|_{2 \rightarrow 2} = \frac{s}{p}. \end{aligned}$$

Hence, one can take $\mu_3(S) = \frac{s}{p}$. Combining all these estimates we get that $\gamma(S) \lesssim \frac{s}{p} \sqrt{\delta \log(s)}$. Therefore, assuming that the lower bound on m in Theorem 3.1 still holds in the case of acquisition by blocks made of Gaussian entries, we need $m = O\left(\frac{s}{p} \log(s) \log(n)\right)$ blocks of measurements to ensure exact recovery, that is an overall number of measurements $q = O(s \log(s) \log(n))$.

C.4 Proof of Proposition 4.5

The proof is divided in two parts. First we show the result for $1 \leq s \leq \sqrt{n}$ and then we show it for $\sqrt{n} < s \leq n$. We let \mathbf{e}_i denote the i -th element of the canonical basis.

Part 1: Fix $s \in \{1, \dots, \sqrt{n}\}$. Let \mathcal{C}_s denote the class of vectors of kind $\mathbf{x} = \alpha \otimes \mathbf{e}_1$, where $\alpha \in \mathbb{R}^{\sqrt{n}}$ is s -sparse. Note that every $\mathbf{x} \in \mathcal{C}_s$ is s -sparse and that

$$\begin{aligned} \mathbf{A}\mathbf{x} &= (\tilde{\Psi}_{K,:} \otimes \Psi) \cdot (\alpha \otimes \mathbf{e}_1) \\ &= (\tilde{\Psi}_{K,:} \alpha) \otimes \Psi \mathbf{e}_1. \end{aligned}$$

In order to identify every s -sparse \mathbf{x} knowing $\mathbf{y} = \mathbf{A}\mathbf{x}$, there should not exist two distinct s -sparse vectors $\alpha^{(1)}$ and $\alpha^{(2)}$ in $\mathbb{C}^{\sqrt{n}}$ such that $\tilde{\Psi}_{K,:} \alpha^{(1)} = \tilde{\Psi}_{K,:} \alpha^{(2)}$. The vector $\alpha^{(1)} - \alpha^{(2)}$ is $\min(2s, \sqrt{n})$ -sparse. Therefore, a necessary condition for recovering all s -sparse vectors with $1 \leq s \leq \sqrt{n}$ is that $\tilde{\Psi}_{K,:} \alpha \neq 0$ for all non-zero $\min(2s, \sqrt{n})$ -sparse vectors α . To finish the first part of the proof it suffices to remark that a necessary condition for a set of $\min(2s, \sqrt{n})$ columns of $\tilde{\Psi}_{K,:}$ to be linearly independent is that $m = |K| \geq \min(2s, \sqrt{n})$, see Lemma 4.4.

Part 2: Assume that $\sqrt{n} < s \leq n$. Consider the class \mathcal{C}_s of s -sparse vectors of kind $\mathbf{x} = \sum_{l=1}^{\sqrt{n}} \alpha^{(l)} \otimes \mathbf{e}_l$, where $\text{supp}(\alpha^{(l)}) = \{1, \dots, \sqrt{n}\}$. For $\mathbf{x} \in \mathcal{C}_s$

$$\mathbf{A}\mathbf{x} = \sum_{l=1}^{\sqrt{n}} (\tilde{\Psi}_{K,:} \alpha^{(l)}) \otimes \Psi \mathbf{e}_l.$$

Similarly to the first part of the proof, in order to identify every s -sparse vectors, there should not exist $\alpha^{(1)}$ and $\alpha^{(1)'}$ with support equal to $\{1, \dots, \sqrt{n}\}$ such that $\tilde{\Psi}_{K,:} \alpha^{(1)} = \tilde{\Psi}_{K,:} \alpha^{(1)'}$. We showed in the previous section that a necessary condition for this condition to hold is $m = \sqrt{n}$.

C.5 Proof of Proposition 4.6

We consider blocks that consist of discrete lines in the 2D Fourier space as in Fig 1(b). We assume that $\sqrt{n} \in \mathbb{N}$ and that \mathbf{A}_0 is the 2D Fourier matrix applicable on $\sqrt{n} \times \sqrt{n}$ images. For all $p_1 \in \{1, \dots, \sqrt{n}\}$,

$$\mathbf{B}_{p_1} = \left[\frac{1}{\sqrt{n}} \exp \left(2i\pi \left(\frac{p_1 \ell_1 + p_2 \ell_2}{\sqrt{n}} \right) \right) \right]_{(p_1, p_2)(\ell_1, \ell_2)} \quad (33)$$

with $1 \leq p_2 \leq \sqrt{n}, 1 \leq \ell_1, \ell_2 \leq \sqrt{n}$. Let $S \subset \{1, \dots, \sqrt{n}\} \times \{1, \dots, \sqrt{n}\}$ denote the support of \mathbf{x} , with $|S| = s$. By definition of the 2D Fourier matrix of size $n \times n$, $\|\mathbf{B}_k^* \mathbf{B}_k\|_{1 \rightarrow \infty} = 1/\sqrt{n}$, for all $k \in \{1, \dots, \sqrt{n}\}$. Thus, Theorem 3.1 leads to

$$m \geq cs \frac{1}{\sqrt{n}} \max_{1 \leq k \leq M} \frac{1}{\pi_k} \log(4n) \log(12\varepsilon^{-1}).$$

Therefore, the choice of an optimal drawing probability, regarding the number of measurements, is given by

$$\pi_k^* = \frac{1}{\sqrt{n}}, \quad \forall k \in \{1, \dots, \sqrt{n}\}$$

and the number of measurements can be written as follows

$$m \geq Cs \log(4n) \log(12\varepsilon^{-1}),$$

which ends the proof of Proposition 4.6.

D An example with overlapping blocks

Let us illustrate the overlapping setting, in the case of blocks that consist in rows and columns in the 2D Fourier domain. Matrix $\mathbf{A}_0 \in \mathbb{C}^{n \times n}$ is the 2D Fourier transform matrix. We set

$$I_k^{\text{row}} = \{i \in \{1, \dots, n\}, (k-1)\sqrt{n} \leq i \leq k\sqrt{n}\} \quad I_k^{\text{col}} = \{k, \sqrt{n} + k, \dots, (\sqrt{n}-1)\sqrt{n} + k\}$$

the sets of indexes of $(\mathbf{a}_i^*)_{i \in \{1, \dots, n\}}$ that respectively correspond to the k -th row and the k -column in the 2D Fourier plane. Then, we can write the blocks as follows:

$$\mathbf{B}_k = \begin{cases} \left(\frac{1}{\sqrt{2}} \mathbf{a}_i^* \right)_{i \in I_k^{\text{row}}} & \text{if } k \in \{1, \dots, \sqrt{n}\} \\ \left(\frac{1}{\sqrt{2}} \mathbf{a}_i^* \right)_{i \in I_{k-\sqrt{n}}^{\text{col}}} & \text{if } k \in \{\sqrt{n} + 1, \dots, 2\sqrt{n}\}. \end{cases}$$

We have chosen the normalization factor equal to $1/\sqrt{2}$, as suggested, since each pixel of the image belongs to two blocks: one row and one column. According to Corollary 4.3, we conclude that the required number of blocks of measurements must satisfy

$$m \geq cs \frac{1}{2\sqrt{n}} \max_{1 \leq k \leq M} \frac{1}{\pi_k} (2 \log(4n) \log(12\varepsilon^{-1}) + \log s \log(12e \log(s)\varepsilon^{-1})). \quad (34)$$

Choosing the uniform probability for Π^* , i.e. $\pi_k^* = \frac{1}{2\sqrt{n}}$ for all $k \in \{1, \dots, 2\sqrt{n}\}$ leads to the following number of blocks of measurements

$$m \geq cs (2 \log(4n) \log(12\varepsilon^{-1}) + \log s \log(12e \log(s)\varepsilon^{-1})), \quad (35)$$

which is the same requirement in the 2D Fourier domain without overlapping, see Proposition 4.6.

References

- [AHPR13] Ben Adcock, Anders C. Hansen, Clarice Poon, and Bogdan Roman. Breaking the coherence barrier: A new theory for compressed sensing. *arXiv preprint arXiv:1302.0561*, 2013.
- [AHR14] Ben Adcock, Anders C. Hansen, and Bogdan Roman. The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing. *arXiv preprint arXiv:1403.6540*, 2014.
- [BLM13] Stéphane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.
- [BSO08] Jérôme Bobin, Jean-Luc Starck, and Roland Ottensamer. Compressed sensing in astronomy. *Selected Topics in Signal Processing, IEEE Journal of*, 2(5):718–726, 2008.

- [BWB14] Claire Boyer, Pierre Weiss, and Jérémie Bigot. An algorithm for variable density sampling with block-constrained acquisition. *SIAM Journal on Imaging Sciences*, 7(2):1080–1107, 2014.
- [Can08] Emmanuel Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9):589–592, 2008.
- [CCKW14] Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, and Pierre Weiss. Variable density sampling with continuous sampling trajectories. *SIAM Journal on Imaging Sciences*, *in press*, 2014.
- [CDD09] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best ℓ_1 -term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.
- [CP11] Emmanuel Candes and Yaniv Plan. A probabilistic and RIPless theory of compressed sensing. *Information Theory, IEEE Transactions on*, 57(11):7235–7254, 2011.
- [CR07] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.
- [CRT06a] Emmanuel Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.
- [CRT06b] Emmanuel Candes, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.
- [DH01] David Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 47(7):2845–2862, 2001.
- [Don06] David Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
- [FR13] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.
- [Gro11] David Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566, 2011.
- [KW14] Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Trans. Image Proc.*, 23(2):612–622, 2014.
- [LDP07] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MRI imaging. *Magnetic resonance in medicine*, 58(6):1182–1195, 2007.
- [LDSP08] Michael Lustig, David Donoho, Juan M. Santos, and John M. Pauly. Compressed sensing MRI. *Signal Processing Magazine, IEEE*, 25(2):72–82, 2008.
- [LKP08] Michael Lustig, Seung-Jean Kim, and John M. Pauly. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. *Medical Imaging, IEEE Transactions on*, 27(6):866–873, 2008.
- [LR10] Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electron. J. Probab.*, 15:no. 41, 1319–1343, 2010.

- [MAAOM10] Marcio M. Marim, Michael Atlan, Elsa Angelini, and Jean-Christophe Olivo-Marin. Compressed sensing with off-axis frequency-shifting holography. *Optics letters*, 35(6):871–873, 2010.
- [PDG14] Adam C. Polak, Marco F. Duarte, and Dennis L. Goeckel. Performance bounds for grouped incoherent measurements in compressive sensing. *arXiv preprint arXiv:1205.2118*, 2014.
- [PMG⁺12] Gilles Puy, Jose P. Marques, Rolf Gruetter, J. Thiran, Dimitri Van De Ville, Pierre Vandergheynst, and Yves Wiaux. Spread spectrum magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 31(3):586–598, 2012.
- [PVGW12] Gilles Puy, Pierre Vandergheynst, Rémi Gribonval, and Yves Wiaux. Universal and efficient compressed sensing by spread spectrum and application to realistic fourier imaging techniques. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–13, 2012.
- [RRT12] Holger Rauhut, Justin Romberg, and Joel A. Tropp. Restricted isometries for partial random circulant matrices. *Applied and Computational Harmonic Analysis*, 32(2):242–254, 2012.
- [RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.
- [Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.
- [WJP⁺09] Yves Wiaux, Laurent Jacques, Gilles Puy, Anna MM. Scaife, and Pierre Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 395(3):1733–1742, 2009.
- [Wri97] Graham A. Wright. Magnetic resonance imaging. *Signal Processing Magazine, IEEE*, 14(1):56–66, 1997.