An analysis of block sampling strategies in compressed sensing

Jérémie Bigot, Claire Boyer, Pierre Weiss

An analysis of block sampling strategies in compressed sensing

Jérémie Bigot[(1)], Claire Boyer[(2)] and Pierre Weiss[(3)]

[(1)] DMIA, Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse, France
jeremie.bigot@isae.fr
[(2)] Institut de Mathématiques de Toulouse, Université de Toulouse, France
claire.boyer@math.univ-toulouse.fr
[(3)] Institut des Technologies Avancées du Vivant, Toulouse, France
pierre.armand.weiss@gmail.com

May 19, 2013

Abstract

Compressed sensing (CS) is a theory which guarantees the exact recovery of sparse signals from a few number of linear projections. The sampling schemes suggested by current CS theories are often of little relevance since they cannot be implemented on practical acquisition systems. In this paper, we study a new random sampling approach that consists in selecting a set of blocks that are predefined by the application of interest. A typical example is the case where the blocks consist in horizontal lines in the 2D Fourier plane. We provide theoretical results on the number of blocks that are required for exact sparse signal reconstruction in a noise free setting. We illustrate this theory for various sensing matrices appearing in applications such as time-frequency bases. A typical result states that it is sufficient to acquire no more than $O\left(s \ln^2(n)\right)$ lines in the 2D Fourier domain for the perfect reconstruction of an s-sparse image of size $\sqrt{n} \times \sqrt{n}$. The proposed results have a large number of potential applications in systems such as magnetic resonance imaging, radio-interferometry or ultra-sound imaging.

Key-words: Compressed Sensing, blocks of measurements, sampling continuous trajectories, exact recovery, ℓ^1 minimization.

1 Introduction

The fundamental Shannon-Nyquist theorem claims that sampling a signal at least twice faster than its bandwidth is sufficient to exactly reconstruct the initial signal. Nevertheless, the resulting number of measurements needed can be so large that the storage becomes impossible and the acquisition time too long. Compressive Sensing (CS) is a new sampling theory, that gives theoretical conditions to ensure exact recovery of signals from a few number of linear projections (below the Nyquist rate). The key property allowing to apply this idea is the sparsity of the signals of interest, i.e. they can be represented by a small number of atoms in a well-chosen basis. We will say that $\boldsymbol{x} \in \mathbb{C}^n$ is s-sparse if

$$\|\boldsymbol{x}\|_0 \leq s,$$

where $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm counting the number of non-zero entries of $\boldsymbol{x}$.

In the seminal papers [Don06], [CRT06a], it has been showed that a sparse signal $\boldsymbol{x}$ can be perfectly reconstructed by solving the following ℓ_1-minimization problem:

$$\min_{\boldsymbol{z} \in \mathbb{C}^n} \|\boldsymbol{z}\|_1 \qquad \text{such that} \qquad \boldsymbol{A}_\Omega \boldsymbol{z} = \boldsymbol{y}, \tag{1}$$

where $\boldsymbol{A}_\Omega \in \mathbb{C}^{q\times n}$ $(q \le n)$ is a sensing matrix, $\boldsymbol{y} = \boldsymbol{A}_\Omega \boldsymbol{x} \in \mathbb{C}^q$ represents the vector of linear measurements, and $\|\boldsymbol{z}\|_1 = \sum_{i=1}^n |z_i|$ for all $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n$. If $\boldsymbol{A}_\Omega$ satisfies an incoherence property described later, then it can be shown that $q = O(s\ln(n))$ measurements are sufficient to perfectly reconstruct $\boldsymbol{x}$.

The construction of good sensing matrices $\boldsymbol{A}_\Omega$ 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 [Rau10, PVGW12, RRT12] may be used in the CS context. With this structure a matrix-vector product can be efficiently implemented on a computer by convolving the signal $\boldsymbol{x}$ 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 physics of the acquisition device, 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, presents the major advantage of allowing to implement CS strategies on almost 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 independently drawn at random. In this paper, we assume that we have access to a fixed projection set $(\boldsymbol{a}_k^*)_{k\in\{1,\ldots,n\}}$ with $\boldsymbol{a}_k^* \in \mathbb{C}^n$. Let us present a typical non-uniform recovery result in this setting. Let $\mathcal{P} = (p_1, \ldots, p_n)$ denote a probability distribution on $\{1, \ldots, n\}$ and define a random set Ω constituted of q indices in $\{1, \ldots, n\}$ resulting from i.i.d. drawings from $\mathcal{P}$.

Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^* \\ \vdots \\ \boldsymbol{a}_n^* \end{bmatrix}$ denote a unitary matrix and $\boldsymbol{A}_\Omega$ be the $q \times n$ matrix obtained by extracting the rows of $\boldsymbol{A}$ corresponding to the indices in Ω. Then, using the above notation, Theorem 4.2 in [Rau10] can be written as follows:

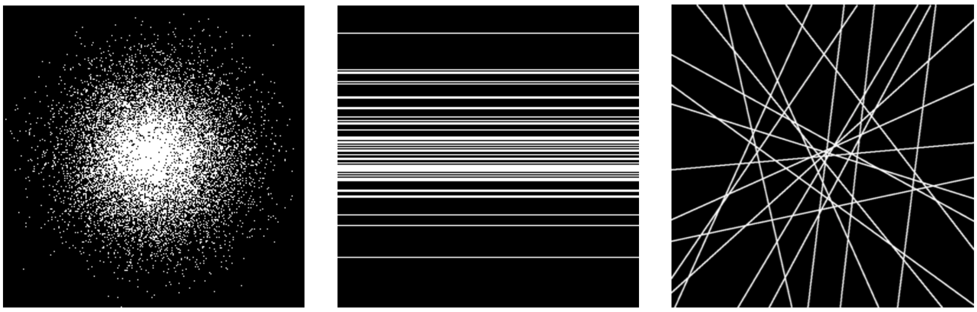
Theorem 1.1. *[Rau10, Theorem 4.2] Let $S \subset \{1, \ldots, n\}$ be of cardinality s and let $\boldsymbol{\epsilon} = (\epsilon_\ell)_{\ell\in S} \in \mathbb{C}^s$ be a sequence of independent Rademacher (or Steinhaus) random variables, see Definitions A.4 and A.5 for more details. Let $\boldsymbol{x}$ be an s-sparse vector of $\mathbb{C}^n$ with support S and* $\operatorname{sign}\left(\boldsymbol{x}^S\right) = \boldsymbol{\epsilon}$*, with $\boldsymbol{x}^S = (x_\ell)_{\ell\in S}$. Let $\boldsymbol{A}_\Omega$ be the sampling matrix. Assume that*

$$q \ge C_R \ln^2\left(\frac{6n}{\varepsilon}\right) s \max_{k\in\{1,\ldots,n\}} \frac{\|\boldsymbol{a}_k\|_\infty^2}{p_k} \tag{2}$$

with $C_R \simeq 26.25$, where $\|\boldsymbol{a}_k^*\|_\infty = \max_{1\le i\le n} |\boldsymbol{a}_k^*(i)|$ for all $k \in \{1, \ldots, n\}$. Set $\boldsymbol{y} = \boldsymbol{A}_\Omega \boldsymbol{x}$. Then with probability at least $1 - \varepsilon$ the vector $\boldsymbol{x}$ is the unique solution to the ℓ_1-minimization problem (1).

For instance, if $\boldsymbol{A}$ represents the matrix associated to the discrete Fourier transform, then $\|\boldsymbol{a}_k\|_\infty^2 = 1/n$ for all $k \in \{1, \ldots, n\}$. Therefore if $\mathcal{P}$ is the uniform distribution, Theorem 1.1 states that $C_R s \ln^2\left(\frac{6n}{\varepsilon}\right)$ measurements are sufficient for perfect recovery with probability $1 - \varepsilon$. Note that Theorem 1.1 gives a non-uniform recovery result, in the sense that a given s-sparse vector $\boldsymbol{x}$ can be reconstructed with high probability. There also exist uniform results that ensure a perfect reconstruction of all s-sparse vectors. Such results are obtained by estimating the restricted isometry constants of $\boldsymbol{A}_\Omega$, see e.g. [Can08, Rau10] for further details. Uniform recovery results are somewhat stronger than non-uniform ones. However they lead to a required number of measurements much larger than those given by Inequality (2).

In this paper, we concentrate on obtaining non-uniform recovery results in the case of sampling schemes that are different from those usually considered in CS. Indeed, the result of



(a) (b) (c)

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 measure Π having a 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.

Theorem 1.1 is valid for the case of isolated measurements of the form $\boldsymbol{a}_i^* \boldsymbol{x}$ for $i \in \Omega$. In the case of measurements in the 2D Fourier domain, this corresponds to sampling points on a square, according to the distribution $\mathcal{P}$, see Figure 1 (a) in which $\mathcal{P}$ is chosen to be a radial probability distribution. Unfortunately implementing such a strategy is impossible or impractical on many devices. Indeed, for the majority of acquisition systems, the number of measurements is of minimal importance relative to the path the sensor must take to collect the measurements. For instance in MRI, images are probed by measuring Fourier coefficients along curves parameterized by smooth functions see e.g. [Wri97, LKP08], see Figure 1 (b) for an example of measurements along lines in the 2D Fourier domain. In [HPH$^+$11], the authors consider the case of mission design strategies for mobile robots whose task is to perform spatial sampling of a static environmental field. The robots are subject to kinematic constraints and the objective is to minimize the energy spent to cross the acquisition path. In ultrasound imaging, images are sampled along lines in the space domain [Sza04].

We consider a sampling strategy consisting of randomly choosing blocks of measurements, and not only isolated measurements. Each block corresponds to a set of rows of an orthogonal sensing matrix. We deal with the case where the blocks are predefined. As far as we know, only reference [Gan07] deals with a similar strategy. In [Gan07], the author proposes to sample images by collecting contiguous pixel blocks of identical size using random matrices. The interest of this work is mostly numerical since no theoretical guarantee on the reconstruction quality is provided.

We study the problem of exact non-uniform sparse recovery in a noise-free setting. This strategy raises various questions. How can we set an optimal drawing probability in the case of blocks of measurements? How many blocks of measurements are needed to ensure exact reconstruction? Is the required number of blocks compatible with faster acquisition? We provide some answers to these questions. Our first contribution is the extension of Theorem 1.1 to the case of blocks of measurements. This result allows us to compute drawing probability distributions that minimize the number of blocks of measurements needed for exact recovery. These results provide a theoretical basis to the use of compressed sensing for many practical devices such as MRI, echography, computed tomography scanners, ...

The remaining of the paper is organized as follows. Section 2 describes the notation. Section 3 contains the main results about acquisition by non-overlapping blocks, along with some examples illustrating their relevance in practical settings. Section 4 contains the proofs of the main results contained in Section 3. In Section 5 , we discuss the extension of our results to the acquisition by overlapping blocks (see Figure 1 (c)), which is of significant importance for many application fields.

2 Notation and problem setting

We consider a full unitary sensing matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, i.e. $\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^* = \mathrm{Id}_n$, with $\boldsymbol{A}^*$ denoting the adjoint matrix of $\boldsymbol{A}$. Let $(\boldsymbol{a}_i^*)_{1\le i\le n}$ be the rows of $\boldsymbol{A}$, such that

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^* \\ \vdots \\ \boldsymbol{a}_n^* \end{bmatrix}.$$

The matrix $\boldsymbol{A}$ is given a block structure, as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B}_1 \\ \vdots \\ \boldsymbol{B}_M \end{bmatrix},$$

where the blocks $(\boldsymbol{B}_j)_{1\le j\le M}$ are made of rows of $\boldsymbol{A}$. In Section 3, we assume that the blocks $(\boldsymbol{B}_j)_{1\le j\le M}$ form a partition of the set of rows $\{\boldsymbol{a}_i^*\}_{1\le i\le n}$ and are such that $\boldsymbol{B}_j \in \mathbb{C}^{b_j\times n}$ with $\sum_{j=1}^M b_j = n$. We will consider the case of overlapping blocks in Section 5.

Let $\Pi = (\pi_1, \ldots, \pi_M)$ be a discrete probability distribution on the set of integers $\{1, \ldots, M\}$. Throughout the paper, $(J_k)_{1\le k\le m}$ denotes a sequence of i.i.d. discrete random variables taking their value in $\{1, \ldots, M\}$ with distribution Π.

Let $S \subset \{1, \ldots, n\}$ be a set of cardinality s. For a matrix $\boldsymbol{M} \in \mathrm{C}^{m\times n}$, we define

$$\boldsymbol{M}^S = (\boldsymbol{M}_{ij})_{1\le i\le m, j\in S}.$$

In this paper, we consider the following sampling strategy. We randomly select m blocks among $(\boldsymbol{B}_j)_{1\le j\le M}$, according to the discrete probability distribution Π, which leads to consider the sequence of i.i.d. random blocks $(\boldsymbol{X}_k)_{1\le k\le m}$ defined by

$$\boldsymbol{X}_k = \frac{1}{\sqrt{\pi_{J_k}}}\boldsymbol{B}_{J_k},\ k = 1\ldots m.$$

We can notice that $\mathrm{E}\left[\boldsymbol{X}_k^*\boldsymbol{X}_k\right] = \mathrm{Id}_n$ and $\mathrm{E}\left[\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S\right] = \mathrm{Id}_s$. This condition is a generalization of the isotropy property defined in [CP11] for the construction of sensing matrices in the case of isolated measurements.

We consider the following random under-sampling matrix

$$\widetilde{\boldsymbol{A}_m} = \frac{1}{\sqrt{m}}\begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix}. \tag{3}$$

Let $\boldsymbol{y} = \widetilde{\boldsymbol{A}_m}\boldsymbol{x}$ denote a set of blocks measurements of a signal $\boldsymbol{x}$. Finally, in order to reconstruct $\boldsymbol{x}$, the following standard ℓ_1-minimization problem is solved:

$$\min_{\boldsymbol{z}\in \mathrm{C}^n} \|\boldsymbol{z}\|_1 \quad \text{subject to} \quad \widetilde{\boldsymbol{A}_m}\boldsymbol{z} = \boldsymbol{y}. \tag{4}$$

3 Main results

3.1 Statement

Let us first introduce a new quantity of interest that will be shown to be of primary importance to obtain exact recovery results.

Definition 3.1. *For all $k \in \{1, \dots, M\}$, we define*

$$\|\boldsymbol{B}_k^* \boldsymbol{B}_k\|_\infty = \max_{\substack{1 \le i \le b_k \\ 1 \le j \le n}} \left| (\boldsymbol{B}_k^* \boldsymbol{B}_k)_{i,j} \right|.$$

We call this quantity the block-coherence.

The block-coherence is somewhat related to the *sub-coherence* of a block already used in [EKB10] in a different context.

The following theorem is the main result of the paper. It gives a set of sufficient conditions for exact recovery of $\boldsymbol{x}$ with large probability.

Theorem 3.2. *Let $S \subset \{1 \dots n\}$ be a set of cardinality $Card(S) = s$ and let $\boldsymbol{\epsilon} = (\epsilon_\ell)_{\ell \in S} \in \mathbb{C}^s$ be a sequence of independent random variables that are uniformly distributed on $\{-1; 1\}$ (or on the torus $\{z \in \mathbb{C}\,,\ |z| = 1\}$). Let $\boldsymbol{x}$ be a sparse vector with support S and $\mathrm{sign}(\boldsymbol{x}^S) = \boldsymbol{\epsilon}$. Let $\widetilde{\boldsymbol{A}_m}$ be the sampling matrix defined in (3). Assume that*

$$m \ge Cs \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) \max_{1 \le k \le M} \frac{\|\boldsymbol{B}_k^* \boldsymbol{B}_k\|_\infty}{\pi_k}, \tag{5}$$

where $C = 200\kappa^2$, and $\kappa^2 = \left(\frac{\sqrt{17}+1}{4} \right)^2$.

Then with probability at least $1 - \varepsilon$, the vector $\boldsymbol{x}$ is the unique solution of the ℓ_1-minimization problem (4).

The proof of Theorem 3.2 is detailed in Section 4.1. The approach is inspired by the results in [Rau10]. To derive Theorem 3.2, probabilistic tools such as symmetrization and Rudelson's lemma have to be extended from the vectorial case [Rau10] to the matricial one, see Lemmas A.6, A.8.

Theorem 3.2 and Condition (5) on the required number m of blocks of measurements can be used to derive an optimal drawing probability.

Proposition 3.3. *The drawing probability distribution Π^* minimizing the right hand side of Inequality (5) on the required number of measurements is defined by*

$$\pi_k^* = \frac{\|\boldsymbol{B}_k^* \boldsymbol{B}_k\|_\infty}{\sum_{\ell=1}^M \left\|\boldsymbol{B}_\ell^* \boldsymbol{B}_\ell\right\|_\infty}, \qquad \forall k \in \{1, \dots, M, \}\,. \tag{6}$$

For this particular choice of Π^*, the right hand side of Inequality (5) can be written as follows

$$m \ge Cs \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) \sum_{\ell=1}^M \|\boldsymbol{B}_\ell^* \boldsymbol{B}_\ell\|_\infty\,. \tag{7}$$

In Section 3.2, we discuss different examples leading to different choices of the drawing probability Π^*.

3.2 Examples

In this section, we illustrate Theorem 3.2 and Proposition 3.3 on various examples of practical interest.

3.2.1 Blocks made of a single row - 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. This is the standard case considered e.g. by [CRT06a], [Rau10], leading to the reconstruction result recalled in Theorem 1.1. According to Theorem 3.2 the number of isolated measurements sufficient to obtain perfect reconstruction with probability $1-\varepsilon$ must verify the following inequality

$$m \geq C s \ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right) \max_{1\leq k\leq M} \frac{\|\boldsymbol{a}_k \boldsymbol{a}_k^*\|_\infty}{\pi_k}. \tag{8}$$

By Theorem 1.1, the required number of isolated measurements is

$$q \geq C_R s \ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right) \max_{1\leq k\leq n} \frac{\|\boldsymbol{a}_k\|_\infty^2}{p_k}. \tag{9}$$

Noting that,

$$\|\boldsymbol{a}_k\|_\infty^2 = \|\boldsymbol{a}_k \boldsymbol{a}_k^*\|_\infty, \qquad \forall k \in \{1,\ldots,n\},$$

it follows that Condition (8) is comparable to (9) up to a multiplicative constant. This difference is not too prejudicial since Theorem 3.2 should be mainly considered as a guide to construct sampling schemes and not as a requirement for perfect recovery.

In addition, by Theorem 3.2, choosing Π^* as suggested by Proposition 3.3 such that

$$\pi_k^* = \frac{\|\boldsymbol{a}_k \boldsymbol{a}_k^*\|_\infty}{\sum_{\ell=1}^n \|\boldsymbol{a}_\ell \boldsymbol{a}_\ell^*\|_\infty}, \qquad \forall k \in \{1,\ldots,n\},$$

leads to the following required number of measurements:

$$m \geq C s \ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right) \sum_{\ell=1}^n \|\boldsymbol{a}_\ell \boldsymbol{a}_\ell^*\|_\infty. \tag{10}$$

Contrary 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. We consider that $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{F}_{n-1} \end{pmatrix}$, 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 (2) is $O(sn\ln^2(n))$. This shows that the CS is not applicable for this sensing matrix. Note that $\|\boldsymbol{A}\|_\infty = 1$, which is the worst possible for orthogonal matrices. On the contrary, if the optimal drawing distribution is chosen, i.e.

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

then, the right hand side of (2) is $O(2s\ln^2(n))$. In this setting, CS remains a relevant sampling strategy. 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 $\boldsymbol{A}$, and uniformly picking the $m-1$ remaining isolated measurements.

3.2.2 Acquisition in the 2D Fourier domain and sparsity in the 2D spatial domain

We now turn to a realistic setting where signals are sparse in the spatial basis and blocks of frequencies are probed in the 2D Fourier domain. We consider blocks that consist of discrete lines in the 2D Fourier space as in Fig 1(b). This scenario is close to what can be encountered in MRI or for some tomographic devices. We assume that $\sqrt{n} \in \mathbb{N}$ and that $\boldsymbol{A}$ is the 2D Fourier matrix applicable on $\sqrt{n} \times \sqrt{n}$ images. For all $p_1 \in \{1, \dots, \sqrt{n}\}$,

$$\boldsymbol{B}_{p_1} = \left[\frac{1}{\sqrt{n}} \exp\left(2i\pi \left(\frac{p_1 \ell_1 + p_2 \ell_2}{\sqrt{n}} \right) \right) \right]_{(p_1, p_2)(\ell_1, \ell_2)} \tag{11}$$

with $1 \leq p_2 \leq \sqrt{n}, 1 \leq \ell_1, \ell_2 \leq \sqrt{n}$. The number of blocks is thus $M = \sqrt{n}$.

Proposition 3.4. *Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ denote the 2D discrete Fourier matrix and consider a partition in $M = \sqrt{n}$ blocks that consist of lines in the 2D Fourier domain. In the same setting as in Theorem 3.2, the required number of measurements for perfect recovery with probability greater than $1 - \varepsilon$ is*

$$m \geq C \ln^2 \left(\frac{2^{3/4} n}{\varepsilon} \right) \frac{s}{\sqrt{n}} \max_{1 \leq k \leq M} \frac{1}{\pi_k}. \tag{12}$$

The drawing probability minimizing (12) *is given by*

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

and for this particular choice, the required number m of blocks of measurements must satisfy the following inequality

$$m \geq C s \ln^2 \left(\frac{2^{3/4} n}{\varepsilon} \right).$$

We refer to Section 4.2.1 for the proof of Proposition 3.4. Here, an optimal drawing probability Π^* is thus the uniform one over the set $\{1, \dots, \sqrt{n}\}$. It is independent of the horizontal frequency $\ell \in \{1, \dots, \sqrt{n}\}$. Moreover, this result shows that it is sufficient to acquire no more than $m = O\left(s \ln^2(n)\right)$ lines in the 2D Fourier domain for the perfect reconstruction of an s-sparse image of size $\sqrt{n} \times \sqrt{n}$.

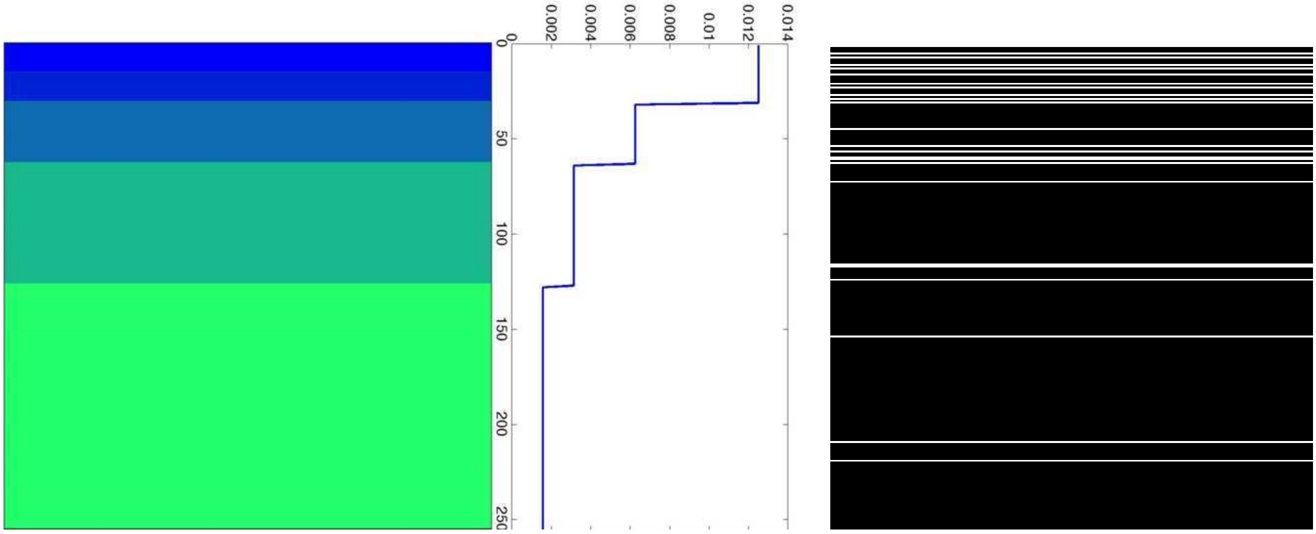
3.2.3 Acquisition in the 2D Fourier domain and sparsity in the 2D wavelet domain

Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ be the 2D Fourier-Wavelet transform, with $\sqrt{n} = 2^J$, $J \in \mathbb{N}$. In this case, the signal is sparse in the wavelet domain, and the acquisition is done in the Fourier domain. This is one of the most common scenario in imaging application. Typical examples include MR imaging [LDSP08] or radio interferometry. In order to simplify the calculations, we consider the case of Shannon wavelets [Mal99]. The proposed reasoning could also be applied to Meyer's wavelets since they are band-limited. Let $\widehat{\Psi_{j,k}}$ be the Fourier transform of the 1D Shannon wavelet at scale j and location parameter k, with $j \in \{L+1, \dots, J\}$, and $k \in \{1, \dots, n_j\}$ where $n_j = 2^{j-1}$ is the number of wavelets at the j-th scale for $j > L$, and $L \in \{0, \dots, J-1\}$ is the coarsest level of resolution (i.e. the maximum level of decomposition of the 2D wavelet transform is $J - L$). Note that at scale $j = L$, we denote by $\widehat{\Psi_{L,k}}$ the Fourier transform of the scaling functions whose number is $n_L = 2^L$. For $\ell \in \{1, \dots, \sqrt{n}\}$, the ℓ-th Fourier coefficient of $\widehat{\Psi_{j,k}}$ is denoted by $\widehat{\Psi_{j,k}}(\ell)$. Since Shannon wavelets have disjoint frequency support, for each $\ell \in \{1, \dots, \sqrt{n}\}$, there exists a unique scale $j = j_\ell$ such that

$$\widehat{\Psi_{j_\ell,k}}(\ell) \neq 0.$$

As in Section 3.2.2 we consider $M = \sqrt{n}$ blocks $(\boldsymbol{B}_\ell)_{\ell \in \{1, \dots, \sqrt{n}\}} \in \mathbb{C}^{\sqrt{n} \times n}$ that are lines in the 2D Fourier domain.

Proposition 3.5. *Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ denote the 2D discrete Fourier-Shannon wavelets matrix and consider a partition in $M = \sqrt{n}$ blocks that consist of lines in the 2D Fourier domain. In the*



(a) (b)

Figure 2: For $J = 8$ and $L = 4$, (a): On the left hand side, the 2D Fourier acquisition plane is laid out. The various colours denote different sets of rows having the same drawing probability. The corresponding probability density with respect to the row indexes is outlined on the middle plot. (b) Example of 25 draws according to the probability density described in (a).

same setting as in Theorem 3.2, the required number of measurements for perfect recovery with probability greater than $1 - \varepsilon$ *is*

$$m \geq C s \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) \max_{1 \leq \ell \leq \sqrt{n}} \frac{1}{n_{j_\ell} \pi_\ell} \tag{13}$$

The drawing probability minimizing (28) *is given by*

$$\pi^*_\ell = \frac{1}{n_{j_\ell} (J - L + 1)}, \quad \forall \ell \in \left\{1, \ldots, \sqrt{n}\right\}$$

and for this particular choice, the required number m *of blocks of measurements must satisfy the following inequality*

$$m \geq C s \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) (J - L + 1).$$

The proof of Proposition 3.5 is presented in Section 4.2.2. One can remark that the optimal drawing probability Π^* is not the uniform one, as illustrated in Figure 2.

3.2.4 Acquisition in the Dirac basis and sparsity in the 2D Haar domain

In this section, we propose to study sparse signals in the 2D Haar domain, sampled in the 2D Dirac basis, which could illustrate the photography setting. Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ be the inverse 2D Haar transform, with $\sqrt{n} = 2^J$, $J \in \mathbb{N}$. We focus on the case where the blocks $(\boldsymbol{B}_\ell)_{\ell \in \left\{1, \ldots, \sqrt{n}\right\}}$ consist in a horizontal line of the 2D spatial domain (i.e. in the Dirac basis). Let $\Psi_{j,k}$ be the 1D Haar wavelet at scale j and location parameter k, with $j \in \{L+1, \ldots, J\}$, and $k \in \{1, \ldots, n_j\}$ where $n_j = 2^{j-1}$ is the number of wavelets at the j-th scale for $j > L$, and $L \in \{0, \ldots, J-1\}$ is the coarsest level of resolution (i.e. the maximum level of decomposition of the 2D wavelet transform is $J - L$). Note that at scale $j = L$, we denote by $\Psi_{L,k}$ the scaling functions whose number is $n_L = 2^L$.

Proposition 3.6. *Let $\boldsymbol{A}$ denote the inverse 2D Haar wavelets matrix and consider a partition in $M = \sqrt{n}$ blocks that consist of lines in the 2D Dirac domain. In the same setting as in Theorem 3.2, the required number of measurements for perfect recovery with probability greater than $1 - \varepsilon$ is*

$$m \geq Cs \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) \frac{1}{2} \max_{1 \leq \ell \leq \sqrt{n}} \frac{1}{\pi_\ell}. \tag{14}$$

The drawing probability minimizing (14) *is given by*

$$\pi_\ell^* = \frac{1}{\sqrt{n}} = \frac{1}{2^J}, \quad \forall \ell \in \{1, \ldots, \sqrt{n}\}$$

and for this particular choice, the required number m of blocks of measurements must satisfy the inequality

$$m \geq Cs \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) 2^{J-1}.$$

The proof of Proposition 3.6 is presented in Section 4.2.3. Now, let us consider the case of acquisition by isolated measurements. On the one hand, the optimal drawing probability distribution $\mathcal{P}^*$ in Theorem 1.1 is given by

$$p_k^* = \frac{\|\boldsymbol{a}_k^*\|_\infty}{\sum_{\ell=1}^n \|\boldsymbol{a}_\ell^*\|_\infty} = \frac{1}{n}, \quad \forall k \in \{1, \ldots, n\}.$$

It leads to the following required number of measurements

$$q \geq C_R s \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) 2^{2J-1}. \tag{15}$$

see Section 4.2.3 for computation details. On the other hand, by Proposition 3.6, it can be seen that our approach requires an equivalent number of linear measurements

$$q' := m\sqrt{n} \geq Cs \ln^2 \left(\frac{2^{3/4} 3n}{\varepsilon} \right) 2^{2J-1}.$$

In this example, both approaches, based either on blocks acquisition or on isolated measurements, still lead to comparable results in terms of linear measurements. Note that the Haar wavelet matrix is little adapted to the use of compressed sensing since $2^{2J} = n$. Equivalence in terms of number of isolated measurements by both approaches could be also observed in the case where $\boldsymbol{A}$ is the Identity matrix Id_n. Indeed, sampling lines in the 2D plane will lead to $\|\boldsymbol{B}_k^* \boldsymbol{B}_k\|_\infty = 1$ for all $k \in \{1, \ldots, \sqrt{n}\}$, and $\|\boldsymbol{a}_\ell^*\|_\infty^2 = 1$ for all $\ell \in \{1, \ldots, n\}$ as well.

4 Proofs of the main results

4.1 Proof of Theorem 3.2

We present the proof in the case where $\boldsymbol{\epsilon}$ is a Rademacher sequence, but it can be easily extended to the case of a Steinhaus sequence at the price of changing some constants.

First, let us recall the three following results, that are basics to obtain reconstruction results in CS, see e.g. [Rau10].

Proposition 4.1. *[Rau10, Proposition 7.1] Let* $\boldsymbol{M} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n) \in \mathrm{C}^{q\times n}$ *and let* $S \subset \{1\ldots n\}$ *of size* $Card(S) = s$*. Assume that* $\boldsymbol{M}^S$ *is injective and*

$$\left\|\boldsymbol{M}^{S\dagger}\boldsymbol{m}_\ell\right\|_2 \leq \alpha < \frac{1}{\sqrt{2}} \quad \forall \ell \notin S, \tag{16}$$

where $\boldsymbol{M}^{S\dagger}$ *is the Moore-Penrose pseudo-inverse of* $\boldsymbol{M}^S$*, see Definition A.1. Let* $\boldsymbol{\epsilon} = (\epsilon_j)_{j\in S} \in \mathrm{C}^s$ *be a random Rademacher or Steinhaus sequence, see Definitions A.4 and A.5. Then with probability at least*

$$1 - 2^{3/4}(n-s)\exp\left(-\frac{\alpha^{-2}}{2}\right),$$

every vector $\boldsymbol{x} \in \mathrm{C}^n$ *with support* S *and* $\operatorname{sign}\left(\boldsymbol{x}^S\right) = \boldsymbol{\epsilon}$ *is the unique solution to the* ℓ_1*-minimization problem* (4)*.*

Definition 4.2 (Coherence)**.** *Let* $\boldsymbol{M} = \left(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n\right) \in \mathrm{C}^{m\times n}$*. In order to evaluate the quality of the measurement matrix* $\boldsymbol{M}$*, the coherence is defined by*

$$\mu(\boldsymbol{M}) := \max_{j\neq k} \frac{|\langle \boldsymbol{m}_j, \boldsymbol{m}_k\rangle|}{\|\boldsymbol{m}_j\|_2 \|\boldsymbol{m}_k\|_2}$$

Proposition 4.3. *[Rau10, Proposition 7.2] Let* $\boldsymbol{M} \in \mathrm{C}^{q\times n}$ *with coherence* $\mu(\boldsymbol{M})$ *and let* $S \subset \{1\ldots n\}$ *of size* $Card(S) = s$*. Assume that*

$$\left\|\left(\boldsymbol{M}^S\right)^* \boldsymbol{M}^S - \mathrm{Id}_s\right\|_2 \leq \delta, \tag{17}$$

for some $\delta \in]0,1[$*. Then,*

$$\left\|\boldsymbol{M}^{S\dagger}\boldsymbol{m}_\ell\right\|_2 \leq \frac{\sqrt{s}\mu\left(\boldsymbol{M}\right)}{1-\delta} \quad \forall \ell \notin S.$$

By Proposition 4.1, we have to control

$$\left\|\widetilde{\boldsymbol{A}_m}^{S\dagger}\widetilde{\boldsymbol{a}}_\ell\right\|_2 \quad \forall \ell \notin S \tag{18}$$

where $\widetilde{\boldsymbol{a}}_\ell$ are the columns of $\widetilde{\boldsymbol{A}_m}^S$. To do so, we use Proposition 4.3, which requires an estimate on $\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2$ and on $\mu\left(\widetilde{\boldsymbol{A}_m}\right) = \mu$ to ensure exact reconstruction. The key point in the proof is to verify the inequality

$$\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2 \leq \delta$$

for small δ. To obtain such a result, we use symmetrization, Khintchine's inequality and Rudelson's lemma, that we need to extend to the matricial case. Then, we derive a probabilistic estimate of the coherence of $\widetilde{\boldsymbol{A}_m}$. Finally, we complete the proof by combining all the previous developments, and by adjusting all the parameters.

Theorem 4.4. *Let* $\widetilde{\boldsymbol{A}_m}$ *be the sampling matrix* (3)*, let* $S \subset \{1\ldots n\}$ *be a set of cardinality* $Card(S) = s \geq 2$*. Let* $\delta \in]0,1/2]$*.*

Then with probability at least

$$1 - 2^{3/4}s\exp\left(-\frac{m\delta^2}{8\kappa^2 \max_{1\leq k\leq M}\frac{\rho_k^S}{\pi_k}}\right)$$

for some constant universal $\kappa = \frac{\sqrt{17}+1}{4}$*, where* ρ_k^S *is any quantity satisfying* $\rho_k^S \geq \left\|\left(\boldsymbol{B}_k^S\right)^* \boldsymbol{B}_k^S\right\|_2$*, for all* $k \in \{1, \ldots, M\}$*, the sensing matrix* $\widetilde{\boldsymbol{A}_m} = \frac{1}{\sqrt{m}}\boldsymbol{A}_m$ *satisfies*

$$\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2 \leq \delta.$$

Proof. First, we define H for any $p \geq 2$,

$$\begin{aligned} H &:= \mathrm{E}\left[\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2^p\right] \\ &= \mathrm{E}\left[\left\|\frac{1}{m}\sum_{k=1}^m \boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S - \mathrm{Id}_s\right\|_2^p\right] \\ &= \mathrm{E}\left[\left\|\frac{1}{m}\sum_{k=1}^m \left(\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S - \mathrm{E}\left(\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S\right)\right)\right\|_2^p\right]. \end{aligned} \tag{19}$$

We aim at giving an upper bound on H and the result will follow by Markov's inequality. First by Lemma A.6, one can deduce that

$$H \leq \left(\frac{2}{m}\right)^p \mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S\right)\right\|_2^p\right],$$

where, $(\epsilon_k)_{k=1,\ldots,m}$ is a Rademacher sequence, and we get

$$\mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S\right)\right\|_2^p\right] \leq \mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S\right)\right\|_{S_p}^p\right],$$

where $\|.\|_{S_p}$ is the Schatten p-norm, see Definition A.2. Indeed, by Proposition A.3 (i), for any matrix $\boldsymbol{M}$,

$$\|\boldsymbol{M}\|_2^p \leq \|\boldsymbol{M}\|_{S_p}^p, \qquad \forall p \geq 2.$$

From now on, we do calculations by conditioning to $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, so as to apply Rudelson's lemma. Let E_X denote the expectation conditionally to $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$.

Since the matrix $\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S$ is a matrix of size $s \times s$, its rank is at most s. Applying Rudelson's lemma A.8 to the sensing matrix $\widetilde{\boldsymbol{A}_m}$ brings

$$\left(\mathrm{E}_X\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{X}_k^{S*}\boldsymbol{X}_k^S\right)\right\|_2^p\right]\right)^{1/p} \leq 2^{3/4p} s^{1/p}\sqrt{p}e^{-1/2}\left\|\boldsymbol{A}_m^S\right\|_2 \max_{k=1..M}\sqrt{\left\|\frac{\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S}{\pi_k}\right\|_2}, \tag{20}$$

where $\boldsymbol{A}_m = \sqrt{m}\widetilde{\boldsymbol{A}_m} = \begin{pmatrix}\boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m\end{pmatrix}$.

In order to finish the proof, we have to decondition on $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, to finally obtain an upper bound on H. By applying expectation on both sides of (20) and then Cauchy-Schwarz's inequality, one can deduce that

$$\begin{aligned} H &\leq \left(\frac{2}{m}\right)^p 2^{3/4} s p^{p/2} e^{-p/2} \mathrm{E}\left[\left\|\boldsymbol{A}_m^S\right\|_2^p \max_{k=1..M}\sqrt{\left\|\frac{\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S}{\pi_k}\right\|_2^p}\right] \\ &\leq \left(\frac{2}{m}\right)^p 2^{3/4} s p^{p/2} e^{-p/2}\sqrt{\mathrm{E}\left[\|\boldsymbol{A}_m^S\|_2^{2p}\right]\mathrm{E}\left[\max_{k=1..M}\left\|\frac{\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S}{\pi_k}\right\|_2^{2p/2}\right]}. \end{aligned} \tag{21}$$

We recall that

$$\rho_k^S \geq \left\|\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S\right\|_2, \quad \forall k \in \{1, \ldots, M\}. \tag{22}$$

Then, one can write

$$
\begin{aligned}
H &\le \left(\frac{2}{m}\right)^p 2^{3/4} s p^{p/2} e^{-p/2} \sqrt{\mathrm{E}\left[\|\boldsymbol{A}_m^S\|_2^{2p}\right]\left(\max_{k=1..M}\frac{\rho_k^S}{\pi_k}\right)^{2p/2}} \\
&\le \left(\frac{2}{m}\right)^p 2^{3/4} s p^{p/2} e^{-p/2} \sqrt{\mathrm{E}\left[\left\|\sqrt{m}\widetilde{\boldsymbol{A}_m}^S\right\|_2^{2p}\right]\left(\max_{k=1..M}\frac{\rho_k^S}{\pi_k}\right)^{2p/2}} \\
&\le \left(\frac{2}{m}\right)^p 2^{3/4} s p^{p/2} e^{-p/2} \sqrt{m}^p \left(\max_{k=1..M}\frac{\rho_k^S}{\pi_k}\right)^{2p/2} \sqrt{\mathrm{E}\left[\left\|\widetilde{\boldsymbol{A}_m}^S\right\|_2^{2p}\right]} \\
&\le \underbrace{\left(2\sqrt{\frac{\max_{k=1..M}\frac{\rho_k^S}{\pi_k}}{m}}\right)^p 2^{3/4} s p^{p/2} e^{-p/2}}_{D^p} \sqrt{\mathrm{E}\left[\left(\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2 + 1\right)^p\right]} \\
&\le D^p \left(\left(\mathrm{E}\left[\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2^p\right]\right)^{1/p} + 1\right)^{p/2},
\end{aligned}
$$

which implies

$$
\begin{aligned}
&\Rightarrow H^{1/p} \le D\sqrt{H^{1/p}+1} && \Rightarrow H^{2/p} \le D^2\left(H^{1/p}+1\right) \\
&\Rightarrow H^{2/p} - H^{1/p}D^2 \le D^2 && \Rightarrow \left(H^{1/p} - \frac{1}{2}D^2\right)^2 - \frac{1}{4}D^4 \le D^2 \\
&\Rightarrow H^{1/p} \le \sqrt{D^2 + D^4/4} + D^2/2.
\end{aligned}
$$

If $D \le 1/2$, this comes to

$$
H^{1/p} \le \sqrt{1+1/16}D + \frac{1}{4}D = \kappa D,
$$

with $\kappa = \frac{\sqrt{17}+1}{4}$. Hence, we can write to cover both cases : $D \le 1/2$ and $D > 1/2$, that

$$
\begin{aligned}
\left(\mathrm{E}\min\left((1/2)^p, \left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2^p\right)\right)^{1/p} &\le \min\left(1/2, \left(\mathrm{E}\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2^p\right)^{1/p}\right) \\
&\le \kappa D.
\end{aligned}
$$

By Proposition A.9, we can control the tail of random variables by the mean of their moments. Considering (21), we set

$$
\alpha = 2e^{-1/2}\kappa\sqrt{\frac{\max_{k=1..M}\frac{\rho_k^S}{\pi_k}}{m}}, \qquad \beta = 2^{3/4}s,
$$

and

$$
\gamma = 2, \qquad p_0 = 2.
$$

Then, Proposition A.9 with the previous values leads to, for all $u \ge \sqrt{2}$,

$$
\mathrm{P}\left(\min\left(1/2, \left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^* \widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2\right) \ge 2\kappa\sqrt{\frac{\max_{k=1..M}\frac{\rho_k^S}{\pi_k}}{m}}u\right) \le 2^{3/4}se^{-u^2/2}.
$$

Then for $2\kappa\sqrt{2\frac{\max_{k=1..M}\frac{\rho_k^S}{\pi_k}}{m}} \leq \delta \leq 1/2$, we have that

$$\mathrm{P}\left(\left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^*\widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2 \geq \delta\right) \leq 2^{3/4}s\exp\left(\frac{m\delta^2}{8\kappa^2\max_{k=1..M}\frac{\rho_k^S}{\pi_k}}\right). \tag{23}$$

The right hand side of (23) is less than ε if

$$m \geq \frac{8\kappa^2}{\delta^2}\ln\left(2^{3/4}s/\varepsilon\right)\max_{k=1..M}\frac{\rho_k^S}{\pi_k}. \tag{24}$$

One can notice that (24) directly implies that $2\kappa\sqrt{2\frac{\max_{k=1..M}\frac{\rho_k^S}{\pi_k}}{m}} \leq \delta$, if ε is relevant, i.e. $\varepsilon < 1$. We can conclude that (23) is valid also for $0 < \delta < 2\kappa\sqrt{2\frac{\max_{k=1..M}\frac{\rho_k^S}{\pi_k}}{m}}$. ■

To go forward in the proof of Theorem 3.2, an estimation of the coherence $\mu = \mu\left(\widetilde{\boldsymbol{A}_m}\right)$ is still missing. To overcome it, we derive a corollary of Theorem 4.4.

Corollary 4.5. *Let $\widetilde{\boldsymbol{A}_m}$ be the sampling matrix. Then the coherence $\mu = \mu\left(\widetilde{\boldsymbol{A}_m}\right)$ satisfies*

$$\mu \leq \sqrt{\frac{16\kappa^2\max_{k=1..M}\frac{\|\boldsymbol{B}_k^*\boldsymbol{B}_k\|_\infty}{\pi_k}\ln(2^{3/4}n^2/\varepsilon)}{m}}$$

with probability at least $1-\varepsilon$ provided the right hand side is at most $1/2$.

Proof. Let $(\widetilde{\boldsymbol{a}}_j)_{1\leq j\leq n}$ be the set of columns of $\widetilde{\boldsymbol{A}_m}$. Let $S = (j,k) \subset \{1,\ldots,n\}$ be of cardinality 2. Then the matrix $\left(\widetilde{\boldsymbol{A}_m}^S\right)^*\widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s$ contains $\langle\widetilde{\boldsymbol{a}}_j,\widetilde{\boldsymbol{a}}_k\rangle$ as a matrix entry off the diagonal. We have

$$|\langle\widetilde{\boldsymbol{a}}_j,\widetilde{\boldsymbol{a}}_k\rangle| \leq \left\|\left(\widetilde{\boldsymbol{A}_m}^S\right)^*\widetilde{\boldsymbol{A}_m}^S - \mathrm{Id}_s\right\|_2.$$

By Theorem 4.4, the probability that the operator norm on the right is not bounded by $t \in]0,1/2]$ is at most

$$2^{3/4}2\exp\left(-\frac{mt^2}{8\kappa^2\max_{1\leq k\leq M}\frac{\rho_k^2}{\pi_k}}\right),$$

where $\rho_k^2 = \max_{\substack{S\subset\{1..n\}\\ Card(S)=2}}\left\|\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S\right\|_2 \leq 2\left\|\boldsymbol{B}_k^*\boldsymbol{B}_k\right\|_\infty$, for all $k \in \{1,\ldots,M\}$. Taking the union bound over all $n(n-1)/2 \leq n^2/2$ two element sets $S \subset \{1..n\}$ shows that

$$\mathrm{P}(\mu \geq t) \leq 2^{3/4}n^2\exp\left(-\frac{mt^2}{8\kappa^2 2\max_{1\leq k\leq M}\frac{\|\boldsymbol{B}_k^*\boldsymbol{B}_k\|_\infty}{\pi_k}}\right).$$

Requiring that the right hand side is at most ε leads to the desired solution. ■

Now we can end the proof of Theorem 3.2. Set $\alpha = \frac{\sqrt{s}t}{1-\delta}$, for some $t,\delta \in]0,1/2]$ to be chosen later. By Propositions 4.1 and 4.3, the probability that the recovery by ℓ_1-minimization fails is bounded from above by

$$2^{3/4}(n-s)e^{-\alpha^{-2}/2} + \mathrm{P}\left(\max_{j\in\{1..n\}\ S}\left\|\widetilde{\boldsymbol{A}_m}^{S\dagger}\boldsymbol{a}_\ell\right\|_2 \geq \alpha\right)$$

$$\leq 2^{3/4}(n-s)e^{-\alpha^{-2}/2} + \mathrm{P}\left(\left\|\left(\widetilde{\boldsymbol{A}_m}^{S}\right)^*\widetilde{\boldsymbol{A}_m}^{S} - \mathrm{Id}_s\right\|_2 \geq \delta\right) + \mathrm{P}\left(\mu \geq t\right) \quad (25)$$

By bounding above the right hand side in (25) by 3ε, one can write

$$\begin{cases} \delta & \in]0,1/2] \\ t & \in]0,1/2] \\ \alpha & = \frac{\sqrt{st}}{1-\delta} \leq \frac{1}{\sqrt{2}} \\ \mathrm{P} & (\mu \geq t) \leq \varepsilon \\ \mathrm{P} & \left(\left\|\left(\widetilde{\boldsymbol{A}_m}^{S}\right)^*\widetilde{\boldsymbol{A}_m}^{S} - \mathrm{Id}_s\right\|_2 \geq \delta\right) \leq \varepsilon \\ 2^{3/4} & (n-s)e^{-\alpha^{-2}/2} \leq \varepsilon \end{cases}$$

By Theorem 4.4, we get $\mathrm{P}\left(\left\|\left(\widetilde{\boldsymbol{A}_m}^{S}\right)^*\widetilde{\boldsymbol{A}_m}^{S} - \mathrm{Id}_s\right\|_2 \geq \delta\right)$ is less than ε provided that

$$m \geq \frac{1}{\delta^2}\ln\left(\frac{2^{3/4}s}{\varepsilon}\right) 8\kappa^2 \max_{1\leq k\leq M}\frac{\rho_k^S}{\pi_k}, \quad (26)$$

and Corollary 4.5 asserts that $\mathrm{P}(\mu \geq t) \leq \varepsilon$ provided that

$$m \geq \frac{1}{t^2}\ln\left(\frac{2^{3/4}n^2}{\varepsilon}\right) 16\kappa^2 \max_{1\leq k\leq M}\frac{\|\boldsymbol{B}_k^*\boldsymbol{B}_k\|_\infty}{\pi_k}.$$

We can bound from above the right hand side of Inequation (26), by

$$\frac{1}{\delta^2}\ln\left(\frac{2^{3/4}s}{\varepsilon}\right) 8\kappa^2 s \max_{1\leq k\leq M}\frac{\left\|\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S\right\|_\infty}{\pi_k}$$

Therefore, the probability that the exact recovery by ℓ_1-minimization fails is less than 3ε, if

$$\begin{cases} \delta & \in]0,1/2] \\ t & \in]0,1/2] \\ \alpha & = \frac{\sqrt{st}}{1-\delta} \leq \frac{1}{\sqrt{2}} \\ m & \geq \frac{1}{t^2}\ln\left(\frac{2^{3/4}n^2}{\varepsilon}\right) 16\kappa^2 \max_{1\leq k\leq M}\frac{\left\|\boldsymbol{B}_k^*\boldsymbol{B}_k\right\|_\infty}{\pi_k} \\ m & \geq \frac{1}{\delta^2}\ln\left(\frac{2^{3/4}s}{\varepsilon}\right) 8\kappa^2 s \max_{1\leq k\leq M}\frac{\left\|\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S\right\|_\infty}{\pi_k} \\ \alpha^{-2} & \geq 2\ln\left(\frac{2^{3/4}(n-s)}{\varepsilon}\right). \end{cases}$$

By choosing $t = \frac{(1-\delta)}{\sqrt{s}\sqrt{2\ln\left(\frac{2^{3/4}(n-s)}{\varepsilon}\right)}}$, the required number of measurements should satisfy

$$\begin{cases} \delta & \in]0,1/2] \\ m & \geq \frac{1}{(1-\delta)^2} s \ln\left(\frac{2^{3/4}(n-s)}{\varepsilon}\right)\ln\left(\frac{2^{3/4}n^2}{\varepsilon}\right) 32\kappa^2 \max_{1\leq k\leq M}\frac{\left\|\boldsymbol{B}_k^*\boldsymbol{B}_k\right\|_\infty}{\pi_k} \\ m & \geq \frac{1}{\delta^2}\ln\left(\frac{2^{3/4}s}{\varepsilon}\right) 8\kappa^2 s \max_{1\leq k\leq M}\frac{\left\|\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S\right\|_\infty}{\pi_k}. \end{cases}$$

Considering that

$$\ln\left(\frac{2^{3/4}(n-s)}{\varepsilon}\right)\left(\ln\left(\frac{2^{3/4}n}{\varepsilon}\right) + \ln(n)\right) \leq 2\ln^2\left(\frac{2^{3/4}n}{\varepsilon}\right),$$

leads to

$$\left\{\begin{array}{ll}\delta & \in]0,1/2] \\ m & \geq \frac{1}{(1-\delta)^2} s \ln^2\left(\frac{2^{3/4}n}{\varepsilon}\right) 64\kappa^2 \max_{1\leq k\leq M} \frac{\left\|\boldsymbol{B}_k^*\boldsymbol{B}_k\right\|_\infty}{\pi_k} \\ m & \geq \frac{1}{\delta^2} \ln\left(\frac{2^{3/4}s}{\varepsilon}\right) 8\kappa^2 s \max_{1\leq k\leq M} \frac{\left\|\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S\right\|_\infty}{\pi_k}.\end{array}\right.$$

By choosing $\delta = 1/5$, the bounds become

$$\left\{\begin{array}{ll} m & \geq 200\kappa^2 s \ln^2\left(\frac{2^{3/4}n}{\varepsilon}\right) \max_{1\leq k\leq M} \frac{\left\|\boldsymbol{B}_k^*\boldsymbol{B}_k\right\|_\infty}{\pi_k} \\ m & \geq 200\kappa^2 s \ln\left(\frac{2^{3/4}s}{\varepsilon}\right) \max_{1\leq k\leq M} \frac{\left\|\boldsymbol{B}_k^{S*}\boldsymbol{B}_k^S\right\|_\infty}{\pi_k}.\end{array}\right.$$

Therefore, if

$$m \geq C s \ln^2\left(\frac{2^{3/4}n}{\varepsilon}\right) \max_{1\leq k\leq M} \frac{\left\|\boldsymbol{B}_k^*\boldsymbol{B}_k\right\|_\infty}{\pi_k}$$

with $C = 200\kappa^2$, the exact recovery by ℓ_1-minimization fails with probability $1-3\varepsilon$. We end the proof of Theorem 3.2 replacing ε by $\varepsilon/3$.

The proof of Proposition 3.3 easily follows the result of Theorem 3.2.

4.2 Proofs of the results in 3.2

4.2.1 Acquisition in the 2D Fourier domain and sparsity in the 2D spatial domain

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

$$\boldsymbol{B}_{p_1} = \left[\frac{1}{\sqrt{n}} \exp\left(2i\pi\left(\frac{p_1\ell_1 + p_2\ell_2}{\sqrt{n}}\right)\right)\right]_{(p_1,p_2)(\ell_1,\ell_2)} \tag{27}$$

with $1 \leq p_2 \leq \sqrt{n}, 1 \leq \ell_1, \ell_2 \leq \sqrt{n}$. Let $S \subset \{1,\ldots,\sqrt{n}\} \times \{1,\ldots,\sqrt{n}\}$ denote the support of $\boldsymbol{x}$, with $Card\,(S) = s$. By definition of the 2D Fourier matrix of size $n\times n$, $\left\|\boldsymbol{B}_k^*\boldsymbol{B}_k\right\|_\infty = 1/\sqrt{n}$, for all $k \in \{1,\ldots,\sqrt{n}\}$. Thus, Theorem 3.2 leads to

$$m \geq C s \ln^2\left(\frac{2^{3/4}n}{\varepsilon}\right) \frac{1}{\sqrt{n}} \max_{1\leq k\leq M} \frac{1}{\pi_k}.$$

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

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

and the number of measurements can be written as follows

$$m \geq C s \ln^2\left(\frac{2^{3/4}n}{\varepsilon}\right),$$

which ends the proof of Proposition 3.4.

4.2.2 Acquisition in the 2D Fourier domain, sparsity in the 2D wavelet domain

Let $\mathcal{F}_{2D}$ be the 2D Fourier matrix, let $\mathcal{F}$ be the 1D Fourier matrix, let Ψ be the inverse of the Shannon wavelet decomposition matrix, and let Ψ_{2D} be the inverse of the 2D Shannon wavelet decomposition matrix. We have

$$\mathcal{F}_{2D} = \mathcal{F} \otimes \mathcal{F} \qquad \text{and} \qquad \Psi_{2D} = \Psi \otimes \Psi$$

with $\otimes$ denoting the Kronecker product. We recall that $\sqrt{n} = 2^J$, and that L denotes the coarsest level of resolution. Therefore, we have that

$$\begin{aligned} \boldsymbol{A} &= \mathcal{F}_{2D}^* \Psi_{2D} \\ &= (\mathcal{F} \otimes \mathcal{F})^* (\Psi \otimes \Psi) \\ &= (\mathcal{F}^* \otimes \mathcal{F}^*) (\Psi \otimes \Psi) \\ &= (\mathcal{F}^* \Psi) \otimes (\mathcal{F}^* \Psi) \end{aligned}$$

We define $\boldsymbol{\phi} = \mathcal{F}^* \Psi \in \mathbb{C}^{\sqrt{n} \times \sqrt{n}}$, note that $\boldsymbol{\phi}^* \boldsymbol{\phi} = \boldsymbol{\phi} \boldsymbol{\phi}^* = \mathrm{Id}_{\sqrt{n}}$. Thus, we can write

$$\boldsymbol{A} = \begin{pmatrix} \phi_{1,1} \boldsymbol{\phi} & \dots & \phi_{1,\sqrt{n}} \boldsymbol{\phi} \\ \vdots & \ddots & \vdots \\ \phi_{\sqrt{n},1} \boldsymbol{\phi} & \dots & \phi_{\sqrt{n},\sqrt{n}} \boldsymbol{\phi} \end{pmatrix}$$

Drawing horizontal lines in the 2D Fourier space corresponds to forming the blocks of measurements

$$\boldsymbol{B}_\ell = \begin{pmatrix} \phi_{\ell,1} \boldsymbol{\phi} & \dots & \phi_{\ell,\sqrt{n}} \boldsymbol{\phi} \end{pmatrix}, \quad \forall \ell \in \left\{1, \dots, \sqrt{n}\right\}.$$

Note that $\boldsymbol{B}_\ell \in \mathbb{C}^{\sqrt{n} \times n}$. We may also consider that $\phi_{\ell,i}$ can be renumbered using two indexes $i = (j,k)$ where (j,k) specifies the wavelet at scale $j \in \{L+1, \dots, J\}$ or the scaling function at scale $j = L$ with location parameter k, i.e.

$$(\phi_{\ell,i})_{i=1..\sqrt{n}} = \left(\phi_{\ell,(j,k)}\right)_{\substack{j=L..J \\ k=1...n_j}}.$$

We recall that n_j is defined as

$$n_j = \begin{cases} 2^{j-1} & \text{if} \quad j > L, \\ 2^L & \text{if} \quad j = L. \end{cases}$$

Now let us remark that

$$\phi_{\ell,(j,k)} = \widehat{\psi_{j,k}}(\ell), \qquad \ell \in \left\{1, \dots, \sqrt{n}\right\},$$

is the ℓ-th Fourier coefficient of the Fourier transform of the wavelet $\psi_{j,k}$ (with the convention that $\left(\psi_{(L,k)}\right)_{k=1..n_L}$ denote the scaling functions). Since Shannon's wavelets have compact supports in the Fourier domain that are disjoint at different scales, one has that

$$\operatorname{supp} \widehat{\psi_{j,k}} = \begin{cases} [n_j + 1 \; ; \; 2n_j] & \text{if } j \geq L+1 \\ [1 \; ; \; n_L] & \text{if } j = L. \end{cases}$$

For a given $\ell \in \{1, \dots, \sqrt{n}\}$,

$$Card\left(\{(j,k), \qquad \widehat{\psi_{j,k}}(\ell) \neq 0\}\right) = n_{j_\ell}$$

where j_ℓ is the unique scale verifying

$$\widehat{\psi_{j_\ell,k}}(\ell) \neq 0.$$

Furthermore, since $\boldsymbol{\phi}$ is an orthogonal matrix, and thanks to the fact that

$$\left|\phi_{\ell,(j_\ell,k)}\right| = \frac{1}{\sqrt{n_{j_\ell}}}, \quad k \in \{1,\dots,n_{j_\ell}\}, \quad \text{and} \quad \phi_{\ell,(j,k)} = 0 \quad \text{for} \quad j \neq j_\ell,$$

it follows that

$$\boldsymbol{B}_\ell^* \boldsymbol{B}_\ell = \begin{pmatrix} \phi_{\ell,1}^* \boldsymbol{\phi}^* \\ \vdots \\ \phi_{\ell,\sqrt{n}}^* \boldsymbol{\phi}^* \end{pmatrix} \begin{pmatrix} \phi_{\ell,1}\boldsymbol{\phi} & \dots & \phi_{\ell,\sqrt{n}}\boldsymbol{\phi} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \boldsymbol{\chi}_{j_\ell} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with

$$\boldsymbol{\chi}_{j_\ell} = \begin{pmatrix} \left|\phi_{\ell,(j_\ell,1)}\right|^2 \mathrm{Id}_{\sqrt{n}} & \phi_{\ell,(j_\ell,2)}^* \phi_{\ell,(j_\ell,1)} \mathrm{Id}_{\sqrt{n}} & \dots & \phi_{\ell,(j_\ell,1)}^* \phi_{\ell,(j_\ell,n_{j_\ell})} \mathrm{Id}_{\sqrt{n}} \\ \phi_{\ell,(j_\ell,2)}^* \phi_{\ell,(j_\ell,1)} \mathrm{Id}_{\sqrt{n}} & \ddots & & \phi_{\ell,(j_\ell,2)}^* \phi_{\ell,(j_\ell,n_{j_\ell})} \mathrm{Id}_{\sqrt{n}} \\ & & \ddots & \\ \phi_{\ell,(j_\ell,n_{j_\ell})}^* \phi_{\ell,(j_\ell,1)} \mathrm{Id}_{\sqrt{n}} & & & \left|\phi_{\ell,(j_\ell,n_{j_\ell})}\right|^2 \mathrm{Id}_{\sqrt{n}} \end{pmatrix}$$

which is a hermitian matrix of size $n_{j_\ell}\sqrt{n} \times n_{j_\ell}\sqrt{n}$, satisfying $\|\boldsymbol{\chi}_{j_\ell}\|_\infty = \frac{1}{n_{j_\ell}}$. Then

$$\|\boldsymbol{B}_\ell^* \boldsymbol{B}_\ell\|_\infty = \|\boldsymbol{\chi}_{j_\ell}\|_\infty = \frac{1}{n_{j_\ell}}, \qquad \forall \ell \in \left\{1,\dots,\sqrt{n}\right\}.$$

Therefore, by Theorem 3.2, we can deduce that the required number of blocks of measurements should satisfy

$$m \geq Cs \ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right) \max_{1\leq \ell \leq \sqrt{n}} \frac{1}{n_{j_\ell}\pi_\ell}. \tag{28}$$

Consequently, by Proposition 3.3, the optimal drawing probability Π^* is given by

$$\pi_\ell^* = \frac{1/n_{j_\ell}}{\sum_{\ell'=1}^{\sqrt{n}} 1/n_{j_{\ell'}}} = \frac{1}{n_{j_\ell}(J-L+1)}, \quad \forall \ell \in \left\{1,\dots,\sqrt{n}\right\}.$$

Such a choice for the drawing probability distribution implies the following requirement on the number m of blocks of measurements

$$m \geq Cs \ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right)(J-L+1),$$

which ends the proof of Proposition 3.5.

4.2.3 Acquisition in the dirac basis and sparsity in the 2D Haar domain

Let $\boldsymbol{\phi} \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$ be the inverse 1D Haar wavelet decomposition matrix, and let $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ be the inverse 2D Haar wavelet decomposition matrix. We have

$$\boldsymbol{A} = \boldsymbol{\phi} \otimes \boldsymbol{\phi}$$

with $\otimes$ denoting the Kronecker product. Note that $\boldsymbol{\phi}^*\boldsymbol{\phi} = \boldsymbol{\phi}\boldsymbol{\phi}^* = \mathrm{Id}_{\sqrt{n}}$. Thus

$$\boldsymbol{A} = \begin{pmatrix} \phi_{1,1}\boldsymbol{\phi} & \dots & \phi_{1,\sqrt{n}}\boldsymbol{\phi} \\ \vdots & \ddots & \vdots \\ \phi_{\sqrt{n},1}\boldsymbol{\phi} & \dots & \phi_{\sqrt{n},\sqrt{n}}\boldsymbol{\phi} \end{pmatrix}$$

By drawing horizontal lines in the 2D Dirac space, we form the blocks $\forall \ell \in \{1, \ldots, \sqrt{n}\}$

$$\boldsymbol{B}_\ell = \begin{pmatrix} \phi_{\ell,1}\boldsymbol{\phi} & \ldots & \phi_{\ell,\sqrt{n}}\boldsymbol{\phi} \end{pmatrix}$$

where $\boldsymbol{B}_\ell \in \mathrm{C}^{\sqrt{n}\times n}$.

As in the previous section, we may also consider that $\phi_{\ell,i}$ can be renumbered using two indexes $i = (j,k)$ where (j,k) specifies the wavelet at scale j (or the scaling functions) and with translation parameter k, such that

$$(\phi_{\ell,i})_{i=1..\sqrt{n}} = \left(\phi_{\ell,(j,k)}\right)_{\substack{j=L..J \\ k=1...n_j}}.$$

We recall that n_j is defined as follows

$$n_j = \begin{cases} 2^{j-1} & \text{if} \quad j > L, \\ 2^L & \text{if} \quad j = L. \end{cases}$$

It should be remarked that

$$\operatorname{supp}\left(\boldsymbol{\phi}_{:,(j,k)}\right) = \left[(k-1)\frac{\sqrt{n}}{n_j} + 1 \ , \ k\frac{\sqrt{n}}{n_j}\right],$$

with $\boldsymbol{\phi}_{:,(j,k)}$ denoting the discrete wavelet at scale j (or the scaling functions for $j = L$), location parameter k. By the orthogonality of $\boldsymbol{\phi}$ and the fact that

$$\left|\phi_{\ell,(j,k)}\right| = \frac{\sqrt{n_j}}{n^{1/4}} \quad \text{if } \ell \in \operatorname{supp}\left(\boldsymbol{\phi}_{:,(j,k)}\right), \text{ and } 0 \text{ otherwise,}$$

it follows

$$\boldsymbol{B}_\ell^* \boldsymbol{B}_\ell = \begin{pmatrix} |\phi_{\ell,1}|^2 \,\mathrm{Id}_{\sqrt{n}} & \phi_{\ell,1}^*\phi_{\ell,2}\mathrm{Id}_{\sqrt{n}} & \cdots & \phi_{\ell,1}^*\phi_{\ell,\sqrt{n}}\mathrm{Id}_{\sqrt{n}} \\ \phi_{\ell,2}^*\phi_{\ell,1}\mathrm{Id}_{\sqrt{n}} & \ddots & & \phi_{\ell,2}^*\phi_{\ell,\sqrt{n}}\mathrm{Id}_{\sqrt{n}} \\ & & \ddots & \\ \phi_{\ell,\sqrt{n}}^*\phi_{\ell,1}\mathrm{Id}_{\sqrt{n}} & & & \left|\phi_{\ell,\sqrt{n}}\right|^2 \mathrm{Id}_{\sqrt{n}} \end{pmatrix}$$

from which we can deduce that $\|\boldsymbol{B}_\ell^*\boldsymbol{B}_\ell\|_\infty = \frac{n_J}{\sqrt{n}} = \frac{2^{J-1}}{2^J} = \frac{1}{2}$ for all $\ell \in \{1, \ldots, \sqrt{n}\}$, Theorem 3.2 gives

$$m \geq Cs\ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right)\frac{1}{2}\max_{1\leq\ell\leq\sqrt{n}}\frac{1}{\pi_\ell}. \tag{29}$$

Therefore, by Proposition 3.3, the optimal drawing probability Π^* is defined such that

$$\pi_\ell^* = \frac{1}{\sqrt{n}} = \frac{1}{2^J}, \quad \forall \ell \in \left\{1, \ldots, \sqrt{n}\right\}.$$

Such a choice for the drawing probability distribution implies the following requirement on the number of blocks of measurements

$$m \geq Cs\ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right)2^{J-1},$$

which ends the proof of Proposition 3.6.

Since for all $k \in \{1, \ldots, n\}$, $\|\boldsymbol{a}_k^*\|_\infty^2 = \frac{n_J}{\sqrt{n}} = 1/2$, Theorem 1.1 leads to the following number of isolated measurements

$$\begin{aligned} q &\geq C_R s\ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right)\sum_{k=1}^{n}\|\boldsymbol{a}_k^*\|_\infty^2, \\ &\geq C_R s\ln^2\left(\frac{2^{3/4}3n}{\varepsilon}\right)2^{2J-1}, \end{aligned}$$

which provides Inequality (15).

5 Extension to overlapping blocks

5.1 Statement

Theorem 3.2 only holds for non-overlapping blocks. Allowing blocks to overlap allows to design sampling patterns with much more degrees of freedom and might be valuable in applications. For instance the crossing paths in Figure 1 (c) could have an important practical interest in MRI.

In this section we assume that blocks are defined $\forall k \in \{1, \ldots, M\}$ by

$$\boldsymbol{B}_k = \left(\alpha_i^k \boldsymbol{a}_i^*\right)_{i \in I_k}$$

where $(I_k)_{k=1,\ldots,M}$ is a cover of $\{1, \ldots, n\}$. The coefficients $\alpha_i^k \in \mathbb{C}$ are weights such that $\alpha_i^k = 0$ if $i \notin I_k$ and $\alpha_i^k \neq 0$ if $i \in I_k$.

We randomly select m blocks among $(\boldsymbol{B}_j)_{1 \leq j \leq M}$, according to the discrete probability distribution $\Pi = (\pi_1, \ldots, \pi_M)$. This leads to consider the sequence of i.i.d. random blocks $(\boldsymbol{X}_k)_{1 \leq k \leq m}$ defined by

$$\boldsymbol{X}_k = \frac{1}{\sqrt{\pi_{J_k}}} \boldsymbol{B}_{J_k}, \; k = 1 \ldots m$$

where $(J_k)_{k=1\ldots m}$ are i.i.d. random variables of probability distribution Π. We consider the following random sampling matrix

$$\widetilde{\boldsymbol{A}_m} = \frac{1}{\sqrt{m}} \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix}. \tag{30}$$

The following condition

$$\sum_{\ell=1}^{M} \left|\alpha_i^\ell\right|^2 = 1, \quad \forall i \in \{1, \ldots, n\} \tag{31}$$

ensures that

$$\mathbb{E}\left[\left(\boldsymbol{X}_k^S\right)^* \boldsymbol{X}_k^S\right] = \mathrm{Id}_s, \quad \forall k \in \{1, \ldots, m\}, \tag{32}$$

which is the crucial assumption to write (19) and prove Theorem 3.2. Indeed,

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{X}_k^* \boldsymbol{X}_k\right] &= \sum_{\ell=1}^{M} \boldsymbol{B}_\ell^* \boldsymbol{B}_\ell = \sum_{\ell=1}^{M} \sum_{i \in I_\ell} \left|\alpha_i^\ell\right|^2 \boldsymbol{a}_i \boldsymbol{a}_i^* \\ &= \sum_{i=1}^{n} \sum_{\ell=1}^{M} \left|\alpha_i^\ell\right|^2 \boldsymbol{a}_i \boldsymbol{a}_i^* = \mathrm{Id}_n. \end{aligned}$$

For instance, condition (31) is verified if

$$\alpha_i^\ell = \frac{1}{\sqrt{m_i}}, \quad \forall i \in I_\ell, \tag{33}$$

where $m_i = Card\left(\{k, i \in I_k\}\right)$ is the multiplicity of $\boldsymbol{a}_i^*$, i.e. the number of occurrences of $\boldsymbol{a}_i^*$ in different blocks. Under the above hypotheses we can derive the following theorem.

Theorem 5.1. *Let $S \subset \{1 \ldots n\}$ be a set of cardinality $Card\,(S) = s$ and let $\boldsymbol{\epsilon} = (\epsilon_\ell)_{\ell \in S} \in \mathbb{C}^s$ be a sequence of independent random variables that are uniformly distributed on $\{-1; 1\}$ (or on the torus $\{z \in \mathbb{C}, \, |z| = 1\}$). Let $\boldsymbol{x}$ be an s-sparse vector with support S and $\mathrm{sign}(\boldsymbol{x}^S) = \boldsymbol{\epsilon}$. Let $\widetilde{\boldsymbol{A}_m}$ be the sampling matrix defined in (30) verifying Condition (31).*

Assume that

$$m \geq C s \ln^2\left(\frac{2^{3/4} 3n}{\varepsilon}\right) \max_{1\leq k\leq M} \frac{\|\boldsymbol{B}_k^* \boldsymbol{B}_k\|_\infty}{\pi_k}, \tag{34}$$

where $C = 200\kappa^2$*, and* $\kappa^2 = \left(\frac{\sqrt{17}+1}{4}\right)^2$.

Then with probability at least $1-\varepsilon$ *the vector* $\boldsymbol{x}$ *is the unique solution of the* ℓ_1*-minimization problem* (4).

The proof of Theorem 5.1 is the same as that of Theorem 3.2. Indeed, Condition (31) implies Condition (32), which allows to write the key step (19) and to unwind the proof.

5.2 Example: sampling by overlapping blocks in the 2D Fourier domain

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

$$I_k^{\mathrm{row}} = \left\{i \in \{1,\dots,n\},\ (k-1)\sqrt{n} \leq i \leq k\sqrt{n}\right\} \qquad I_k^{\mathrm{col}} = \left\{k, \sqrt{n}+k, \dots, (\sqrt{n}-1)\sqrt{n}+k\right\}$$

the sets of indexes of $(\boldsymbol{a}_i^*)_{i\in\{1,\dots,n\}}$ 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:

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

We have chosen the normalization factor equal to $1/\sqrt{2}$, as suggested in (33), since each pixel of the image belongs to two blocks: one row and one column. According to Theorem 5.1, we conclude that the required number of blocks of measurements must satisfy

$$m \geq C s \ln^2\left(\frac{2^{3/4} 3n}{\varepsilon}\right) \frac{1}{2\sqrt{n}} \max_{1\leq k\leq M} \frac{1}{\pi_k}. \tag{35}$$

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

$$m \geq C s \ln^2\left(\frac{2^{3/4} 3n}{\varepsilon}\right), \tag{36}$$

which is the same requirement in the 2D Fourier domain without overlapping, see Proposition 3.4.

6 Conclusion

We have introduced new sensing matrices that are constructed by randomly selecting pre-defined blocks of measurements. Such matrices are amenable to implementation in many imaging systems. We have derived theorems that guarantee exact reconstruction using these matrices via ℓ_1-minimization algorithms and outlined the crucial role of one quantity: the block-coherence introduced in Definition 3.1. Our theorems state that if this quantity is low, then CS can be successfully applied. We have illustrated the wide applicability of this theory on time-frequency bases. A typical result states that acquiring lines of the 2D Fourier domain drawn uniformly at random requires no more than $O\left(s\ln^2(n)\right)$ blocks of measurements for perfect reconstruction. Similar results hold for Fourier-Wavelet or Dirac-Wavelet bases. These are probably the most common sampling bases in imaging applications.

A Technical background

In this appendix, we recall tools that are necessary to prove our results. We gather well-known results in algebra and in probability theory. Furthermore, we prove the extension of some probabilistic tools set up in the vector case to the matrix one.

A.1 Matrix tools

Definition A.1 (The Moore-Penrose pseudo-inverse). *If $\boldsymbol{M}$ is a matrix of full rank (i.e. injective) then its Moore Penrose pseudo-inverse is*

$$\boldsymbol{M}^{\dagger} = (\boldsymbol{M}^{*}\boldsymbol{M})^{-1}\,\boldsymbol{M}^{*}.$$

Definition A.2 (Schatten p-norm). *For a matrix $\boldsymbol{M}$ we let $\sigma(\boldsymbol{M}) = (\sigma_1(\boldsymbol{M}), \ldots, \sigma_n(\boldsymbol{M}))$ be its sequence of singular values. Then the Schatten p-norm is defined as*

$$\|\boldsymbol{M}\|_{S_p} := \|\sigma(\boldsymbol{M})\|_p \qquad 1 \le p \le \infty.$$

We can show triangle inequalities for Schatten p-norm [Bha97].

Proposition A.3 (Nice inequalities on matrix norms). *For any matrix $\boldsymbol{M}$*

(i)

$$\|\boldsymbol{M}\|_2 \le \|\boldsymbol{M}\|_{S_p} \qquad 1 \le p \le \infty,$$

(ii)

$$\|\boldsymbol{M}\|_{S_p} \le \mathrm{rk}\,[\boldsymbol{M}]^{1/p}\,\|\boldsymbol{M}\|_2\,.$$

For any square symmetric matrix $\boldsymbol{M} \in \mathrm{C}^{n\times n}$,

(iii)

$$\|\boldsymbol{M}\|_2 \le \|\boldsymbol{M}\|_1\,,$$

where $\|\boldsymbol{M}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|\boldsymbol{M}_{i,j}|$. Note that $\|\boldsymbol{M}\|_1 = \max_{1\le i\le n}\sum_{j=1}^{n}|\boldsymbol{M}_{i,j}|$ when the matrix is symmetric.

Proof. (i) Since a singular value is always positive, and the 2-norm corresponds to the largest singular value, we obtain the first inequality by definition of $\|.\|_{S_p}$.

(ii) Let $\sigma(\boldsymbol{M}) = (\sigma_1(\boldsymbol{M}), \ldots, \sigma_n(\boldsymbol{M}))$ be the sequence of singular values of $\boldsymbol{M}$, we know that if $r = \mathrm{rk}\,[\boldsymbol{M}]$ then $\boldsymbol{M}$ has r non-zero singular values. We deduce that

$$\|\boldsymbol{M}\|_{S_p} = (\sigma_1(\boldsymbol{M})^p + \ldots + \sigma_n(\boldsymbol{M})^p)^{1/p} \le \mathrm{rk}\,[\boldsymbol{M}]^{1/p}\,\|\boldsymbol{M}\|_2\,.$$

■

A.2 Probabilistic tools

Definition A.4 (Rademacher random variable). *A Rademacher random variable is uniformly distributed on $\{-1; 1\}$.*

Definition A.5 (Steinhaus random variable). *A Steinhaus random variable is uniformly distributed on the torus $\{z \in \mathrm{C}; |z| = 1\}$.*

Now, we focus on symmetrization, which allow to derive nice estimates on Rademacher sums. Initial symmetrization Lemma can be found in [LT11], [DlPG99]. Here, we develop a symmetrization lemma in the case of matrices.

Lemma A.6 (Symmetrization). *Assume that $\boldsymbol{\xi} = (\boldsymbol{\xi}_j)_{j=1}^m$ is a sequence of independent random matrices in $\mathrm{C}^{s\times s}$ equiped with a norm $\|.\|$ and having expectation $\boldsymbol{M}_j = \mathrm{E}(\boldsymbol{\xi}_j)$. Then for $1 \le p < \infty$,*

$$\left(\mathrm{E}\left[\left\|\sum_{j=1}^m (\boldsymbol{\xi}_j - \boldsymbol{M}_j)\right\|^p\right]\right)^{1/p} \le 2\left(\mathrm{E}\left\|\sum_{j=1}^m \epsilon_j \boldsymbol{\xi}_j\right\|^p\right)^{1/p}$$

where $\boldsymbol{\epsilon} = (\epsilon_j)_{j=1}^m$ is a Rademacher sequence.

Proof. Let $\boldsymbol{\xi}'$ denote an independent copy of $\boldsymbol{\xi}$. Since $\mathrm{E}\left[\boldsymbol{\xi}'_j\right] = \boldsymbol{M}_j$, an application of Jensen's inequality leads to

$$\begin{aligned}\mathrm{E}\left[\left\|\sum_{j=1}^m (\boldsymbol{\xi}_j - \boldsymbol{M}_j)\right\|^p\right] &= \mathrm{E}\left[\left\|\sum_{j=1}^m (\boldsymbol{\xi}_j - \mathrm{E}\left[\boldsymbol{\xi}'_j\right])\right\|^p\right], \\ &\le \mathrm{E}\left[\left\|\sum_{j=1}^m (\boldsymbol{\xi}_j - \boldsymbol{\xi}'_j)\right\|^p\right]. \end{aligned} \tag{37}$$

Note that $\left(\boldsymbol{\xi}_j - \boldsymbol{\xi}'_j\right)_{j=1}^m$ is a vector of independent symmetric random variables. Hence, it has the same distribution as $\left(\epsilon_j\left(\boldsymbol{\xi}_j - \boldsymbol{\xi}'_j\right)\right)_{j=1}^m$. By (37) and the triangle inequality, we can deduce that

$$\begin{aligned}\left(\mathrm{E}\left[\left\|\sum_{j=1}^m (\boldsymbol{\xi}_j - \boldsymbol{M}_j)\right\|^p\right]\right)^{1/p} &\le \left(\mathrm{E}\left\|\sum_{j=1}^m \epsilon_j\left(\boldsymbol{\xi}_j - \boldsymbol{\xi}'_j\right)\right\|^p\right)^{1/p} \\ &\le \left(\mathrm{E}\left\|\sum_{j=1}^m \epsilon_j \boldsymbol{\xi}_j\right\|^p\right)^{1/p} + \left(\mathrm{E}\left\|\sum_{j=1}^m \epsilon_j \boldsymbol{\xi}'_j\right\|^p\right)^{1/p} \\ &\le 2\left(\mathrm{E}\left\|\sum_{j=1}^m \epsilon_j \boldsymbol{\xi}_j\right\|^p\right)^{1/p}.\end{aligned}$$

■

The next statement is due to Buchholz [Buc01], and see [Tro08] to link the first result to our context.

Lemma A.7 (Non-commutative Khintchine's inequality for matrix valued Rademacher sums)**.** *Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)$ be a Rademacher sequence and let $\boldsymbol{M}_j$, $j = 1\ldots m$ be complex matrices of the same dimension. Choose $n \in \mathrm{N}$, then*

$$\begin{aligned}&\mathrm{E}\left[\left\|\sum_{j=1}^m \epsilon_j \boldsymbol{M}_j\right\|_{S_{2n}}^{2n}\right] \\ &\le \frac{2n}{2^n n!}\max\left(\left\|\left(\sum_{j=1}^m \boldsymbol{M}_j \boldsymbol{M}_j^*\right)^{1/2}\right\|_{S_{2n}}^{2n}, \left\|\left(\sum_{j=1}^m \boldsymbol{M}_j^* \boldsymbol{M}_j\right)^{1/2}\right\|_{S_{2n}}^{2n}\right).\end{aligned} \tag{38}$$

The next statement is an extension of Rudelson's lemma [Rud99] in the case of vectors. In this paper, we extend it to the case of block-structure matrices.

Lemma A.8 (Rudelson's lemma). *Let $\boldsymbol{M}$ be a matrix with a block structure, such that $\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}_1 \\ \vdots \\ \boldsymbol{M}_m \end{pmatrix}$. Let $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_m)$ be a Rademacher sequence. Then for $2 \leq p < \infty$,*

$$
\begin{aligned}
&\mathrm{E}\left(\left\|\sum_{k=1}^{m} \epsilon_k \boldsymbol{M}_k^* \boldsymbol{M}_k\right\|_2^p\right)^{1/p} \\
&\leq 2^{3/4p} r^{1/p} \sqrt{p} e^{-1/2} \left\|\boldsymbol{M}\right\|_2 \max_{k=1..m} \sqrt{\left\|\boldsymbol{M}_k^* \boldsymbol{M}_k\right\|_2},
\end{aligned} \tag{39}
$$

with

$$
r = \mathrm{rk}\left[\left(\sum_{k=1}^{m} \left\|\boldsymbol{M}_k^* \boldsymbol{M}_k\right\|_2 \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right)^{1/2}\right].
$$

Proof. We consider that $p = 2q + 2\theta$, with $q \in \mathrm{N}^*$ and $\theta \in [0,1]$, and we apply Holder's inequality for $\theta \in [0,1]$, then the majoration of the 2-norm by the Schatten p-norm for any $p \geq 1$, see Proposition A.3(ii). Therefore,

$$
\begin{aligned}
\mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^p\right] &= \mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^{2q+2\theta}\right] \\
&= \mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^{(1-\theta)2q} \left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^{\theta(2q+2)}\right] \\
&\leq \left(\mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^{2q}\right]\right)^{1-\theta} \left(\mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^{2q+2}\right]\right)^{\theta}.
\end{aligned} \tag{40}
$$

Using Khintchine's inequality (38) in Lemma A.7 leads to

$$
\begin{aligned}
&\mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^{2q}\right] \\
&\leq \frac{(2q)!}{2^q q!} \max\left(\left\|\left(\sum_{k=1}^{m} \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)^* \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right)^{1/2}\right\|_{S_{2q}}^{2q}, \left\|\left(\sum_{k=1}^{m} \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right) \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)^*\right)^{1/2}\right\|_{S_{2q}}^{2q}\right) \\
&\leq \frac{(2q)!}{2^q q!} \left\|\left(\sum_{k=1}^{m} \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)^2\right)^{1/2}\right\|_{S_{2q}}^{2q}, \qquad q \in \mathrm{N}^*.
\end{aligned} \tag{41}
$$

Let $\boldsymbol{W}_k$ be the matrix such that $\boldsymbol{W}_k = \boldsymbol{M}_k^* \boldsymbol{M}_k$. Matrix $\boldsymbol{W}_k$ is hermitian and then, diagonalisable by a unitary matrix $\boldsymbol{U}_k$ into a real diagonal matrix $\boldsymbol{D}_k$. Therefore,

$$
\begin{aligned}
\left(\boldsymbol{W}_k\right)^2 &= \boldsymbol{U}_k \boldsymbol{D}_k \boldsymbol{U}_k^{-1} \boldsymbol{W}_k \\
&\preceq \boldsymbol{U}_k \left(\left\|\boldsymbol{W}_k\right\|_2 \mathrm{Id}\right) \boldsymbol{U}_k^{-1} \boldsymbol{W}_k \\
&\preceq \left\|\boldsymbol{W}_k\right\|_2 \boldsymbol{W}_k
\end{aligned} \tag{42}
$$

where $\preceq$ represents the partial order on the set of Hermitian matrices, by writing

$$
\boldsymbol{A} \preceq \boldsymbol{B} \iff \boldsymbol{B} - \boldsymbol{A} \text{ is positive semidefinite.}
$$

Step (42) can be rewritten as

$$\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)^2 \preceq \lambda_{\max}\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right).$$

Since $\|\boldsymbol{M}_k^*\boldsymbol{M}_k\|_2\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right) - \left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)^2$ is a semi-definite positive matrix, and $\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)^2$ and $\lambda_{\max}\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)$ are two diagonalisable matrices into the same bases, we can assert that for a given eigenvector, the associated (positive) eigenvalue of $\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)^2$ is inferior to the associated (positive) eigenvalue of $\lambda_{\max}\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)$. Then, we can deduce that

$$\left\|\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)^2\right\|_{S_p} \leq \left\|\|\boldsymbol{M}_k^*\boldsymbol{M}_k\|_2\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right\|_{S_p}, \qquad \forall p \in [1,\infty[\,. \tag{43}$$

Applying (43) to (41), we can write

$$\begin{aligned}
\mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right\|_2^{2q}\right] &\leq \frac{2q!}{2^q q!}\left\|\left(\sum_{k=1}^m \left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)^2\right)^{1/2}\right\|_{S_{2q}}^{2q} \\
&\leq \frac{2q!}{2^q q!}\left\|\left(\sum_{k=1}^m \|\boldsymbol{M}_k^*\boldsymbol{M}_k\|_2\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right)^{1/2}\right\|_{S_{2q}}^{2q}, \qquad q \in \mathrm{N}^*.
\end{aligned}$$

By Proposition A.3(ii)), for a matrix $\boldsymbol{M}$ of rank $\mathrm{rk}\left[\boldsymbol{M}\right]$,

$$\|\boldsymbol{M}\|_{S_p} \leq \mathrm{rk}\left[\boldsymbol{M}\right]^{1/p}\|\boldsymbol{M}\|_2, \qquad \forall p \in [1,\infty[\,.$$

Then, by defining

$$r = \mathrm{rk}\left[\left(\sum_{k=1}^m \|\boldsymbol{M}_k^*\boldsymbol{M}_k\|_2\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right)^{1/2}\right],$$

it follows that

$$\begin{aligned}
&\mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right\|_2^{2q}\right] \\
&\leq \frac{(2q)!}{2^q q!}\mathrm{rk}\left[\left(\sum_{k=1}^m \|\boldsymbol{M}_k^*\boldsymbol{M}_k\|_2\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right)^{1/2}\right]\left\|\left(\sum_{k=1}^m \|\boldsymbol{M}_k^*\boldsymbol{M}_k\|_2\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right)^{1/2}\right\|_2^{2q} \\
&\leq \frac{(2q)!}{2^q q!}r\left\|\sum_{k=1}^m \boldsymbol{M}_k^*\boldsymbol{M}_k\right\|_2^q \max_{k=1..m}\left(\sqrt{\left\|\boldsymbol{M}_k^*\boldsymbol{M}_k\right\|_2}\right)^{2q} \\
&\leq \frac{(2q)!}{2^q q!}r\,\|\boldsymbol{M}^*\boldsymbol{M}\|_2^q \max_{k=1..m}\left(\sqrt{\left\|\boldsymbol{M}_k^*\boldsymbol{M}_k\right\|_2}\right)^{2q} \\
&\leq \frac{(2q)!}{2^q q!}r\,\|\boldsymbol{M}\|_2^{2q} \max_{k=1..m}\left(\sqrt{\left\|\boldsymbol{M}_k^*\boldsymbol{M}_k\right\|_2}\right)^{2q}.
\end{aligned} \tag{44}$$

Therefore, combining (40) and (44) gives

$$\begin{aligned}
\mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right\|_2^{p}\right] &\leq \left(\mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right\|_2^{2q}\right]\right)^{1-\theta}\left(\mathrm{E}\left[\left\|\sum_{k=1}^m \epsilon_k\left(\boldsymbol{M}_k^*\boldsymbol{M}_k\right)\right\|_2^{2q+2}\right]\right)^{\theta} \\
&\leq \left(\frac{(2q)!}{2^q q!}\right)^{1-\theta}\left(\frac{(2q+2)!}{2^{q+1}(q+1)!}\right)^{\theta} r^{1-\theta}r^{\theta}\,\|\boldsymbol{M}\|_2^{2q(1-\theta)}\,\|\boldsymbol{M}\|_2^{2(q+1)\theta} \\
&\quad \max_{k=1..m}\left(\sqrt{\left\|\boldsymbol{M}_k^*\boldsymbol{M}_k\right\|_2}\right)^{2q(1-\theta)} \max_{k=1..m}\left(\sqrt{\left\|\boldsymbol{M}_k^*\boldsymbol{M}_k\right\|_2}\right)^{2(q+1)\theta} \\
&\leq \left(\frac{(2q)!}{2^q q!}\right)^{1-\theta}\left(\frac{(2q+2)!}{2^{q+1}(q+1)!}\right)^{\theta} r\,\|\boldsymbol{M}\|_2^{p} \max_{k=1..m}\sqrt{\left\|\boldsymbol{M}_k^*\boldsymbol{M}_k\right\|_2}^{\,p},
\end{aligned}$$

since $p = 2q + 2\theta$. We recall Stirling's formula

$$q! = \sqrt{2\pi q} q^q e^{-q\lambda_q}$$

with $\frac{1}{12q+1} \leq \lambda_q \leq \frac{1}{12q}$, from which it can be deduced that

$$\frac{(2q)!}{2^q q!} \leq \sqrt{2}\left(\frac{2}{e}\right)^q q^q.$$

Combining the two previous inequalities leads to

$$\mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^p\right] \leq 2^{3/4}\left(\frac{2}{e}\right)^{q+\theta} (q+\theta)^{q+\theta}\, r \left\|\boldsymbol{M}\right\|_2^p \max_{k=1..m} \sqrt{\left\|\boldsymbol{M}_k^* \boldsymbol{M}_k\right\|_2}^p,$$

which implies

$$\left(\mathrm{E}\left[\left\|\sum_{k=1}^{m} \epsilon_k \left(\boldsymbol{M}_k^* \boldsymbol{M}_k\right)\right\|_2^p\right]\right)^{1/p} \leq 2^{3/4p} r^{1/p} \sqrt{p} e^{-1/2} \left\|\boldsymbol{M}\right\|_2 \max_{k=1..m} \sqrt{\left\|\boldsymbol{M}_k^* \boldsymbol{M}_k\right\|_2}$$

since $p = 2q + 2\theta$. ■

Proposition A.9 (Subgaussian tail bound [LT11], [Tro08])**.** *Let Z be a random variable such that*

$$\boldsymbol{E}\left(|Z|^p\right)^{1/p} \leq \alpha\beta^{1/p} p^{1/\gamma} \qquad \forall p \geq p_0$$

for some constants $\alpha, \beta, \gamma, p_0 > 0$, then

$$\mathrm{P}\left(|Z| \geq e^{1/\gamma}\alpha u\right) \leq \beta e^{-\frac{u^\gamma}{\gamma}} \qquad \forall u \geq p_0^{1/\gamma}$$

Proof. By Markov's inequality, and for some arbitrary K, we have

$$\mathrm{P}\left(|Z| \geq e^{1/\gamma}\alpha u\right) \leq \frac{\mathrm{E}\left(|Z|^p\right)}{\left(e^K \alpha u\right)^p} \leq \beta\left(\frac{\alpha p^{1/\gamma}}{e^K \alpha u}\right)^p$$

$p = u^\gamma$ and $K = 1/\gamma$ gives the result. ■

Acknowledgement : The authors would like to thank Nicolas Chauffert and Philippe Ciuciu for fruitful discussion.

References

[Bha97] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Verlag, 1997.

[Buc01] Artur Buchholz. Operator khintchine inequality in non-commutative probability. *Mathematische Annalen*, 319(1):1–16, 2001.

[Can08] Emmanuel J Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[CP11] Emmanuel J. 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 J 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 J Candes, Justin K 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 L Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 47(7):2845–2862, 2001.

[DlPG99] Víctor De la Peña and Evarist Giné. *Decoupling: from dependence to independence*. Springer Verlag, 1999.

[Don06] David Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[EKB10] Yonina C Eldar, Patrick Kuppinger, and Helmut Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *Signal Processing, IEEE Transactions on*, 58(6):3042–3054, 2010.

[Gan07] Lu Gan. Block compressed sensing of natural images. In *Digital Signal Processing, 2007 15th International Conference on*, pages 403–406. IEEE, 2007.

[HPH+11] Robert Hummel, Sameera Poduri, Franz Hover, Urbashi Mitra, and Guarav Sukhatme. Mission design for compressive sensing with mobile robots. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2362–2367. IEEE, 2011.

[LDP07] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine*, 58(6):1182–1195, 2007.

[LDSP08] M. Lustig, D.L. Donoho, J.M. Santos, and J.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.

[LT11] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer, 2011.

[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.

[Mal99] Stéphane Mallat. *A wavelet tour of signal processing*. Academic press, 1999.

[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.

[Rau10] Holger Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[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.

[Rud99] Mark Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164(1):60–72, 1999.

[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.

[Sza04] T.L. Szabó. *Diagnostic Ultrasound Imaging: Inside Out: Inside Out*. Academic Press Series in Biomedical Engineering. Academic Press, 2004.

[Tro08] Joel A Tropp. On the conditioning of random subdictionaries. *Applied and Computational Harmonic Analysis*, 25(1):1–24, 2008.

[WJP+09] Yves Wiaux, Laurent Jacques, Gilles Puy, AMM 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.