



# An analysis of block sampling strategies in compressed sensing

Jérémie Bigot, Claire Boyer, Pierre Weiss

## ► To cite this version:

Jérémie Bigot, Claire Boyer, Pierre Weiss. An analysis of block sampling strategies in compressed sensing. IEEE Transactions on Information Theory, 2016, 62 (4), pp.2125–2139. hal-00823711v4

**HAL Id: hal-00823711**

**<https://hal.science/hal-00823711v4>**

Submitted on 11 Apr 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



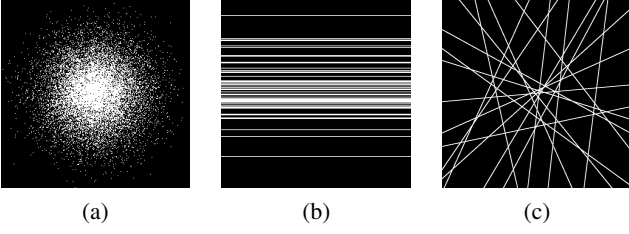


Figure 1. An example of MRI sampling schemes in the  $k$ -space (the 2D Fourier plane where low frequencies are centered) (a): Isolated measurements drawn from a probability distribution with radial distribution. (b): Sampling scheme in the case of non-overlapping blocks of measurements that correspond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

measurements. Our formalism, is based on and extends the work [17], in which only isolated measurements are considered. For instance, this setting 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). The notion of block of measurements allows to encompass structured acquisition, well-spread in various application fields. We study the problem of exact non-uniform recovery of  $s$ -sparse signals in a noise-free setting. This sampling strategy raises various questions. How many blocks of measurements are needed to ensure exact reconstruction of an  $s$ -sparse signal? Is the required number of blocks compatible with faster acquisition?

Our second contribution is to provide preliminary answers to these questions. We extend the theorems proposed in [17] to the case of blocks of measurements for the recovery of  $s$ -sparse signals, only when the degree  $s$  of sparsity is considered. We then show that our result is sharp in a few practical examples and extends the best currently known results in compressed sensing. This work provides some insight on many currently used sampling patterns in MRI, echography, computed tomography scanners, ... by proposing theoretical foundations of block-constrained acquisition.

Our third contribution is to emphasize the limitations of block-constrained acquisition strategies for the recovery of *any*  $s$ -sparse signal: we 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. This result highlights that the standard CS setting focusing on  $s$ -sparse recovery is not appropriate when the acquisition is constrained: the degree  $s$  of sparsity may not be the relevant feature to consider for the signal to reconstruct, when block-structure is imposed in the acquisition.

Overall, we believe that the presented results give good theoretical foundations to the use of blocks of measurements in compressed sensing and show the limitations of this setting for  $s$ -sparse recovery.

*b) Related work:* After submitting the first version of this paper, the authors of [18] 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 IV-C3.

*c) Outline of the paper:* The remaining of the paper is organized as follows. In Section II, we first describe the notation

and the main assumptions necessary to derive a general theory for the acquisition of blocks of measurements. We present the main result of this paper about  $s$ -sparse recovery with block-sampling acquisition in Section III. In Section IV, 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.

## II. PRELIMINARIES

### A. Notation

Let  $S = (S_1, \dots, S_s)$  be a subset of  $\{1, \dots, n\}$  of cardinality  $|S| = s$ . We denote by  $P_S \in \mathbb{C}^{n \times s}$  the matrix with columns  $(e_i)_{i \in S}$  where  $e_i$  denotes the  $i$ -th vector of the canonical basis of  $\mathbb{C}^n$ . For given  $M \in \mathbb{C}^{n \times n}$  and  $v \in \mathbb{C}^n$ , we also define  $M_S = MP_S$ , and  $v_S = P_S^* v$ . We denote by  $\|\cdot\|_p$  the  $\ell_p$ -norm for  $p \in [0, \infty]$ . We will also use  $\|\cdot\|_{p \rightarrow q}$  to denote the operator norm defined by

$$\|M\|_{p \rightarrow q} = \sup_{\|v\|_p \leq 1} \|Mv\|_q.$$

### B. Main assumptions

Recall that we consider the following  $\ell_1$ -minimization problem:

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

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

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} B_1 \\ \vdots \\ B_m \end{pmatrix}, \quad (3)$$

where  $B_1, \dots, B_m$  are independent copies of a random matrix  $B$ , meaning that  $B_1, \dots, B_m$  are independently drawn from the same distribution, satisfying

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

where  $\text{Id}$  is the  $n \times n$  identity matrix. This condition is the extension of the isotropy property described in [17] in a block-constrained acquisition setting.

In most cases studied in this paper, the random matrix  $B$  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  $B$  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 II.1.** Let  $(\mu_i)_{1 \leq i \leq 3}$  denote the smallest positive reals such that the following bounds deterministically hold

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

in which the supremum is taken over all subsets  $S \subset \{1, \dots, n\}$  of cardinality at most  $s$ . Define

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

The quantities introduced in Definition II.1 can be interpreted as follows. The number  $\mu_1$  can be seen as an *intra-support* block coherence, whereas  $\mu_2$  and  $\mu_3$  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  $\mu_2$  and  $\mu_3$  ensure homogeneity between all of these quantities.

### C. Application examples

The number of applications of the proposed setting is large. For instance, it encompasses those proposed in [17]. Let us provide a few examples of new applications below.

1) *Partition of orthogonal transforms:* Let  $A_0 \in \mathbb{C}^{n \times n}$  denote an orthogonal transform. Blocks can be constructed by partitioning the rows  $(a_i^*)_{1 \leq i \leq n}$  from  $A_0$ :

$$B_j = (a_i^*)_{i \in I_j} \quad \text{for } I_j \subset \{1, \dots, n\} \quad \text{s.t.}$$

$$\bigsqcup_{j=1}^M I_j = \{1, \dots, n\},$$

where  $\bigsqcup$  stands for the disjoint union. This case is the one studied in [18].

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

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 be also handled. To do so, define the blocks

$(B_j)_{1 \leq j \leq M}$  as follows:  $B_j = \left( \frac{1}{\sqrt{\alpha_i}} a_i^* \right)_{i \in I_j}$ , where  $\bigcup_{j=1}^M I_j =$

$\{1, \dots, n\}$ , and  $\alpha_i$  denotes the multiplicity of the row  $a_i^*$ , 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  $\mathbb{E}[B^* B] = \text{Id}_n$  where  $B_k$  is defined similarly to the previous example. See Appendix E for an illustration of this setting in the case of 2D Fourier measurements.

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 [5], [17].

3) *Blocks issued from tight or continuous frames:* 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). In this case, the bounds presented in Definition II.1 should be adapted to hold with high probability. 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 [19].

## III. MAIN RESULT

Our main result reads as follows.

**Theorem III.1.** Let  $S \subset \{1, \dots, n\}$  be a set of indices of cardinality  $s$  and suppose that  $x \in \mathbb{C}^n$  is an  $s$ -sparse vector supported on  $S$ . Fix  $\varepsilon \in (0, 1)$ . Suppose that the sampling matrix  $A$  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

$$m \geq c\gamma(s) \left( 2 \ln(4n) \ln(12\varepsilon^{-1}) + \ln s \ln(12e \ln(s) \varepsilon^{-1}) \right)$$

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

The proof of Theorem III.1 is detailed in Section A. It is based on the so-called golfing scheme introduced in [20] for matrix completion, and adapted by [17] for compressed sensing from isolated measurements. Note that Theorem III.1 is a non uniform result in the sense that reconstruction holds for a given support  $S$  of size  $s$  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 a larger number of required measurements.

**Remark III.2** (Improvement of Theorem III.1). By assuming that  $\ln(s) \ln(\ln(s)) \leq c' \ln(n)$ , one can simplify the previous bound in Theorem III.1, by

$$m \geq c'' \gamma(s) \ln(4n) \ln(12\varepsilon^{-1}), \quad (6)$$

for some constants  $c'$  and  $c''$ . Note that this bound can be also obtained considering the trick presented in [21], [22]. In the sequel, for the sake of clarity, we will assume that the condition  $\ln(s) \ln(\ln(s)) \leq c' \ln(n)$  is satisfied and therefore consider the inequality (6).

**Remark III.3** (The case of stochastic bounds). *In Definition II.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 III.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 II.1 could stochastically hold, meaning that the inequalities (5) are satisfied with large probability. This extended setting was actually also proposed in the paper [17]. The proof of the main result can be modified by conditioning the deviation inequalities in the Lemmas of Appendix A to the event that the bounds in Definition II.1 hold. Therefore, even though we do not provide a detailed proof, the lower bound on the sufficient number of blocks in Theorem III.1 remains accurate. Hence, we will propose in Section IV-B some estimates of the quantities (5) in the case of Gaussian measurements.*

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  $O(s\mu \ln(n))$  isolated measurements are sufficient to reconstruct  $\mathbf{x}$  with high probability. The coherence  $\mu$  is the smallest number such that  $\|\mathbf{a}\|_\infty^2 \leq \mu$ . The quantity  $\gamma$  in Theorem III.1 therefore replaces the standard factor  $s\mu$ . The coherence  $\mu$  is usually much simpler to evaluate than  $\gamma$  which depends on three properties of the random matrix  $\mathbf{B}$ : the intra-support coherence  $\mu_1$  and the inter-support coherences  $\mu_2$  and  $\mu_3$ . As will be seen in Section IV, it is important to keep all those quantities in order to obtain tight reconstruction results. Nevertheless, a rough upper bound of  $\gamma$ , reminiscent of the coherence, can be used as shown in Proposition III.4.

**Proposition III.4.** *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

$$\gamma \leq s\mu_4. \quad (7)$$

The proof of Proposition III.4 is given in Appendix B. The bound given in Proposition III.4 is an upper bound on  $\gamma$  that should not be considered as optimal. For instance, for Gaussian measurements, it is important to precisely evaluate the three quantities  $(\mu_i)_{1 \leq i \leq 3}$ .

**Remark III.5** (Noisy setting). *In this paper, we concentrate on a noiseless setting. Noise can be taken into account quite easily by mimicking the proofs in [17]. We do not include such results to clarify the presentation.*

#### IV. RELEVANCY OF THE MAIN RESULT

In this section, we discuss the relevancy of the lower bound given by Theorem III.1 for sparse recovery when *only* the degree of sparsity  $s$  is known. First, we show that Theorem III.1 allows to recover the best known results in compressed sensing: (i) from isolated measurements of an orthogonal transform and (ii) from Gaussian measurements, the results

derived from Theorem III.1 comply with the state-of-the-art results (up to logarithmic factors). Secondly, we study a new setting of interest based on structured measurements drawn from a separable orthogonal transform. We show that the bound on the sufficient number of blocks of measurements cannot be improved modulo logarithmic factors. Finally, we compare our results to a related work [18]. In all these examples, we show that bounds on  $\gamma$  derived from Theorem III.1 are not too restrictive to ensure exact  $s$ -sparse recovery.

##### A. 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 [2], [5], [17]. 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 III.4, for a support  $S$  of cardinality  $s$  the following upper bound holds:

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

Therefore, according to Theorem III.1, it is sufficient that

$$q \geq cs \max_{1 \leq j \leq M} \frac{\|\mathbf{a}_j \mathbf{a}_j^*\|_{1 \rightarrow \infty}}{p_j} \ln(4n) \ln(12\varepsilon^{-1}). \quad (8)$$

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 (8) is the same (up to a multiplicative constant) to that of [17].

In addition, choosing  $\mathcal{P}^*$  in order to minimize the right-hand side of (8) 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 sufficient condition on the number of measurements:

$$q \geq cs \sum_{k=1}^n \|\mathbf{a}_k^*\|_\infty^2 \ln(4n) \ln(12\varepsilon^{-1}). \quad (9)$$

Contrarily to common belief, the probability distribution minimizing the sufficient number of measurements is not the uniform one, but the one depending on the  $\ell_\infty$ -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 (8) 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 (8) 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 [23] which promote non-uniform sampling strategies in compressed sensing.

### B. 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 bounds introduced in Definition II.1 can be shown to hold with high probability, see Section C. With a same argument as in [17], one can show that this stochastic control is enough to derive recovery guarantees with high probability as in the following proposition. The matrix  $\mathbf{A}$  constructed by concatenating those 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 III.1 is sharp, one can expect that  $q = O(s \ln(n))$  measurements are enough to perfectly reconstruct  $\mathbf{x}$ . In what follows, we show that this is indeed the case.

**Proposition IV.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,  $\gamma = O\left(\frac{s \ln(s)}{p}\right)$ . Therefore,  $O\left(\frac{s \ln(s) \ln(n)}{p}\right)$  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 [1]. A proof of this result is presented in Section C.

### C. 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 [24] or products of Fourier and wavelet transforms met in MRI [25] or radio-interferometry [12]. 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 [14]. The authors propose to undersample the 2D  $k_x$ - $k_y$  plane and sample continuously along the orthogonal direction  $k_z$  (see Figure 2).

The remaining of this Section is as follows. We first introduce the notation. We then provide theoretical results about

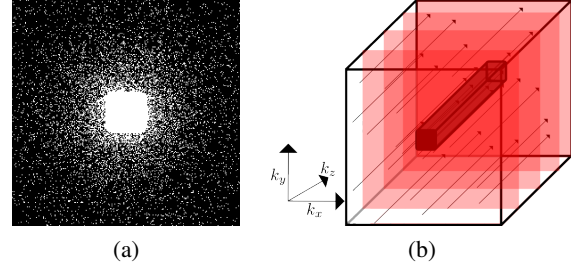


Figure 2. Example of sampling pattern used in MRI [14]. (a) Visualization in the  $k_x$ - $k_y$  plane. (b) Visualization in 3D.

the minimal amount of blocks necessary to reconstruct all  $s$ -sparse vectors. Next, we show that Theorem III.1 is sharp in this setting since the sufficient amount of blocks to reconstruct  $s$ -sparse vectors coincides with the necessary minimal amount. Finally, we perform a comparison with the results in [18].

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

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

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

$$\mathbf{B}_k = \Psi_{k,:} \otimes \Psi \quad (10)$$

$$= [\Psi_{k,1} \Psi, \dots, \Psi_{k,\sqrt{n}} \Psi] \in \mathbb{C}^{\sqrt{n} \times n}. \quad (11)$$

For instance, if  $\Psi$  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 [14]. Similarly to Section II-C1, a sensing matrix  $\mathbf{A}$  can be constructed by drawing  $m$  i.i.d. blocks with distribution II. Letting  $K = (k_1, \dots, k_m) \in \{1, \dots, \sqrt{n}\}^m$  denote the drawn blocks indexes,  $\mathbf{A}$  reads:

$$\begin{aligned} \mathbf{A} &= \frac{1}{\sqrt{m}} \begin{bmatrix} \frac{\mathbf{B}_{k_1}}{\sqrt{\pi_{k_1}}} \\ \vdots \\ \frac{\mathbf{B}_{k_m}}{\sqrt{\pi_{k_m}}} \end{bmatrix} \\ &= \left( \frac{D(\pi)^{-1/2}}{\sqrt{m}} \cdot \Psi_{K,:} \right) \otimes \Psi \\ &= \tilde{\Psi}_{K,:} \otimes \Psi \end{aligned} \quad (12)$$

where  $D(\pi) := \text{diag}(\pi_{k_1}, \dots, \pi_{k_m})$  and  $\tilde{\Psi}_{K,:} := \frac{D(\pi)^{-1/2}}{\sqrt{m}} \cdot \Psi_{K,:}$ . By combining the results in Theorem III.1 and Proposition III.4, we easily get the following reconstruction guarantees.

**Proposition IV.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*

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



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 IV.3.** *The drawing probability distribution  $\Pi^*$  minimizing the right hand side of Inequality (13) on the sufficient 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  $\Pi^*$ , the right hand side of Inequality (13) can be written as follows

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

The sharpness of the bounds on the sufficient number of measurements in Corollary IV.3 will be discussed in the following paragraph.

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  $m = O(s/p \ln(n))$  blocks of measurements to perfectly recover all  $s$ -sparse vectors. Indeed, it is known since [2] that  $O(s \ln(n))$  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 *any*  $s$ -sparse vectors. It means that this specific block structure is inadequate to obtain strong reconstruction guarantees. This result also shows that Corollary IV.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  $\Delta$  is not necessarily a linear mapping.

**Lemma IV.4.** [26, Lemma 3.1] *Set  $\Sigma_s$  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  $\Delta$  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 IV.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 IV.5.

**Proposition IV.5.** *Assume that the sensing matrix  $\mathbf{A}$  has the special block structure described in (12). If  $m < \min(2s, \sqrt{n})$ , then there exists no decoder  $\Delta$  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  $\min(2s, \sqrt{n})$ .*

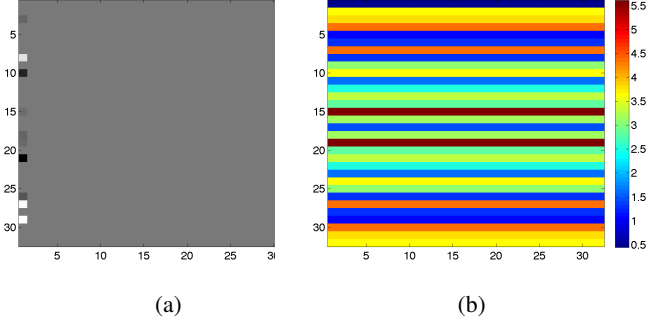


Figure 3. A pathological case where  $n = 32 \times 32$  (a): The signal is  $s$ -sparse for  $s = 10$  and its support is concentrated on its first column. (b) Its 2D Fourier transform is constant along horizontal lines in the Fourier plane.

Proposition IV.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 (10). 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 IV.3, one can derive the following result.

**Proposition IV.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

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

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

3) *Relation to previous work:* To the best of our knowledge, the only existing compressed sensing results based on blocks of measurements appeared in [18]. 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 [18] only covers the case described in Section II-C1 (i.e. partitions of orthogonal transforms), our work covers the case of overlapping blocks of measurements (see Section II-C2), subsampled tight or continuous frames (see Section II-C3), and it can also be extended to the case of randomly generated blocks (see Section II-C4). Last but not least, the work [18] 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 Section IV-A for an edifying example).

Apart from those contextual differences, the comparison between the results in [18] and the ones in this paper is not straightforward. The criterion in [18] 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 [18] is

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

which should be compared to our result

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

As shown in the previous paragraphs, the bound (17) is tight 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 [18].

## V. 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  $\ell_1$ -minimization algorithms and outlined the crucial role of two properties: the *extra* and *intra* support block-coherences introduced in Definition II.1. We have shown that our main result (Theorem III.1) coincides with the best so far results for isolated measurements and is tight for a few sampling schemes used in actual applications.

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 shown in Section IV-C2 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  $\gamma$  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 IV.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 [27], [28] highlight the central role of *structured sparsity* to explain the practical success of compressed sensing. In Figure 5, we aim at reconstructing 2D sparse signals in the spatial domain using a same and unique sampling scheme based on horizontal lines in the Fourier domain and presented at the top of the Figure. We consider three  $s$ -sparse signals with different kinds of sparsity: in (a) the sparsity structure is supported on a column. This pattern is pathological for such a sampling setting as Proposition IV.5 suggests; in (d), the sparsity is uniformly distributed in the spatial domain; in (g), the sparsity structure is supported on a row and it is actually a rotated version of (a). We run an  $\ell_1$ -based reconstruction algorithm and the reconstructed signals are displayed in (b,e,h).

- In (b), we do not reconstruct at all the signal with the “worst” support for such acquisition constraints. This was predicted by Proposition IV.5: there are not enough sensed horizontal lines in the Fourier domain to reconstruct a signal with such a sparsity structure.
- In (e), we are able to partially reconstruct the signal with an “unstructured” sparsity. In (f), we show the difference image between the reconstructed and the original images.
- In (h), we perfectly recover the original image: the structured sparsity presented in (g) seems very adapted to these sampling modalities.

This short experiment highlights that the reconstruction quality does not only depend on the structure in the acquisition but also on how the structured sparsity of the signal to reconstruct is adapted to it.

A very promising perspective is therefore to couple the ideas of structured sparsity in [27], [28] 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. We have proposed new strategies to develop such a theory in [29].

## ACKNOWLEDGMENTS

The authors would like to warmly thank Anders Hansen for his careful reading and comments on the manuscript.



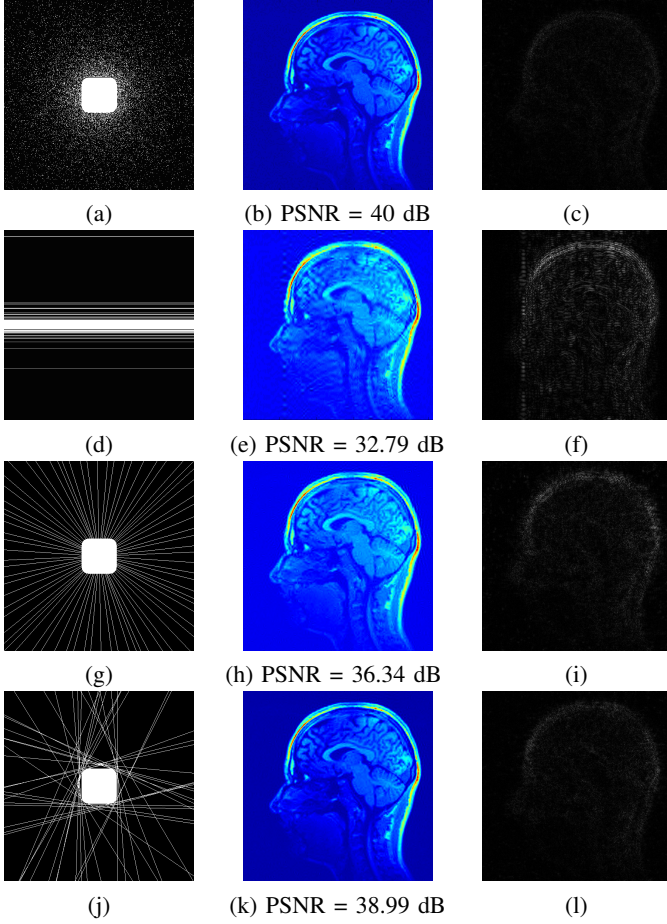


Figure 4. Reconstruction results using different sampling strategies. Each sampling pattern contains 10% of the total number of possible measurements. From top to bottom: measurements drawn independently at random with a radial distribution - horizontal lines in the Fourier domain - deterministic radial sampling - heuristic method proposed in [30]. From left to right: sampling scheme - corresponding reconstruction - difference with the reference (the same colormap is used in every experiment).

Claire Boyer was funded by the CIMI (Centre International de Mathématiques et d'Informatique) Excellence program.

#### APPENDIX

**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). [31, 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*

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

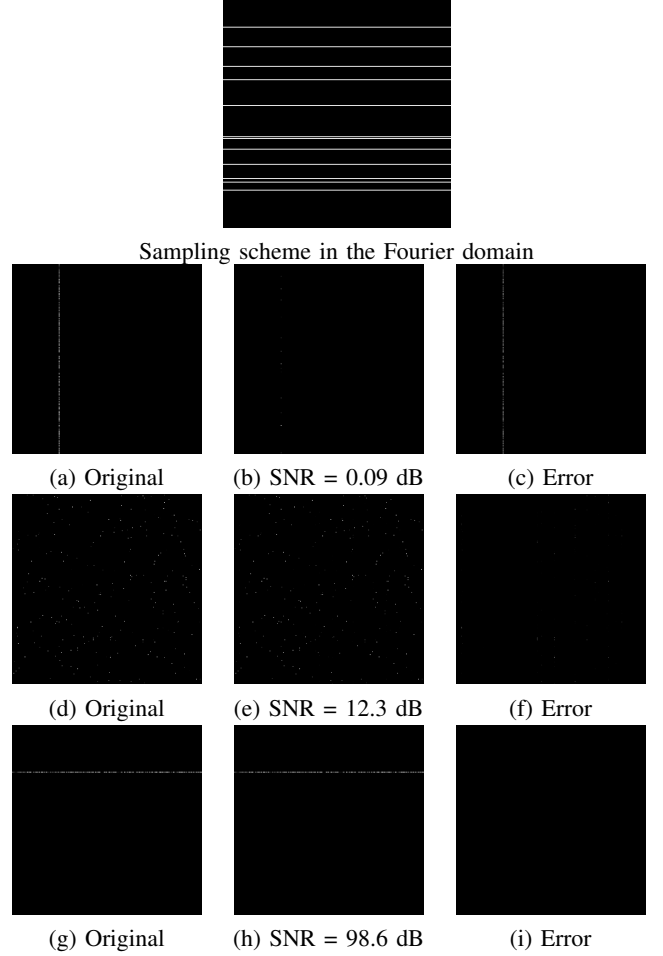


Figure 5. Illustration of a possible missing key feature for structured acquisition: the structured sparsity. In (a,d,g) we present 3 signals with the same degree of sparsity in the spatial domain, but (a) corresponds to a pathological vector introduced in Section IV-C, (d) has an "unstructured" sparsity and (g) is the rotation of 90° of (a). The same sampling scheme, based on horizontal lines in the Fourier domain, is used for all the reconstructions and it is presented at the top. In (b)(e)(h), we display the corresponding reconstructions. In (c,f,i), we display the difference images between reconstructed and original signals (on a same gray scale).

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)). [17, 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*

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

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)). [5, 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

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

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  $\sigma^2$  where  $\sigma^2 \geq \mathbb{E}Z^2$ .

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  $i \in S^c$ ,  $\mathbb{E}(\mathbf{B}_S^*\mathbf{B}\mathbf{e}_i) = 0$ .

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

**Lemma A.5.** Let  $S \subset \{1, \dots, n\}$  be of cardinality of  $s$ . Then, for any  $\delta > 0$ , one has that

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

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

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

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  $\|\mathbf{B}_{k,S}^*\mathbf{B}_{k,S}\|_{2 \rightarrow 2} \leq \mu_1$ , we have that

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

Lastly, we remark that

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

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

**Lemma A.6.** 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

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

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

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

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

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

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

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

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

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

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

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

Using the above upper bounds, namely  $\|\frac{1}{m}\mathbf{y}_k\|_2 \leq \frac{\mu_1}{m}$  and  $\mathbb{E}[Z^2] \leq \frac{\mu_1 - 1}{m}$ , 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. ■

**Lemma A.7.** Let  $S \subset \{1, \dots, n\}$ , such that  $|S| = s$ . Let  $\mathbf{v}$  be a vector of  $\mathbb{C}^s$ . Then we have

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

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

$$\begin{aligned}\|A_{S^c}^* A_S v\|_\infty &= \max_{i \in S^c} \langle e_i, A^* A_S v \rangle \\ &= \max_{i \in S^c} \frac{1}{m} \sum_{k=1}^m \langle e_i, B_k^* B_{k,S} v \rangle.\end{aligned}$$

Let us define  $Z_k = \frac{1}{m} \langle e_i, B_k^* B_{k,S} v \rangle$ . Note that  $\mathbb{E} Z_k = 0$ . From the Cauchy-Schwarz inequality, we get

$$\begin{aligned}|Z_k| &= \left| \frac{1}{m} \langle e_i, B_k^* B_{k,S} v \rangle \right| = \left| \frac{1}{m} v^* B_{k,S}^* (B_k e_i) \right| \\ &\leq \frac{1}{m} \|v\|_2 \|B_{k,S}^* (B_k e_i)\|_2 \\ &\leq \frac{1}{m} \frac{\mu_2}{\sqrt{s}}.\end{aligned}$$

Furthermore,

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

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

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

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

**Lemma A.8.** Let  $S$  be a subset of  $\{1, \dots, n\}$ . Then, for any  $0 < t < \frac{\mu_1}{\mu_2}$ , one has that

$$\begin{aligned}\mathbb{P} \left( \max_{i \in S^c} \|A_S^* A e_i\|_2 \geq t \right) \\ \leq n \exp \left( -\frac{\left( \sqrt{m/\mu_1} t - 1 \right)^2}{4} \right).\end{aligned}\tag{E4}$$

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

$$x_k := \frac{1}{m} B_{k,S}^* B_k e_i.$$

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

$$\|A_S^* A e_i\|_2 = \left\| \frac{1}{m} \sum_{k=1}^m B_{k,S}^* B_k e_i \right\|_2 = \left\| \sum_{k=1}^m x_k \right\|_2.$$

It follows that

$$\|x_k\|_2 = \frac{1}{m} \|B_{k,S}^* B_k e_i\|_2 \leq \frac{1}{m} \frac{\mu_2}{\sqrt{s}}.$$

Furthermore, using Cauchy-Schwarz inequality, one has that

$$\begin{aligned}\mathbb{E} \|x_1\|_2^2 &= \frac{1}{m^2} \mathbb{E} \|B_{1,S}^* B_1 e_i\|_2^2 \\ &\leq \frac{1}{m^2} \mathbb{E} \|B_{1,S}^*\|_{2 \rightarrow 2}^2 \|B_1 e_i\|_2^2 \\ &\leq \frac{1}{m^2} \mu_1 \mathbb{E} \|B_1 e_i\|_2^2 = \frac{1}{m^2} \mu_1 \|e_i\|_2^2 \\ &\leq \frac{1}{m^2} \mu_1.\end{aligned}$$

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

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

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

#### A. Proof of Theorem III.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  $x$  is the unique minimizer of (2), see [17]. 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  $M|_R$  denotes the restriction of a square matrix  $M$  to its range  $R$ , and we define

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

as the operator norm of the inverse of  $M|_R$  restricted to its range.

**Lemma A.9** (Inexact duality [17]). Suppose that  $x \in \mathbb{R}^n$  is supported on  $S \subset \{1, \dots, n\}$ . Then, assume that

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

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

$$\|v_S - \operatorname{sign}(x_S)\|_2 \leq 1/4 \quad \text{and} \quad \|v_{S^c}\|_\infty \leq 1/4, \tag{19}$$

Then, the vector  $x$  is the unique solution of the minimization problem (2).

For ease of reading, we will use the shorthand notation  $\mu_i$  for  $i = 1, 2, 3$  or  $\gamma$  instead of  $\mu_i(s)$  or  $\gamma(s)$ . First, let us focus on Conditions (18). We can remark that

$$\begin{aligned}\|(A_S^* A_S)|_S^{-1}\|_{2 \rightarrow 2} &= \left\| \sum_{k=1}^{\infty} (A_S^* A_S - \operatorname{Id}_S)^k \right\|_{2 \rightarrow 2} \\ &\leq \sum_{k=1}^{\infty} \|A_S^* A_S - \operatorname{Id}_S\|_{2 \rightarrow 2}^k.\end{aligned}$$

Therefore, if the condition  $\|A_S^* A_S - \operatorname{Id}_S\|_{2 \rightarrow 2} \leq 1/2$  is satisfied, then  $\|(A_S^* A_S)|_S^{-1}\|_{2 \rightarrow 2} \leq 2$ . Hence, by Lemma A.5,

it is clear that  $\|(\mathbf{A}_S^* \mathbf{A}_S)^{-1}\|_{2 \rightarrow 2} \leq 2$  with probability at least  $1 - \varepsilon$ , provided that

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

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

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

Furthermore, using Lemma A.8, 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

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

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

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

Conditions (19) remain to be verified. The rest of the proof of Theorem III.1 relies on the construction of a vector  $\mathbf{v}$  satisfying the conditions described in Lemma A.9 with high probability. To do so, we adapt the so-called golfing scheme introduced by Gross [20] and adapted by [17] to  $\ell_1$ -reconstruction, to our setting. More precisely, we will iteratively construct a vector that converges to a vector  $\mathbf{v}$  satisfying (19) with high probability. The main differences with the work in [17] are

- we catch the block-structured acquisition in the estimates of Section A,
- we modify the golfing scheme by partitioning the sensing matrix  $\mathbf{A}$  into blocks of blocks of measurements. By doing so, we can deduce conditions on a sufficient number of blocks of measurements to ensure exact recovery.

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 A 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 A.9 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  $1 \leq \ell \leq L$ ,

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

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  $\ell$  increases. Therefore,  $\mathbf{v}_S^{(\ell)}$  becomes close to  $\mathbf{e}$  as  $\ell$  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  $\varepsilon$ ).

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

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

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

The values of the quantities  $t_\ell$  and  $r_\ell$ , introduced in the above equations, will be specified later in the proof. Note that using (25), we can write that

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

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 respective probability that the upper bounds (25) and (26) do not hold. Now, let us set the number of blocks of blocks  $L$ , the number of blocks  $m_\ell$  in each  $\mathbf{A}^{(\ell)}$  and the values of the parameters  $t_\ell$  and  $r_\ell$  that have been introduced above. We propose to make the following choices :

- (i)  $L = 2 + \left\lceil \frac{\ln(s)}{2\ln 2} \right\rceil$ ,
- (ii)  $m_1, m_2 \geq c\gamma \ln(4n) \ln(2\varepsilon^{-1})$   
 $m_\ell \geq c\gamma \ln(2L\varepsilon^{-1})$ , for  $\ell = 3, \dots, L$ , for some sufficiently large  $c \geq 1$ ,
- (iii)  $r_1, r_2 = \frac{1}{4\sqrt{\ln 4n}}$ ,  
 $r_\ell = \frac{1}{4}$ , for  $\ell = 3, \dots, L$ ,
- (iv)  $t_1, t_2 = \frac{1}{8\sqrt{s}}$ ,  
 $t_\ell = \frac{\ln(4n)}{8\sqrt{s}}$ , for  $\ell = 3, \dots, L$ .

With such choices, we obtain that

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

and

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

Furthermore, using (27), we obtain that

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

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

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

For such a choice of parameters, and by Lemmas A.6 and A.7, 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 A.6 and A.7 we finally obtain that if the overall number  $m$  of blocks samples obeys the condition

$$\begin{aligned} m &= \sum_{\ell=1}^L m_\ell \\ &\geq c\gamma (2\ln(4n) \ln(2\varepsilon^{-1}) + (L-2) \ln(2L\varepsilon^{-1})), \end{aligned}$$

which can be simplified into

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

then the random vector  $v$ , defined by (24), satisfies Assumptions 19 of Lemma A.9 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 A.9 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 III.1 by replacing  $\varepsilon$  by  $\varepsilon/6$ . The final result on the sufficient number of blocks measurements reads as follows

$$m \geq c\gamma(s) \left( 2\ln(4n) \ln\left(\frac{12}{\varepsilon}\right) + \ln s \ln\left(\frac{12e \ln(s)}{\varepsilon}\right) \right),$$

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 [21], and reused in [22].

### B. Proof of Proposition III.4

By Definition II.1, 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:

$$\begin{aligned} \|B_S^* B_S\|_{2 \rightarrow 2} &\leq \|B_S^* B_S\|_{\infty \rightarrow \infty} \leq s \|B^* B\|_{1 \rightarrow \infty} \\ &\leq s\mu_4. \end{aligned}$$

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

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

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

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

### C. Proof of Proposition IV.1

Let us evaluate the quantities  $(\mu_i)_{1 \leq i \leq 3}$  introduced in Definition II.1 to upper bound  $\gamma$  with high probability. For this purpose, using Theorem 2 in [32], we get that for any  $0 < t < 1$

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

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

$$\begin{aligned} \max_{i \in S^c} \|B_S^* B e_i\|_2 &\leq \max_{i \in S^c} \|B_S^* \|_{2 \rightarrow 2} \|B e_i\|_2 \\ &\leq \sqrt{\|B_S^* B_S\|_{2 \rightarrow 2}} \max_{i \in S^c} \sqrt{\|B e_i\|_2^2}. \end{aligned}$$

We already know that the first term  $\sqrt{\|B_S^* B_S\|_{2 \rightarrow 2}}$  in the above inequality is bounded by  $\sqrt{\frac{s}{p}}$  (up to a constant) with high probability, thanks to the previous discussion on  $\mu_1$ . As for the second term, we use a union bound and the sub-gamma property of the chi-squared distribution, see [33, p.29], to derive that

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

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

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

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

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

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

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

Hence, one can take  $\mu_3 = \frac{s}{p}$ . Combining all these estimates we get that  $\gamma \lesssim \frac{s}{p} \sqrt{\delta \ln(s)}$ . Therefore, assuming that the lower bound on  $m$  in Theorem III.1 still holds in the case of acquisition by blocks made of Gaussian entries, we need  $m = O\left(\frac{s}{p} \ln(s) \ln(n)\right)$  blocks of measurements to ensure exact recovery, that is an overall number of measurements  $q = O(s \ln(s) \ln(n))$ .

#### D. Proof of Proposition IV.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  $e_i$  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  $\mathbf{x} = \alpha \otimes e_1$ , where  $\alpha \in \mathbb{R}^{\sqrt{n}}$  is  $s$ -sparse. Note that every  $\mathbf{x} \in \mathcal{C}_s$  is  $s$ -sparse and that

$$\begin{aligned} A\mathbf{x} &= (\tilde{\Psi}_{K,:} \otimes \Psi) \cdot (\alpha \otimes e_1) \\ &= (\tilde{\Psi}_{K,:} \alpha) \otimes \Psi e_1. \end{aligned}$$

In order to identify every  $s$ -sparse  $\mathbf{x}$  knowing  $\mathbf{y} = A\mathbf{x}$ , 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  $\alpha$ . 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 IV.4.

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

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

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  $m = \sqrt{n}$ .

#### E. Proof of Proposition IV.6

We consider blocks that consist of discrete lines in the 2D Fourier space as in Figure 1(b). We assume that  $\sqrt{n} \in \mathbb{N}$  and that  $A_0$  is the 2D Fourier matrix applicable on  $\sqrt{n} \times \sqrt{n}$  images. For all  $p_1 \in \{1, \dots, \sqrt{n}\}$ ,

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

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  $n \times n$ ,  $\|B_k^* B_k\|_{1 \rightarrow \infty} = 1/\sqrt{n}$ , for all  $k \in \{1, \dots, \sqrt{n}\}$ . Thus, Theorem III.1 leads to

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

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

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

which ends the proof of Proposition IV.6.

Let us illustrate the overlapping setting, in the case of blocks that consist in rows and columns in the 2D Fourier domain. Matrix  $A_0 \in \mathbb{C}^{n \times n}$  is the 2D Fourier transform matrix. We set

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

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:

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

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 IV.3, we conclude that the number of blocks of measurements must satisfy

$$\begin{aligned} m &\geq cs \frac{1}{2\sqrt{n}} \max_{1 \leq k \leq M} \frac{1}{\pi_k} \\ &\quad \times \left( 2 \ln(4n) \ln \left( \frac{12}{\varepsilon} \right) + \ln s \ln \left( \frac{12e \ln(s)}{\varepsilon} \right) \right). \end{aligned} \quad (34)$$



Choosing the uniform probability for  $\Pi^*$ , 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

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

which is the same requirement in the 2D Fourier domain without overlapping, see Proposition IV.6.

## REFERENCES

- [1] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.
- [3] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [4] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathématique*, vol. 346, no. 9, pp. 589–592, 2008.
- [5] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.
- [6] G. Puy, P. Vandergheynst, R. Gribonval, and Y. Wiaux, "Universal and efficient compressed sensing by spread spectrum and application to realistic fourier imaging techniques," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–13, 2012.
- [7] H. Rauhut, J. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 242–254, 2012.
- [8] G. Puy, J. P. Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux, "Spread spectrum magnetic resonance imaging," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 3, pp. 586–598, 2012.
- [9] M. Rudelson and R. Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.
- [10] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [11] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, 2007.
- [12] Y. Wiaux, L. Jacques, G. Puy, A. M. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 3, pp. 1733–1742, 2009.
- [13] M. M. Marim, M. Atlan, E. Angelini, and J.-C. Olivo-Marin, "Compressed sensing with off-axis frequency-shifting holography," *Optics letters*, vol. 35, no. 6, pp. 871–873, 2010.
- [14] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [15] G. A. Wright, "Magnetic resonance imaging," *Signal Processing Magazine, IEEE*, vol. 14, no. 1, pp. 56–66, 1997.
- [16] M. Lustig, S.-J. Kim, and J. M. Pauly, "A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 6, pp. 866–873, 2008.
- [17] E. Candès and Y. Plan, "A probabilistic and ripples theory of compressed sensing," *Information Theory, IEEE Transactions on*, vol. 57, no. 11, pp. 7235–7254, 2011.
- [18] A. C. Polak, M. F. Duarte, and D. L. Goeckel, "Performance bounds for grouped incoherent measurements in compressive sensing," *arXiv preprint arXiv:1205.2118*, 2014.
- [19] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss, "Variable density sampling with continuous sampling trajectories," *SIAM Journal on Imaging Sciences*, in press, 2014.
- [20] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1548–1566, 2011.
- [21] B. Adcock and A. C. Hansen, "Generalized sampling and infinite-dimensional compressed sensing," *Foundations of Computational Mathematics*, to appear, 2015.
- [22] D. Gross, F. Krahmer, and R. Kueng, "A partial derandomization of phaselift using spherical designs," *Journal of Fourier Analysis and Applications*, pp. 1–38.
- [23] F. Krahmer and R. Ward, "Stable and robust sampling strategies for compressive imaging," *IEEE Trans. Image Proc.*, vol. 23, no. 2, pp. 612–622, 2014.
- [24] J. Bobin, J.-L. Starck, and R. Ottensmeyer, "Compressed sensing in astronomy," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 5, pp. 718–726, 2008.
- [25] M. Lustig, D. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing mri," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 72–82, 2008.
- [26] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.
- [27] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," *arXiv preprint arXiv:1302.0561*, 2013.
- [28] B. Adcock, A. C. Hansen, and B. Roman, "The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing," *arXiv preprint arXiv:1403.6540*, 2014.
- [29] C. Boyer, J. Bigot, and P. Weiss, "Compressed sensing with structured sparsity and structured acquisition," *arXiv:1505.01619*, 2015.
- [30] C. Boyer, P. Weiss, and J. Bigot, "An algorithm for variable density sampling with block-constrained acquisition," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1080–1107, 2014.
- [31] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, pp. 389–434, 2012.
- [32] M. Ledoux and B. Rider, "Small deviations for beta ensembles," *Electron. J. Probab.*, vol. 15, pp. no. 41, 1319–1343, 2010.
- [33] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

**Jérémy Bigot** received a PhD in applied mathematics in 2003 from the University Joseph Fourier in Grenoble. From 2003 to 2004 he was a postdoc researcher at the Institute of Statistics, UCL, Belgium. From 2004 to 2014, he was an Assistant Professor at the Institute of Mathematics of Toulouse and the Paul Sabatier University. Since 2014 he is a Professor of Statistics within the Institute of Mathematics of Bordeaux from the University of Bordeaux. His research interests deal mainly with high-dimensional statistics and the use of geometry tools for nonparametric inference for applications in signal processing.

**Claire Boyer** graduated from INSA de Toulouse and Paul Sabatier University in Toulouse in 2012. She received a PhD degree in applied mathematics from Paul Sabatier University in 2015. Since then, she is a teaching and research assistant in the maths department of INSA Toulouse. Her research interests are in the field of applied mathematics and signal processing.

**Pierre Weiss** graduated from INSA de Toulouse in 2004. He received a PhD degree in mathematics and signal processing from the University of Nice-Sophia Antipolis in 2008 and did a post-doc at Hong Kong Baptist University in 2009. From 2010 to 2013 he served as an associate professor at INSA de Toulouse. Since 2013, he is a CNRS researcher at ITAV and IMT. His research interests are in the field of applied mathematics and biomedical imaging.