



HAL
open science

Anisotropic compressed sensing for non-Cartesian MRI acquisitions

Philippe Ciuciu, Anna Kazeykina

► **To cite this version:**

Philippe Ciuciu, Anna Kazeykina. Anisotropic compressed sensing for non-Cartesian MRI acquisitions. [Research Report] CEA Paris Saclay; INRIA; Université Paris Sud, Université Paris Saclay. 2019. hal-02339689

HAL Id: hal-02339689

<https://hal.science/hal-02339689v1>

Submitted on 30 Oct 2019

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.

Anisotropic compressed sensing for non-Cartesian MRI acquisitions

Philippe CIUCIU *

Anna KAZEYKINA †

Abstract

In the present note we develop some theoretical results in the theory of anisotropic compressed sensing that allow to take structured sparsity and variable density structured sampling into account. We expect that the obtained results will be useful to derive explicit expressions for optimal sampling strategies in the non-Cartesian (radial, spiral, etc.) setting in MRI.

Keywords: compressed sensing, MRI, anisotropy, non-Cartesian

1 Introduction

The mathematical problem of compressed sensing (CS) consists in recovering a sparse signal from a small number of measurements. More precisely, we wish to recover a vector $x \in \mathbb{C}^n$ from a vector of measurements $y = Ax$, where $A \in \mathbb{C}^{m \times n}$ is the sensing matrix and $m \ll n$. The signal is said to be s -sparse if it has at most s non-zero entries. The recovery is usually performed by solving the following minimization problem called basis pursuit (BP):

$$\min_{x \in \mathbb{C}^n, y = Ax} \|x\|_{\ell^1}. \quad (1)$$

One of the classical CS results can be formulated as follows [7, 6, 10]. Let $A_0 \in \mathbb{C}^{n \times n}$ satisfy the *isotropy* condition: $A_0^* A_0 = I$, and suppose that the measurement matrix A is constructed by drawing m random rows of A_0 in an independent uniform manner. Define the coherence of matrix A to be $\mu(A) = n \max_i \|a_i^*\|_{\infty}^2$, where a_i^* are the rows of matrix A . This quantity represents the coherence between the sensing and the sparsifying bases (low coherence meaning that a vector in the sparsifying basis is approximately uniformly spread in the sensing basis). If $m \gtrsim \mu(A) s \ln(n/\varepsilon)$, then an s -sparse vector x can be exactly recovered by solving (1) with probability at least $1 - \varepsilon$.

In many applications (including MRI), the sensing matrix A is coherent, meaning that $\mu(A)$ is large. It can be shown that incoherence is typically met between the standard basis and the Fourier basis. However, natural images x have sparse representations not in the pixel basis directly, but rather in wavelet bases, i.e. $x = \Psi z$ with z sparse, which are not incoherent with the Fourier basis.

In practice, uniformly drawn measurements lead to very poor reconstructions. It was observed, however, that MR image reconstruction from undersampled frequencies could be significantly improved by drawing measurements according to variable densities strategies (VDS), preferring low to high frequencies.

*Neurospin, CEA Saclay, Parietal, INRIA, 91191 Gif-sur-Yvette, France, philippe.ciuciu@cea.fr

†Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France, anna.kazeykina@math.u-psud.fr

VDS strategies have received a justification in the CS literature [6, 8, 9]. If the measurements are drawn independently with the probability to draw the j -th measure equal to $\pi_j = \frac{\|a_j^*\|_\infty^2}{\sum_{j=1}^n \|a_j^*\|_\infty^2}$, then an s -sparse vector x can be reconstructed exactly from m measurements with probability at least $1 - \varepsilon$ provided that $m \gtrsim \sum_{j=1}^n \|a_j^*\|_\infty^2 s \ln(n/\varepsilon)$.

It was shown experimentally, however, that this result is not sufficient to explain the success of CS in applications such as MRI [1]. It is in particular due to the fact that in the above result we do not assume any structure (apart from sparsity) in the signals to be recovered. A natural extension would be to consider the structured sparsity approach, where one assumes that some prior information on the support S is known, e.g. sparsity by level in the wavelet domain (see [1] for a comprehensive theory for Fourier sampling, based on isolated measurements under a sparsity-by-levels assumption in the wavelet domain). This strategy allows to incorporate any kind of prior information on the structure of S and to study its influence on the quality of CS reconstructions.

Another obstacle to applying classical CS results in a large number of practical settings is that the isolated measurements are incompatible with the physics of acquisition. For this reason, more recent relevant contributions [5, 2] have addressed structured VDS i.e. over sampling trajectories. This approach allows to give recovery guarantees for block-structured acquisition with an explicit dependency on the support of the vector to reconstruct and it provides many possibilities such as optimizing the drawing probability π or identifying the classes of supports recoverable with block sampling strategies.

Instead of drawing rows of A_0 , which corresponds to probing isolated points in the frequency domain (k-space) in the context of MRI, it was proposed in [5, 2] to draw blocks of rows, which corresponds to drawing independent Cartesian lines in k-space. This framework does not cover, however, the case of non-Cartesian acquisition, e.g. acquisition along radial spokes, whose intersection is given by the center of k-space, or more complex trajectories often used in MRI (spiral or non-parametric SPARKLING trajectories, see Figure 1).

One important aspect of the non-Cartesian setting from the CS theory viewpoint is that when frequencies of the Fourier transform are not taken to be in \mathbb{Z}^d , then the corresponding matrix A_0 no longer fulfills the condition $A_0^* A_0 = I$. The isotropy condition is violated in the non-Cartesian setting leading to the necessity to develop a theory for *anisotropic* CS.

Some classical CS results were extended to the anisotropic setting in [12]. The authors provided a theoretical bound on the number of measurements necessary for the exact reconstruction of a sparse vector x in the case of uniform isolated measurements. Another recent paper [3] on anisotropic CS extends the results of [12] to the infinite-dimensional setting.

In the present work we propose to combine the approaches of [12] and [5, 2] to develop anisotropic CS results that take structured sparsity and variable density structured acquisition into account.

The present note is organised as follows. In section 2 we introduce the notation. In section 3 we present the main result. In section 4 we give the proof of the main theorem. In section 5 we present formulas for optimal sampling densities in the case of isolated measurements and block-structured sampling. The Appendix contains some classical results of probability theory and compressed sensing theory that are used in the proofs of section 4.

Acknowledgements. This work was carried out during a one-year Inria delegation of A. Kazeykina in the Parietal team of Inria at NeuroSpin, CEA Saclay.

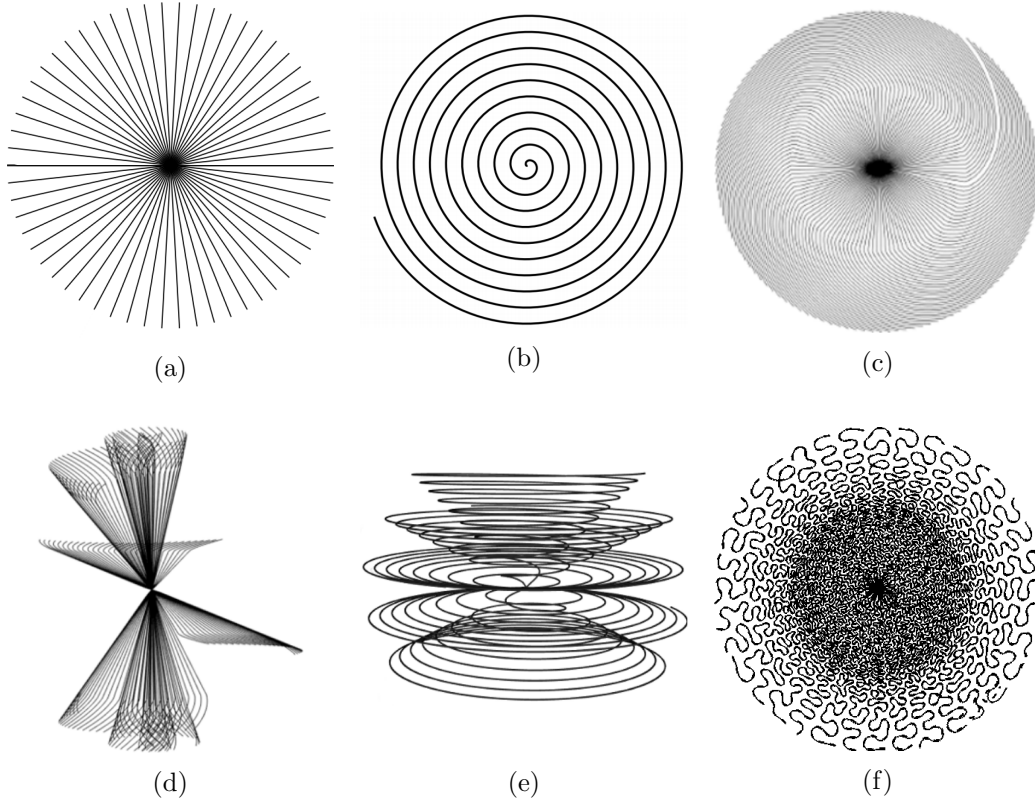


Figure 1: Examples of some non-Cartesian trajectories used in CS to accelerate the acquisition: (a) radial trajectory [13], (b) spiral trajectory, (c) TWIRL: the first combination of radial and spiral trajectories for 2D acquisitions [11], (d) TWIST: a 3D extension of the TWIRL method [4], (e) FLORET: a 3D non-Cartesian sampling pattern based on the combination of radial and spiral, (f) SPARKLING: a trajectory approximating a target variable density and satisfying physical constraints [14].

2 Notation

Let $n \in \mathbb{N}$ and $d_k \in \mathbb{N}$, $k = 1, \dots, M$, such that $\sum_{k=1}^M d_k = n$. Let $\mathcal{B}_k \in \mathbb{C}^{d_k \times n}$ and construct matrix $A_0 \in \mathbb{C}^{n \times n}$ by stacking the blocks \mathcal{B}_k on top of each other: $A_0 = (\mathcal{B}_k)_{k=1}^M$. Matrix A_0 represents the set of possible measurements imposed by a specific sensor device. We will assume that $A_0^* A_0$ is invertible.

The sensing matrix A is constructed by drawing randomly blocks of rows of matrix A_0 . More precisely, let B be a random variable taking values $\mathcal{B}_k / \sqrt{\pi_k}$ with probabilities π_k , $k = 1, \dots, M$. For $m \leq M$ let B_1, \dots, B_m be i.i.d. copies of the random block B . The random sensing matrix is constructed as follows:

$$A = \frac{1}{\sqrt{m}} (B_l)_{l=1}^m. \quad (2)$$

Define

$$X = (\mathbb{E}[B^* B])^{-1}.$$

Note that X exists since $\mathbb{E}[B^* B] = A_0^* A_0$ that we assumed to be invertible. Note also that if A_0 satisfies the isotropy condition $A_0^* A_0 = I_n$, where $I_n \in \mathbb{C}^{n \times n}$ is the identity matrix, then $X = I_n$.

Let $S \subset \{1, 2, \dots, n\}$. Denote $S^c = \{1, \dots, n\} \setminus S$. Define P_S to be the matrix of the linear projection $x \mapsto P_S x$, where $P_S x$ is the restriction of x to the components in S . Define quantities

Θ_S, Λ_S to be positive numbers such that

$$\Theta_S \geq \sqrt{\|B^*BX P_S^*\|_{\infty \rightarrow \infty} \|P_SXB^*B\|_{\infty \rightarrow \infty}} \text{ a.s.} \quad (3)$$

$$\Lambda_S \geq \|P_SXB^*BP_S^*\|_{2 \rightarrow 2} \text{ a.s.} \quad (4)$$

Note that

$$\begin{aligned} \|P_SXB^*BP_S^*\|_{2 \rightarrow 2} &\leq \sqrt{\|P_SXB^*BP_S^*\|_{1 \rightarrow 1} \|P_SXB^*BP_S^*\|_{\infty \rightarrow \infty}} = \\ &\sqrt{\|P_SB^*BX P_S^*\|_{\infty \rightarrow \infty} \|P_SXB^*BP_S^*\|_{\infty \rightarrow \infty}} \leq \sqrt{\|B^*BX P_S^*\|_{\infty \rightarrow \infty} \|P_SXB^*B\|_{\infty \rightarrow \infty}} \leq \Theta_S, \end{aligned}$$

and thus, if Λ_S is taken as the least upper-bound, then

$$\Lambda_S \leq \Theta_S. \quad (5)$$

Note that in the isotropic case ($X = I_n$) Θ_S defined by (3) does not coincide with the quantity $\Theta(S)$ introduced in [5, 2]. That is due to the fact that a straightforward generalisation of the definition used in [5, 2] to the anisotropic case does not preserve the relation (5) verified in the isotropic case. To preserve this relation in the anisotropic case we prefer to consider a symmetrised version of Θ .

We will denote by $(e_i)_{i=1}^n$ the canonical basis of \mathbb{R}^n .

Finally, for a number $x \in \mathbb{C}$ we denote

$$\text{sgn}(x) = \begin{cases} \frac{x}{|x|}, & x \neq 0, \\ 0, & x = 0 \end{cases}$$

and for a vector $x \in \mathbb{C}^n$ we define $\text{sgn}(x) = (\text{sgn}(x_j))_{j=1}^n$.

3 Main result

Theorem 1. *Let $x \in \mathbb{R}^n$ or \mathbb{C}^n be a vector supported on S , such that $\text{sgn}(x_S)$ forms a Rademacher or a Steinhaus sequence. Let A be the random sensing matrix defined by (2) associated with parameter Θ_S . Suppose we are given the data $y = Ax$. Then, given $0 < \varepsilon < 1$ and provided that*

$$m > c \Theta_S (\Theta_S + 2) \ln^2 \left(\frac{8n}{\varepsilon} \right) \quad (6)$$

for c a numerical constant, the vector x is the unique minimizer of the basis pursuit problem (1) with probability at least $1 - \varepsilon$.

Remark 1. Note that the analogous result in the isotropic case (see Theorem 3.3 of [2]) ensures the exact reconstruction of x provided that $m \gtrsim \Theta \ln^2(6n/\varepsilon)$. The possibility to obtain a bound in Θ rather than in Θ^2 is due to the presence of additional symmetries that can be efficiently exploited in the isotropic case (see also Remark 2).

4 Proof

The proof of Theorem 1 is based on the following proposition.

Proposition 1. *For $x \in \mathbb{C}^N$ with support S if*

(i) $P_S X A^* A P_S^*$ is injective,

(ii) $|\langle (P_S X A^* A P_S^*)^{-1} P_S X A^* A e_l, \text{sgn}(x_S) \rangle| < 1 \forall l \in S^c$,

then the vector x is the unique solution of (1).

Proof This proposition is a corollary of Theorem 4.26 of [10] that we formulate as Theorem 2 of the Appendix (see also Corollary 4.28 of [10]).

Indeed, consider the condition (ii) of Theorem 2. First of all, we note that the fact that $P_S X A^* A P_S^*$ is injective implies that $A P_S^*$ is injective. Next, take

$$h = A X P_S^* (P_S A^* A X P_S^*)^{-1} \text{sgn}(x_S)$$

where $(P_S A^* A X P_S^*)^{-1}$ exists due to assumption (i) of the proposition (indeed $P_S A^* A X P_S^*$ is the adjoint of the matrix $P_S X A^* A P_S^*$ which is invertible because it is square and injective).

Then the condition $P_S A^* h = \text{sgn}(x_S)$ is satisfied. Further, the condition $|(A^* h)_l| < 1, l \in S^c$ is rewritten as $|\langle A e_l, h \rangle| < 1, l \in S^c$, which is satisfied if (ii) of Proposition 1 is satisfied. \square

We now formulate and prove two Lemmas that will be used in the proof of Theorem 1.

Lemma 1. For every $S \subset \{1, 2, \dots, n\}$ with $|S| = s$ and for every $\delta > 0$, the following holds

$$\mathbb{P}(\|P_S X A^* A P_S^* - I_s\|_{2 \rightarrow 2} \geq \delta) \leq 2s \exp\left(-\frac{m\delta^2}{4\Lambda_S(2\Lambda_S + \delta/3)}\right)$$

Proof. Let

$$M_i = P_S X B_i^* B_i P_S^* \text{ and } X_i := \frac{1}{m} (M_i - \mathbb{E}M_i), \quad i = 1, 2, \dots, m.$$

Then $P_S X A^* A P_S^* - I_s = \sum_{i=1}^m X_i$.

We have the following estimate:

$$\|X_i\|_{2 \rightarrow 2}^2 = \frac{1}{m^2} \sup_{\|x\|_2 \leq 1} \|(M_i - \mathbb{E}M_i)x\|_2^2 \leq \frac{4\Lambda_S^2}{m^2} =: K^2.$$

Consider $\sum_{i=1}^m \mathbb{E}X_i^* X_i$, this matrix being self-adjoint we have

$$\sigma_1^2 := \left\| \sum_{i=1}^m \mathbb{E}X_i^* X_i \right\|_{2 \rightarrow 2} = \sup_{\|x\|_2 \leq 1} \sum_{i=1}^m \langle x, \mathbb{E}X_i^* X_i x \rangle.$$

Since

$$\langle x, X_i^* X_i x \rangle \leq \frac{4\Lambda_S^2}{m^2},$$

we have that $\sigma_1^2 \leq \frac{4\Lambda_S^2}{m}$. In a similar way,

$$\sigma_2^2 := \left\| \sum_{i=1}^m \mathbb{E}X_i X_i^* \right\|_{2 \rightarrow 2} \leq \frac{4\Lambda_S^2}{m}.$$

Finally, the required result follows from Proposition 4 of Appendix; it suffices to set $t = \delta$, $B = K = \frac{2\Lambda_S}{m}$. \square

Lemma 2. Let $S \subset \{1, 2, \dots, n\}$. Then, for every $t > 0$

$$\mathbb{P}\left(\max_{i \in S^c} \|P_S X A^* A e_i\|_2 \geq \Theta_S / \sqrt{m} + t\right) \leq n \exp\left(-\frac{mt^2/2}{4\Theta_S^2 + 4\Theta_S^2/\sqrt{m} + 2\Theta_S t/3}\right).$$

Proof. Fix $i \in S^c$ and define $N_j = P_S X B_j^* B_j$, $Y_j = \frac{1}{m}(N_j e_i - \mathbb{E}N_j e_i)$ and $Z = \|\sum_{j=1}^m Y_j\|_2$. Note that $\mathbb{E}Y_j = 0$.

First, due to the definition of Θ_S , we can estimate

$$\|N_j e_i\|^2 = |\langle N_j^* N_j e_i, e_i \rangle| \leq \|N_j^* N_j e_i\|_2 \leq \|N_j^* N_j e_i\|_1 \leq \|N_j^* N_j\|_{\infty \rightarrow \infty} \leq \|N_j^*\|_{\infty \rightarrow \infty} \|N_j\|_{\infty \rightarrow \infty} \leq \Theta_S^2, \quad (7)$$

which is due to the following:

$$\|M\|_{\infty \rightarrow \infty} = \max_i \|e_i^* M\|_1.$$

Thus the following estimate is true: $\|Y_j\|_2 \leq 2\Theta_S/m =: K$.

The next required estimate is obtained by using the Cauchy-Schwarz inequality and (7):

$$\begin{aligned} \sup_{\|x\|_2 \leq 1} \sum_{j=1}^m \mathbb{E}|\langle x, Y_j \rangle|^2 &= \frac{1}{m^2} \sup_{\|x\|_2 \leq 1} \sum_{j=1}^m \mathbb{E}|\langle x, N_j e_i - \mathbb{E}N_j e_i \rangle|^2 \leq \\ &\leq \frac{1}{m^2} \sup_{\|x\|_2 \leq 1} \sum_{j=1}^m \mathbb{E}\|x\|_2^2 (\|N_j e_i\|_2 + \mathbb{E}\|N_j e_i\|_2)^2 \leq \frac{4\Theta_S^2}{m} =: \sigma^2. \end{aligned}$$

We use the independence of the vectors Y_j and the fact that they have zero mean value to get the following estimate:

$$(\mathbb{E}Z)^2 \leq \mathbb{E}Z^2 = \mathbb{E}\left\|\sum_{j=1}^m Y_j\right\|_2^2 = \sum_{j=1}^m \mathbb{E}\|Y_j\|_2^2 + \sum_{j=1}^m \sum_{k \neq j} \langle \mathbb{E}Y_j, \mathbb{E}Y_k \rangle = \sum_{j=1}^m \mathbb{E}\|Y_j\|_2^2.$$

Now we use (7) again to estimate:

$$\sum_{j=1}^m \mathbb{E}\|Y_j\|_2^2 = \frac{1}{m^2} \sum_{j=1}^m \mathbb{E}\|N_j e_i - \mathbb{E}N_j e_i\|_2^2 = \frac{1}{m^2} \sum_{j=1}^m (\mathbb{E}\|N_j e_i\|_2^2 - \|\mathbb{E}N_j e_i\|_2^2) \leq \frac{1}{m^2} \sum_{j=1}^m \mathbb{E}\|N_j e_i\|_2^2 \leq \frac{\Theta_S^2}{m},$$

which implies

$$\mathbb{E}Z \leq \sqrt{\Theta_S^2/m} =: \mu.$$

The result then follows from Proposition 5 of Appendix and a union bound. \square

Remark 2. Note that for $X = I_n$ the estimates of Lemmas 1, 2 are looser than those obtained for the isotropic case in [2] (see Lemmas C.1, C.2 of [2]; note Λ^2 instead of Λ and Θ^2 instead of Θ in our results). That is due to the fact that in the isotropic case it is possible to exploit some extra symmetries to obtain tighter estimates.

Proof of Theorem 1. We follow the reasoning proposed in [2].

By Lemma 1, condition (i) of Proposition 1 fails with probability not higher than $2s \exp\left(-\frac{m}{4\Lambda_S(2\Lambda_S+1/3)}\right)$. The latter expression is bounded by $\frac{\varepsilon}{4}$ provided that

$$m \geq 4\Lambda_S(2\Lambda_S + \frac{1}{3}) \ln\left(\frac{8s}{\varepsilon}\right).$$

Now let us study when condition (ii) of Proposition 1 fails. Denote $A_S^\dagger = (P_S X A^* A P_S^*)^{-1} P_S X A^*$. Then, by union bound,

$$\begin{aligned}
\mathbb{P}(\text{(ii) fails}) &= \mathbb{P}\left(\exists l \in S^c: \left|\langle A_S^\dagger A e_l, \text{sgn}(x_S) \rangle\right| \geq 1\right) \\
&\leq \mathbb{P}\left(\exists l \in S^c: \left|\langle A_S^\dagger A e_l, \text{sgn}(x_S) \rangle\right| \geq 1 \text{ and } \max_{l \in S^c} \|A_S^\dagger A e_l\|_2 \leq \alpha\right) + \mathbb{P}\left(\max_{l \in S^c} \|A_S^\dagger A e_l\|_2 \geq \alpha\right) \\
&\leq \sum_{l \in S^c} \mathbb{P}\left(\left|\langle A_S^\dagger A e_l, \text{sgn}(x_S) \rangle\right| \geq \alpha^{-1} \|A_S^\dagger A e_l\|_2 \text{ and } \max_{l \in S^c} \|A_S^\dagger A e_l\|_2 \leq \alpha\right) + \mathbb{P}\left(\max_{l \in S^c} \|A_S^\dagger A e_l\|_2 \geq \alpha\right) \\
&\leq 2n \exp\left(-\frac{1}{2\alpha^2}\right) + \mathbb{P}\left(\max_{l \in S^c} \|A_S^\dagger A e_l\|_2 \geq \alpha\right),
\end{aligned}$$

where the last bound is due to Hoeffding type inequality for Rademacher or Steinhaus sequence (see Propositions 6, 7: we take a_l to be equal to $(P_S X A^* A P_S^*)^{-1} P_S X A^* A e_l$ and we set $u = \alpha^{-1}$, $\lambda = \frac{1}{2}$).

Now we study the second term. Note that

$$\|A_S^\dagger A e_l\|_2 = \|(P_S X A^* A P_S^*)^{-1} P_S X A^* A e_l\|_2 \leq \|(P_S X A^* A P_S^*)^{-1}\|_{2 \rightarrow 2} \|P_S X A^* A e_l\|_2.$$

Take $0 < \delta < 1$ and $\tilde{t} > 0$. Denote $s = |S|$. Let \mathcal{A} be the event that $\|P_S X A^* A P_S^* - I_s\|_{2 \rightarrow 2} < \delta$ and let \mathcal{B} be the event that $\max_{i \in S^c} \|P_S X A^* A e_i\|_2 < \tilde{t}$. Note that \mathcal{A} implies that $\|(P_S X A^* A P_S^*)^{-1}\| < \frac{1}{1-\delta}$.

Set $\alpha = \frac{\tilde{t}}{1-\delta}$. Then $\mathcal{A} \cap \mathcal{B}$ implies that $\max_{l \in S^c} \|A_S^\dagger A e_l\|_2 < \alpha$, and so $\max_{l \in S^c} \|A_S^\dagger A e_l\|_2 \geq \alpha$ means that $\mathcal{A}^c \cup \mathcal{B}^c$ holds, where \mathcal{E}^c denotes the event complementary to \mathcal{E} . Thus we get the following estimate

$$\mathbb{P}(\max_{l \in S^c} \|A_S^\dagger A e_l\|_2 \geq \alpha) \leq \mathbb{P}(\|P_S X A^* A P_S^* - I_s\|_{2 \rightarrow 2} \geq \delta) + \mathbb{P}(\max_{i \in S^c} \|P_S X A^* A e_i\|_2 \geq \tilde{t}).$$

Define

$$P_1 = 2n \exp\left(-\frac{1}{2\alpha^2}\right), \quad P_2 = \mathbb{P}(\|P_S X A^* A P_S^* - I_s\|_{2 \rightarrow 2} \geq \delta), \quad P_3 = \mathbb{P}(\max_{i \in S^c} \|P_S X A^* A e_i\|_2 \geq \tilde{t}).$$

By Lemma 1 the probability P_2 is bounded by $2s \exp\left(-\frac{m\delta^2}{4\Lambda_S(2\Lambda_S + \delta/3)}\right)$. Thus it can be majorised by $\frac{\varepsilon}{4}$ if

$$m \geq \frac{4\Lambda_S(2\Lambda_S + \delta/3)}{\delta^2} \ln\left(\frac{8s}{\varepsilon}\right). \quad (8)$$

Now take $\tilde{t} = \frac{\Theta_S}{\sqrt{m}} + t$ for some $t > 0$. By Lemma 2 probability P_3 is bounded by $\frac{\varepsilon}{4}$, if

$$m \geq \frac{8}{t^2} \Theta_S (\Theta_S + \Theta_S/\sqrt{m} + t/6) \ln\left(\frac{4n}{\varepsilon}\right).$$

If we assume that $m \geq \Theta_S^2$, then we can write that P_3 is bounded by $\frac{\varepsilon}{4}$, if

$$m \geq \frac{8}{t^2} \Theta_S (\Theta_S + 1 + t/6) \ln\left(\frac{4n}{\varepsilon}\right). \quad (9)$$

Finally, set $t = \delta$, assume

$$m \geq c\Theta_S^2 \ln\left(\frac{8n}{\varepsilon}\right), \text{ for some constant } c > 0, \quad (10)$$

and choose $\delta = \sqrt{\frac{1}{c' \ln(\frac{8n}{\varepsilon})}}$ with $\min(c', c) \geq 16$. Then P_1 is bounded by $\frac{\varepsilon}{4}$ if

$$2n \exp\left(-\frac{(1-\delta)^2}{2(\Theta_S/\sqrt{m} + \delta)^2}\right) \leq \frac{\varepsilon}{4} \Leftrightarrow \frac{(1-\delta)^2}{2(\Theta_S/\sqrt{m} + \delta)^2} \geq \ln\left(\frac{8n}{\varepsilon}\right).$$

The latter condition is satisfied due to assumption (10) and the choice of δ . Indeed, due to (10) we have that

$$\frac{\Theta_S}{\sqrt{m}} \leq \frac{1}{\sqrt{c \ln(\frac{8n}{\varepsilon})}}$$

and thus, due to the definition of δ ,

$$\left(\frac{\Theta_S}{\sqrt{m}} + \delta\right)^2 \leq \frac{4}{\min(c, c') \ln(\frac{8n}{\varepsilon})},$$

which implies that

$$\frac{(1-\delta)^2}{2(\Theta_S/\sqrt{m} + \delta)^2} \geq \frac{\min(c, c')}{8} \ln\left(\frac{8n}{\varepsilon}\right) (1-\delta)^2 \geq \ln\left(\frac{8n}{\varepsilon}\right).$$

Plugging the chosen value of δ into (8), we obtain that it suffices to take

$$\begin{aligned} m &\geq 64\Lambda_S(2\Lambda_S + 1) \ln\left(\frac{8n}{\varepsilon}\right) \ln\left(\frac{8s}{\varepsilon}\right), \\ m &\geq 128\Theta_S(\Theta_S + 2) \ln^2\left(\frac{8n}{\varepsilon}\right) \end{aligned}$$

for (BP) to have a unique solution with probability $1 - \varepsilon$. If we choose Λ_S in the definition (4) to be the least upper-bound, then the inequality (5) implies that it suffices to choose m verifying (6) to guarantee the result of Theorem 1. \square

5 Optimal sampling strategies

In this Section we derive formulas for probabilities π_k minimising the quantity Θ_S defined by (3) and arising in the bound (6).

We consider the following two principal sampling strategies.

- **Isolated measurements**

Let $(a_i^*)_{1 \leq i \leq n} \in \mathbb{C}^n$ be a set of row vectors. Set $M = n$, $d_k = 1$ for all k and $\mathcal{B}_k = a_k^*$. This setting represents isolated measurements in MRI.

- **Block-structured sampling** Let $(a_i^*)_{1 \leq i \leq n} \in \mathbb{C}^n$ be a set of row vectors and let $(\mathcal{I}_k)_{1 \leq k \leq M}$ denote a partition of the set $\{1, \dots, n\}$. The rows (a_i^*) are then partitioned into blocks $(D_k)_{1 \leq k \leq M}$: $D_k = (a_i^*)_{i \in \mathcal{I}_k}$. In this setting $d_k = |\mathcal{I}_k|$ and $\mathcal{B}_k = D_k$. This setting corresponds to sampling blocks of measurements in MRI.

Define

$$c_{S,k}^1 = \|a_k\|_\infty \|a_k^* X P_S^*\|_1, \quad c_{S,k}^2 = \|P_S X a_k\|_\infty \|a_k^*\|_1, \quad k = 1, \dots, n.$$

Proposition 2. *In the setting of isolated measurements the probability minimising Θ_S is*

$$\pi_k^{\Theta_S} = \frac{\sqrt{c_{S,k}^1 c_{S,k}^2}}{\sum_{k=1}^n \sqrt{c_{S,k}^1 c_{S,k}^2}}, \quad k = 1, 2, \dots, n.$$

The corresponding Θ_S is given by

$$\Theta_S = \sum_{k=1}^n \sqrt{c_{S,k}^1 c_{S,k}^2}.$$

Define

$$C_{S,k}^1 = \|D_k^* D_k X P_S^*\|_{\infty \rightarrow \infty}, \quad C_{S,k}^2 = \|P_S X D_k^* D_k\|_{\infty \rightarrow \infty}, \quad k = 1, \dots, M.$$

Proposition 3. *In the setting of block-structured sampling the probability minimising Θ_S is*

$$\pi_k^{\Theta_S} = \frac{\sqrt{C_{S,k}^1 C_{S,k}^2}}{\sum_{k=1}^M \sqrt{C_{S,k}^1 C_{S,k}^2}}, \quad k = 1, 2, \dots, M.$$

The corresponding Θ_S is given by

$$\Theta_S = \sum_{k=1}^M \sqrt{C_{S,k}^1 C_{S,k}^2}.$$

The proof of Propositions 2 and 3 follows from Lemma D1 of [2].

Remark 3. Note that for $X = I_n$ we do not find the same expressions for π^{Θ_S} as those presented in [2] for the isotropic case, but rather their symmetrised versions. That difference is due to the difference in the definition of Θ explained in Section 2.

6 Conclusion

In the anisotropic setting we provided a theoretical bound on the number of measurements that are needed to reconstruct a sparse signal with large probability. The bound is given in terms of a quantity that allows to take into account a priori information on the sparsity structure of the signal and to apply variable density block-structured sampling strategies. We have provided general formulas for probability distributions optimising the obtained theoretical bound. We hope that these formulas can be further analysed to derive explicit expressions for optimal sampling densities in the context of non-Cartesian MRI.

Appendix

Proposition 4. *(Matrix Bernstein inequality [15])*

Consider a finite sequence $\{M_k\} \in \mathbb{C}^{d \times d}$ of independent random matrices. Assume that each random matrix satisfies $\mathbb{E}[M_k] = 0$ and $\|M_k\| \leq B$ a.s. and define

$$\sigma^2 = \max \left\{ \left\| \sum_k \mathbb{E}(M_k M_k^*) \right\|_{2 \rightarrow 2}, \left\| \sum_k \mathbb{E}(M_k^* M_k) \right\|_{2 \rightarrow 2} \right\}.$$

Then for all $t \geq 0$,

$$\mathbb{P} \left(\left\| \sum_k M_k \right\|_{2 \rightarrow 2} \geq t \right) \leq 2d \exp \left(-\frac{t^2/2}{\sigma^2 + Bt/3} \right).$$

Proposition 5. (Vector Bernstein inequality [2])

Consider a set of independent random vectors Y_1, Y_2, \dots, Y_m such that

$$\mathbb{E}Y_i = 0, \quad \|Y_i\|_2 \leq K \text{ a.s. } \forall i = 1, \dots, m,$$

and let $\sigma, \mu > 0$ such that

$$\sup_{\|x\|_2 \leq 1} \sum_{i=1}^m \mathbb{E}|\langle x, Y_i \rangle|^2 \leq \sigma^2, \quad \mathbb{E}Z \leq \mu, \quad \text{where } Z := \left\| \sum_{i=1}^m Y_i \right\|_2.$$

Then for every $t > 0$ the following holds:

$$\mathbb{P}(Z > \mu + t) \leq \exp \left(-\frac{t^2/2}{\sigma^2 + 2K\mu + tK/3} \right)$$

Proposition 6. (Hoeffding's bound for Rademacher sequence [10]) Let $a \in \mathbb{C}^M$ and $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_M)$ be a Rademacher sequence. Then

$$\mathbb{P} \left(\sum_{i=1}^M |\epsilon_i a_i| \geq u \|a\|_2 \right) \leq 2 \exp(-u^2/2) \quad \forall u > 0.$$

Proposition 7. (Hoeffding-type bound for Steinhaus sequence [10]) Let $a \in \mathbb{C}^M$ and $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_M)$ be a Steinhaus sequence. Then for any $0 < \lambda < 1$

$$\mathbb{P} \left(\sum_{i=1}^M |\epsilon_i a_i| \geq u \|a\|_2 \right) \leq \frac{1}{1-\lambda} \exp(-\lambda u^2) \quad \forall u > 0.$$

Theorem 2. (Theorem 4.26 of [10]) Given a matrix $A \in \mathbb{C}^{m \times N}$, a vector $x \in \mathbb{C}^N$ with support S is the unique minimizer of $\|z\|_1$ subject to $Az = Ax$ if one of the following equivalent conditions holds:

(i)

$$\left| \sum_{j \in S} \overline{\text{sgn}(x_j)} v_j \right| < \|v_{S^c}\|_1 \text{ for all } v \in \ker A \setminus \{0\},$$

(ii) AP_S^* is injective and there exists a vector $h \in \mathbb{C}^m$ such that

$$(A^*h)_j = \text{sgn}(x_j), \quad j \in S, \quad |(A^*h)_l| < 1, \quad l \in S^c.$$

References

- [1] Adcock B., Hansen A., Poon C., Roman B. Breaking the coherence barrier: A new theory for compressed sensing. Forum of Mathematics, Sigma. 5 (2017)
- [2] Adcock B., Boyer C., Brugiapaglia S. On oracle-type local recovery guarantees in compressed sensing. arXiv:1806.03789 (2018)

- [3] Alberti G.S., Santacesaria M. Infinite dimensional compressed sensing from anisotropic measurements and applications to inverse problems in PDE. *Appl. Comput. Harmon. Anal.* <https://doi.org/10.1016/j.acha.2019.08.002> (2019)
- [4] Boada F., Gillen J., Shen G., Chang S., Thulborn K. Fast three dimensional sodium imaging. *Magnetic resonance in medicine.* 37(5), 706–715 (1997)
- [5] Boyer C., Bigot J., Weiss P. Compressed sensing with structured sparsity and structured acquisition. *Appl. Comput. Harmon. Anal.* 46(2), 312–350 (2019)
- [6] Candes E., Plan Y. A probabilistic and RIPless theory of compressed sensing. *IEEE Trans. Inf. Theory.* 57(11) 7235–7254 (2011)
- [7] Candes E., Romberg J., Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory.* 52(2), 489–509 (2006)
- [8] Chauffert N., Ciuciu P., Weiss P. Variable density compressed sensing in MRI. Theoretical vs heuristic sampling strategies. *IEEE 10th International Symposium on Biomedical Imaging.* 298–301 (2013)
- [9] Chauffert N., Ciuciu P., Kahn, J. Weiss P. Variable density sampling with continuous trajectories. Application to MRI. *SIAM J. Imaging Sci.* 7(4), 1962–1992, (2014)
- [10] Foucart S., Rauhut S. A mathematical introduction to compressive sensing. Birkhäuser, New York (2013)
- [11] Jackson J., Meyer C., Nishimura D., Macovski A. Selection of a convolution function for Fourier inversion using gridding (computerised tomography application). *IEEE Trans. Med. Imag.* 10(3), 473–478 (1991)
- [12] Kueng R., Gross D. RIPless compressed sensing from anisotropic measurements. *Linear Alg. Appl.* 441, 110-123 (2014)
- [13] Lauterbur P. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature.* 242, 190–191 (1973)
- [14] Lazarus C., Weiss P., Chauffert N., Mauconduit F., El Gueddari L., Destrieux C., Zemmoura I., Vignaud A., Ciuciu P. SPARKLING: variable-density k-space filling curves for accelerated T2*-weighted MRI. *Magn. Reson. Med.* 81(6), 3643–3661 (2019)
- [15] Tropp J. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* 12(4), 389–434 (2012)