THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *23/06/2015* **par :**

CLAIRE BOYER

Block-constrained Compressed Sensing

JURY

JÉRÉMIE BIGOT	Professeur des Universités	Directeur
PHILIPPE CIUCIU	Directeur de Recherche	Examinateur
ALBERT COHEN	Professeur des Universités	Examinateur
JALAL FADILI	Professeur des Universités	Examinateur
FABRICE GAMBOA	Professeur des Universités	Examinateur
ANDERS HANSEN	Professeur des Universités	Rapporteur
ANATOLI IOUDITSKY	Professeur des Universités	Rapporteur
PIERRE WEISS	Chargé de Recherche	Co-directeur

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeur(s) de Thèse :
Jérémie BIGOT et *Pierre WEISS*
Rapporteurs :
Anders HANSEN et *Anatoli IOUDITSKY*

À mes professeurs,
dont M. Petit,
et le pari qu'il avait pris.

Remerciements

Deux ans et demi de travail sont résumés dans ce manuscrit, mais je sais que nombre d'entre vous n'iront pas plus loin que ces quelques lignes. Lourde tâche qui m'incombe de satisfaire le lecteur par les élucubrations non-mathématiques qui vont suivre. Et pour cause, si soutenance il y a aujourd'hui, le mérite revient également à ceux qui m'ont soutenue jusqu'ici. Ces quelques mots leur sont dédiés.

Je tiens à remercier en premier lieu mes directeurs de thèse, Jérémie et Pierre, de m'avoir fait confiance pour se lancer dans ce projet triennal. Après de nombreuses hésitations pré-thèse, je ne regrette aucunement que nous nous soyons choisis : votre fraîcheur d'âge n'aura été synonyme que de dynamisme, de vivacité, d'entrain et de travail convivial. Merci à Pierre de m'avoir mis un pied à l'étrier en recherche, de m'avoir encouragée toujours, rassurée souvent, et mise face à moi-même parfois. Au delà de l'aspect scientifique, la thèse est aussi une aventure humaine parsemée de bons souvenirs, parmi lesquels je garderai le doux sobriquet Boyerovitch de la part de Jérémie pour mes prédispositions à l'écriture d'articles (et moins pour la non-trivialité de mes théorèmes, dommage) ou les nombreux rires incoercibles avec Pierre (et Google).

Je souhaite dire ma sincère gratitude à Anders Hansen et Anatoli Juditsky pour avoir accepté d'être rapporteur. Outre les formalités de soutenance, merci à Anders Hansen pour ses travaux qui ont inspiré les résultats de ce deuxième chapitre ; merci à Anatoli Juditsky pour ses remarques et conseils lors de nos régulières rencontres.

Je suis également reconnaissante à Fabrice Gamboa, de participer à ce jury et d'avoir toujours répondu oui au soutien financier de mes pérégrinations durant ces trois années. Merci à Albert Cohen de me faire l'honneur de faire partie de ce jury. Mes remerciements vont aussi à Jalal Fadili dont j'ai de nombreuses fois croisé le chemin ici ou dans des contrées lointaines et qui m'a toujours abreuvée de références. Merci à lui de prendre part à ce jury. Merci également à Philippe Ciuciu de s'y joindre et de m'avoir offert mes premiers pas en recherche.

Je tiens à témoigner mon amitié aux membres de l'institut qui ont bien souvent apaisé les affres de la vie d'une thésarde[1] : merci à Pierre pour sa sempiternelle sérénité et sa constante distanciation, merci à Laurent de me faire aimer toujours davantage mon pays, merci à Benoît d'être aussi éclairé et lumineux, merci à Claire pour ses certitudes rassurantes, son écoute et son enthousiasme, merci à Nil pour son impétuosité et les quelques passes de Lindy, merci à Claire pour sa franchise et son inépuisable remise en question de la mode, merci à Anton de douter avec moi et d'être un excellent représentant des doctorants auprès de n instances pour n tendant vers l'infini[2] ; je remercie aussi Malika et Mélanie respectivement pour leurs capacités culinaires et organisationnelles que je n'ai pas, Diane pour partager les mémoires de jeunes filles pas très rangées de l'INSA autour d'un bon café, Anne-Charline pour avoir donné le ton de la doctorante modèle,

[1] ne prêtez pas attention à l'ordre des noms, la liste a été établie aléatoirement.

[2] l'infini est donc égal à 2.

Mathieu pour ramener à tribord les discussions. Merci à Fabien, je te dois bien une demi-agrégation, mais surtout ces nombreux éclats de rire qu'Etienne Bezout et Claude Shannon auraient sûrement partagé avec nous, malgré nous. Sans oublier Antoine, Ioana, Guillaume, Anne-Claire, Fanny, Sofiane, Stéphane, Raphaël, Yuriy, Sylvain, Brendan, Adil, Thibaut, Hélène, Gaëlle, Loïc, Tatiana, Olfa et tous ceux que je n'aurais pas cités nommément ; qu'ils ne me tiennent pas rigueur, je leur adresse ma sincère affection. Plus largement, j'ai été flattée et heureuse de partager la vie du laboratoire et d'apprécier son excellence scientifique.

Je n'omets pas non plus la PRIMO team de la place Pierre Potier qui a su arborer parfois le préfixe 'dé-', constituée de mon frère de thèse Paul et des intermittents Nicolas et Jonas. Merci pour votre aide qu'elle eût été théorique, numérique ou humaine.

La danse restant la réponse la plus saine que j'ai su donner au questionnement rémanent de ces années de thèse, merci à Mélissa Delfau et Raphaël Olive de m'avoir extirpée hebdomadairement de ce que je croyais parfois inextricable. Merci aux guerrières vétéranes du mardi soir, Laure, Audrey, Sabrina et Mathilde.

Une pensée sincère pour Arnold[3] et son goût prononcé pour l'imagerie et les restaurants américains, pour Benjamin, sa torpeur et son goût très prononcé pour la mécanique en général, pour Sophie et son goût trop prononcé pour les mathématiciens[4]. Merci aux frisées Camille, Hélène, Marie pour leur allégresse quand je manquais d'allant.

Au-delà de la formule attendue et quelque peu surannée, je voudrais exprimer ma reconnaissance à ma famille pour leur indéfectible soutien, leur constante égide et leur conscience du cosmos qu'ils m'ont transmise. Et parce qu'une thèse finit parfois sans les personnes qui l'ont vu débuter, une pensée émue pour celui qui n'a plus le temps de voir son aboutissement, et qui aurait su ponctuer cette soutenance avec sagacité.

Le dernier mot sera à l'égard de celui avec qui je partage mon pain, qui n'est qu'ultime nécessité, qui a souvent cru plus que moi-même, et qui m'apprend et me nourrit chaque jour.

[3]et Marielle !

[4]Puis aussi pour m'avoir hébergée et offert l'eau chaude quelques fois quand la recherche me menait à Paris.

Abstract

This Ph.D. thesis is dedicated to combine compressed sensing with block structured acquisition. Compressed Sensing (CS) is an appealing framework in many fields including signal and image processing. One of the most successful construction of sensing matrices in CS is based on randomly subsampling a full deterministic transform $A_0 \in \mathbb{C}^{n\times n}$, describing the physics of acquisition. A typical example is Magnetic Resonance Imaging (MRI) in which samples are probed in the Fourier domain. However, current CS theories lead to sampling patterns that cannot be implemented on real devices. Indeed, standard CS sampling schemes consist in sets of isolated measurements. In contrast, in various applications such as MRI, the samples should lie on a continuous trajectory due to electromagnetic constraints. Therefore, sampling patterns used in practice are very structured and they depart from theoretically justified sampling schemes. Yet, they lead to satisfactory reconstruction results. In this work, we aim at bridging the gap between theory and practice. To do so, two perspectives have been considered.

In the first part of this work, we provide theoretical recovery results using structured random matrices implementable in practice. The full deterministic matrix $A_0 \in \mathbb{C}^{n\times n}$ is structured by blocks of measurements $B_k \in \mathbb{C}^{p\times n}$ for $1 \leq k \leq M$. The block structure embodies acquisition constraints. For instance, blocks can stand for straight lines of the 2D or 3D Fourier space. The sensing matrix is then constructed by concatenating blocks randomly drawn according to a probability distribution π. In such a setting, the main questions we address are: is this structured acquisition compatible with perfect reconstruction? If so, how many blocks of measurements are required? How to choose the drawing probability π? We answer these questions. A typical result derived in this Ph.D. is the following. Let x denote an S-sparse vector supported on $S \subseteq \{1, \dots, n\}$. We show that the required number m of blocks of measurements needed to guarantee exact recovery is of the order

$$m \gtrsim \Gamma(S, \pi) \ln(n), \tag{1}$$

where $\Gamma(S, \pi)$ is a new quantity, different in many regards from the usual notions of coherence or restricted isometry constant. This result brings out an explicit dependency on (i) the whole support S of the signal to reconstruct and (ii) on the way π of drawing the blocks of measurements. It generalizes standard CS results since bound (1) coincides with the state-of-the-art when only $s = |S|$ is known. Finally, we show that structured acquisition can be successfully used in a CS framework, provided that the signal to reconstruct presents an additional structure in its sparsity, adapted to the sampling constraints. Overall, this work justifies the use of CS in cases that were not relevant before.

In order to generate efficient sampling schemes, a natural idea consists in minimizing $\Gamma(S, \pi)$ with respect to π, given some assumptions on the support S. Unfortunately, this problem turns out to be very difficult. The second part of this work is therefore dedicated to the numerical design of block sampling schemes using heuristic rules. The main idea

is to emulate a sampling scheme based on i.i.d. drawings with blocks of measurements. This problem can be cast as a projection of a probability distribution on a set of measures supported on admissible shapes (for instance, straight lines in a 2D setting). Two numerical approaches have been considered.

The first one is based on the minimization of a tailored dissimilarity measure between a probability distribution defined on the set of isolated measurements and a probability distribution defined on a set of blocks of measurements. This problem turns out to be convex and solvable in high dimension. To solve it, an efficient minimization algorithm is defined, based on Nesterov's accelerated gradient descent in metric spaces. A careful study is proposed on the choice of the metrics and of the prox function, in order to optimize the theoretical convergence rate.

The second one is based on the approximation of a target probability distribution defined on the set of isolated measurements by a probability distribution uniformly supported on N admissible shapes, for N fixed. The selected sampling patterns are optimized simultaneously in order to best fit a target density.

Résumé

Combiner de l'échantillonnage compressé (CS) avec une acquisition structurée par blocs de mesures fait l'objet de ce manuscrit. Ces dernières années, le CS a eu un grand retentissement pour les applications telles que le traitement d'image et de signal. La construction de matrices d'échantillonnage en CS peut être réalisée en sous-échantillonnant aléatoirement une matrice A_0 déterministe et dite pleine, décrivant la physique d'acquisition. Par exemple, en Imagerie par Résonance Magnétique (IRM), les échantillons sont mesurés dans l'espace de Fourier. Cependant, les théories CS actuelles mènent à des schémas d'échantillonnage difficiles à implémenter sur de vrais systèmes d'acquisition : un schéma d'échantillonnage CS typique repose sur un ensemble de mesures isolées. Or, dans de nombreuses applications, telles que l'IRM, les échantillons doivent être mesurés le long de trajectoires continues, de par les contraintes électromagnétiques de l'appareil. Les schémas d'échantillonnage sont donc très structurés en pratique et s'éloignent des schémas typiques CS justifiés théoriquement. Ils donnent pourtant de bons résultats de reconstruction.

Dans ce travail, nous cherchons à combler le fossé entre les résultats théoriques CS et le succès de l'acquisition structurée en pratique. Pour ce faire, deux stratégies ont été considérées.

Dans une première partie, nous proposons des résultats théoriques CS compatibles avec une acquisition structurée. De façon explicite, la matrice déterministe A_0 est structurée en sous-matrices par blocs de mesures $B_k \in \mathbb{C}^{p\times n}$ pour $1 \leq k \leq M$. Le cadre de blocs de mesures permet de prendre en compte les contraintes physiques de l'acquisition. Par exemple, les blocs peuvent représenter des lignes droites dans un espace d'acquisition 2D ou 3D, qui sont des motifs d'échantillonnage répandus dans les applications. La matrice d'échantillonnage est alors issue de la concaténation de blocs tirés aléatoirement selon une certaine loi de probabilité π. Nous répondons aux questions suivantes : l'acquisition structurée est-elle compatible avec une reconstruction exacte ? Si oui, combien de blocs de mesures sont nécessaires ? Et comment choisir la loi de tirage π ? Notamment, un des principaux résultats de cette thèse suit. Soit x un signal parcimonieux supporté par $S \subseteq \{1, \ldots, n\}$. Nous montrons que le nombre m de blocs de mesures requis afin de garantir une reconstruction exacte est de l'ordre de

$$m \gtrsim \Gamma(S, \pi) \ln(n),$$

où $\Gamma(S, \pi)$ est une nouvelle quantité d'intérêt, différente à bien des égards des notions usuelles de cohérence ou de constante d'isométrie restreinte. On remarquera que le nombre m de blocs de mesures nécessaires dépend (i) du support S du signal à reconstruire, et (ii) de la manière de tirer les blocs via π. Ce résultat généralise les résultats CS standards quand seul le degré de parcimonie $s = |S|$ est connu. Nous montrons également que l'acquisition structurée, combinée à du CS, peut mener à de bons résultats de reconstruction à condition que le signal à reconstruire présente une certaine structure de parcimonie,

adaptée aux contraintes d'échantillonnage. Enfin, ce travail vient justifier l'utilisation des théories CS dans des cas qui ne pouvaient être traités auparavant.

Afin de générer des schémas d'échantillonnage efficaces, une idée naturelle est de chercher à minimiser $\Gamma(S, \pi)$ par rapport à π, pour des hypothèses sur le support S données. Malheureusement, ce problème s'avère délicat. Nous proposons alors des méthodes numériques de génération de schémas d'échantillonnage structurés par blocs de mesures. L'idée est d'imiter au mieux un schéma reposant sur des échantillons i.i.d. à l'aide de blocs de mesures. L'approche consiste donc à projeter une distribution de probabilité sur un ensemble de mesures de probabilité à valeurs dans un dictionnaire de formes admissibles (lignes droite par exemple). Deux approches numériques ont été menées. La première repose sur la minimisation d'une mesure de dissimilarité entre une distribution supportée par un ensemble de mesures isolées et une distribution supportée par un ensemble de blocs de mesures. Ce problème s'avère être convexe et soluble, même en grande dimension. Nous proposons d'utiliser un algorithme rapide de minimisation reposant sur les idées de descente de gradient accélérée de Nesterov dans les espaces métriques. Une étude approfondie du choix de la métrique est menée afin d'optimiser le taux de convergence théorique. La deuxième méthode s'appuie sur l'approximation d'une mesure de probabilité cible définie sur des points isolés par une loi de probabilité uniformément supportée par N formes admissibles, pour N fixé. Les formes d'échantillonnage sélectionnées sont optimisées simultanément afin de correspondre au mieux à la densité cible.

Contents

Introduction

Contents

1 Compressed sensing: motivations and setting

In many fields of applications, one has to solve linear systems of the form

$$y = Ax,$$

where $y \in \mathbb{C}^m$ is the vector of collected data, $x \in \mathbb{C}^n$ is the signal to reconstruct and $A \in \mathbb{C}^{m \times n}$ is the matrix modeling the data acquisition. For reasons of acquisition costs or availability of data, in practice, it is often considered that $m \ll n$. Therefore, we aim at recovering x with an underdetermined system of linear equations. Such systems then admit infinitely many solutions. To compensate the lack of data, additional information is thus necessary to select an appropriate solution candidate. In this direction, a common extra assumption on the signal to reconstruct is its sparsity: there exists a (sparsity) basis

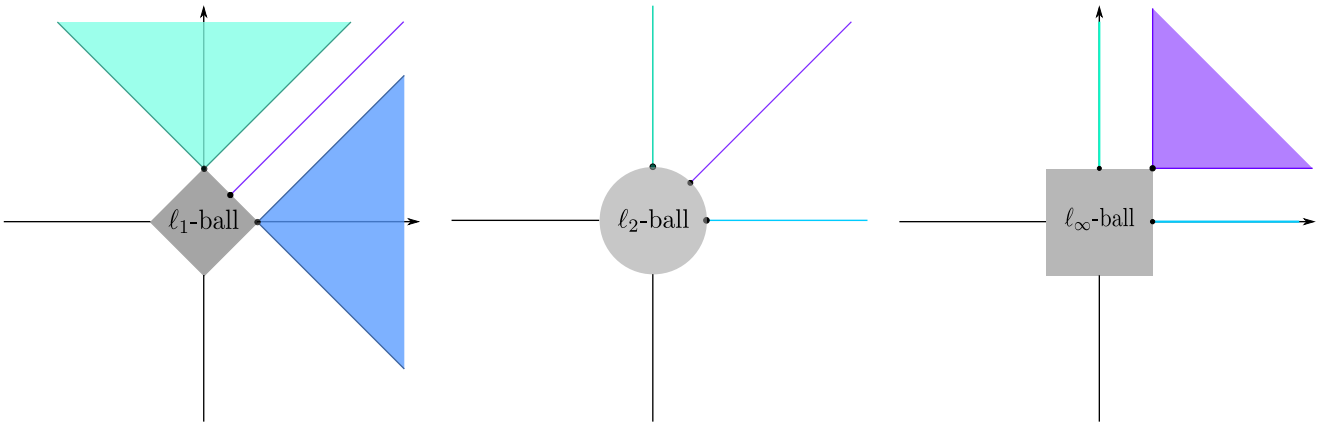


Figure 1: In (a), we illustrate that the projection onto the ℓ_1-ball is mostly directed towards vertices (which are sparse). In (b), we illustrate that the projection onto the ℓ_2-ball is isotropic. In (c), we illustrate that the projection onto the ℓ_∞-ball encourages anti-sparsity (all components are non zero and equal). In fact, the sets represented in (a,b,c) are the normal cones to the balls evaluated at points on their boundaries (for visual clarity, they have been shifted to originate at the points of interest).

in which x can be written as a combination of a few elements of this basis. We say that x is s-sparse when it has at most s non-zero entries. Therefore, the problem of recovering x becomes

$$\min_{z\in\mathbb{C}^n} \|z\|_0 \qquad \text{s.t.} \qquad y = Az,$$

with $\|\cdot\|_0$ the function counting the non-zero entries of z. Unfortunately, this combinatorial problem is NP-hard and unsolvable for high dimensions n. That is why, in practice, the recovery of x is based on a convexified version of the previous problem:

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

with $\|\cdot\|_1$ the ℓ_1-norm defined by $\|z\|_1 = \sum_{i=1}^n |z_i|$. Minimizing the ℓ_1-norm, convex surrogate of the ℓ_0 one, will still enforce the sparsity of the solution, see Figure 1. Indeed, the ℓ_1-norm can be seen as the gauge of the convex hull of the unit one-sparse vectors, from which combinations of a few elements give sparse vectors. Problem (2) is also known as *Basis Pursuit* (BP) introduced in [33]. The main advantage of the ℓ_1-minimization problem is that exact recovery is often possible by simple convex programming methods. It can be solved using the simplex algorithm [39], interior-points methods [14]. All the simulations presented in this manuscript are based on the Douglas-Rachford splitting algorithm, see [37] for a brief history and applications to image processing.

Beside the sparsity of the unknown signal, the success of sparse recovery also relies on the sampling matrix A. Indeed, the input sparse vector being very localized, a good sampling matrix must spread out the information through the data y. Therefore, if x is sparse, the rows of A should not, so that each component of y is a linear combination of *all* the components of x. Another way to see this is that the sampling vectors a_k, i.e. the rows of A, do not correlate with the signal x, so that information about components of x can be contained in each component of y. The ability of the sampling matrix A to spread out information is called the *incoherence* property. An easy way to obtain an

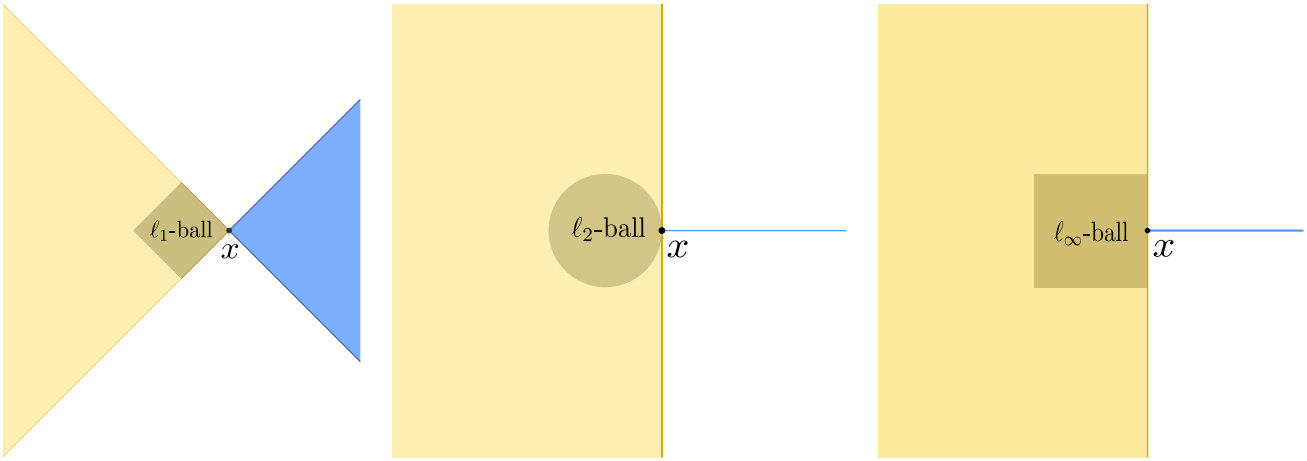


Figure 2: The tangent cones are represented in yellow, while the normal cones are represented in blue. For visual clarity, they have been shifted to originate at the point of interest. As it can be seen, the tangent cone of the ℓ_1 norm is the narrowest at sparse vectors.

incoherent measurement matrix is to consider random matrices with i.i.d. random entries, or to uniformly draw at random the rows of a deterministic orthogonal matrix with small components in absolute value, such as the Fourier matrix for instance.

Assuming an incoherent sampling, since the signal to reconstruct is sparse with at most s significant components, a typical result in Compressed Sensing is that the necessary amount m of measurements should be of the order of the amount of genuine information in x, i.e. of the order of s, up to log factors, in order to ensure exact recovery via ℓ_1-minimization.

Hence, Compressed Sensing (or any combination of the words compressed/compressive with sensing/sampling) is the term overarching such techniques aiming at reconstructing sparse vectors based on non-linear resolution from a few linear measurements sensed with an incoherent transform.

But how can this work? How, with $y = Ax$ a system of $s\ln(n)$ random measurements, can ℓ_1-minimization give the unique s-sparse solution? We know that all the points z satisfying condition $y = Az$ can be written as follows $z = x + h$ with $h \in \mathrm{Ker}\,(A)$. Therefore, x is the solution of our problem if $\|x+h\|_1 > \|x\|_1$ for all $h \in \mathrm{Ker}\,(A)$. We can therefore introduce the descent cone $\mathcal{C}(x) = \{d : \exists c > 0, \quad \|x + cd\|_1 \leq \|x\|_1\}$, giving the descent directions d of $\|\cdot\|_1$ at x. As intuited but more formally, a necessary and sufficient condition to have x as a unique solution of the ℓ_1 minimization problem is that the null space of A misses the cone of descent at x, i.e. $\mathcal{C}(x) \cap \mathrm{Ker}\, A = \{0\}$. We give in Figure 2 an illustration of this cone for various norms. As it can be seen, the tangent cone of the ℓ_1 norm is the most pinched at sparse vectors. This feature will ensure that, even if the null space of A is of dimension m, assuming m large enough, it will miss the tangent cone. This characteristic is clearly related to the *Nullspace Property* introduced later. The next paragraph gathers others properties on the sampling matrix A ensuring exact recovery via ℓ_1 minimization.

2 Compressed sensing results

CS results can be classified into two main categories: uniform or non-uniform approaches.

2.1 Uniform approach: reconstructing any s-sparse signal

We say that a matrix verifies the *restricted isometry property* (RIP) if there exists a small enough $\delta \geq 0$ such that

$$(1-\delta)\,\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\,\|x\|_2^2,$$

for all s-sparse vectors $x \in \mathbb{C}^n$, [23]. This condition is sufficient to ensure the reconstruction of any s-sparse vector via ℓ_1-minimization, as the following theorem shows.

Theorem ([23, 19]). *Assume that the restricted isometry constant of $A \in \mathbb{C}^{m\times n}$ satisfies*

$$\delta_{2s} < 4/\sqrt{41} \sim 0.62,$$

then ℓ_1-minimization recovers every s-sparse vector $x \in \mathbb{C}^n$ from $y = Ax$.

The RIP condition requires that all s-column submatrices of A are well-conditioned. For instance Bernoulli or Gaussian matrix, i.e. which entries are i.i.d. Bernoulli/Gaussian random variables, can be shown to verify the RI property.

Other famous conditions, in the same flavor as the RIP, ensuring sparse reconstruction are the restricted eigenvalue property [10], or the compatibility condition [18], or even the condition $H_{s,1}$ of [59]. They will not be detailed here, but quoted for information purposes only. Indeed, these properties are characteristic of uniform recovery in the sense that, with high probability on the choice of a random matrix, *all* sparse signals can be recovered using the same matrix. The RIP condition can be very hard to verify, especially for matrices considered later in this manuscript and in real applications.

Related to the condition $\mathcal{C}(x) \cap \operatorname{Ker} A = \{0\}$ introduced previously, there also exists a necessary and sufficient condition to reconstruct every s-sparse vector given the sampling matrix A: the *null-space property* (NSP) [43, 35], which reads as follows

$$\|h_S\|_1 \leq \|h_{S^c}\|_1$$

for every $h \in \operatorname{Ker} A$ and for every indexes subset S of cardinality s, where h_S denotes the restriction of h to the components indexed by S. This feature means that components of h representing the half of $\|h\|_1$ cannot be concentrated on a set of small cardinality. Again this condition can be hard to verify in practice.

2.2 Non-uniform approach: reconstructing a given s-sparse signal

In order to crystallize the main difference between uniform and non-uniform approaches, we can write it in mathematical term as in [48]: a uniform recovery result provides a lower probability estimate of the form

$$\mathbb{P}\left(\forall x \quad s\text{-sparse}, \quad \text{recovery of } x \text{ is successful given } A\right) \geq 1-\varepsilon,$$

whereas non-uniform recovery leads to

$$\forall x \quad s\text{-sparse}, \quad \mathbb{P}\left(\text{recovery of } x \text{ is successful given } A\right) \geq 1-\varepsilon,$$

where both probabilities are evaluated over the random draw of A. The main advantage of non-uniform approaches is that they provide a way to prove results of reconstruction for structured random matrices often considered in applications. Moreover, non-uniform approaches provide reconstruction guarantees with fewer samples.

In the literature, in order to develop nonuniform recovery results, the proof is based on the construction of a *dual certificate*. By writing the optimality condition of the ℓ_1-minimization problem, we can show that x supported on S is the unique solution, if there exists v in the row space of A, such that $v_S = \operatorname{sign}(x_S)$ and $\|v_{S^c}\|_\infty < 1$. The reader may recognize in the last two conditions that v is in fact in the subdifferential $\partial\|\cdot\|_1(x)$. Such a vector v is called an exact *dual certificate*. A natural candidate as a dual certificate is

$$v = A^* A_S \left(A_S^* A_S\right)^{-1} \operatorname{sign}(x), \tag{3}$$

where A_S is the restriction of A to the linear span of vectors supported on the support S of x. One can note that $A_S^* A_S$ is invertible if and only if A_S is injective, i.e. $S \cap \operatorname{Ker}(A) = \{0\}$. The interest for such a dual certificate is twofold: (i) it has a closed form expression and can be studied analytically, (ii) v defined in (3) is a dual certificate minimizing the ℓ_2-norm, and we can hope that v is a good candidate as well for minimizing the ℓ_∞-norm. Indeed, the proof of the existence of a dual certificate defined in (3) can be done by assuming random sign patterns for x, see for instance [21]

Another strategy, exposed in [20] and inspired by the work of D. Gross [53], is to construct an inexact dual certificate, obeying to the relaxed versions of the previous conditions, i.e.

$$\|v_S - \operatorname{sign}(x_S)\|_2 \leq 1/4 \qquad \|v_{S^c}\|_\infty \leq 1/4.$$

Since this strategy will be central for this thesis work, here is a sketch of such a construction. A key element is that the random sensing matrix A has independent rows and verify $\mathbb{E}A^*A = m\mathrm{Id}$ and $\mathbb{E}A_S^*A_S = m\mathrm{Id}_S$. Note that since the rank of A^*A is at most $m \ll n$, A^*A cannot be close to the identity. It turns out that the restricted matrix $A_S^*A_S$ is reasonably close to its expectation. We call this the concentration phenomenon. Now partition A into L submatrices, so that $A^{(1)}$ is the submatrix with the first m_1 rows, $A^{(2)}$ with the next m_2 rows, and so on, so that $m_1 + m_2 + \ldots + m_L = m$. Note that the matrices $\left(A^{(\ell)}\right)_{\ell=1,\ldots,L}$ are distributed independently. The golfing scheme initializes $v_0 = 0$ and then iteratively sets

$$v_\ell = \frac{1}{m_\ell}\left(A^{(\ell)}\right)^* A_S^{(\ell)}\left(\operatorname{sign}(x) - v_{\ell-1}\right) + v_{\ell-1},$$

for $\ell = 1, \ldots, L$ and defines $v = v_L$. By definition, v is orthogonal to $\operatorname{Ker}(A)$ since it is in the rowspace of A. One can notice that at the first iteration we have $\mathbb{E}v_1 = \operatorname{sign}(x)$, but v_1 is only $\operatorname{sign}(x)$ in expectation. The next step then tries to correct this delta by approximating $(\operatorname{sign}(x) - v_1)$ via

$$v_2 = v_1 + \frac{1}{m_2}\left(A^{(2)}\right)^* A_S^{(2)}\left(\operatorname{sign}(x) - v_1\right).$$

Proceeding to the next iterations, the golfing scheme stops when v_L is close enough from the target verifying the relaxed conditions of an inexact dual vector. The latter is attained

after a finite number of steps, giving rise to the term "golfing scheme": in golf, the hole is reached after (hopefully) a finite number of strokes. Let us rewrite this scheme to understand the role of concentration inequalities. Set $q_\ell = P_S\left(\operatorname{sign}(x) - v_\ell\right)$, where P_S is the operator restricting its following vector to components that are indexed by S. Note that

$$q_\ell = \left(\mathrm{Id}_S - \frac{1}{m_\ell}\left(A_S^{(\ell)}\right)^* A_S^{(\ell)}\right) q_{\ell-1}.$$

Suppose that the row size of the submatrices $(A_\ell)_{\ell=1,\dots,L}$ are large enough to ensure that

$$\left\|\mathrm{Id}_S - \frac{1}{m_\ell}\left(A_S^{(\ell)}\right)^* A_S^{(\ell)}\right\|_{2\to 2} \leq 1/2$$

via concentration inequalities, then the error $\|q_\ell\|_2$ exponentially decreases to zero. In fact, the final inexact dual vector can be written as

$$v = \sum_{\ell=1}^{L} \frac{1}{m_\ell}\left(A^{(\ell)}\right)^* A^{(\ell)} q_{\ell-1}.$$

By construction, $\left(A^{(\ell)}\right)^* A^{(\ell)}$ and $q_{\ell-1}$ are *stochastically independent*, and one can use large deviation inequalities to control

$$\left\|\frac{1}{m_\ell}\left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} q_{\ell-1}\right\|_\infty \leq t_\ell \|q_{\ell-1}\|_\infty,$$

with high probability. Then, this entails the construction of an inexact dual certificate such that

$$\|v_S - \operatorname{sign}(x_S)\|_2 \leq 1/4 \qquad \|v_{S^c}\|_\infty \leq 1/4.$$

2.3 Non-uniform vs. uniform

In this work, we will favor non-uniform strategies for three main reasons:

- Bounds on the number of measurements required by non-uniform strategies are (i) less demanding than those required by uniform approaches and (ii) closer to what needed in practice. Indeed, differences about the amount of measurements will be discussed in the following section.
- In real applications, the sampling matrices very seldom verify RIP, though Compressed Sensing is successfully used.
- To mirror with the previous item, in real applications, it is usually not important to reconstruct *any* s-sparse vector. One can be even optimistic by considering that worst-case signals (that are dependent on the random matrix A) will not erupt in applications. For instance, one can assume some structure in the signal support. Then, in order to exploit this sparsity structure, one clearly has to eliminate RIP strategies and to favor non-uniform strategies.

In order to illustrate the last two points, one can take into account the striking fact presented in [4] called the *flip test*, and reproduced in Figure 3. We consider the case of MR image reconstruction. In Figure 3(a), we show the target brain image $x \in \mathbb{C}^n$. The image x is considered sparse in the wavelet domain. We denote by $\alpha \in \mathbb{C}^n$ the

decomposition of x in a given wavelet basis Ψ. Due to the sparsity of α, we aim at reconstructing α and applying an inverse wavelet transform will lead to the reconstructed brain image. In (b), we present a typical CS-MRI sampling scheme, where the acquisition is performed in the 2D Fourier domain. Recall that if the sensing matrix satisfies the RIP property, then we can reconstruct *any* s-sparse vector, independently of the non-zero entries locations of the signal. Let $\sigma : \alpha \in \mathbb{C}^n \mapsto \sigma(\alpha) \in \mathbb{C}^n$ denote a random permutation of the wavelet coefficients α. The flip test consists in reconstructing $\widetilde{\alpha}$ and $\widetilde{\alpha^\sigma}$, with $\alpha^\sigma = \sigma(\alpha)$, respectively using the data $y = A\alpha$ and $y^\sigma = A\alpha^\sigma$ and via a ℓ_1-minimization problem. In (c), we present the image $\Psi^*\widetilde{\alpha}$, reconstructed from the data y. In (d), we present the image $\Psi^*\sigma^{-1}\left(\widetilde{\alpha^\sigma}\right)$, reconstructed from the data y^σ. Obviously, the sensing matrix used in MRI cannot be used to reconstruct *any* s-sparse vector, and thus it does not verify the RIP property.

2.4 Noisy case in non-uniform approach

So far, we have presented the noiseless recovery problem (2). This setting will be the main purpose of study in this manuscript. It is however possible to derive stable recovery guarantees in a noisy setting by considering a different reconstruction method:

$$\min_{z\in\mathbb{C}^n} \|z\|_1 \qquad \text{s.t.} \qquad \|Az - y\|_2 \leq \eta, \tag{4}$$

where the data y are noisy measurements such that $y = Ax + \beta$, and the noisy component $\beta \in \mathbb{C}^m$ verifies $\|\beta\|_2 \leq \eta$ for some $\eta > 0$. For the same required number of measurements as in the noiseless case, a typical non-uniform ℓ_2-stability result states that

$$\|x - z^*\|_2 \leq C_1\sigma_s(x)_1 + \left(C_2 + C_3\sqrt{s}\right)\eta, \tag{5}$$

with C_1, C_2, C_3 some constants, z^* a minimizer of Problem (4), $\sigma_s(x)_1$ the error of the best s-term approximation of x given by

$$\sigma_s(x)_1 := \inf_{\|z\|_0\leq s} \|z - x\|_1.$$

The strategy of proof is based on the same construction of the dual certificate v as in the noiseless non-uniform case. Recall that v must be in the rowspace of A, therefore there exists h such that $v = A^*h$. To derive (5), one has to improve the golfing scheme by verifying the extra assumption $\|h\|_2 \leq \varrho\sqrt{s}$ for some $\varrho > 0$.

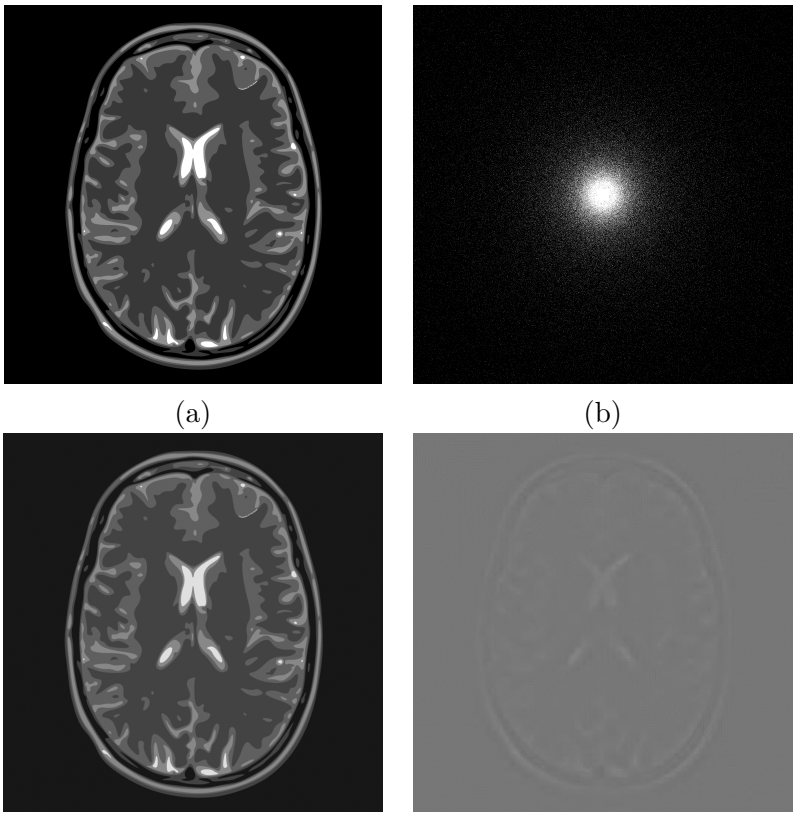
3 On the art of constructing the sensing matrix

The main challenge in Compressed Sensing is to find sensing matrices, ensuring good reconstructions results.

3.1 Sampling matrices & CS

The CS literature offers different ways of choosing or constructing sensing matrices with good incoherence properties. Here is a non-exhaustive list.

Matrices with i.i.d. random entries. Historically, Gaussian or Bernoulli matrices (with i.i.d. random Gaussian or Bernoulli entries) have played a central role in compressed sensing: they capitalize the first results about uniform recovery.



(a) (b)

(c) SNR = 24.2 dB (d) SNR = 3 dB

Figure 3: Compressed sensing and MRI: the flip test. In (a) the reference image x is displayed, it is considered sparse in the wavelet basis. We will note α its decomposition by the wavelet transform. The sampling pattern in the Fourier domain used for the ℓ_1 reconstruction of α is presented in (b), this is a typical CS scheme used in MRI. In (c), we show the reconstructed image. In (d), this is an illustration of the flip test : the data are generated with the same sampling scheme (b) and a permutation $\sigma(\alpha)$ of the wavelet coefficients α of x in (a). If the sensing matrix, denoted by A, used in MRI was verifying the RIP, then it should reconstruct α as well as $\sigma(\alpha)$. However in (d) we present the reconstruction of image corresponding to the data $A\sigma(\alpha)$. Clearly, the ordering of the wavelet coefficients plays a crucial role in the reconstruction quality: therefore, the RIP does not hold.

Theorem ([25, 6, 74]). *Let $A \in \mathbb{C}^{m\times n}$ be a Bernoulli or Gaussian matrix, and assume that*

$$m \geq C\delta^{-2}\left(s\ln(en/s) + \ln(2\varepsilon^{-1})\right),$$

for a universal constant $C > 0$. Then with probability at least $1-\varepsilon$, the restricted isometry constant of A satisfies $\delta_s \leq \delta$.

Note that the ℓ_1-minimization problem recovers x with probability exceeding $1-e^{-cm}$ provided that

$$m \geq C \cdot s \cdot \ln(n/s). \tag{6}$$

This bound has been proved to be optimal, see [49, 47]. Clearly, leaving aside the difficulty of their numerical use for storage or computation, such matrices are the crème de la crème for uniform reconstruction. However they are rare to pop up in real applications...

Partial random circulant matrices allow an efficient matrix-vector multiplication via the fast Fourier transform. Indeed, products with such matrices correspond to convolutions with a random vector. Furthermore, the use of such matrices for CS reconstructions presents the advantage of good reconstruction results with uniform or non-uniform approaches. On the one hand, as a RIP estimate, one can cite the work of [62] in which they show that for $\psi(\epsilon)$ a convolution with a Rademacher vector ϵ, if

$$m \geq C\delta^{-2}s\ln^2(s)\ln^2(n),$$

then with probability $1-n^{-\ln^2(s)\ln(n)}$ the restricted constant of $\psi(\epsilon)$ satisfies $\delta_s \leq \delta$. On the other hand, as a non-uniform result, one can cite the work of [86] ensuring exact reconstruction based on partial random circulant matrices with probability $1-\varepsilon$ if

$$m \geq Cs\ln^2(n/\varepsilon).$$

As the reader may notice, the bound on the number of measurements required by non-uniform results are less demanding than those obtained with a uniform approach. Nevertheless, even if this kind of matrices presents numerical assets, the constants in the previous bound may be large. Using such matrices often requires changing the acquisition protocols, which may be uneasy.

Subsampled Orthogonal Deterministic Transforms. "Completely random" matrices such as Gaussian or Bernoulli matrices are not widespread for practical purposes, especially their storage remains difficult in large dimension. Moreover, physical constraints due to the physics of the acquisition may introduce structure or organization in the sampling matrix. This setting will be mainly regarded in this manuscript, and special attention will be given to bounded orthogonal transforms. Indeed, this sampling technique is quite widespread in many fields of application: magnetic resonance imaging (Fourier-wavelets), ultrasound imaging (Fourier), sampling spatial field (identity on the ambient space), tomography (line integrals), radio-interferometry (correlation in the Fourier domain)... Moreover, structure in the measurement matrix may give fast matrix-vector products, which can be precious for implementing recovery algorithms, for instance the Fast Fourier transform (FFT) or more generally time-frequency transforms such that wavelets transform.

Let us take an example with A being the non-equispaced Fourier matrix defined by $A_{k,\ell} = \frac{1}{\sqrt{m}} e^{2\pi i k \cdot t_\ell}$ where k lives in some subset of $\mathbb{Z}^d$ and $(t_\ell)_{\ell=1\dots m}$ are uniformly chosen at random in $[0,1]^d$. This means that the acquisition is done in the Fourier domain, and the signal is a sparse trigonometric polynomial. In order to ensure the reconstruction of every sparse trigonometric polynomials, one can show that, in such a case, the restricted constant δ_s of $A \in \mathbb{C}^{m\times n}$ satisfies $\delta_s < \delta$ with probability $1 - N^{-\ln^3(n)}$, provided that

$$m \geq C\delta^{-2} s \ln^4(n),$$

see [25, 91]. However, adopting a non-uniform strategy leads to the following and weaker requirement

$$m \geq Cs \ln(n/\varepsilon),$$

see [20]. Again, the bound on the number m of measurements is more affordable in the case of non-uniform approach.

Nevertheless, the Fourier transform is a rare case where RIP can hold. Indeed, in practice, it is unrealistic to derive uniform results for sensing matrices. For instance in [7] for level based reconstruction basis, there can be proved that RIP condition is not satisfied.

Let us then recall a non-uniform result of Candès and Plan [20] in the case of any subsampled orthogonal transforms including the last example of the Fourier transform.

Theorem. *Let $A \in \mathbb{C}^{m\times n}$ to be the sensing matrix resulting from drawing the rows $(a_i^*)_{i=1\dots n}$ of an orthogonal transform $A_0 \in \mathbb{C}^{n\times n}$. Set $(\pi_i)_{i=1\dots n}$ to be the drawing probability distribution, i.e. the probability to draw the row a_i is π_i. Suppose the signal x to reconstruct to be s-sparse. If*

$$m \geq C \cdot s \max_{1\leq i\leq n} \frac{\|a_i\|_\infty^2}{\pi_i} \ln(n/\varepsilon), \tag{7}$$

then x is the unique solution of the Problem 2.

Clearly, we find the stated result for the case of the Fourier transform: for any i, $\|a_i\|_\infty^2 = 1/n$ and picking uniformly at random its rows, i.e. $\pi_i = 1/n$ for every i leads to a bound on m of the order $s \cdot \ln(n/\varepsilon)$.

3.2 Variable density sampling

In order to minimize the bound given in the last theorem, one can choose π such that

$$\pi_i = \frac{\|a_i\|_\infty^2}{\sum_{k=1}^n \|a_k\|_\infty^2}, \qquad \text{for } i = 1,\dots,n. \tag{8}$$

This suggests that π is not necessarily the uniform probability distribution. Note that *variable density sampling* was so far an efficient heuristic to shorten acquisition time in practice. Indeed, Magnetic Resonance Imaging has resorted a lot to spiral or radial trajectories for acquisition [69], but see also holography acquisition [87, 73]. Mention should also be made of variable density sampling in the fields of radio interferometry or tomographic modalities (e.g. X-ray) where sensing is made along fixed sets of measurements [93, 103].

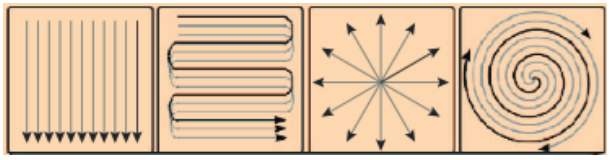


Figure 4: [70] Illustration of typical sampling schemes in MRI: the acquisition relies on continuous trajectories.

This heuristic has recently found a partial justification in the compressed sensing (CS) literature, notably with the choice of π presented in (8). Subsequently, it is worth to note that a theoretical formalization of the concept of *variable density sampling* has been proposed in [28].

However, drawing independent rows of A_0, and thereby justifying the need for variable density sampling, is interesting from a theoretical perspective; yet, it has little practical relevance, as it will be emphasized in the next section.

4 Limitations of current CS results

Subsampling techniques based on structured acquisition are very popular in applications since they lead to good reconstruction results. So far, CS theories do not take into account the possibility of structured acquisition. In this section, we emphasize the need for new CS results compatible with structured acquisition.

4.1 Structured acquisition in applications

Here, we succinctly present various acquisition modalities.

MRI In Magnetic Resonance Imaging (MRI), data are acquired in the Fourier domain, but MR images are considered sparse in the wavelet domain. Therefore the full acquisition can be modeled by the linear transform A_0 product of the Fourier matrix and the inverse wavelet transform. Due to electromagnetic constraints, MR sampling trajectories must lie on continuous patterns such as straight lines or smooth curves [31]. For instance, pattern based on radial or straight lines are very popular in practice for MR application, see Figure 4.

Parallel MRI Parallel MRI (pMRI) relies on multiple receiver coils [15]. The goal is then to reconstruct MR images from multichannel data.

CT tomography In Computed Tomography (CT), each measurement represents the summation or line integral of the the object along a particular ray path: each sample corresponds to a point of the Radon transform. These measurements are collected along different angles and different distances. Again, the choice of the sensing matrix is dictated by the physics of acquisition. Sharp variations in CT images are usually confined to the borders of internal tissue structure so that images have a sparse gradient representation. During CT acquisition, we collect a set of views, see Figure 5. As it can be seen, the

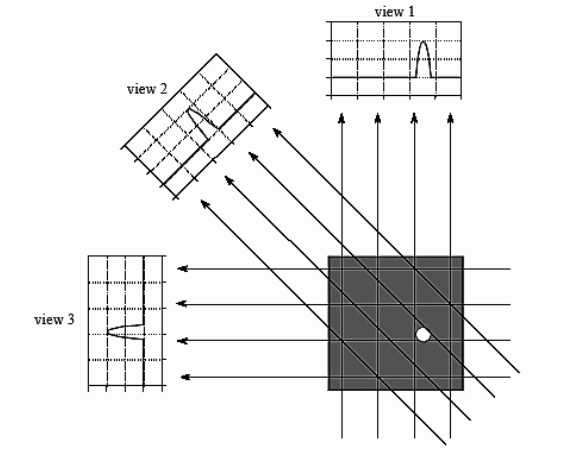


Figure 5: [94] Illustration of CT acquisition. Each sample corresponds to the line integral of the object along a particular line. The samples are acquired in groups, called 'views' on the Figure.

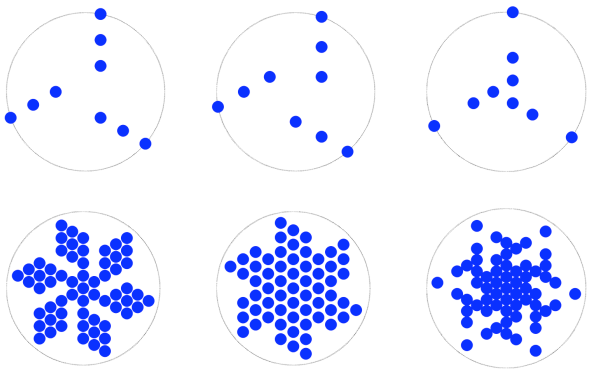


Figure 6: [67] At the top, we present the disposal of antennas for various interferometers. At the bottom, we respectively present the Fourier domain filling, i.e. the corresponding measurements acquired in the Fourier domain.

acquisition is intrinsically structured by sets of measurements: the projection views are usually taken along a circular orbit uniformly distributed on $[0, 2\pi]$ [34].

Radio-interferometry The principle of aperture synthesis is to combine signals from at least two antennas that are observing the same scene. The produced images have the same angular resolution as an instrument of the size of the entire antennas collection would give. The combination of measurements by all pairs of telescopes give access to spatial frequencies. Any interferometer is thus simply identified by a binary mask in Fourier equal to 1 for each spatial frequency probed and 0 otherwise [102]. Here there is two strategies: using a fixed telescopes network or vary the distance between telescopes. Note that the relative antennas disposal conditions the probed frequencies and thus structures the acquisition, see Figure 6. Again, the sensing matrix is fixed by the physics of acquisition and by the instrument geometry.

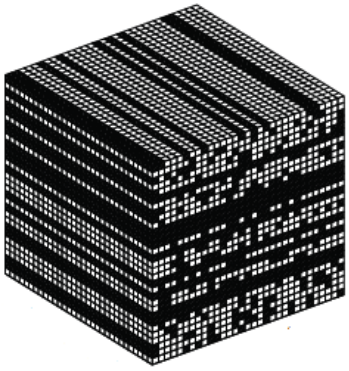


Figure 7: [85] Example of 3D data (RF lines) used in ultrasound imaging for the signal recovery.

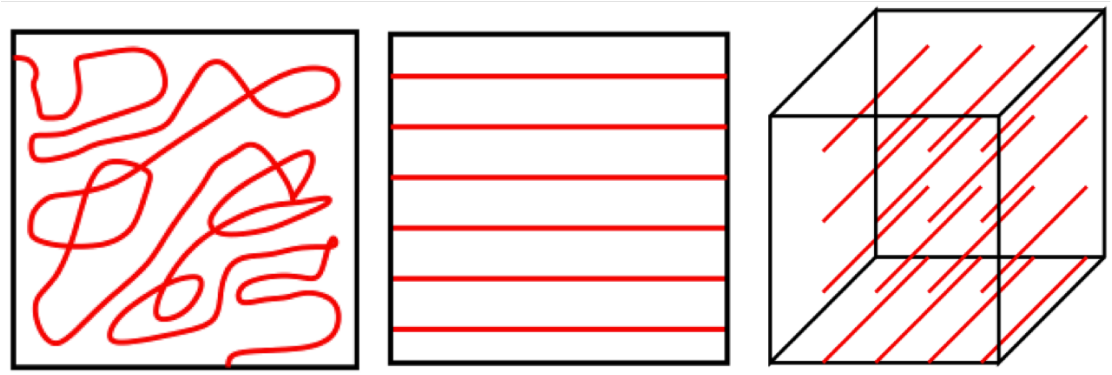


Figure 8: [98] Example of acquisition trajectories for mobile sensing in 2D and 3D settings.

Echography Ultrasound imaging (US) is generally used for obstetrics or cardiology purposes. In US imaging, the initial acquisition is performed in the spatial domain by several transducers. US images are considered sparse in the Fourier domain [41]. In the 3D setting, the measured spatial samples generally consist of whole radio-frequency (RF) lines in the lateral direction, see Figure 7 for a typical US sampling scheme. However, for providing real-time imaging by reducing the acquisition time, entire RF lines can be skipped in the lateral direction of the sampling scheme, see Figure 7.

Mobile sensing In mobile sensing, we aim at reconstructing bandlimited spatial fields using mobile sensors [98, 99, 100, 52]. The mobile sensors are usually constrained to move along continuous paths, see Figure 8 for examples of acquisition trajectories.

4.2 On the (non)-applicability of CS in applications

One may ask about the relevancy of current CS theories in the applications previously described. Let us focus on applying standard CS strategy to the MRI case. Note that the same conclusions could be drawn for other fields of applications, as those mentioned earlier. Since the full MR sampling matrix is fixed by the physics of MR acquisition, one can consider the construction of CS sensing matrices based on the random subsampling of a deterministic orthogonal transform A_0. Recall that A_0 is the product of the Fourier

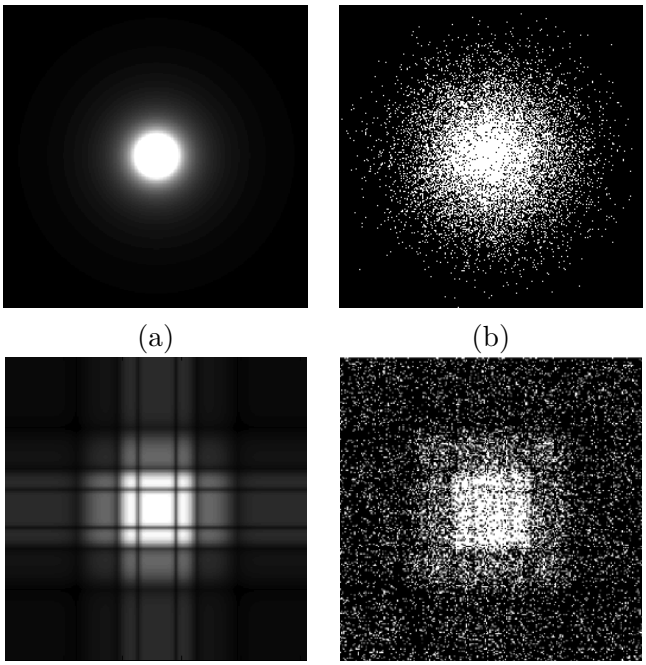


Figure 9: Illustration of 2D variable density sampling in MRI. In (a) (c), examples of drawing distributions π of particular interest in MRI are given: the probability is concentrated in the low frequencies localized in the center of the image. In (b) and (d), examples of sampling schemes resulting from drawing isolated measurements according to π as in (a) and (c) respectively.

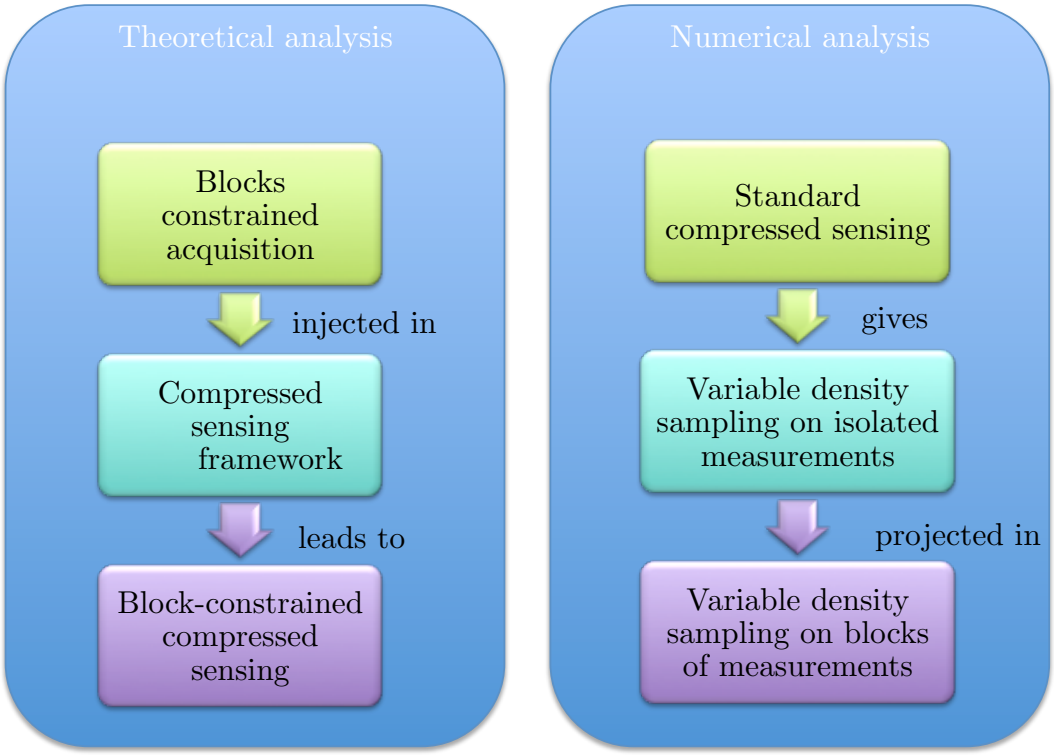
transform with a given inverse wavelets transform. Therefore, in the MR case, a typical CS sensing scheme is made of random locations in the Fourier domain. The sensing matrix A thus consists in concatenating the rows of A_0 corresponding to the drawn frequencies samples. It can be shown that such a transform does not verify the RIP, see [7]. Nevertheless, using (7), current non-uniform CS strategies require $O(s \ln(n) \ln(n/\varepsilon))$ isolated measurements to ensure exact reconstruction, with a choice for the drawing probability π depicted in Figure 9.

CS strategies are then an appealing framework for MR imaging. Nevertheless, the MR sensing schemes recommended by CS are made of isolated measurements, see Figure 9 (b) (d), which can be physically hard to implement. Obviously, standard CS sensing schemes depart from the structured acquisition made in practice and depicted in the last section. Moreover, current CS theories miss to legitimate such structured sampling schemes.

4.3 What is missing in CS: structured acquisition

This inability of CS to provide sampling schemes easy to implement in real applications has been the incentive for my Ph.D. work. Different notions of structure could be considered, the most general one being sampling with generic stochastic processes [27]. In this Ph.D. we concentrate on a more restricted setting based on blocks of measurements. This setting is interesting since it allows using existing concentration inequalities. It is currently unclear to us how to derive similar results for general stochastic processes.

A block of measurements is a set of isolated measurements, and therefore it can repre-



(a) (b)

Figure 10: Illustrated résumé of contributions

sent any arbitrary sampling pattern. This notion allows to reflect the acquisition structure given by the physics of the acquisition for a given sensing modality.

When my Ph.D. started, no work was existing in that direction. Very recently, a similar work was published [81]. It seems clear to us that the results in this Ph.D. manuscript go well beyond from theoretical, numerical and practical point of view.

Another inspiring work about structured acquisition but still based on isolated measurements can be found in [2]. Even if it is restricted to sampling with isolated measurements, this is a pioneering paper questioning the empirical success of CS in real-life applications, managing to capture the necessity for variable density sampling.

Note that despite the structure introduced in the acquisition, the problem of recovering x remains the ℓ_1-minimization problem (2). On the one hand, this setting is compatible with various assumptions on the signal to reconstruct. On the other hand, it seems to be more efficient to exploit structure in the sampling strategy than introducing structure in the recovery algorithm, see e.g. [4, 89].

5 Contributions

The contributions presented in this manuscript can be organized into two parts: the first one is pictorially summarized in Figure 10(a), and the second one is illustrated in Figure 10(b).

5.1 Theoretical CS contributions

In a first stage, the idea is to incorporate block structured acquisition in a noiseless compressed sensing setting, see Figure 10(a). Instead of sensing isolated measurements as standard CS theories incite, we introduce the notion of blocks of measurements. A block of measurements is a submatrix composed of several rows of the full sampling matrix modeling the acquisition features. This new concept can easily reflect the physical constraints of the acquisition: for instance a block of measurements can correspond to a straight line, which is a widespread pattern in practical sampling. The proposed block constrained sensing matrices enrich the family of CS sensing matrices.

5.1.1 Recovering *any* s-sparse vector: the limits of the structured acquisition

In order to reconstruct s-sparse vectors, we derive an analysis of CS strategies based on blocks of measurements by constructing an inexact dual certificate using the so-called golfing scheme. A new quantity of interest $\gamma(s)$ is introduced and will play the role of the usual coherence in the case of blocks of measurements, see Definition 1.1. We show that the number m of blocks measurements needed to ensure exact recovery should satisfy

$$m \geq c \cdot \gamma(s) \cdot \ln(n). \tag{9}$$

The tightness of bound (9) is studied. We show that (9) cannot be improved up to logarithmic factors in many cases such as Gaussian measurements, isolated measurements or blocks measurements in time-frequency bases. For instance, we show that acquiring blocks of $\mathbb{R}^{p\times n}$ made of i.i.d. Gaussian measurements leads to $\gamma(s) = O(s/p)$. Therefore, (9) can be rewritten as follows

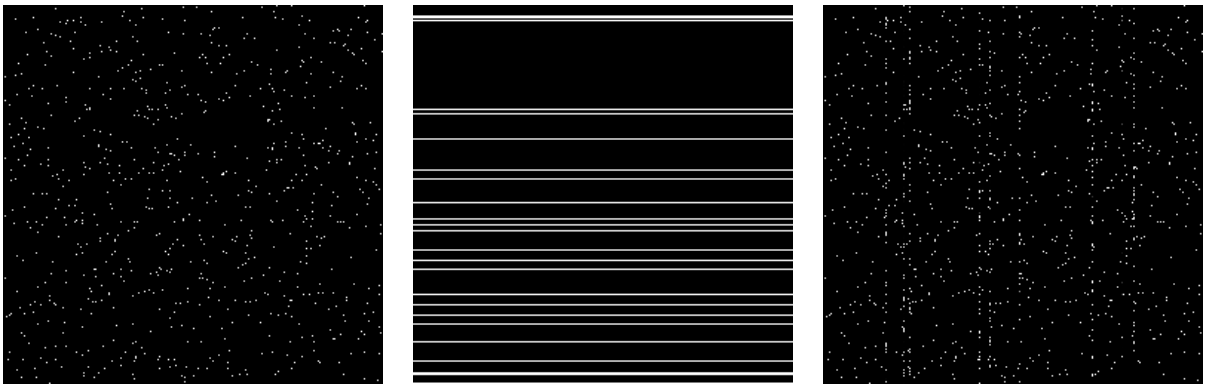
$$m \geq c \cdot \frac{s}{p} \ln(n).$$

This means that the corresponding number of isolated measurements is $O\left(p \times \frac{s}{p}\right) = O(s)$, which is similar to (6).

Moreover, we highlight the limitations of blocks sampling strategies to recover *any* s-sparse vector using separable transforms. For instance, a typical result of this work states that

$$m \geq 2s \tag{10}$$

horizontal lines in a 2D acquisition space are *necessary* to reconstruct any s-sparse vector of $\mathbb{C}^n$, when the initial sampling matrix is a 2D separable transform. In Figure 11, we illustrate the limitations of blocks sampling strategies with the 2D Fourier transform (note that it is a separable transform). In Figure 11(a), the image to reconstruct is sparse in the Dirac basis, with a degree of sparsity of 1.2%. The acquisition is done in the Fourier domain by sampling random horizontal lines. The sampling scheme in Figure 11(b) is made of 9.38% of measurements. Note that a number m of acquired horizontal lines satisfying $m \geq 2s$ would lead to a full sampling strategy in this case. In Figure 11(c), we present the reconstructed image which is far different from Figure 11(a). This inability for reconstructing the signal in Figure 11(a) can be explained by the violation of the condition $m \geq 2s$ in Figure 11(b).



(a) Image to reconstruct $s/n = 1.2\%$

(b) Sampling scheme in the Fourier domain (9.38%)

(c) Reconstructed image SNR = 16.37 dB

Figure 11: Toy example: illustration of the limitations of structured acquisition when no structured sparsity is considered. In (a), the signal to reconstruct is (1.2%)-sparse in the Dirac basis. In (b), the sampling scheme consists in horizontal lines in the Fourier plane. It is made of 9.38% of measurements. Note that the bound $m \geq 2s$ would lead to a full acquisition of the Fourier plane since a block represents $\sqrt{n}$ isolated measurements. In (c), the reconstructed image via ℓ_1-minimization fails to be faithful to the target image (a).

5.1.2 Recovering a vector supported on S: breaking the limits of block structured acquisition

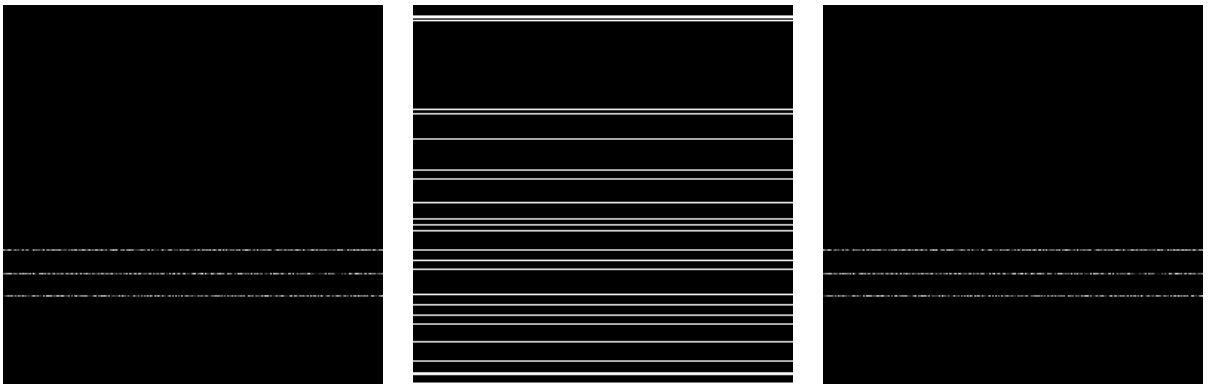
In order to break the previous limitations of blocks sampling strategies, we modify the standard golfing scheme of [20]. By doing so, we derive CS reconstruction guarantees of a signal $x \in \mathbb{C}^n$ with a given support S when the acquisition is structured by blocks. The concepts of RIP or coherence are not sound anymore. They are replaced by a new quantity $\Gamma(S, \pi)$ which explicitly depends on the support S and the distribution π from which sampled blocks are drawn, see Definition 2.3. In order to ensure exact recovery, we show that a lower bound on the required number of blocks measurements is

$$m \geq c \cdot \Gamma(S, \pi) \cdot \ln(n).$$

The explicit dependency on the support S allows making various assumptions on the support structure, but also provides reconstruction guarantees for random signals with known distribution. The structured sparsity turns out to overcome the limits of blocks sampling strategies previously described. A striking example is the reconstruction of a 2D signal, sparse in the Dirac basis, where the samples are horizontal lines in the 2D Fourier domain. Assuming that the support of x is concentrated on q horizontal lines of the spatial plane, the new CS theory ensures exact reconstruction if the number m of sensed horizontal lines verifies

$$m \geq c \cdot q \cdot \ln(n). \tag{11}$$

Since $q \ll s$, a reasonable number m of acquired lines will lead to exact recovery. An illustration of this result is given in Figure 12. In Figure 12(a), we aim at reconstructing an image which is sparse in the Dirac basis, and with support concentrated on 3 horizontal lines in a 2D representation. The acquisition is performed in the 2D Fourier domain and it is made of random horizontal lines. We use the same sampling scheme as in Figure



(a) Image to reconstruct $s/n = 1.2\%$ (b) Sampling scheme in the Fourier domain (9.38%) (c) Reconstructed image SNR = 58.4 dB

Figure 12: Toy example: illustration of the crucial role of structured sparsity when blocks-sampling strategies are considered. In (a), the signal to reconstruct is (1.2%)-sparse in the Dirac basis but its support is concentrated on 3 horizontal lines. In (b), the sampling scheme consists in horizontal lines in the Fourier plane. It is made of 9.38% of measurements and it is the same as in Figure 11(b). In (c), we perfectly reconstruct the image via ℓ_1-minimization.

11(b). Whereas this sampling scheme cannot be used to reconstruct the "unstructured" sparse image of Figure 11(a), it leads to exact recovery of the structured sparse image in Figure 12(a). This result was predicted by the theoretical bound (11).

Let us now turn to the case of MR sampling. MR images are considered sparse in the wavelet domain. Note that we consider the standard form of the 2D wavelet transform which consists of atoms constructed as tensor product of 1D wavelets at different scales. We denote by $(\Omega_{j,j'})_{0\leq j,j'\leq J}$ the dyadic partition of the plane corresponding to the 2D levels of wavelet decomposition, see Figure 13. When the acquisition is done along horizontal lines of the 2D Fourier domain, we introduce a new quantity of interest $(s_j^c)_{0\leq j\leq J}$. The latter characterizes the degree of sparsity of x restricted to its columns and restricted to the levels $(\Omega_{j,j'})_{0\leq j'\leq J}$. In Figure 13, we illustrate in blue the restriction to the k-th column and to the decomposition levels $(\Omega_{J,j'})_{0\leq j'\leq J}$. We show that if the number m of acquired horizontal lines satisfy

$$m \geq c \cdot \sum_{j=0}^{J} s_j^c \cdot \ln(n)$$

then we can perfectly recover x with high probability. In many practical cases,

$$\sum_{j=0}^{J} s_j^c \ll s.$$

This new result therefore improves the bound (10) on the necessary number of acquired horizontal lines for separable transform. For example, in Figure 14(a), the black dots represent the support of a sparse vector. The quantities s_j^c are the maximal number of dots per vertical line on given scales $(\Omega_{j,j'})_{0\leq j'\leq J}$. In Figure 14(b), we show a sampling scheme made of horizontal lines in the 2D Fourier domain which is adapted to the structured degrees of sparsity $(s_j^c)_{0\leq j\leq J}$: the number of acquired horizontal lines in the j-th horizontal domain is proportional to s_j^c. To show the relevancy of the quantities $(s_j^c)_{0\leq j\leq J}$ in MR

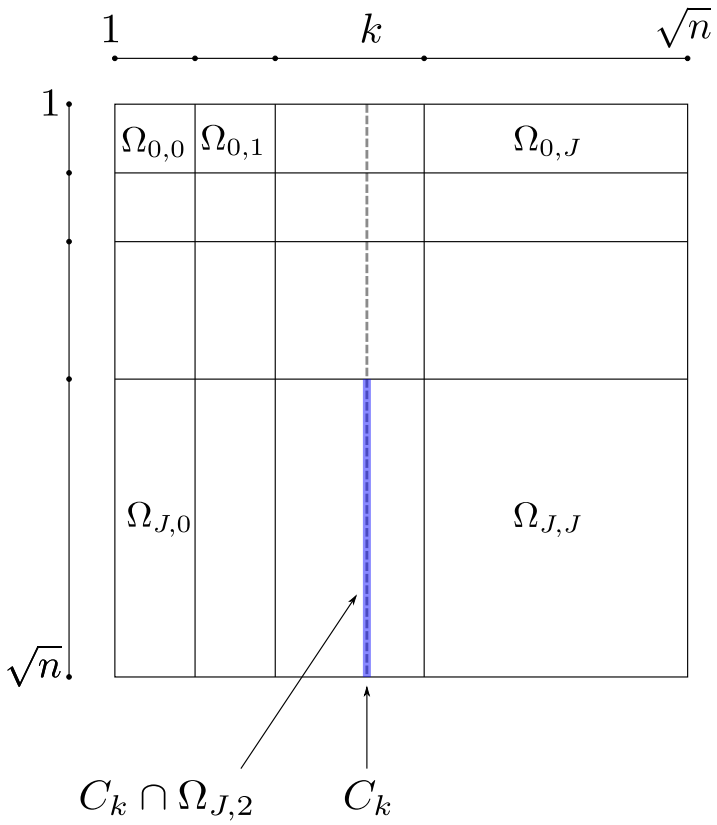
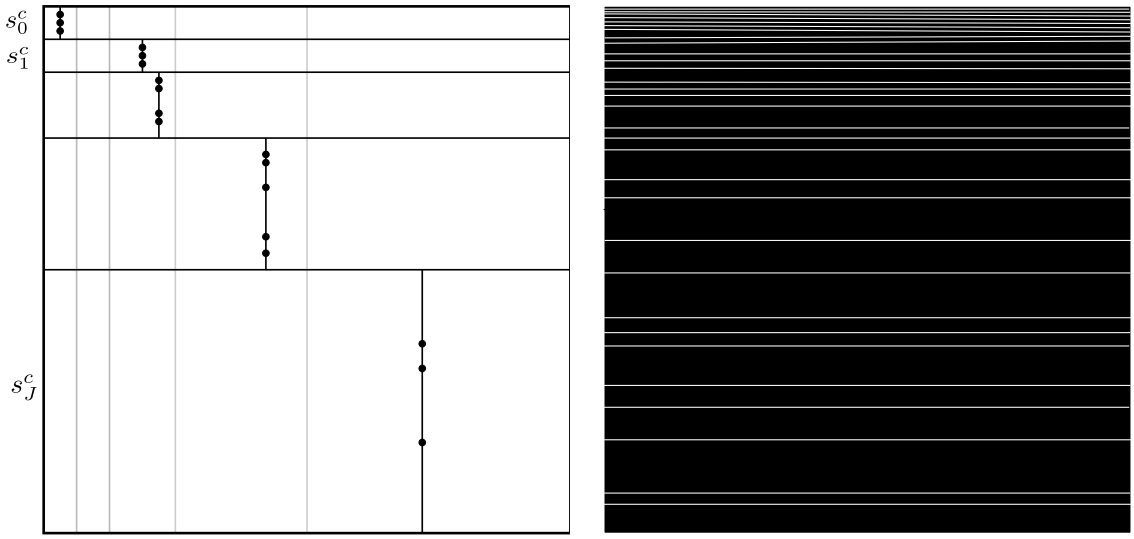


Figure 13: Illustration of the decomposition levels $(\Omega_{j,j'})_{0 \leq j,j' \leq J}$ for a separable wavelet transform.



(a) (b)

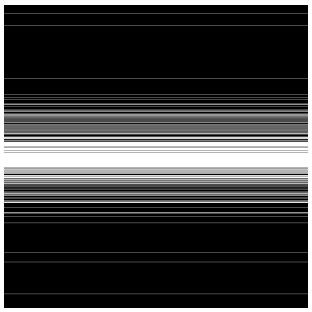
Figure 14: Illustration of MR sampling with horizontal lines in the 2D Fourier domain. In (a), we present the wavelet domain in which the image to reconstruct is considered sparse and supported on S. We illustrate the quantities $\left(s_j^c\right)_{0\leq j\leq J}$ by the black points: they show where the maximal cardinality of S, restricted to columns and to the j-th horizontal decomposition levels, is reached. In (b), we display an efficient sampling scheme for MR reconstruction, based on horizontal lines in the 2D Fourier domain where the number of acquired lines in each horizontal decomposition level is proportional to the corresponding quantity s_j^c.

sampling, we illustrate in Figure 15 the reconstruction of a real image. In Figure 15(a)(c), the images to reconstruct are considered sparse in the Haar domain. Note that the image in (c) is the same as in (a) but rotated of 90°. For both images, we evaluate the quantities $(s_j^c)_{0\leq j\leq J}$ which are larger in the case of Figure 15(c). As in MR sampling, the acquisition is done in the Fourier domain, using horizontal lines. The same sampling scheme, presented at the top of Figure 15, is used for both reconstructions. In Figure 15(b)(d), the corresponding reconstructions are displayed and show the influence of the quantities $(s_j^c)_{0\leq j\leq J}$ on the reconstruction quality.

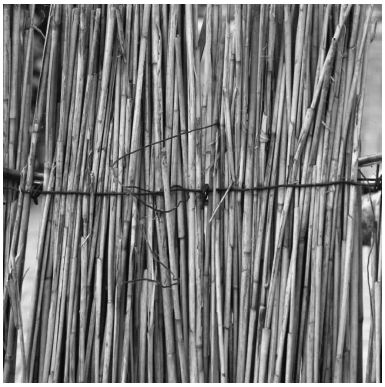
Ultimately, we highlight that a block structured acquisition can be used, only if the support structure is adapted to it. Furthermore, the proposed theory allows envisioning the use of CS in situations that were not possible before: the use of incoherent transforms is not necessary anymore, given that the support S has some good properties.

5.2 Numerical algorithms for generating blocks-constrained sampling schemes

In a second step, we propose numerical approaches to generate blocks-constrained sampling schemes. This is based on variable density sampling given by standard compressed sensing theories, see Figure 10(b). Indeed, CS results give target probability distributions p from which ideally, isolated measurements should be drawn. The proposed strategy is to project these target probability distributions on the set of probability distributions supported on admissible shapes, see Figure 16. The latter will represent the physical



Sampling scheme



(a) Original image

$s^c = (16, 16, 32, 59, 81, 75, 48)$

(b) SNR = 27.8 dB

(c) Zoom

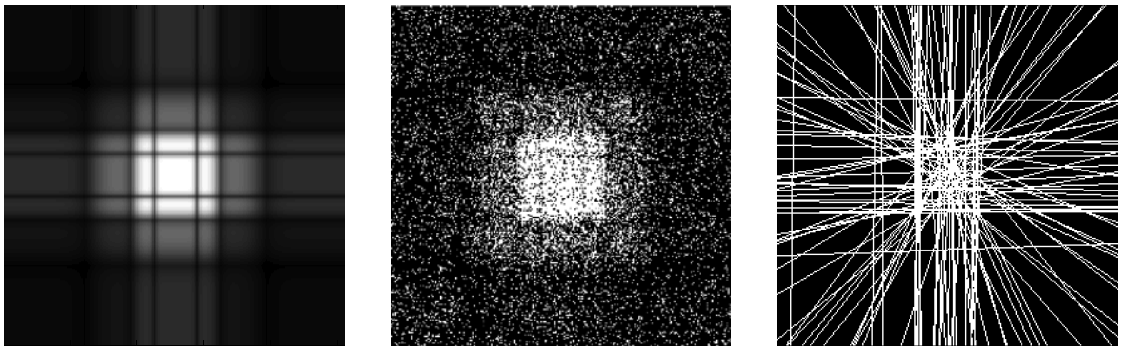
(c) Original image

$s^c = (16, 16, 32, 64, 124, 240, 411)$

(d) SNR = 14.7 dB

(e) Zoom

Figure 15: An example of reconstruction of a 2048×2048 real image from MR sensing. In (a) (c), Reference images to reconstruct: the image in (c) is the same image as (a) but rotated of 90°. We precise the value of the (s_j^c) for both images which are sparse in the Haar domain. Note that the quantities (s_j^c) are larger in the case of image (b). For the reconstruction, we use the sampling scheme at the top of the Figure in which low frequencies are in the center. It corresponds to 9.8 % of measurements. In (b) (d), corresponding reconstruction via ℓ_1-minimization. We have rotated the image in (d) to facilitate the comparison between both. Clearly, the value of s^c impacts the quality of reconstruction when the acquisition is made of horizontal lines in the Fourier domain.



(a) Target distribution p (b) Sampling scheme made of isolated measurements drawn from p (c) Sampling scheme made of N blocks

Figure 16: Standard CS strategies provide a target distribution p in (a) from which isolated measurements should be drawn. A typical resulting sampling scheme is displayed in (b). The question raised in this part is "How to choose the blocks in (c) (which are straight lines here) in order to mimic standard CS sampling schemes such as in (b) ?"

constraints imposed by the sensing modality.

To do so, we devise two strategies.

The first one generates blocks sampling schemes by (i) projecting a given target probability distribution p defined on the set of isolated measurements on the set of distributions defined on a blocks dictionary $\mathcal{I}$ and call π the projected distribution; (ii) then draw N blocks according to the distribution π. To do so, we construct a dissimilarity measure $\mathcal{D}(\pi, p, \mathcal{I})$ that depends on the blocks dictionary $\mathcal{I}$. The projection problem turns out to be a minimization problem of $\mathcal{D}(\pi, p, \mathcal{I})$, which is strongly convex, but non-differentiable. We propose an efficient strategy based on the numerical optimization of the dual problem, and on the use of Nesterov's ideas [77]. Contrarily to most first order methods proposed recently in the literature which are based on Hilbert space formalisms, Nesterov's algorithm is stated in a (finite dimensional) normed space. We thus perform the minimization of the dual problem on a metric space, and we carefully study the optimal choice of the norms in the primal and dual spaces. We show that depending on the blocks length, the optimal choice might well be different from the standard ℓ^2-norm. Subsequently, we propose a blocks dictionary made of straight lines, which is of particular interest in MRI, and we illustrate the relevancy of the obtained probability distribution π for MR reconstruction, over standard sampling schemes. Indeed, such strategies give high quality reconstruction in MRI.

In the second approach, we propose to quantify a target distribution p defined on the set of isolated measurements by a distribution π uniformly supported on N blocks, for a given integer N. In order to simultaneously optimize the chosen N blocks, the proposed resolution is inspired by halftoning techniques. We derive a minimization problem of an objective function decomposable into an attraction and a repulsion term. We propose an efficient way to solve it using fast summation methods based on non-uniform fast Fourier transform. The resulting blocks sampling schemes are still in a trial period for MR reconstruction, but some are presented in Figure 17.

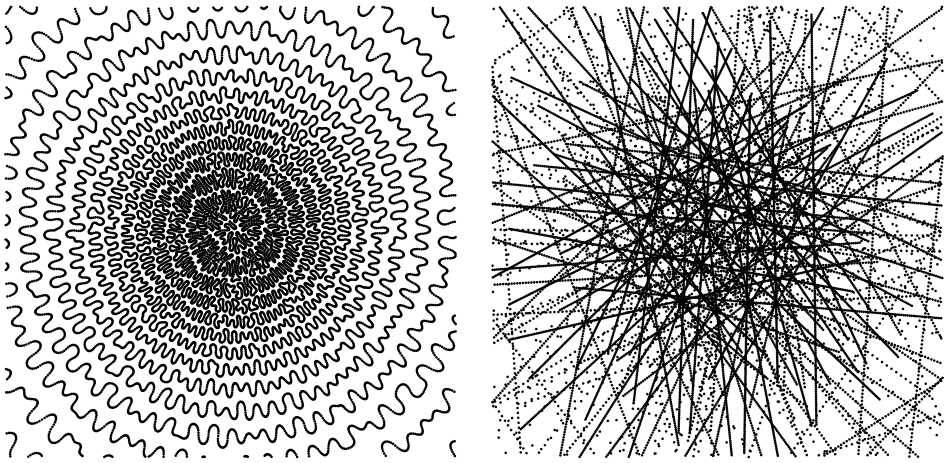


Figure 17: Illustration of sampling schemes obtained in Chapter 4

6 Organization

As pointed out in Figure 10, the manuscript is organized into two parts. Theoretical results on blocks-constrained compressed sensing, corresponding to Figure 10(a), are presented in Chapters 1 and 2. Numerical approaches for generating blocks-constrained sampling schemes via measure projection, see Figure 10(b) are presented in Chapters 3 and 4. Note that each chapter is self-contained.

Chapter 1 In this chapter, we propose a first analysis of blocks sampling CS strategies. We focus on the exact recovery of an s-sparse vector $x \in \mathbb{C}^n$ supported on S. We exhibit a quantity $\gamma(s)$, which will play the role of the well-known coherence. We then show through a series of examples including Gaussian measurements, isolated measurements or blocks in time-frequency bases, that the main result is sharp in the sense that the minimum amount of blocks necessary to reconstruct sparse signals cannot be improved up to a multiplicative logarithmic factor. We also highlight the limitations of CS blocks sampling strategies. In particular, in the case of a 2D separable transform, we show that $2s$ blocks as horizontal lines are needed to identify any s-sparse vector. This theoretical result seems at odds with the good reconstruction results observed in practice with blocks sampling strategies, for instance in magnetic resonance imaging, radio-interferometry or ultra-sound imaging. This last observation suggests that a key feature is missing in this first study to fully understand the potential of block sampling in applications. A very promising perspective is therefore to couple the ideas of structured sparsity with block sampling, which is a tackled issue in the following chapter.

Chapter 2 Structured acquisition with blocks of measurements are easy to implement, and they give good reconstruction results in practice [69]. However, very few results exist on the theoretical guarantees of CS reconstructions in this setting. In this chapter, we fill the gap between CS theory and acquisitions made in practice. To this end, the key feature to consider is the structured sparsity of the signal to reconstruct. In this chapter, we derive new CS results for structured acquisitions and signal satisfying an *a priori* structured sparsity. The obtained results are RIPless, in the sense that they do not hold for any s-sparse vector, but for sparse vectors with a given support S. Our results

are thus support-dependent, and they offer the possibility for flexible assumptions on the structure of S. Moreover, our results are also drawing-dependent, since we highlight an explicit dependency between the probability of reconstructing a sparse vector and the way of choosing the blocks of measurements.

Chapter 3 In this chapter, we propose a numerical method to perform variable density sampling with block constraints. The first contribution is to propose a new way to draw the blocks in order to mimic CS strategies based on isolated measurements. The basic idea is to minimize a tailored dissimilarity measure between a probability distribution defined on the set of isolated measurements and a probability distribution defined on a set of blocks of measurements. This problem turns out to be convex and solvable in high dimension. The second contribution is to define an efficient minimization algorithm based on Nesterov's accelerated gradient descent in metric spaces. We study carefully the choice of the metrics and of the prox function. We show that the optimal choice may depend on the type of blocks under consideration. Finally, we show that we can obtain better MRI reconstruction results using our sampling schemes than standard strategies such as equiangularly distributed radial lines.

Chapter 4 In this chapter, we propose another numerical method to perform variable density sampling with block constraints. It is inspired by recent results on electrostatic halftoning [92]. A target measure is projected on a set of pushforward measures supported on N blocks, for a fixed N. This algorithm highly generalizes techniques such as Poisson disc sampling, used in CS-MRI [101]. In contrast with the first method, all the blocks that support the resulting measure are optimized simultaneously. We show some applications to MRI sampling.

Part I

Theory: compressed sensing with structured acquisition

1 A first analysis of blocks sampling strategies

Contents

In this chapter, we consider a sensing matrix A that is constructed by stacking blocks of measurements and not just isolated measurements. In the proposed formalism, the blocks can be nearly arbitrary random matrices. For instance, our main result covers the case of blocks made of groups of rows of a deterministic sensing matrix (e.g. lines in the Fourier domain) or blocks with random entries (e.g. Gaussian blocks). We study the problem of exact non-uniform sparse recovery in a noise-free setting. This sampling strategy raises various questions. How many blocks of measurements are needed to ensure exact reconstruction? Is the required number of blocks compatible with faster acquisition?

Our first contribution is to extend the standard compressed sensing theorems to the case of blocks of measurements. We then show that our result is sharp in a few practical examples and extends the best currently known results in compressed sensing for the

This chapter is a sligthly different form of the paper [11], as joint work with Pierre Weiss and Jérémie Bigot.

recovery of s-sparse vectors. We also prove that in many cases, imposing a block structure has a dramatic effect on the recovery guarantees since it strongly impoverishes the variety of admissible sampling patterns. Overall, we believe that the presented results give a good theoretical basis to the use of block compressed sensing and show the limits to this setting.

1 Preliminaries

1.1 Notation

Let $S = (S_1, \dots, S_s)$ be a subset of $\{1, \dots, n\}$ of cardinality s. We denote by $P_S \in \mathbb{C}^{n \times s}$ the matrix with columns $(e_i)_{i \in S}$ 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 $v_S = P_S^* v$.

1.2 Main assumptions

Recall that we consider the following ℓ_1-minimization problem:

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \qquad \text{s.t.} \quad y = Az, \tag{1.1}$$

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 chapter, we assume that the sensing matrix A can be written as

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} B_1 \\ \vdots \\ B_m \end{pmatrix}, \tag{1.2}$$

where $B_1, \dots, B_m$ are i.i.d. copies of a random matrix B, satisfying

$$\mathbb{E}\,(B^* B) = \mathrm{Id}, \tag{1.3}$$

where Id is the $n \times n$ identity matrix. This condition is the extension of the isotropy property described in [20] in a blocks-constrained acquisition setting.

In most cases studied in this chapter, 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 1.1. *Let $S \subset \{1, \dots, n\}$ be a set of cardinality s. We denote by $(\mu_i(S))_{1 \le i \le 3}$ the smallest positive reals such that the following bounds hold either deterministically or stochastically (in a sense discussed later)*

$$\begin{aligned} \|B_S^* B_S\|_{2 \to 2} \le \mu_1(S), \qquad & \sqrt{s} \max_{i \in S^c} \|B_S^* B e_i\|_2 \le \mu_2(S), \\ s \max_{i \in S^c} \left\| \mathbb{E}\left[B_S^* \left(Be_i\right) \left(Be_i\right)^* B_S \right] \right\|_{2 \to 2} & \le \mu_3(S). \end{aligned} \tag{1.4}$$

Define

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

The quantities introduced in Definition 1.1 can be interpreted as follows. The number $\mu_1(S)$ can be seen as an *intra-support* block coherence, whereas $\mu_2(S)$ and $\mu_3(S)$ are related to the *inter-support* block coherence, that is the coherence between blocks restricted to the support of the signal and blocks restricted to the complementary of this support. Note that the factors $\sqrt{s}$ and s involved in the definition of $\mu_2(S)$ and $\mu_3(S)$ ensure homogeneity between all of these quantities.

1.3 Application examples

The number of applications of the proposed setting is very large. For instance, it encompasses those proposed in [20]. Let us provide a few examples of new applications below.

1.3.1 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:

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

where $\bigsqcup$ stands for the disjoint union. This case is the one studied in [80].

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 that the isotropy condition $\mathbb{E}\,[B^*B] = \mathrm{Id}_n$ is verified.

1.3.2 Overlapping blocks issued from orthogonal transforms

In the latter example, we concentrated on partitions, i.e. non-overlapping blocks of measurements. The case of overlapping blocks can also be handled. To do so, define the blocks $(B_j)_{1\leq j\leq M}$ 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 α_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 $\mathbb{E}\,[B^*B] = \mathrm{Id}_n$ where B_k is defined similarly to the previous example. See Section 5.4 for an illustration of this setting in the case of 2D Fourier measurements.

1.3.3 Random blocks

In the previous examples, the blocks were predefined and extracted from deterministic matrices or systems. The proposed theory also applies to random blocks. For instance, one could consider blocks with i.i.d. Gaussian entries since these blocks satisfy the isotropy condition (1.3). 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 of potential interest is that of blocks generated randomly using random walks over the acquisition space [28].

2 Main result

The main result reads as follows.

Theorem 1.2. *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 (1.2), and that the isotropy condition (1.3) holds. Suppose that the bounds (1.4) hold deterministically. If the number of blocks m satisfies the following inequality*

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

then x is the unique solution of (1.1) with probability at least $1-\varepsilon$. The constant c can be taken equal to 534.

The proof of Theorem 1.2 is detailed in Section 5.3.1. It is based on the so-called golfing scheme introduced in [53] for matrix completion, and adapted by [20] for compressed sensing from isolated measurements. Note that Theorem 1.2 is a non uniform result in the sense that reconstruction holds for a given s-sparse vector and not for all s-sparse signals. It is likely that uniform results could be derived by using the so-called Restricted Isometry Property. However, this strong property is usually harder to prove and leads to narrower classes of admissible matrices and to larger number of required measurements.

Remark 1.3 (Improvement of Theorem 1.2)**.** *By assuming that $\ln(s)\ln(\ln(s)) \leq c'\ln(n)$, one can simplify the previous bound in Theorem 1.2, by*

$$m \geq c''\gamma(s)\ln(4n)\ln\left(12\varepsilon^{-1}\right), \tag{1.5}$$

for some constants c' and c''. Note that this bound can be also obtained considering the trick presented in [1, 54]. In the sequel, for the sake of clarity, we will assume that the bound (1.5) holds for Theorem 1.2.

Remark 1.4 (Noisy setting)**.** *In this work, we concentrate on a noiseless setting. It is likely that noise can be accounted for mimicking the proofs in [20] for instance.*

Remark 1.5 (The case of stochastic bounds)**.** *In Definition 1.1, we say that the bounds* deterministically hold *if the inequalities (1.4) are satisfied almost surely. This assumption is convenient to simplify the proof of Theorem 1.2. Obviously, it is not satisfied in the setting where the entries of B are i.i.d. Gaussian variables. To encompass such cases, the bounds in Definition 1.1 could* stochastically hold*, meaning that the inequalities (1.4) are satisfied with large probability. The proof of the main result can be modified by conditioning the deviation inequalities in the Lemmas of Section 5.3.1 to the event that the bounds in Definition 1.1 hold. Therefore, even though we do not provide a detailed proof, the lower bound on the required number of blocks in Theorem 1.2 remains accurate. Hence, we will propose in Section 3.2 some estimates of the quantities (1.4) in the case of Gaussian measurements.*

The lower bound on the number m of blocks of measurements in Theorem 1.2 depends on $\gamma(s)$ and thus on the support S of the vector x to reconstruct. In the usual compressed sensing framework, the matrix A is constructed by stacking realizations of a random vector a. The best known results state that $O(s\mu\ln(n))$ isolated measurements are sufficient to

reconstruct x with high probability. The coherence μ is the smallest number such that $\|a\|_\infty^2 \leq \mu$. The quantity $\gamma(s)$ in Theorem 1.2 therefore replaces the standard factor $s\mu$. The coherence μ is usually much simpler to evaluate than $\gamma(s)$ which depends on three properties of the random matrix B: the intra-support coherence μ_1 and the inter-support coherences μ_2 and μ_3. As will be seen in Section 3, it is important to keep all those quantities in order to obtain tight reconstruction results. Nevertheless, a rough upper bound of $\gamma(s)$, reminiscent of the coherence, can be used as shown in Proposition 1.6.

Proposition 1.6. *Let S be a subset of $\{1,\dots,n\}$ of cardinality s. Assume that the following inequality holds either deterministically or stochastically*

$$\|B^*B\|_{1\to\infty} \;\leq\; \mu_4$$

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

$$\gamma(s) \leq s\mu_4. \tag{1.6}$$

The proof of Proposition 1.6 is given in Section 5.3.2. The bound given in Proposition 1.6 is an upper bound on $\gamma(s)$ that should not be considered as optimal. For instance, for Gaussian measurements, it is important to precisely evaluate the three quantities $(\mu_i(S))_{1\leq i\leq 3}$.

3 Sharpness of the main result

In this section, we discuss the sharpness of the lower bound given by Theorem 1.2 by comparing it to the best known results in compressed sensing.

3.1 The case of isolated measurements

First, let us show that our result matches the standard setting where the blocks are made of only one row, that is $p = 1$. This is the standard compressed sensing framework considered e.g. by [22, 48, 20]. Consider that $A_0 = (a_i^*)_{1\leq i\leq n}$ is a deterministic matrix, and that the sensing matrix A is constructed by drawing m rows of A_0 according to some probability distribution $\mathcal{P} = (p_1,\dots,p_n)$, i.e. one can write A as follows:

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

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 1.6, for a support S of cardinality s the following upper bound holds:

$$\gamma(s) \leq s \max_{1\leq j\leq M} \frac{\|a_j a_j^*\|_{1\to\infty}}{p_j}.$$

Therefore, according to Theorem 1.2 with bound in (1.5), it is sufficient that

$$q \geq c'' s \max_{1\leq j\leq M} \frac{\|a_j a_j^*\|_{1\to\infty}}{p_j} \ln(4n) \ln\left(12\varepsilon^{-1}\right). \tag{1.7}$$

to obtain perfect reconstruction with probability $1-\varepsilon$. Noting that $\|a_j\|_\infty^2 = \|a_j a_j^*\|_{1\to\infty}$, for all $j \in \{1,\dots,n\}$, it follows that Condition (1.7) is the same (up to a multiplicative and logarithmic constant) to that of [20].

In addition, choosing $\mathcal{P}^\star$ in order to minimize the right-hand side of (1.7) leads to

$$p_j^\star = \frac{\|a_j a_j^*\|_{1\to\infty}}{\sum_{k=1}^n \|a_k a_k^*\|_{1\to\infty}}, \qquad \forall k \in \{1,\dots,n\},$$

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

$$q \geq c'' s \sum_{k=1}^n \|a_k^*\|_\infty^2 \ln(4n) \ln\left(12\varepsilon^{-1}\right). \tag{1.8}$$

Contrarily to common belief, the probability distribution minimizing the required number of measurements is not the uniform one, but the one depending on the ℓ_∞-norm of the considered row. Let us highlight this fact. Consider that $A_0 = \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{F}_{n-1} \end{pmatrix}$, 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 (1.7) is $O(sn\ln^2(n))$. This shows that uniform random sampling is not interesting for this sensing matrix. Note that the coherence $\|A_0\|_{1\to\infty}^2$ of 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.

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

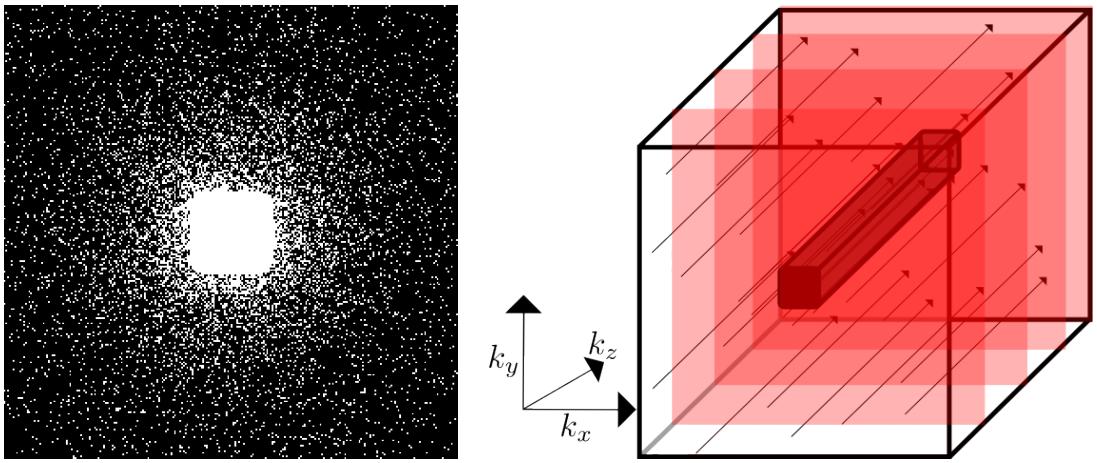
then, the right hand side of (1.7) becomes $O(2s\ln^2(n))$. Using this sampling strategy, compressed sensing therefore remains relevant. Furthermore, note that the latter bound could be easily reduced by a factor 2 by systematically sampling the location associated to the first row of A_0, and uniformly picking the $q-1$ remaining isolated measurements. Similar remarks were formulated in [63] which promote non-uniform sampling strategies in compressed sensing.

3.2 The case of Gaussian measurements

We suppose that the entries of $B \in \mathbb{R}^{p\times n}$ are i.i.d. Gaussian random variables with zero-mean and variance $1/p$. This assumption on the variance ensures that the isotropy condition (1.3) is satisfied. The matrix A constructed by concatenating such blocks is also a Gaussian random matrix with i.i.d. entries and does not differ from an acquisition setting based on isolated measurements. Therefore, if Theorem 1.2 is sharp, one can expect that $q = O(s\ln(n))$ measurements are enough to perfectly reconstruct x. In what follows, we show that this is indeed the case.

Proposition 1.7. *Assume that the entries of $B \in \mathbb{R}^{p\times n}$ are i.i.d. Gaussian random variables with zero-mean and variance $1/p$. Then, $\gamma(s) = O\left(\frac{s\ln(s)}{p}\right)$. Therefore, $O\left(\frac{s\ln(s)\ln(n)}{p}\right)$ Gaussian blocks are sufficient to ensure perfect reconstruction with high probability.*

The proof can be found in Section 5.3.3. This is similar to an acquisition based on isolated Gaussian measurements and this is optimal up to a logarithmic factor, see [42].



(a) (b)

Figure 1.1: Example of sampling pattern used in MRI [69]. (a) Visualization in the k_x-k_y plane. (b) Visualization in 3D.

3.3 The case of separable transforms

In this section, we consider d-dimensional deterministic transforms obtained as Kronecker products of orthogonal one-dimensional transforms. This setting is widespread in applications. Indeed, separable transforms include d-dimensional Fourier transforms met in astronomy [12] or products of Fourier and wavelet transforms met in MRI [70] or radio-interferometry [102]. 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 [69]. The authors propose to undersample the 2D k_x-k_y plane and sample continuously along the orthogonal direction k_z (see Figure 1.1).

The remaining of this Section is as follows. We first introduce the notation. We then provide theoretical results about the minimal amount of blocks necessary to reconstruct all s-sparse vectors. Next, we show that Theorem 1.2 is sharp in this setting since the amount of blocks required to reconstruct s-sparse vectors coincides with the minimal amount. Finally, we perform a comparison with the results in [80].

3.3.1 Preliminaries

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

$$A_0 = \Psi \otimes \Psi = \begin{bmatrix} \Psi_{1,1}\Psi & \dots & \Psi_{1,\sqrt{n}}\Psi \\ \Psi_{2,1}\Psi & \dots & \Psi_{2,\sqrt{n}}\Psi \\ \vdots & \ddots & \vdots \\ \Psi_{\sqrt{n},1}\Psi & \dots & \Psi_{\sqrt{n},\sqrt{n}}\Psi \end{bmatrix} \in \mathbb{C}^{n\times n},$$

where $\otimes$ denote the Kronecker product. Note that A_0 is also orthogonal. We define groups of measurements from A_0 as follows:

$$B_k = \Psi_{k,:} \otimes \Psi \tag{1.9}$$

$$= \left[\Psi_{k,1}\Psi, \dots, \Psi_{k,\sqrt{n}}\Psi\right] \in \mathbb{C}^{\sqrt{n}\times n}. \tag{1.10}$$

For instance, if Ψ is the 1D discrete Fourier transform, this strategy consists in constructing $\sqrt{n}$ blocks as horizontal discrete lines of the discrete Fourier plane. This is similar to

the blocks used in [69]. Similarly to paragraph 1.3.1, a sensing matrix A can be constructed by drawing m i.i.d. blocks with distribution π. Letting $K = (k_1, \dots, k_m) \in \{1, \dots, \sqrt{n}\}^m$ denote the drawn blocks indexes, A reads:

$$A = \frac{1}{\sqrt{m}} \begin{bmatrix} \frac{B_{k_1}}{\sqrt{\pi_{k_1}}} \\ \vdots \\ \frac{B_{k_m}}{\sqrt{\pi_{k_m}}} \end{bmatrix} \tag{1.11}$$

$$= \left(\frac{D(\pi)^{-1/2}}{\sqrt{m}} \cdot \Psi_{K,:} \right) \otimes \Psi$$
$$= \widetilde{\Psi}_{K,:} \otimes \Psi \tag{1.12}$$

where $D(\pi) := \mathrm{diag}(\pi_{k_1}, \dots, \pi_{k_m})$ and $\widetilde{\Psi}_{K,:} := \frac{D(\pi)^{-1/2}}{\sqrt{m}} \cdot \Psi_{K,:}$. By combining the results in Theorem 1.2 with bound (1.5) and Proposition 1.6, we easily get the following reconstruction guarantees.

Proposition 1.8. *Let $S \subset \{1, \dots, n\}$ be the support of cardinality s of the signal $x \in \mathbb{C}^n$ to reconstruct. Under the above hypotheses, if*

$$m \geq c'' s \max_{1 \leq j \leq M} \frac{\|B_j^* B_j\|_{1 \to \infty}}{\pi_j} \ln(4n) \ln\left(12\varepsilon^{-1}\right), \tag{1.13}$$

then the vector x is the unique solution of (1.1) with probability at least $1 - \varepsilon$.

Using the above result we also obtain the following Corollary.

Corollary 1.9. *The drawing probability distribution $\pi^\star$ minimizing the right hand side of Inequality (1.13) on the required number of measurements is defined by*

$$\pi_j^\star = \frac{\left\|B_j^* B_j\right\|_{1 \to \infty}}{\sum_{k=1}^M \|B_k^* B_k\|_{1 \to \infty}}, \qquad \forall j \in \{1, \dots, M, \}. \tag{1.14}$$

For this particular choice of $\pi^\star$, the right hand side of Inequality (1.13) can be written as follows

$$m \geq c'' s \sum_{j=1}^M \|B_j^* B_j\|_{1 \to \infty} \ln(4n) \ln\left(12\varepsilon^{-1}\right). \tag{1.15}$$

The sharpness of the bounds on the required number of measurements in Proposition 1.9 will be discussed in the following paragraph.

3.3.2 The limits of separable transforms

Considering a 2D discrete Fourier transform and a dictionary of blocks made of horizontal lines in the discrete Fourier domain, one could hope to only require $m = O(s/p \ln(n))$ blocks of measurements to perfectly recover all s-sparse vectors. Indeed, it is known since [22] 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 all s-sparse vectors. It means that this

specific block structure is inadequate to obtain strong reconstruction guarantees. This result also shows that Theorem 1.9 is nearly optimal.

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

Lemma 1.10. *[35, Lemma 3.1] Set Σ_s to be the set of s-sparse vectors in $\mathbb{C}^n$. If A is any $m \times n$ matrix, then the following propositions are equivalent:*

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

Looking at (iii) of Lemma 1.10, since the rank of 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 x, we cannot reconstruct x. This property is an important step to prove Proposition 1.11.

Proposition 1.11. *Assume that the sensing matrix A has the special block structure described in (1.11). If $m < \min(2s, \sqrt{n})$, then there exists no decoder Δ such that $\Delta(Ax) = x$ for all s-sparse vector $x \in \mathbb{C}^n$. In other words, the minimal number m of distinct blocks required to identify every s-sparse vectors is necessarily larger than $\min(2s, \sqrt{n})$.*

Proposition 1.11 shows that there is no hope to reconstruct any s-sparse vectors with less than $m = O(s)$ blocks of measurements, using sensing matrices A made of blocks such as (1.9). Moreover, since the blocks are of length $p = \sqrt{n}$, it follows that whenever $s \geq \frac{\sqrt{n}}{2}$, the full matrix A_0 should be used to identify every s-sparse x. Let us illustrate this result on a practical example. Set 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 $x \in \mathbb{R}^{32\times 32}$ to be 10-sparse in the spatial domain and only supported on the first column as illustrated in Figure 1.2(a). Due to this specific signal structure, the Fourier coefficients of x are constant along horizontal lines, see Figure 1.2(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 1.9, one can derive the following result.

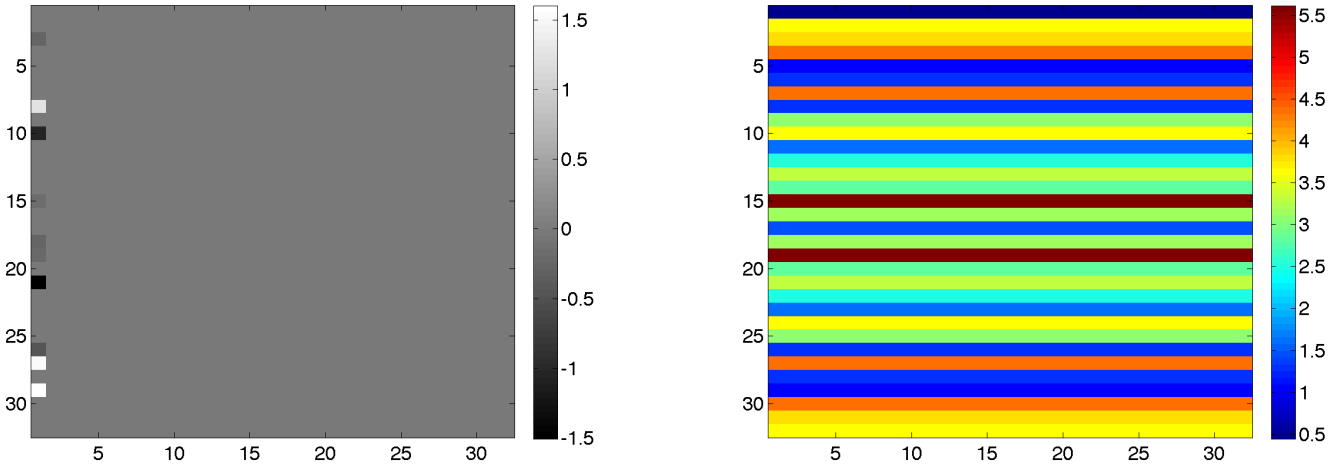
Proposition 1.12. *Let $A_0 \in \mathbb{C}^{n\times n}$ denote the 2D discrete Fourier matrix and consider a partition in $M = \sqrt{n}$ blocks that consist of lines in the 2D Fourier domain. Assume that $x \in \mathbb{C}^n$ is s-sparse. The drawing probability minimizing the right hand side of (1.13) is given by*

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

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

$$m \geq c'' s \ln(4n) \ln\left(12\varepsilon^{-1}\right).$$

This result is disappointing but optimal up to a logarithmic factor, due to Proposition 1.11. We refer to Section 5.3.5 for the proof. This Proposition indicates that $O(s \ln(n))$ blocks are sufficient to reconstruct x which is similar to the minimal number given in Proposition 1.11 up to a logarithmic factor.



(a) (b)

Figure 1.2: 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.

3.3.3 Relation to previous work

To the best of our knowledge, the only existing compressed sensing results based on blocks of measurements appeared in [80]. 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 [80] only covers the case described in Paragraph 1.3.1 (i.e. partitions of orthogonal transforms), our work covers the case of overlapping blocks of measurements (see Paragraph 1.3.2), and it can be also extended to the case of randomly generated blocks (see Paragraph 1.3.3). Last but not least, the work [80] only deals with *uniform* sampling densities which is well known to be of little interest when dealing with partially coherent matrices (see e.g. end of Paragraph 3.1 for an edifying example).

Apart from those contextual differences, the comparison between the results in [80] and the ones in this chapter is not straightforward. The criterion in [80] that controls the overall number of measurements q depends on the following quantity:

$$\Upsilon(A_0, S, B) := \|\overline{B_S}\|_{2\to 1},$$

where $\overline{B_S}$ stands for the block restricted to the columns in S with renormalized rows. The total number of measurements required in the approach [80] is

$$q_{PDG} \geq C\Upsilon(A_0, S, B) \max_{i,j} |A_0(i,j)|^3 n^{3/2} \ln(n) \tag{1.16}$$

which should be compared to our result

$$q \geq c'' p\gamma(s) \ln(4n) \ln\left(12\varepsilon^{-1}\right). \tag{1.17}$$

As shown in the previous paragraphs, the number (1.17) is sharp in various settings of interest, while (1.16) is usually hard to explicitly compute or too large in the case of

patially incoherent transforms. It therefore seems that our results should be preferred over those of [80].

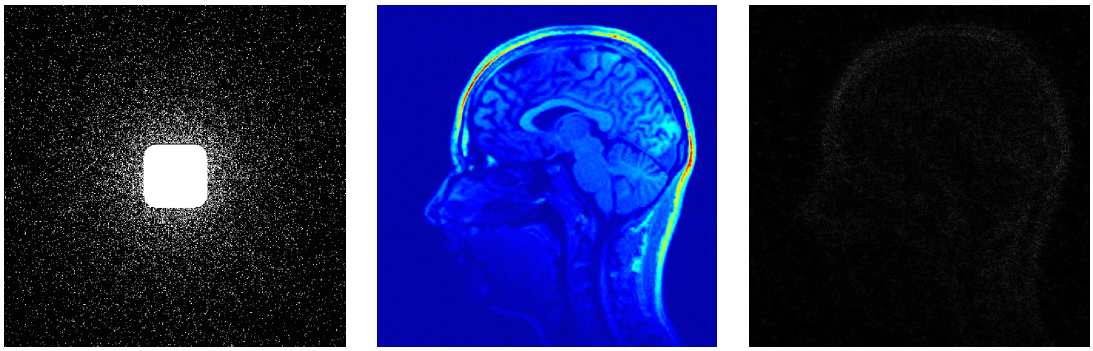
4 Outlook

We have introduced new sensing matrices that are constructed by stacking random blocks of measurements. Such matrices play an important role in applications since they can be implemented easily on many imaging devices. We have derived theorems that guarantee exact reconstruction using these matrices via ℓ_1-minimization algorithms and outlined the crucial role of two properties: the *extra* and *intra* support block-coherences introduced in Definition 1.1. We have showed that our main result (Theorem 1.2) is sharp in a few settings of practical interest, suggesting that it cannot be improved in the general case up to logarithmic factors.

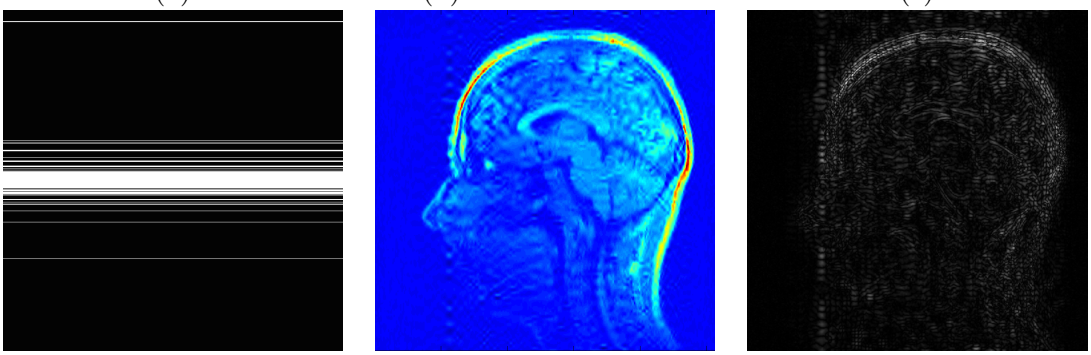
Apart from those positive results, this work also reveals some limits of block sampling approaches. First, it seems hard to evaluate the *extra* and *intra* support block-coherences - except in a few particular cases - both analytically and numerically. This evaluation is however central to derive optimal sampling approaches. More importantly, we have showed in Paragraph 3.3.2 that not much could be expected from this approach in the specific setting where separable transforms and blocks consisting of lines of the acquisition space are used. Despite the peculiarity of such a dictionary, we believe that this result might be an indicator of a more general weakness of block sampling approaches. Since the best known compressed sensing strategies heavily rely on randomness (e.g. Gaussian measurements or uniform drawings of Fourier atoms), one may wonder whether the more rigid sampling patterns generated by block sampling approaches have a chance to provide decent results. It is therefore legitimate to ask the following question: is it reasonable to use variable density sampling with pre-defined blocks of measurements in compressed sensing?

Numerical experiments indicate that the answer to this question is positive. For instance, it is readily seen in Figure 1.3(a)(b)(c) and (j)(k)(l), that block sampling strategies can produce comparable results to acquisitions based on isolated measurements. The first potential explanation to this phenomenon is that $\gamma(s)$ is low for the dictionaries chosen in those experiments. However, even acquisitions based on horizontal lines in the Fourier domain (see Figure 1.3(d)(e)(f)) produce rather good reconstruction results while Proposition 1.12 seems to indicate that this strategy is doomed.

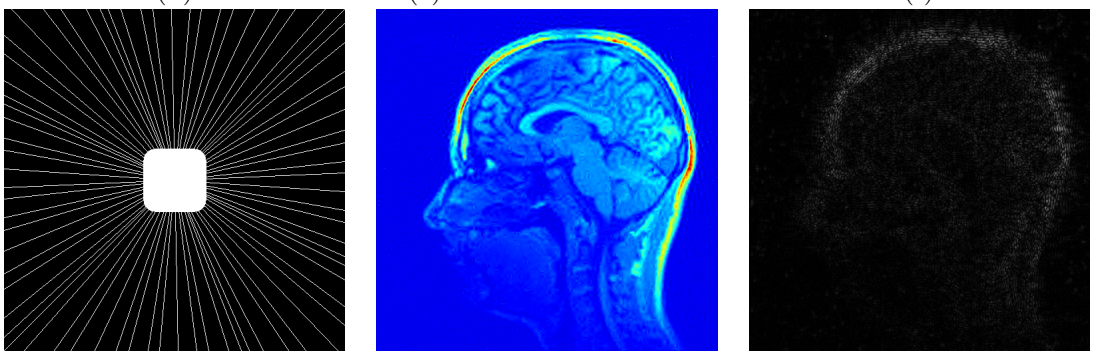
This last observation suggests that a key feature is missing in our study to fully understand the potential of block sampling in applications. Recent papers [2, 4] highlight the central role of *structured sparsity* to explain the practical success of compressed sensing. A very promising perspective is therefore to couple the ideas of structured sparsity in [2, 4] and the ideas of block sampling proposed in this chapter to finely understand the results in Figure 1.3 and perhaps design new optimal and applicable sampling strategies. This will be the goal of the next chapter.



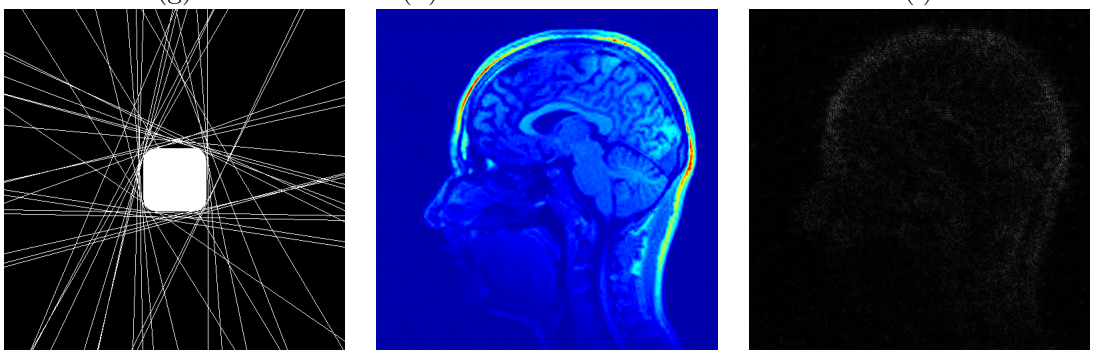
(a) (b) PSNR = 40 dB (c)



(d) (e) PSNR = 32.79 dB (f)



(g) (h) PSNR = 36.34 dB (i)



(j) (k) PSNR = 38.99 dB (l)

Figure 1.3: Reconstruction results using different sampling strategies. Each sampling pattern contains 10% of 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 [16]. From left to right: sampling scheme - corresponding reconstruction - difference with the reference (the same colormap is used in every experiment).

5 Proof of the main results

5.1 Bernstein's inequalities

Theorem 1.13 (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 1.14 (Rectangular Matrix Bernstein Inequality). *[97, Theorem 1.6]*

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

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

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

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

Theorem 1.15 (Vector Bernstein Inequality (V1)). *[20, Theorem 2.6] Let $(y_k)_{1\leq k\leq m}$ be a finite sequence of independent and identically distributed random vectors of dimension n. Suppose that $\mathbb{E}y_1 = 0$ and $\|y_1\|_2 \leq K$ a.s. for some constant $K > 0$ and set $\sigma^2 \geq \sum_k \mathbb{E}\|y_k\|_2^2$. Let $Z = \|\sum_{k=1}^{m} y_k\|_2$. Then, for any $0 < t \leq \sigma^2/K$, we have that*

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

where $\mathbb{E}Z^2 = \sum_{k=1}^{m} \mathbb{E}\|y_k\|_2^2 = m\mathbb{E}\|y_1\|_2^2$.

Theorem 1.16 (Vector Bernstein Inequality (V2)). *[48, Corollary 8.44] Let $(y_k)_{1\leq k\leq m}$ be a finite sequence of independent and indentically distributed random vectors of dimension n. Suppose that $\mathbb{E}y_1 = 0$ and $\|y_1\|_2 \leq K$ a.s. for some constant $K > 0$. Let $Z = \|\sum_{k=1}^{m} y_k\|_2$. Then, for any $t > 0$, we have that*

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

where $\mathbb{E}Z^2 = \sum_{k=1}^{m} \mathbb{E}\|y_k\|_2^2 = m\mathbb{E}\|y_1\|_2^2$. Note that the previous inequality still holds by replacing $\mathbb{E}Z^2$ by σ^2 where $\sigma^2 \geq \mathbb{E}Z^2$.

5.2 Estimates: auxiliary results

Let S be the support of the signal to be reconstructed such that $|S| = s$. Note that the isotropy condition (1.3) ensures that the following properties hold

1. $\mathbb{E}(B^*B) = \mathrm{Id}_n$ and $\mathbb{E}(B_S^* B_S) = \mathrm{Id}_s$.
2. for any vector $w \in \mathbb{C}^s$, $\mathbb{E}[B_S w]^2 = \|w\|_2^2$.
3. for any $i \in S^c$, $\mathbb{E}(B_S^* B e_i) = 0$.

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

Lemma 1.17. *Let $S \subset \{1, \dots, n\}$ be of cardinality of s. Then, for any $\delta > 0$, one has that*

$$\mathbb{P}\left(\|A_S^* A_S - Id_s\|_{2\to 2} \geq \delta\right) \leq 2s \exp\left(-\frac{m\delta^2/2}{\mu_1(S) + \max(\mu_1(S) - 1, 1)\delta/3)}\right). \tag{E1}$$

Proof. We decompose the matrix $A_S^* A_S - \mathrm{Id}_s$ as

$$A_S^* A_S - \mathrm{Id}_s = \frac{1}{m}\sum_{k=1}^m \left(B_{k,S}^* B_{k,S} - \mathrm{Id}_s\right) = \frac{1}{m}\sum_{k=1}^m X_k,$$

where $X_k := \left(B_{k,S}^* B_{k,S} - \mathrm{Id}_s\right)$. It is clear that $\mathbb{E}X_k = 0$, and since $\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2} \leq \mu_1(S)$, we have that

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

Lastly, we remark that

$$0 \quad \preceq \quad \mathbb{E}X_k^2 = \mathbb{E}\left[B_{k,S}^* B_{k,S}\right]^2 - \mathrm{Id}_s \quad \preceq \quad \mathbb{E}\left\|B_{k,S}^* B_{k,S}\right\|_2 B_{k,S}^* B_{k,S} \preceq \mu_1(S)\mathrm{Id}_s.$$

Therefore, $\sum_{k=1}^m \mathbb{E}X_k^2 \preceq m\mu_1(S)\mathrm{Id}_s$ which implies that $\|\sum_{k=1}^m \mathbb{E}X_k^2\|_2 \leq m\mu_1(S)$. Hence, inequality (E1) follows immediately from Bernstein's inequality for random matrices (see Therorem 1.14). □

Lemma 1.18. *Let $S \subset \{1, \dots, n\}$, such that $|S| = s$. Let w be a vector in $\mathbb{C}^s$. Then, for any $t > 0$, one has that*

$$\mathbb{P}\left(\left\|\left(A_S^* A_S - Id_s\right) w\right\|_2 \geq \left(\sqrt{\frac{\mu_1(S) - 1}{m}} + t\right)\|w\|_2\right) \tag{E2}$$
$$\leq \exp\left(-\frac{mt^2/2}{(\mu_1(S) - 1) + 2\sqrt{\frac{\mu_1(S)-1}{m}}\mu_1(S) + \mu_1(S)t/3}\right).$$

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

$$\left(A_S^* A_S - \mathrm{Id}_s\right) w_S = \frac{1}{m}\sum_{k=1}^m \left(B_{k,S}^* B_{k,S} - \mathrm{Id}_s\right) w = \frac{1}{m}\sum_{k=1}^m y_k,$$

where $y_k = \left(B_{k,S}^* B_{k,S} - \mathrm{Id}_s\right) w$ is a random vector with zero mean. Simple calculations yield that

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

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

$$\begin{aligned}
\mathbb{E}\left[Z^2\right] &= \frac{1}{m}\mathbb{E}\left\|y_1\right\|_2^2 = \frac{1}{m}\mathbb{E}\left[\langle B_S^* B_S w, B_S^* B_S w\rangle - 2\left\langle B_S^* B_S w, w\right\rangle + \langle w, w\rangle\right] \\
&= \frac{1}{m}\mathbb{E}\left[\left\langle \left(B_S^* B_S\right)^2 w, w\right\rangle - 2\left\|B_S w\right\|_2^2 + 1\right].
\end{aligned}$$

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

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

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

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

Using the above upper bounds, namely $\left\|\frac{1}{m} y_k\right\|_2 \leq \frac{\mu_1(S)}{m}$ and $\mathbb{E}\left[Z^2\right] \leq \frac{\mu_1(S)-1}{m}$, the result of the lemma is thus a consequence of the Bernstein's inequality for random vectors (see Theorem 1.16), which completes the proof. □

Lemma 1.19. *Let $S \subset \{1, \ldots, n\}$, such that $|S| = s$. Let v be a vector of $\mathbb{C}^s$. Then we have*

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

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

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

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

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

Furthermore,

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

Using Bernstein's inequality 1.13 for complex random variables, we end to

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

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

Lemma 1.20. *Let S be a subset of $\{1,\dots,n\}$. Then, for any $0 < t < \frac{\mu_1(S)}{\mu_2(S)}$, one has that*

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

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

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

It follows that

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

Furthermore, using Cauchy-Schwarz inequality, one has that

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

Hence, using the above upper bounds, it follows from Bernstein's inequality for random vectors (see Theorem 1.15) that

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

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

5.3 Proofs of the main results

5.3.1 Proof of Theorem 1.2

In this section, we recall an inexact duality formulation of the minimization problem (1.1) in the form of sufficient conditions to guarantee that the vector x is the unique minimizer of (1.1), see [20]. These conditions give the properties that an inexact dual vector must satisfy to ensure the uniqueness of the solution of (1.1). In what follows, the notation $M_{|R}$ denotes the restriction of a square matrix M to its range R, and we define

$$\|M_{|R}^{-1}\|_{2\to 2} = \sup_{x\in R;\ \|x\|_2=1} \|M_{|R}^{-1}x\|_2$$

as the operator norm of the inverse of $M_{|R}$ restricted to its range.

Lemma 1.21 (Inexact duality [20])**.** *Suppose that $x \in \mathbb{R}^n$ is supported on $S \subset \{1,\dots,n\}$. Then, assume that*

$$\|\,(A_S^*A_S)_{|S}^{-1}\,\|_{2\to 2} \le 2 \qquad \text{and} \qquad \max_{i\in S^c} \|A_S^*Ae_i\|_2 \le 1. \tag{1.18}$$

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

$$\|v_S - \operatorname{sign}(x_S)\|_2 \le 1/4 \qquad \text{and} \qquad \|v_{S^c}\|_\infty \le 1/4, \tag{1.19}$$

Then, the vector x is the unique solution of the minimization problem (1.1)

First, let us focus on Conditions (1.18). We can remark that

$$\|\,(A_S^*A_S)_{|S}^{-1}\,\|_{2\to 2} = \left\|\sum_{k=1}^{\infty} (A_S^*A_S - \mathrm{Id}_s)^k\right\|_{2\to 2} \le \sum_{k=0}^{\infty} \|A_S^*A_S - \mathrm{Id}_s\|_{2\to 2}^k\,.$$

Therefore, if the condition $\|A_S^*A_S - \mathrm{Id}_s\|_{2\to 2} \le 1/2$ is satisfied, then $\|\,(A_S^*A_S)_{|S}^{-1}\,\|_{2\to 2} \le 2$. Hence, by Lemma 1.17, it is clear that $\|\,(A_S^*A_S)_{|S}^{-1}\,\|_{2\to 2} \le 2$ with probability at least $1-\varepsilon$, provided that

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

By definition of $\gamma(s)$, the first inequality of Conditions (1.18) is ensured with probability larger than $1-\varepsilon$ if

$$m \ge 8\left(\gamma(s) + \frac{1}{6}\max\left(\gamma(s) - 1, 1\right)\right)\ln\left(\frac{2s}{\varepsilon}\right). \tag{1.20}$$

Furthermore, using Lemma 1.20, we obtain that

$$\max_{i\in S^c} \|A_S^*Ae_i\|_2 \le 1$$

with probability larger than $1-\varepsilon$ if

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

Again by definition of $\gamma(s)$, the second part of Conditions (1.19) is ensured if

$$m \ge 9\gamma(s)\ln\left(\frac{n}{\varepsilon}\right). \tag{1.21}$$

Conditions (1.19) remain to be verified. The rest proof of Theorem 1.2 relies on the construction of a vector v satisfying the conditions described in Lemma 1.21 with high probability.To do so, we adapt the so-called golfing scheme introduced by Gross [53] to our setting. More precisely, we will iteratively construct a vector that converges to a vector v satisfying (1.19) with high probability.

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

$$v^{(\ell)} = \frac{m}{m_\ell} A^{(\ell)*} A_S^{(\ell)} \left(e - v_S^{(\ell-1)}\right) + v^{(\ell-1)}, \tag{1.22}$$

for $\ell = 1, \dots, L$. In the rest of the proof, we set $v = v^{(L)}$. By construction, v is in the row space of A. The main idea of the golfing scheme is then to combine the results from the various Lemmas in Section 5.2 with an appropriate choice of L and the number m of measurements, to show that the random vector v will satisfy the assumptions of Lemma 1.21 with large probability. Using the shorthand notation $v_S^{(\ell)} = P_S^* v^{(\ell)}$, let us define

$$w^{(\ell)} = e - v_S^{(\ell)},\ \ell = 1, \dots, L,$$

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

From the definition of $v_S^{(\ell)}$, it follows that, for any $1 \le \ell \le L$,

$$w^{(\ell)} = \left(\mathrm{Id}_s - \frac{m}{m_\ell} A_S^{(\ell)*} A_S^{(\ell)}\right) w^{(\ell-1)} = \prod_{j=1}^{\ell} \left(\mathrm{Id}_s - \frac{m}{m_j} A_S^{(j)*} A_S^{(j)}\right) e, \tag{1.23}$$

and

$$v = \sum_{\ell=1}^{L} \frac{m}{m_\ell} A^{(\ell)*} A_S^{(\ell)} w^{(\ell-1)}. \tag{1.24}$$

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

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

$$\left\|w^{(\ell)}\right\|_2 \le \underbrace{\left(\sqrt{\frac{\mu_1(S) - 1}{m_\ell}} + r_\ell\right)}_{r'_\ell} \left\|w^{(\ell-1)}\right\|_2 \tag{1.25}$$

$$\left\|\frac{m}{m_\ell}\left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)}\right\|_\infty \le t_\ell \|w^{(\ell-1)}\|_2. \tag{1.26}$$

The values of the quantities t_ℓ and r_ℓ, introduced in the above equations, will be specified later in the proof. Note that using (1.25), we can write that

$$\|e - v_S\|_2 = \|w^{(L)}\|_2 \le \|e\|_2 \prod_{\ell=1}^{L} r'_\ell \le \sqrt{s} \prod_{\ell=1}^{L} r'_\ell. \tag{1.27}$$

Furthermore, Equation (1.26) implies that

$$\begin{aligned}\|v_{S^c}\|_\infty &= \left\|\sum_{\ell=1}^{L} \frac{m}{m_\ell}\left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)}\right\|_\infty \\ &\leq \sum_{\ell=1}^{L} \left\|\frac{m}{m_\ell}\left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)}\right\|_\infty \\ &\leq \sum_{\ell=1}^{L} t_\ell \left\|w^{(\ell-1)}\right\|_\infty \\ &\leq \sqrt{s}\sum_{\ell=1}^{L} t_\ell \prod_{j=1}^{\ell-1} r'_j. \end{aligned} \tag{1.28}$$

We denote by $p_1(\ell)$ and $p_2(\ell)$ the probability that the upper bound (1.25), (1.26) do not hold. Now, let us set the number of blocks of blocks L, the number of blocks m_ℓ in each $A^{(\ell)}$ and the values of the parameters t_ℓ and r_ℓ that have been introduced above. We propose to make the following choices :

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

With such choices, we obtain that

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

and

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

Furthermore, using (1.27), we obtain that

$$\|e - v_S\|_2 \leq \sqrt{s}\prod_{\ell=1}^{L} r'_\ell \leq \frac{\sqrt{s}}{2^L} \leq \frac{1}{4}, \tag{1.29}$$

where the last inequality follows from the previously specified choice on L. Moreover, using (1.28), we have that

$$\begin{aligned}\|v_{S^c}\|_\infty &\leq \sqrt{s}\sum_{\ell=1}^{L} t_\ell \prod_{j=1}^{\ell-1} r'_j = \sqrt{s}\left(t_1 + t_2 r'_1 + t_3 r'_1 r'_2 + \ldots\right) \\ &\leq \left(\frac{1}{8} + \frac{1}{16\sqrt{\ln n}} + \frac{1}{32} + \ldots\right) \\ &\leq \frac{1}{4}. \end{aligned} \tag{1.30}$$

For such a choice of parameters, and by Lemmas 1.18 and 1.19, 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 $\ell = 3, \ldots, L$. 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 1.18 and 1.19 we finally obtain that if the overall number m of blocks samples obeys the condition

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

which can be simplified into

$$m \geq c\gamma(s) \left(2 \ln (4n) \ln \left(2\varepsilon^{-1}\right) + \ln s \ln \left(2e \ln(s)\varepsilon^{-1}\right)\right), \tag{1.31}$$

then the random vector v, defined by (1.24), satisfies Assumptions 1.19 of Lemma 1.21 with probability larger than $1 - 4\varepsilon$.

Hence, we have thus shown that if m satisfies the conditions (1.20), (1.21) and (1.31), then the Assumptions 1.18 and 1.19 of Lemma 1.21 simultaneously hold with probability larger than $1 - 6\varepsilon$. Note that the bound (1.31) is stronger than (1.20) and (1.21). We complete the proof of Theorem 1.2 by replacing ε by $\varepsilon/6$. The final result on the required number of blocks measurements reads as follows

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

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 strenghten the previous result. Yet, we could have used the clever trick used in [1], and reused in [54].

5.3.2 Proof of Proposition 1.6

Since $\gamma(s) = \max(\mu_1, \mu_2, \mu_3)$, it suffices to show that setting $\mu_i = s\mu_4$ for $i \in \{1, 2, 3\}$ is sufficient to ensure the inequalities (1.4).

The first inequality in (1.4) can be shown as follows:

$$\|B_S{}^* B_S\|_{2\to 2} \leq \|B_S{}^* B_S\|_{\infty\to\infty} \leq s\|B^* B\|_{1\to\infty} \leq s\mu_4.$$

The second inequality in (1.4) can be shown as follows:

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

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

$$\begin{aligned} s\mathbb{E} B_S{}^* \left(Be_i\right)\left(Be_i\right)^* B_S &\preceq s\| \left(Be_i\right)\left(Be_i\right)^* \|_{2\to 2} \mathbb{E} B_S{}^* B_S \\ &\preceq s \max_i \|Be_i\|_2^2 \mathrm{Id} \\ &\preceq s \|B^* B\|_{1\to\infty} \mathrm{Id} \\ &\preceq s\mu_4 \mathrm{Id}. \end{aligned}$$

5.3.3 Proof of Proposition 1.7

Let us evaluate the quantities $(\mu_i(S))_{1\leq i\leq 3}$ introduced in Definition 1.1 to upper bound $\gamma(s)$ with high probability. For this purpose, using Theorem 2 in [66], we get that for any $0<t<1$

$$\mathbb{P}\left(\|B_S^*B_S\|_{2\to 2}\geq\left(1+\sqrt{\frac{s}{p}}\right)^2(1+t)\right)\leq C\exp\left(-\sqrt{ps}t^{3/2}\left(\frac{1}{\sqrt{t}}\wedge\left(\frac{s}{p}\right)^{1/4}\right)/C\right), \tag{1.32}$$

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

$$\max_{i\in S^c}\|B_S^*Be_i\|_2\leq\max_{i\in S^c}\|B_S^*\|_{2\to 2}\|Be_i\|_2\leq\sqrt{\|B_S^*B_S\|_{2\to 2}}\max_{i\in S^c}\sqrt{\|Be_i\|_2^2}.$$

We already know that the first term $\sqrt{\|B_S^*B_S\|_{2\to 2}}$ in the above inequality is bounded by $\sqrt{\frac{s}{p}}$ (up to a constant) with high probability, thanks to the previous discussion on $\mu_1(S)$. As for the second term, we use a union bound and the sub-gamma property of the chi-squared distribution, see [13, p.29], to derive that

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

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

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

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

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

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 Be_i for $i\in S^c$, we have that

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

Hence, one can take $\mu_3(S)=\frac{s}{p}$. Combining all these estimates we get that $\gamma(s)\lesssim\frac{s}{p}\sqrt{\delta\ln(s)}$. Therefore, assuming that the lower bound on m in Theorem 1.2 with bound (1.5) 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 $q=O(s\ln(s)\ln(n))$.

5.3.4 Proof of Proposition 1.11

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 $s \in \{1, \ldots, \sqrt{n}\}$. Let $\mathcal{C}_s$ denote the class of vectors of kind $x = \alpha \otimes e_1$, where $\boldsymbol{\alpha} \in \mathbb{R}^{\sqrt{n}}$ is s-sparse. Note that every $x \in \mathcal{C}_s$ is s-sparse and that

$$\begin{aligned} Ax &= (\widetilde{\Psi}_{K,:} \otimes \Psi) \cdot (\boldsymbol{\alpha} \otimes e_1) \\ &= \left(\widetilde{\Psi}_{K,:}\alpha\right) \otimes \Psi e_1. \end{aligned}$$

In order to identify every s-sparse x knowing $y = Ax$, there should not exist two distinct s-sparse vectors $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\alpha}^{(2)}$ in $\mathbb{C}^{\sqrt{n}}$ such that $\widetilde{\Psi}_{K,:}\boldsymbol{\alpha}^{(1)} = \widetilde{\Psi}_{K,:}\boldsymbol{\alpha}^{(2)}$. The vector $\boldsymbol{\alpha}^{(1)} - \boldsymbol{\alpha}^{(2)}$ 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 $\widetilde{\Psi}_{K,:}\boldsymbol{\alpha} \neq 0$ for all non-zero $\min(2s, \sqrt{n})$-sparse vectors α. To finish the first part of the proof it suffices to remark that a necessary condition for a set of $\min(2s, \sqrt{n})$ columns of $\widetilde{\Psi}_{K,:}$ to be linearly independent is that $m = |K| \geq \min(2s, \sqrt{n})$, see Lemma 1.10.

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

$$Ax = \sum_{l=1}^{\sqrt{n}} \left(\widetilde{\Psi}_{K,:}\boldsymbol{\alpha}^{(l)}\right) \otimes \Psi e_l.$$

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

5.3.5 Proof of Proposition 1.12

We consider blocks that consist of discrete lines in the 2D Fourier space. 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 $p_1 \in \{1, \ldots, \sqrt{n}\}$,

$$B_{p_1} = \left[\frac{1}{\sqrt{n}} \exp\left(2i\pi\left(\frac{p_1\ell_1 + p_2\ell_2}{\sqrt{n}}\right)\right)\right]_{(p_1,p_2)(\ell_1,\ell_2)} \tag{1.33}$$

with $1 \leq p_2 \leq \sqrt{n}, 1 \leq \ell_1, \ell_2 \leq \sqrt{n}$. Let $S \subset \{1, \ldots, \sqrt{n}\} \times \{1, \ldots, \sqrt{n}\}$ denote the support of x, with $|S| = s$. By definition of the 2D Fourier matrix of size $n \times n$, $\|B_k^* B_k\|_{1\to\infty} = 1/\sqrt{n}$, for all $k \in \{1, \ldots, \sqrt{n}\}$. Thus, Theorem 1.2 with bound (1.5) leads to

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

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

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

and the number of measurements can be written as follows

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

which ends the proof of Proposition 1.12.

5.4 An example with overlapping blocks

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

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

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

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

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 1.9, we conclude that the required number of blocks of measurements must satisfy

$$m \ge cs \frac{1}{2\sqrt{n}} \max_{1\le k\le M} \frac{1}{\pi_k} \left(2\ln(4n) \ln\left(12\varepsilon^{-1}\right) + \ln s \ln\left(12e\ln(s)\varepsilon^{-1}\right)\right). \tag{1.34}$$

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

$$m \ge cs \left(2\ln(4n) \ln\left(12\varepsilon^{-1}\right) + \ln s \ln\left(12e\ln(s)\varepsilon^{-1}\right)\right), \tag{1.35}$$

which is the same requirement in the 2D Fourier domain without overlapping, see Proposition 1.12.

2 Compressed sensing with structured sparsity and structured acquisition

Contents

1 Introduction

Since its introduction in [23, 42], compressive sampling triggered a massive interest in fundamental and applied research. However, despite recent progresses, existing theories are still insufficient to explain the success of compressed acquisitions in many practical applications. Our aim in this chapter is to extend the applicability of the theory by combining two new ingredients: structured sparsity and acquisition structured by blocks.

1.1 A brief review of existing results

Compressed sensing - as proposed in [22] - consists in recovering a signal $x \in \mathbb{C}^n$, from a vector of measurements $y = Ax$, where $A \in \mathbb{C}^{m \times n}$ is the sensing matrix. Typical theorems state that if x is s-sparse, $m \gtrsim s \ln(n)$ and A have some good features, then x can be recovered exactly from y by solving the following ℓ_1 minimization problem :

$$\min_{x \in \mathbb{C}^n, Ax=y} \|x\|_1. \tag{2.1}$$

Moreover, it can be shown that the recovery is robust to noise if the constraint in (2.1) is penalized. The important fact about this theorem is that the number of measurements mostly depends on the intrinsic dimension s rather than the ambient dimension n.

The first sensing matrices studied were generated by selecting a few Fourier coefficients uniformly at random [22]. The theory was then extended to random matrices with i.i.d. components [25] and sensing vectors selected randomly from orthogonal bases [23] or discrete or continuous frames [20]. An interesting class of sensing matrices for application was introduced in [90] and based on a convolution with a random vector. In the meanwhile, different tools were introduced to analyse sensing matrices such as UUP (uniform uncertainty principle), RIP (restricted isometry property). Many recent works on CS are rather based on coherence or local coherence [63] and RIPless proofs [20]. The book [48] proposes a detailed and self-contained description of most of those concepts.

A common aspect of the above results is that they assume no structure - apart from sparsity - in the signals to recover. Recovering arbitrary sparse vectors is a very demanding property that precludes the use of CS in many practical settings. To the best of our knowledge, the work [2] is the first to consider the recovery of sparse signals with a structured support. To treat such cases, the authors introduce new concepts such as sparsity by levels.

1.2 The need for new results

One of the main current limitations of CS is the small number of sensing matrices studied so far. Let us illustrate this insufficiency with a practical example from Magnetic Resonance Imaging (MRI). This example will be the red thread of the chapter.

In MRI, images are sampled in the Fourier domain and can be assumed to be sparse in the wavelet domain. Under this hypothesis, a byproduct of standard compressed sensing results [20] imply that variable density sampling [84, 30, 63] allows perfect reconstruction with a limited number of measurements. The theory in [2], based on structured sparsity, also leads to the same conclusion. This is illustrated in Figure (2.1). The white dots on the left image indicate which Fourier coefficients are probed. As can be seen, 4.6% of the coefficients are enough to reconstruct a well resolved image.

Unfortunately, probing Fourier coefficients independently at random is infeasible in MRI: the samples have to lie on piecewise smooth trajectories [28, 16]. One of the most successful practical sampling scheme in MRI consists of measuring whole lines of Fourier coefficients at random [69]. The lines are all parallel and drawn indepently, at random, according to a certain distribution. This is illustrated in Figure 2.2 in 2D [1]. As can be seen on this example, compressed acquisitions with a lot of structure make it possible to reconstruct well resolved images. To the best of our knowledge, there currently exists no theory able to explain this favorable behavior. The only works dealing with such an acquisition are [80, 11]. They assume no structure in the sparsity and we showed in [11] that structure was crucially needed to explain results such as those in Figure 2.2. We will recall this result in Section 4.3.1.

1.3 Contributions

The main contributions of this work are the following: (i) we provide recovery guarantees for vectors $x \in \mathbb{C}^n$ with a fixed support $S \subset \{1, \dots, n\}$. This is in strong contrast with the usual works that consider the reconstruction of arbitrary s-sparse vectors. (ii) we provide a theoretical justification to the use of block acquisitions in compressed sensing. By doing so, we enrich the family of sensing matrices available for compressed sensing.

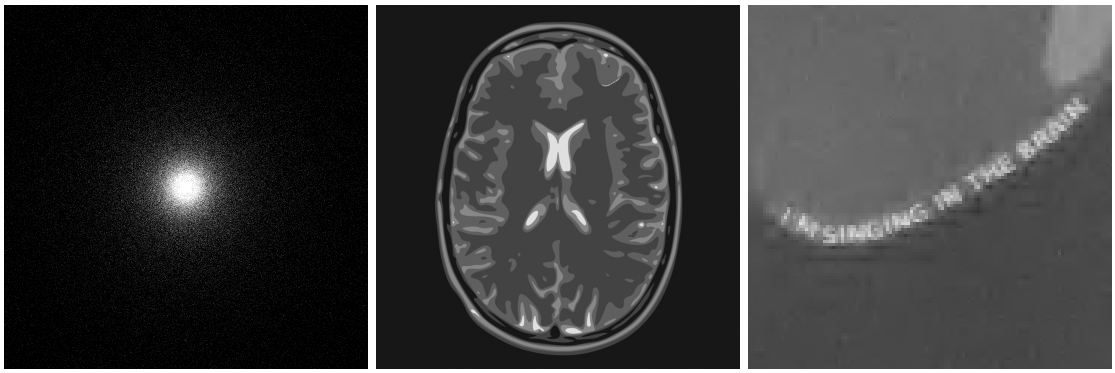
The proposed theory has a few important consequences:

- the concepts of RIP or coherence are not sound anymore. They are replaced by a new quantity $\Gamma(S, \pi)$ which explicitly depends on the support S, the sensing vectors and the drawing probability π of measurements.
- the proposed theory allows envisioning the use of CS in situations that were not possible before. The use of incoherent transforms is not necessary anymore, given that the support S has some good properties.
- The example given in Figure 2.2 can be analyzed precisely. In particular, we show that a block structured acquisition can be used, only if the support structure is adapted to it. The resulting structures are more complex than the sparsity by levels of [2].
- The explicit dependency on the support S allows to provide guarantees of reconstruction for random signals with known distribution.

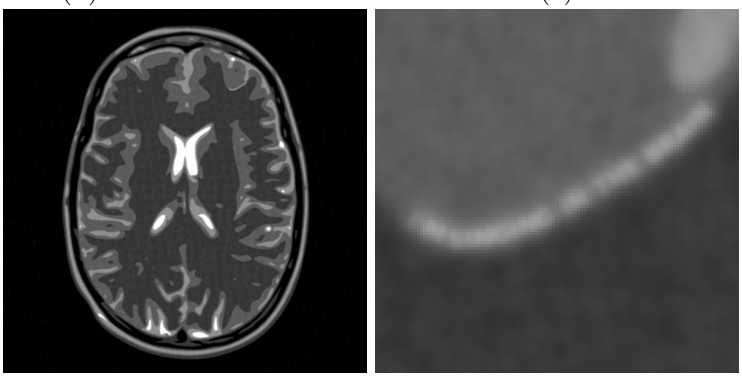
1.4 Related notions in the literature

In this work, structured acquisition denotes the constraints imposed by the physics of the acquisition, that are modeled using blocks of measurements extracted from a full deterministic matrix A_0. This notion of structured acquisition differs from the notion of structured random matrices, as described in [86] and [44]. Indeed, this latter strategy is based on acquiring isolated measurements randomly drawn from the rows of a deterministic matrix. The resulting sensing matrix has thus some inherent structure, which is not the case of random matrices with i.i.d. entries, that were initially considered in CS. In this chapter, the sensing matrix A is even more structured, in the sense that the full sampling matrix A_0 has been partitioned into blocks of measurements.

[1] Paper [69] considers 3D lines.



(a) (b) SNR = 24.2 dB (c)

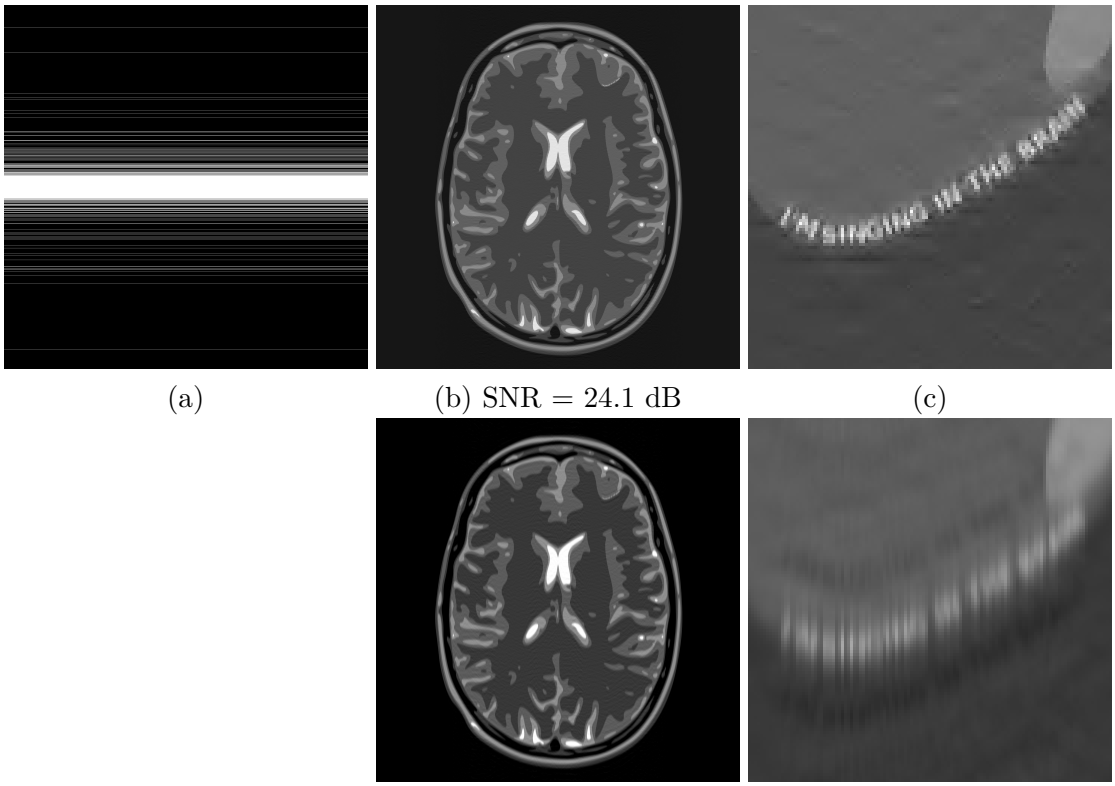


(d) SNR = 21 dB (e)

Figure 2.1: An example of reconstruction of a 2048×2048 MR image from isolated measurements. (a) Sensing pattern from a variable density sampling strategy (with 4.6% measurements). (b) Corresponding reconstruction via ℓ_1-minimization with in (c) a zoom on a part of the reconstructed image. (d) Image obtained by using the pseudo-inverse transform with in (e) a zoom on a part of this image.

We also focus on obtaining RIPless results by combining structured acquisition and structured sparsity. RIPless results [20] refer to CS approaches that are non-uniform in the sense they hold for a given sensing matrix A and a given support S of length s, but not for all s-sparse vectors. Nevertheless, existing RIPless results in the literature are only based on the degree of sparsity $s = |S|$. A main novelty of this work is to develop RIPless results that depend explicitly on the support S (and not only on its length s) of the signal to reconstruct. This strategy allows to incorporate any kind of *a priori* information on the structure of S to study its influence on the quality of CS reconstructions. To the best of our knowledge, this setting has not been considered so far, even if preliminary results have been obtained in [51, 56].

In [44], a more general model on the signal sparsity is also considered. Indeed, it deals with sparse signals that can be represented in a union of subspaces. Nevertheless, in [44], they adapt the recovery algorithm to the chosen assumption on sparsity. In this chapter, any assumption on the support S of the signal to reconstruct can be addressed, including the case of union of subspaces for instance. Moreover, we do not particularize the reconstruction method. Indeed, instead of modifying the recovery algorithm (i.e. the ℓ_1-minimization problem) as in [44, 58], we focus on adapting the sampling scheme to the assumption made on sparsity. Furthermore, as [4] suggests, it seems that exploiting structure in sampling is more efficient that exploiting structure in the recovery algorithm.



(a) (b) SNR = 24.1 dB (c)

(d) SNR = 21 dB (e)

Figure 2.2: An example of reconstruction of a 2048 × 2048 MR image from blocks of measurements. (a) Sampling pattern horizontal lines (13% of measurements). (b) Corresponding reconstruction via ℓ_1-minimization with in (c) a zoom on a part of the reconstructed image. (d) Image obtained by using the pseudo-inverse transform with in (e) a zoom on a part of this image.

2 Preliminaries

2.1 Notation

In this chapter, n denotes the dimension of the signal to reconstruct. The notation $S \subset \{1, \dots, n\}$ refers to the support of the signal to reconstruct. The vectors $(e_i)_{1 \leq i \leq p}$ denote the vectors of the canonical basis of $\mathbb{R}^d$, where d will be equal to n or $\sqrt{n}$, depending on the context. In the sequel, we set $P_S \in \mathbb{R}^{n \times n}$ to be the projection matrix onto $\text{span}\left(\{e_i, i \in S\}\right)$, i.e. the diagonal matrix with the j-th diagonal entry equal to 1 if $j \in S$, and 0 otherwise. We will use the shorthand notation $M_S \in \mathbb{C}^{n \times n}$ and $v_S \in \mathbb{C}^n$ to denote the matrix MP_S and the vector $P_S v$ for $M \in \mathbb{C}^{n \times n}$ and $v \in \mathbb{C}^n$. For any matrix M, for any $1 \leq p, q \leq \infty$, the operator norm $\|M\|_{p \to q}$ is defined as

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

with $\|\cdot\|_p$ and $\|\cdot\|_q$ denoting the standard ℓ_p and ℓ_q norms. Note that for a matrix $M \in \mathbb{R}^{n \times n}$,

$$\|M\|_{\infty \to \infty} = \max_{1 \leq i \leq n} \|e_i^* M\|_1.$$

Finally, the function sign : $\mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$(\text{sign}(x))_i = \begin{cases} 1 & \text{if} \quad x_i > 0 \\ -1 & \text{if} \quad x_i < 0 \\ 0 & \text{if} \quad x_i = 0. \end{cases}$$

2.2 Sampling strategy

In this chapter, we assume that we are given some orthogonal matrix $A_0 \in \mathbb{C}^{n\times n}$ representing the set of possible linear measurements imposed by a specific sensor device. Let $(\mathcal{I}_k)_{1\leq k\leq M}$ a partition of the set $\{1,\dots,n\}$. The rows $(a_i^*)_{1\leq i\leq n} \in \mathbb{C}^n$ of A_0 are partitioned into the following blocks dictionary $(B_k)_{1\leq k\leq M}$, such that

$$B_k = (a_i^*)_{i\in\mathcal{I}_k} \in \mathbb{C}^{|\mathcal{I}_k|\times n} \qquad s.t. \qquad \mathcal{I}_k \subset \{1,\dots,n\},$$

with $\sqcup_{k=1}^M \mathcal{I}_k = \{1,\dots,n\}$. The sensing matrix A is then constructed by randomly drawing blocks as follows

$$A = \frac{1}{\sqrt{m}} \left(\frac{1}{\sqrt{\pi_{K_\ell}}} B_{K_\ell} \right)_{1\leq \ell\leq m}, \tag{2.2}$$

where $(K_\ell)_{1\leq\ell\leq m}$ are i.i.d. copies of a random variable K such that

$$\mathbb{P}(K = k) = \pi_k,$$

for all $1 \leq k \leq M$. Moreover, thanks to the renormalization of the blocks B_{K_ℓ} by the weights $1/\sqrt{\pi_{K_\ell}}$ in model (2.2), the random block B_K satisfies

$$\mathbb{E}\left(\frac{B_K^* B_K}{\pi_K}\right) = \sum_{k=1}^M B_k^* B_k = \text{Id}, \tag{2.3}$$

since A_0 is orthogonal and $(B_k)_{1\leq k\leq M}$ is a partition of the rows of A_0.

Remark 2.1. *The case of overlapping blocks can also be handled. To do so, we may define the blocks $(B_k)_{1\leq k\leq M}$ as follows:*

$$B_k = \left(\frac{1}{\sqrt{\alpha_i}} a_i^* \right)_{i\in\mathcal{I}_k}, \qquad \textit{for} \quad 1 \leq k \leq M,$$

where $\bigcup_{k=1}^M \mathcal{I}_k = \{1,\dots,n\}$. The coefficients $(\alpha_i)_{1\leq i\leq n}$ denotes the multiplicity of the row a_i^*, namely the number of appearances $\alpha_i = |\{k, i \in \mathcal{I}_k\}|$ of this row in different blocks. This renormalization is sufficient to ensure the isotropy condition $\mathbb{E}\left(\frac{B_K^* B_K}{\pi_K}\right) = Id$ where K is defined as above.

Note that our block sampling strategy encompasses the standard acquisition based on isolated measurements. Indeed, isolated measurements can be considered as blocks of measurements consisting of only one row of A_0.

Remark 2.2. *Note that the setting could be extended to the case where the sensing matrix is*

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} B_{K_1} \\ \vdots \\ B_{K_m} \end{pmatrix}$$

where $B_{K_1}, \dots, B_{K_m}$ are i.i.d. copies of a random matrix $B \in \mathbb{C}^{b\times n}$ satisfying

$$\mathbb{E}(B^*B) = Id.$$

The integer b is itself random and Id is the $n \times n$ identity matrix. Assuming that B takes its value in a countable family $(B_k)_{k\in\mathcal{K}}$, this formalism covers a large number of applications described in [11]: (i) blocks with i.i.d. entries, (ii) partition of the rows of orthogonal transforms, (iii) cover of the rows of orthogonal transforms, (iv) cover of the rows from tight frames.

3 Main Results

3.1 Fundamental quantities

Before introducing our main results, we need to define some quantities (reminiscent of the coherence) that will play a key role in our analysis.

Definition 2.3. *Consider a blocks dictionary $(B_k)_{1\le k\le M}$. Let $S \subset \{1, \dots, n\}$ and π be a probability distribution on $\{1, \dots, M\}$. Define*

$$\Theta(S,\pi) := \max_{1\le k\le M} \frac{1}{\pi_k} \|B_k^* B_{k,S}\|_{\infty\to\infty} = \max_{1\le k\le M} \max_{1\le i\le n} \frac{\|e_i^* B_k^* B_{k,S}\|_1}{\pi_k}, \tag{2.4}$$

$$\Upsilon(S,\pi) := \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^{M} \frac{1}{\pi_k} \left|e_i^* B_k^* B_{k,S} v\right|^2, \tag{2.5}$$

$$\Gamma(S,\pi) := \max\left(\Upsilon(S,\pi), \Theta(S,\pi)\right). \tag{2.6}$$

For the sake of readability, we will sometimes use the shorter notation Θ, Υ and Γ to denote $\Theta(S,\pi), \Upsilon(S,\pi)$ and $\Gamma(S,\pi)$. In Definition 2.3, Θ is related to the local coherence and the degree of sparsity, when the blocks are made of only one row (the case of isolated measurements). Indeed, in such a case, Θ reads as follows

$$\Theta(S,\pi) := \max_{1\le k\le n} \frac{\|a_k\|_\infty \|a_{k,S}\|_1}{\pi_k} \le s \cdot \max_{1\le k\le n} \frac{\|a_k\|_\infty^2}{\pi_k}.$$

The quantity $\max_{1\le k\le n} \frac{\|a_k\|_\infty^2}{\pi_k}$ refers to the usual notion of coherence described in [20]. The quantity Υ is new and it is more delicate to interpret. It reflects an inter-block coherence. A rough upper-bound for Υ is

$$\Upsilon(S,\pi) \le \sum_{k=1}^{M} \frac{1}{\pi_k} \|B_k^* B_{k,S}\|_{\infty\to\infty}^2.$$

by switching the maximum and supremum with the sum in the definition of Υ. However, it is important to keep this order (maximum, supremum and sum) to measure interferences between blocks. In Section 4, we give more precise evaluations of $\Theta(S,\pi)$ and $\Upsilon(S,\pi)$ in particular cases.

Remark 2.4 (Support-dependency and drawing-dependency)**.** *In Definition 2.3, the quantities Θ and Υ are drawing-dependent and support-dependent. Indeed, Γ does not only depend on the degree of sparsity $s = |S|$. To the best of our knowledge, existing theories in CS only rely on s, see [22, 20], or on degrees of sparsity structured by levels, see*

[2]. Since Γ is explicitly related to S, this allows to incorporate prior assumptions on the structure of S. Besides, the dependency on π (i.e. the way of drawing the measurements) is also explicit in the definition of Γ. This offers the flexibility to analyze the influence of π on the required number of measurements. We therefore believe that the introduced quantities might play an important role in the future analysis of CS.

3.2 Exact recovery guarantees

Our main result reads as follows.

Theorem 2.5. *Let $S \subset \{1, \dots, n\}$ be a set of indices of cardinality $s \geq 16$ 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 (2.2). Suppose that $\Gamma(S, \pi) \geq 1$. If*

$$m \geq 73 \cdot \Gamma(S, \pi) \ln(64s) \left(\ln\left(\frac{27n}{\varepsilon}\right) + \ln\ln(55s) \right), \tag{2.7}$$

then x is the unique solution of (2.1) with probability larger than $1 - \varepsilon$.

Remark 2.6. *In the sequel, we will simplify condition (2.7) by writing:*

$$m \geq C \cdot \Gamma(S, \pi) \ln(s) \ln\left(\frac{n}{\varepsilon}\right)$$

where C is a universal constant.

The proof of Theorem 2.5 is contained in Appendix 6.1. It relies on the construction of an inexact dual certificate satisfying appropriate properties that are described in Lemma 2.20. Then our proof is based on the so-called golfing scheme introduced in [53] for matrix completion, and adapted by [20] for compressed sensing from isolated measurements. Nevertheless, the methodology in the proof differs from the techniques that are described in [20]. Indeed, a straightforward adaptation of the arguments in [20] does not allow to relate the number of measurements to the support S. In the golfing scheme, the main trick is the control of operator norms of random matrices extracted from the sensing matrix A. In [20], it is proposed to control (in probability) the operator norms $\|\cdot\|_{\infty\to 2}$ and $\|\cdot\|_{2\to 2}$. However, this technique only gives results depending on the degree of sparsity s. In order to include an explicit dependency on the support S, one has to modify the golfing scheme in [20], by controlling the operator norm $\|\cdot\|_{\infty\to\infty}$, instead of controlling the operator norms $\|\cdot\|_{\infty\to 2}$ and $\|\cdot\|_{2\to 2}$. A similar idea has been developed in [2] but our main result is more general than the finite-dimensional setting in [2].

Remark 2.7. *Compared to standard results in compressed sensing, the condition required in Theorem 2.5 involves an extra $\ln(s)$ factor. The latter can be removed at the price of additional technicalities, using an extra hypothesis called the Balancing Property in [2]. In this chapter, this possibility has been put aside for the sake of clarity.*

3.3 Recovery guarantees from noisy measurements

Note that once the exact recovery case has been considered, one can easily deduce a stability result when measurements are corrupted by noise, i.e.

$$y = Ax + h,$$

with $h \in \mathbb{C}^m$ a noise vector satisfying $\|h\|_2 \leq \eta$, with $\eta > 0$. The recovery problem can be thus written as follows

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \qquad \text{s.t.} \qquad \|y - Az\|_2 \leq \eta. \tag{2.8}$$

For such a setting, we can derive the following result.

Theorem 2.8. *Let $S \subset \{1, \dots, n\}$ be a set of indices of cardinality $s \geq 16$ 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 (2.2). Suppose that $\Gamma(S, \pi) \geq 1$. Let $x^\sharp$ be a solution of Problem (2.8). If*

$$m \geq 73 \cdot \Gamma(S, \pi) \ln(64s) \left(\ln\left(\frac{27n}{\varepsilon}\right) + \ln\ln(55s) \right), \tag{2.9}$$

then with probability at least $1 - \varepsilon$, the reconstruction error satisfies

$$\|x - x^\sharp\|_2 \leq c_1 \sigma_s(x)_1 + (c_2 + c_3\sqrt{s})\eta,$$

for $c_1 = 24$, $c_2 = 16\sqrt{3/2}$, $c_3 = 24\sqrt{2}$ and $\sigma_s(x)_1$ is the error of the best s-term approximation given by $\sigma_s(x)_1 := \min_{\|x'\|_0 \leq s} \|x' - x\|_1$.

The proof is proposed in Section 6.2. Note that, due to the extra $\ln(s)$ factor, this result is not as sharp as the one obtained under the ℓ_2-null space property. However, it applies under weaker and verifiable conditions on the sensing matrix A in our case.

In the examples, we will be interested in the required number of measurements. Thus, we will focus on the noiseless recovery case, since the bounds on m are the same either in the noiseless case or in the noisy one.

3.4 Consequences for stochastic signal models

The explicit dependency of Γ in S allows us to consider the case of a random support S.

Proposition 2.9. *Let $S \subset \{1, \dots, n\}$ be a random support of fixed size s. Let ε and ε' be some constants in $[0, 1]$. Suppose that $\Gamma(S, \pi) \leq \gamma := \gamma(s, \pi, n, \epsilon')$ occurs with probability larger than $1 - \varepsilon'$. If $m \gtrsim \gamma \ln(s) \ln(n/\varepsilon)$, then x is the unique solution of Problem 2.1 with probability higher than $1 - \varepsilon - \varepsilon' + \varepsilon\varepsilon'$.*

Proof. Set $m \gtrsim \gamma \ln(s) \ln(n/\varepsilon)$. Define the event R "x is the unique solution of Problem 2.1" where R stands for "reconstruction of the signal". Define also B the event "$\Gamma(S, \pi) \leq \gamma$". The hypothesis of Proposition 2.9 and Theorem 2.5 give that $\mathbb{P}(R|B) \geq 1 - \varepsilon$. To prove Proposition 2.9, we must quantify

$$\begin{aligned} \mathbb{P}(R) &= \mathbb{P}(R \cap B) + \mathbb{P}(R \cap B^c) = \mathbb{P}(R|B)\,\mathbb{P}(B) + \mathbb{P}(R \cap B^c) \\ &\geq (1 - \varepsilon)(1 - \varepsilon') = 1 - \varepsilon - \varepsilon' + \varepsilon\varepsilon', \end{aligned}$$

which concludes the proof. □

3.5 Choice of the drawing probability

The choice of a drawing probability π minimizing the required number of block measurements in Theorem 2.5, is a delicate issue. The distribution $\pi^\star$ minimizing $\Theta(S,\pi)$ in Equation (2.4) can be obtained explicitly:

$$\pi_k^\star = \frac{\|B_k^* B_{k,S}\|_{\infty\to\infty}}{\sum_{\ell=1}^M \|B_\ell^* B_{\ell,S}\|_{\infty\to\infty}}, \quad \text{for} \quad 1 \le k \le M. \tag{2.10}$$

Unfortunately, the minimization of $\Upsilon(S,\pi)$ with respect to π seems much more involved and we leave this issue as an open question in the general case.

Note however that in all the examples treated in this chapter, we derive upper bounds depending on (S,π) for $\Upsilon(S,\pi)$ and $\Theta(S,\pi)$ that coincide. The distribution $\pi^\star$ is then set to minimize the latter upper bound.

Note also that optimizing π independently of S will result in a sole dependence to the degree of sparsity $s = |S|$ which is not desirable if one wants to exploit structured sparsity.

4 Applications

In this section, we first show that Theorem 2.5 can be used to recover state of the art results in the case of isolated measurements [20]. We then show that it allows recovering recent results when a prior on the sparsity structure is available. The proposed setting however applies to a wider setting even in the case of isolated measurements. Finally, we illustrate the consequences of our results when the acquisition is constrained by blocks of measurements. In the latter case, we show that the sparsity structure should be adapted to the sampling structure for exact recovery.

4.1 Isolated measurements with arbitrary support

First, we focus on an acquisition based on isolated measurements which is the most widespread in CS. This case corresponds to choose blocks of form $B_k = a_k^*$ for $1 \le k \le n$ with $M = n$, where a_k^* are the rows of an orthogonal matrix. In such a setting, the sensing matrix can be written as follows

$$A = \frac{1}{\sqrt{m}} \left(\frac{1}{\sqrt{\pi_{K_\ell}}} a_{K_\ell}^* \right)_{1\le \ell \le m}, \tag{2.11}$$

where $(K_\ell)_{1\le\ell\le m}$ are i.i.d. copies of K such that $\mathbb{P}(K=k) = \pi_k$, for $1 \le k \le n$.

We apply Theorem 2.5 when only the degree of sparsity s of the signal to reconstruct is known. This is the setting considered in most CS papers (see e.g. [25, 86, 20]). In this context, our main result can be rewritten as follows.

Corollary 2.10. *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. Fix $\varepsilon \in (0,1)$. Suppose that the sampling matrix A is constructed as in (2.11). If*

$$m \;\ge\; C \cdot s \cdot \max_{1\le k\le n} \frac{\|a_k\|_\infty^2}{\pi_k} \ln(s) \ln\left(\frac{n}{\varepsilon}\right), \tag{2.12}$$

then x is the unique solution of (2.1) with probability at least $1-\varepsilon$.

Moreover, the drawing distribution minimizing (2.12) *is* $\pi_k = \frac{\|a_k\|_\infty^2}{\sum_{\ell=1}^n \|a_\ell\|_\infty^2}$, *which leads to*

$$m \geq C \cdot s \cdot \sum_{k=1}^{n} \|a_k\|_\infty^2 \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

The proof is given in Appendix 6.5.1.

Note that Corollary 2.10 is identical to Theorem 1.1 in [20] up to logarithmic factors. This result is usually used to explain the practical success of variable density sampling. It is the core of papers such as [84, 63, 28].

4.2 Isolated measurements with structured sparsity

When using coherent transforms, meaning that the term $\max_{1\leq k\leq n} \frac{\|a_k\|_\infty^2}{\pi_k}$ in Equation (2.12) is an increasing function of n, Corollary 2.10 is unsufficient to justify the use of CS in applications. In this section, we show that the proposed results allow justifying the use of CS even in the extreme case where the sensing is performed with the canonical basis.

4.2.1 A toy example: sampling isolated measurements from the Identity matrix and knowing the support S

Suppose that the signal x to reconstruct is S-sparse where $S \subseteq \{1, \ldots, n\}$ is a fixed subset. Consider the highly coherent case where $A_0 = \mathrm{Id}$. All current CS theories would give the same unsatisfactory conclusion: it is not possible to use CS since A_0 is a perfectly coherent transform. Indeed, the bound on the required number of isolated measurements given by standard CS theories [20] reads as follows

$$m \geq C \cdot s \cdot \max_{1\leq k\leq n} \frac{\|e_k^*\|_\infty^2}{\pi_k} \ln(n/\varepsilon) = C \cdot s \cdot \max_{1\leq k\leq n} \frac{1}{\pi_k} \ln(n/\varepsilon).$$

Without any assumption on the support S, one can choose to draw the measurements uniformly at random, i.e. $\pi_k = 1/n$ for $1 \leq k \leq n$. This particular choice leads to a required number of measurements of the order

$$m \geq C \cdot s \cdot n \ln(n/\varepsilon),$$

which corresponds to fully sampling the acquisition space several times.

Let us now see what conclusion can be drawn with Theorem 2.5.

Corollary 2.11. *Let* $S \subseteq \{1, \ldots, n\}$ *of cardinality* s. *Suppose that* $x \in \mathbb{C}^n$ *is an* S*-sparse vector. Fix* $\varepsilon \in (0, 1)$. *Suppose that the sampling matrix* A *is constructed as in* (2.11) *with* $A_0 = Id$. *Set* $\pi_k = \frac{\delta_{k,S}}{s}$ *for* $1 \leq k \leq n$ *where* $\delta_{k,S} = 1$ *if* $k \in S$, 0 *otherwise. Suppose that*

$$m \geq C \cdot s \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

then x *is the unique solution of* (2.1) *with probability at least* $1 - \varepsilon$.

With this new result, $O(s \ln(s) \ln(n))$ measurements are sufficient to reconstruct the signal via a totally coherent. The least amount of measurements necessary to recover x is of order $O(s \ln(s))$, by an argument of coupon collector effect [46, p.262]. Therefore, Corollary 2.11 is near-optimal up to logarithmic factors.

Proof. The result ensues from a direct evaluation of Γ. Indeed,

$$\|e_k e_{k,S}^*\|_{\infty\to\infty} = \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} |\left\langle e_i, e_k e_{k,S}^* v\right\rangle| = \sup_{\|v\|_\infty\le 1} |e_{k,S}^* v| = \delta_{k,S},$$

where $\delta_{k,S} = 1$ if $k \in S$, 0 otherwise. Therefore

$$\Theta = \max_{1\le k\le n} \frac{\delta_{k,S}}{\pi_k}.$$

Then, we can write that

$$\begin{aligned}\Upsilon(S,\pi) &= \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* e_k e_{k,S}^* v|^2 = \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \frac{|e_{i,S}^* v|^2}{\pi_i} \\ &= \max_{1\le i\le n} \frac{\delta_{i,S}}{\pi_i}.\end{aligned}$$

To conclude the proof it suffices to apply Theorem 2.5. □

4.2.2 Isolated measurements when the degree of sparsity is structured by levels

In this part, we consider a partition of $\{1,\dots,n\}$ into levels $(\Omega_i)_{i=1,\dots,N} \subset \{1,\dots,n\}$ such that

$$\bigsqcup_{1\le i\le N} \Omega_i = \{1,\dots,n\} \qquad \text{and} \qquad |\Omega_i| = N_i.$$

We consider that x is S-sparse with $|S\cap\Omega_i| = s_i$ for $1 \le i \le N$ meaning that restricted to the level Ω_i, the signal $P_{\Omega_i}x$ is s_i-sparse. This setting is studied extensively in the recent papers [2, 89, 7]. Theorem 2.5 provides the following guarantees.

Corollary 2.12. *Let $S \subset \{1,\dots,n\}$ be a set of indices of cardinality s, such that $|S\cap\Omega_i| = s_i$ for $1 \le i \le N$. Suppose that $x \in \mathbb{C}^n$ is an S-sparse vector. Fix $\varepsilon \in (0,1)$. Suppose that the sampling matrix A is constructed as in (2.11). Set*

$$m \;\ge\; C\left(\max_{1\le k\le n} \frac{\sum_{\ell=1}^N s_\ell \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k}\right) \ln(s)\ln\left(\frac{n}{\varepsilon}\right), \tag{2.13}$$

$$m \;\ge\; C\left(\max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2\right) \ln(s)\ln\left(\frac{n}{\varepsilon}\right), \tag{2.14}$$

then x is the unique solution of (2.1) with probability at least $1-\varepsilon$.

The proof of Corollary 2.12 is given in Appendix 6.5.2. We show in Appendix 6.5.2 that a simple analysis leads to results that are nearly equivalent to those in [2]. It should be noted that the term $\frac{\|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k}$ is related to the notion of local coherence defined in [2]. There are however a few differences making our approach potentially more interesting in the case of isolated measurements:

- Our approach is based on i.i.d. sampling with an arbitrary drawing distribution. This leaves a lot of freedom for generating sampling patterns and optimizing the probability π in order to minimize the upper-bounds (2.13) and (2.14). In contrast, the results in [2] are based on uniform Bernoulli sampling over *fixed* levels. The dependency on the levels is not explicit and it therefore seems complicated to optimize them.

- We can deal with a fixed support S, which enlarges the possibilities for structured sparsity. It is also possible to consider random supports as explained in Proposition 2.9.

4.2.3 Isolated measurements for the Fourier-Haar transform

The bounds in Corollary 2.12 are rather cryptic. They have to be analyzed separately for each sampling strategy. To conclude the discussion on isolated measurements, we provide a practical example with the 1D Fourier-Haar system.

We set $A_0 = \mathcal{F}\phi^*$, where $\mathcal{F} \in \mathbb{C}^{n\times n}$ is the 1D Fourier transform and $\phi^* \in \mathbb{C}^{n\times n}$ is the 1D inverse wavelet transform. To simplify the notation, we assume that $n = 2^J$ and we decompose the signal at the maximum level $J = \ln_2(n) - 1$. In order to state our result, we introduce a dyadic partition $(\Omega_j)_{0\leq j\leq J}$ of the set $\{1,\dots,n\}$. We set $\Omega_0 = \{1\}$, $\Omega_1 = \{2\}$, $\Omega_3 = \{3,4\},\dots,\Omega_J = \{n/2+1,\dots,n\}$. We also define the function $j : \{1,\dots,n\} \to \{0,\dots,J\}$ by $j(u) = j$ if $u \in \Omega_j$.

Corollary 2.13. *Let $S \subset \{1,\dots,n\}$ be a set of indices of cardinality s, such that $|S\cap\Omega_j| = s_j$ for $0 \leq j \leq J$. Suppose that $x \in \mathbb{C}^n$ is an s-sparse vector supported on S. Fix $\varepsilon \in (0,1)$. Suppose that A is constructed from the Fourier-Haar transform A_0. Choose π_k to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$. If*

$$m \;\geq\; C\cdot \max_{0\leq j\leq J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p \cdot \ln(s)\ln\left(\frac{n}{\varepsilon}\right), \tag{2.15}$$

then x is the unique solution of (2.1) with probability at least $1-\varepsilon$.

In particular, the distribution minimizing (2.15) is

$$\tilde{\pi}_j = \frac{2^{-j}\sum_{p=0}^{J} 2^{-|j-p|/2} s_p}{\sum_{\ell=1}^{n} 2^{-j(\ell)} \sum_{p=0}^{J} 2^{-|j(\ell)-p|/2} s_p},$$

which leads to

$$m \;\geq\; C\cdot \sum_{j=0}^{J} \left(s_j + \sum_{\substack{p=0\\ p\neq j}}^{J} 2^{-|j-p|/2} s_p \right) \cdot \ln(s)\ln\left(\frac{n}{\varepsilon}\right). \tag{2.16}$$

The proof is presented in Section 6.5.3. This corollary is once again similar to the results in [4]. The number of measurements in each level j should depend on the degree of sparsity s_j but also on the degree of sparsity of the other levels which is more and more attenuated when the level is far away from the j-th one.

Remark 2.14. *The Fourier-Wavelet system is coherent and the initial compressed sensing theories cannot explain the success of sampling strategies with such a transform. To overcome the coherence, two strategies have been devised. The first one is based on variable density sampling (see e.g. [83, 28, 63]). The second one is based on variable density sampling and an additional structured sparsity assumption (see e.g. [2] and Corollary 2.13). First, note that the results obtained with the latter approach allow recovering signal with arbitrary supports. Indeed, $\sum_{j=0}^{J} s_j + \sum_{\substack{p=0\\ p\neq j}}^{J} 2^{-|j-p|/2} s_p \leq 2s$.*

Second, it is not clear yet - from a theoretical point of view - that the structure assumption allows obtaining better guarantees. Indeed, it is possible to show that the sole variable

density sampling leads to perfect reconstruction from $m \propto s\ln(n)^2$ measurements, which is on par with bound (2.16). It will become clear that structured sparsity is essential when using the Fourier-Wavelet systems with structured acquisition. Morever, the numerical experiments led in [4] let no doubt about the fact that structured sparsity is essential to ensure good reconstruction with a low number of measurements.

4.3 Structured acquisition and structured sparsity

In this paragraph, we illustrate how Theorem 2.5 explains the practical success of structured acquisition in applications. We will mainly focus on the 2D setting: the vector $x \in \mathbb{C}^n$ to reconstruct can be seen as an image of size $\sqrt{n} \times \sqrt{n}$.

4.3.1 The limits of structured acquisition

In [11, 80], the authors provided theoretical CS results when using block-constrained acquisitions. Moreover, the results in [11] are proved to be tight in many practical situations. Unfortunately, the bounds on the number of blocks of measurements necessary for perfect reconstruction are however incompatible with a faster acquisition.

To illustrate this fact, let us recall a typical result emanating from [11]. It shows that the recovery of sparse vectors with an arbitrary support is of little interest when sampling lines of tensor product transforms. This setting is widely used in imaging. It corresponds to the MRI sampling strategy proposed in [69].

Proposition 2.15 ([11])**.** *Suppose that $A_0 = \phi \otimes \phi \in \mathbb{C}^{n\times n}$ is a 2D separable transform, where $\phi \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$ is an orthogonal transform. Consider blocks of measurements made of $\sqrt{n}$ horizontal lines in the 2D acquisition space, i.e. for $1 \leq k \leq \sqrt{n}$*

$$B_k = \left(\phi_{k,1}\phi, \ldots, \phi_{k,\sqrt{n}}\phi\right).$$

If the number of acquired lines m is less than $\min(2s, \sqrt{n})$, then there exists no decoder Δ such that $\Delta(Ax) = x$ for all s-sparse vector $x \in \mathbb{C}^n$.

In other words, the minimal number m of distinct blocks required to identify every s-sparse vectors is necessarily larger than $\min(2s, \sqrt{n})$.

This theoretical bound is quite surprising: it seems to enter in contradiction with the practical results obtained in Figure 2.2 or with one of the most standard CS strategy in MRI [69]. Indeed, the equivalent number of isolated measurements required by Proposition 2.15 is of the order $O(s\sqrt{n})$. This theoretical result means that in many applications, a full sampling strategy should be adopted, when the acquisition is structured by horizontal lines. In the next paragraphs, we show how Theorem 2.5 allows bridging the gap between theoretical recovery and practical experiments.

4.3.2 Breaking the limits with adapted structured sparsity

In this paragraph, we illustrate - through a simple example - that additional assumptions on structured sparsity is the key to explain practical results.

Corollary 2.16. *Let $A_0 \in \mathbb{C}^{n\times n}$ be the 2D Fourier transform. Assume that x is a 2D signal with support S concentrated on q horizontal lines of the spatial plane, i.e.*

$$S \subset \{(j-1)\sqrt{n} + \{1, \ldots, \sqrt{n}\}, j \in J\} \tag{2.17}$$

where $J \subset \{1, \dots, \sqrt{n}\}$ *and* $|J| = q$.

Choose a uniform sampling strategy among the $\sqrt{n}$ *horizontal lines, i.e.* $\pi_k^\star = 1/\sqrt{n}$ *for* $1 \leq k \leq \sqrt{n}$. *The number* m *of sampled horizontal lines sufficient to reconstruct* x *with probability* $1 - \varepsilon$ *is*

$$m \geq C \cdot q \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

The proof is given in Appendix 6.5.4 By Proposition 2.16, we can observe that the required number of sampled lines is of the order of non-zero lines in the 2D signal. In comparison, Proposition 4.6 in [11] (with no structured sparsity) requires

$$m \gtrsim s \cdot \ln(n/\varepsilon),$$

measurements, to get the same guarantees. This means that the required number of horizontal lines to sample is of the order of the non-zero coefficients. By putting aside the logarithmic factors, we see that the gain with our new approach is considerable. Clearly, our strategy is able to take advantage of the sparsity structure of the signal of interest.

4.3.3 Consequences for MRI sampling

We now turn to a real MRI application. We assume that the sensing matrix $A_0 \in \mathbb{C}^{n \times n}$ is the product of the 2D Fourier transform $\mathcal{F}_{2D}$ with the inverse 2D wavelet transform Φ^*. We aim at reconstructing a vector $x \in \mathbb{C}^n$ that can be seen as a 2D wavelet transform with $\sqrt{n} \times \sqrt{n}$ coefficients. Set $J = \ln_2(\sqrt{n}) - 1$ and let $(\tau_j)_{0 \leq j \leq J}$ denote a dyadic partition of the set $\{1, \dots, \sqrt{n}\}$, i.e. $\tau_0 = \{1\}$, $\tau_1 = \{2\}$, $\tau_2 = \{3, 4\}, \dots, \tau_J = \{\sqrt{n}/2 + 1, \dots, \sqrt{n}\}$. Define $j : \{1, \dots, \sqrt{n}\} \to \{0, \dots, J\}$ by $j(u) = j$ if $u \in \tau_j$. Finally, define the sets $\Omega_{\ell,\ell'} = \tau_\ell \times \tau_{\ell'}$, for $0 \leq \ell, \ell' \leq J$. See Figure 2.3 for an illustration of these sets.

Definition 2.17. *Given* $S = supp(x)$, *define the following quantity*

$$s_\ell^c := \max_{0 \leq l' \leq J} \max_{k \in \tau_{\ell'}} |S \cap \Omega_{\ell,\ell'} \cap C_k|, \tag{2.18}$$

where C_k *represents the set corresponding to the* k*-th vertical line (see Figure 2.3).*

The quantity s_ℓ^c represents the maximal sparsity of x restricted to columns (or vertical lines) of $\cup_{1 \leq l' \leq J} \Omega_{\ell,\ell'}$. We have now settled everything to state our result.

As a first step, we will consider the case of Shannon's wavelets, leading to a block-diagonal full sampling matrix A_0.

Corollary 2.18. *Let* $S \subset \{1, \dots, n\}$ *be a set of indices of cardinality* s, *such that* $|S \cap \Omega_{\ell,\ell'}| = s_{\ell,\ell'}$ *for* $0 \leq \ell, \ell' \leq J$. *Suppose that* $x \in \mathbb{C}^n$ *is an* s*-sparse vector supported on* S. *Fix* $\varepsilon \in (0, 1)$. *Suppose that* A_0 *is the product of the 2D Fourier transform with the 2D inverse Shannon's wavelets transform. Consider that the blocks of measurements are the* $\sqrt{n}$ *horizontal lines in the 2D setting. Choose* $(\pi_k)_{1 \leq k \leq \sqrt{n}}$ *to be constant by level, i.e.* $\pi_k = \tilde{\pi}_{j(k)}$. *If the number of horizontal lines to acquire verify*

$$m \gtrsim \max_{0 \leq j \leq J} \frac{1}{\tilde{\pi}_j} 2^{-j} s_j^c \ln(s) \ln\left(\frac{n}{\varepsilon}\right),$$

then x *is the unique solution of Problem 2.1. Furthermore, choosing* $\tilde{\pi}_j = \frac{s_j^c / 2^j}{\sum_{\ell=0}^J s_\ell^c}$, *for* $0 \leq j \leq J$, *leads to the following upper bound*

$$m \gtrsim \sum_{j=0}^{J} s_j^c \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

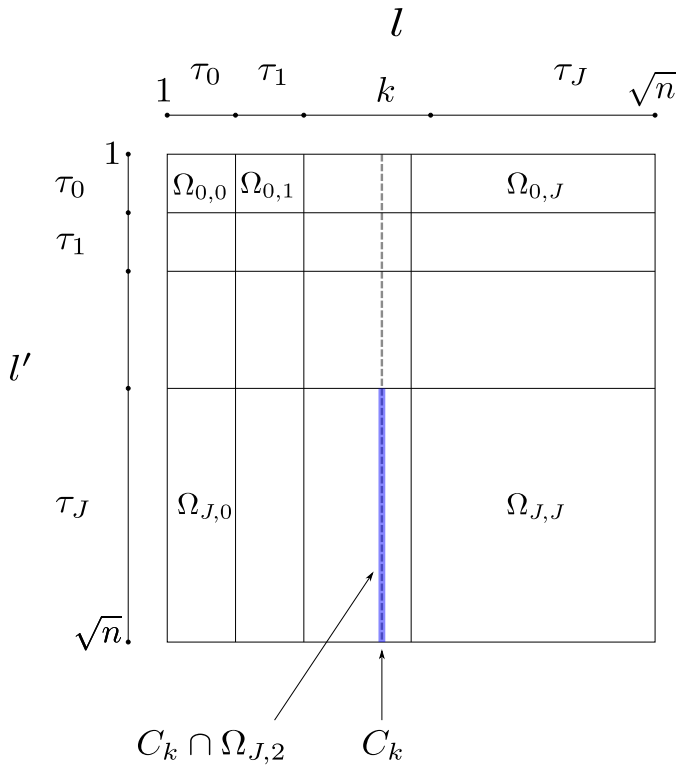


Figure 2.3: 2D view of the signal $x \in \mathbb{C}^n$ to reconstruct. The vector x can be reshaped in a $\sqrt{n} \times \sqrt{n}$ matrix. C_k represents the coefficient indexes corresponding to the k-th vertical column.

The proof is given in Section 6.5.5. Corollary 2.18 shows that the number of lines acquired at level j depends only on an extra-column structure of S restricted to the diagonal set of components indexed by $\Omega_{j,j}$. This is due to the block-diagonality feature of the transform. Now let us present a result when the matrix A_0 is not block-diagonal anymore.

Corollary 2.19. *Suppose that $x \in \mathbb{C}^n$ is an S-sparse vector. Fix $\varepsilon \in (0,1)$. Suppose that A_0 is the product of the 2D Fourier transform with the 2D inverse Haar transform. Consider that the blocks of measurements are the $\sqrt{n}$ horizontal lines. Choose $(\pi_k)_{1 \le k \le \sqrt{n}}$ to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$.*

If the number m of drawn horizontal lines satisfies

$$m \gtrsim \max_{0 \le j \le J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \ln(s) \ln\left(\frac{n}{\varepsilon}\right),$$

then x is the unique solution of Problem 2.1 with probablity $1-\varepsilon$.

In particular, if

$$\pi_k = \frac{2^{-j(k)} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c}{\sum_{\ell=1}^{\sqrt{n}} 2^{-j(\ell)} \sum_{r=0}^{J} 2^{-|j(\ell)-r|/2} s_r^c},$$

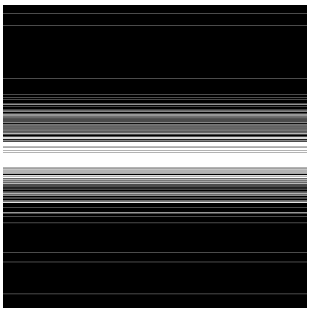
then

$$m \gtrsim \sum_{j=0}^{J} \left(s_j^c + \sum_{\substack{r=0 \\ r \neq j}}^{J} 2^{-|j-r|/2} s_r^c \right) \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right)$$

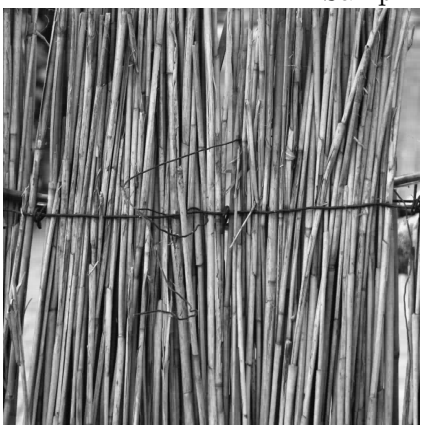
ensures perfect reconstruction with probability $1 - \varepsilon$.

The proof of Corollary 2.19 is given in Section 6.5.6.

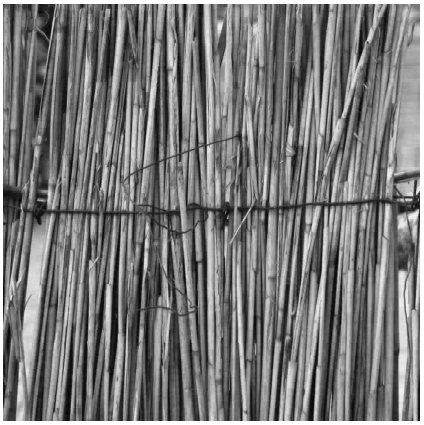
This result indicates that the number of acquired lines in the "horizontal" level j should be chosen depending on the quantities s_j^c. Note that this is very different from the sparsity by levels proposed in [2]. In conclusion, Corollary 2.19 reveals that with a structured acquisition, the sparsity needs to be more structured in order to guarantee exact recovery. To the best of our knowledge, this is the first theoretical result which can explain why sampling lines in MRI as in [69] might work. We illustrate the results in Corollary 2.19 in practical reconstruction of a reeds image, see Figure 2.4.



Sampling scheme

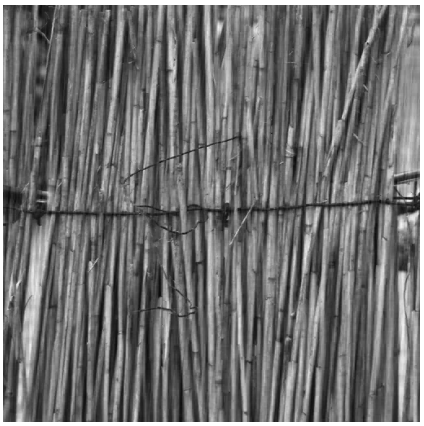


(a) Original image
$s^c = (16, 16, 32, 59, 81, 75, 48)$



(b) SNR = 27.8 dB

(c) Original image
$s^c = (16, 16, 32, 64, 124, 240, 411)$



(d) SNR = 14.7 dB

Figure 2.4: An example of reconstruction of a 2048×2048 real image from MR sensing. In (a) (c), Reference images of to reconstruct: the image in (c) is the same image as (a) but rotated of 90°. We precise the value of the vector $s^c = \left(s_j^c\right)_{1 \leq j \leq 7}$ for both images. Note that the quantities s_j^c are larger in the case of image (b). For the reconstruction, we use the sampling scheme at the top of the Figure. It corresponds to 9.8 % of measurements. In (b) (d), corresponding reconstruction via ℓ_1-minimization. We have rotated the image in (d) to facilitate the comparison between both.

5 Extensions

5.1 The case of Bernoulli block sampling

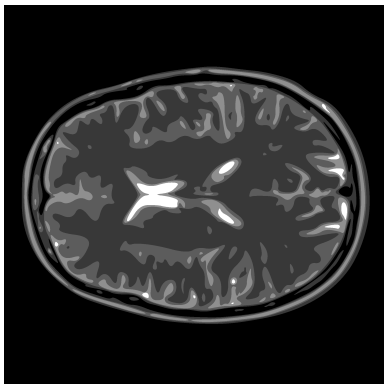
Combining CS strategies with structured acquisition and structured sparsity has been derived for the case of i.i.d. random blocks B_J where J is defined in Section 2. However, these results can be extended to a Bernoulli sampling setting. In such a setting, the sensing matrix is constructed as follows

$$A = \left(\frac{\delta_k}{\sqrt{\pi_k}} B_k \right)_{1 \leq k \leq M},$$

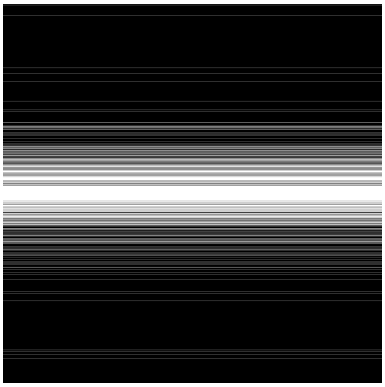
where $(\delta_k)_{1 \leq k \leq M}$ are independent Bernoulli random variables such that $\mathbb{P}(\delta_k = 1) = \pi_k$, for $1 \leq k \leq M$. We may set $\sum_{k=1}^{M} \pi_k = m$ in order to measure m blocks of measurements in expectation. By considering the same definition for $\Gamma(S, \pi)$ with $(\pi_k)_{1 \leq k \leq M}$ the Bernoulli weights, it is possible, for the case of Bernoulli block sampling, to give a reconstruction result that is of the same flavor than Theorem 2.5.

5.2 Adapting the sensing scheme to the structured sparsity

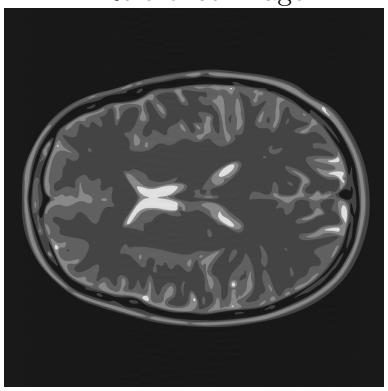
The results in Section 4.3.3 lead to the conclusion that, the more structure you have in the acquisition, the more structure you need in the signal to reconstruct in order to ensure exact recovery. Conversely, one may consider the reciprocal case where one searches for the appropriate structured acquisition with good reconstruction guarantees, given a structured sparsity. Indeed, in Figure 2.5, we aim at reconstructing a MR image considered sparse in the wavelet basis. This MR image has an intrinsic structured sparsity in the wavelet domain. One may search for the best structured sampling to use to ensure good reconstruction results. For instance, in Figure 2.5(a)(d), we propose two different sampling schemes that differently affect the quality of reconstruction. We believe that our main theorem could be used to predict the efficiency of a sampling scheme on the reconstruction of a signal with a given structured sparsity.



Reference image

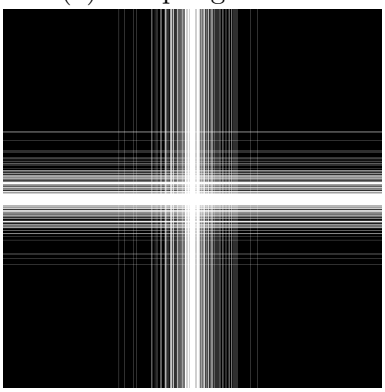


(a) Sampling scheme

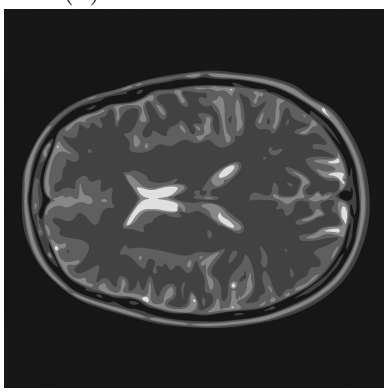


(b) SNR = 24.47 dB

(c)



(d) Sampling scheme



(e) SNR = 26.74 dB

(f)

Figure 2.5: An example of reconstruction of a 2048×2048 brain image from MR sensing. The reference image to reconstruct is presented at the top of the figure. It is considered sparse in the wavelet domain. In (a) (d), we present two kinds of sampling schemes with 20 % of measurements: the samples are acquired in the 2D Fourier domain. In (b) (e), we show the corresponding reconstruction via ℓ_1-minimization. In (c) (f) we enhance the results by zooming on the reconstructed images.

6 Proofs of the main results

6.1 Proof of Theorem 2.5

In this section, we give sufficient conditions to guarantee that the vector x is the unique minimizer of (2.1), using an inexact dual certificate see [20].

Lemma 2.20 (Inexact duality [20]). *Suppose that $x \in \mathbb{R}^n$ is supported on $S \subset \{1, \dots, n\}$. Assume that A_S is full column rank and that*

$$\| (A_S^* A_S)^{-1} \|_{2\to 2} \leq 2 \qquad \text{and} \qquad \max_{i \in S^c} \|A_S^* A e_i\|_2 \leq 1, \tag{2.19}$$

where $(A_S^* A_S)^{-1}$ only makes sense on the set span$\{e_i, i \in S\}$. Morever, suppose that there exists $v \in \mathbb{R}^n$ in the row space of A obeying

$$\|v_S - \operatorname{sign}(x_S)\|_2 \leq 1/4 \qquad \text{and} \qquad \|v_{S^c}\|_\infty \leq 1/4, \tag{2.20}$$

Then, the vector x is the unique solution of the minimization problem (2.1)

First, let us focus on Conditions (2.19). Remark that $A_S^* A_S$ is invertible by assuming that A_S is full column-rank. Moreover,

$$\| (A_S^* A_S)^{-1} \|_{2\to 2} = \left\| \sum_{k=0}^{\infty} (A_S^* A_S - P_S)^k \right\|_{2\to 2} \leq \sum_{k=0}^{\infty} \|A_S^* A_S - P_S\|_{2\to 2}^k .$$

Therefore, if $\|A_S^* A_S - P_S\|_{2\to 2} \leq 1/2$ is satisfied, then $\| (A_S^* A_S)^{-1} \|_{2\to 2} \leq 2$. Moreover, by Lemma 2.25, $\| (A_S^* A_S)^{-1} \|_{2\to 2} \leq 2$ with probability at least $1 - \varepsilon$, provided that

$$m \geq \frac{28}{3} \Theta(S, \pi) \ln\left(\frac{2s}{\varepsilon}\right).$$

By definition of $\Gamma(S, \pi)$, the first inequality of Conditions (2.19) is therefore ensured with probability larger than $1 - \varepsilon$ if

$$m \geq \frac{28}{3} \Gamma(S, \pi) \ln\left(\frac{2s}{\varepsilon}\right). \tag{2.21}$$

Furthermore, using Lemma 2.29, we obtain that

$$\max_{i \in S^c} \|A_S^* A e_i\|_2 \leq 1$$

with probability larger than $1 - \varepsilon$ if

$$m \geq \Theta(S, \pi) \left(1 + 4\sqrt{\ln\left(\frac{n}{\varepsilon}\right)} + 4 \ln\left(\frac{n}{\varepsilon}\right)\right).$$

Again by definition of $\Gamma(S, \pi)$, the second part of Conditions (2.20) is ensured if $n \geq 3$ and

$$m \geq 9 \Gamma(S, \pi) \ln\left(\frac{n}{\varepsilon}\right). \tag{2.22}$$

Conditions (2.20) remain to be verified. The rest of the proof of Theorem 2.5 relies on the construction of a vector v satisfying the conditions described in Lemma 2.20 with high probability. To do so, we adapt the so-called golfing scheme introduced by Gross [53] to our setting. More precisely, we will iteratively construct a vector that converges to a vector v satisfying (2.20) with high probability.

Let us first partition the sensing matrix A into blocks of blocks so that, from now on, we denote by $A^{(1)}$ the first m_1 blocks of A, $A^{(2)}$ the next m_2 blocks, and so on. The L random matrices $\left\{A^{(\ell)}\right\}_{\ell=1,\dots,L}$ are independently distributed, and we have that $m = m_1 + m_2 + \ldots + m_L$. As explained before, $A_S^{(\ell)}$ denotes the matrix $A^{(\ell)} P_S$.

The golfing scheme starts by defining $v^{(0)} = 0$, and then it iteratively defines

$$v^{(\ell)} = \frac{m}{m_\ell} A^{(\ell)*} A_S^{(\ell)} \left(\operatorname{sign}(x) - v^{(\ell-1)} \right) + v^{(\ell-1)}, \tag{2.23}$$

for $\ell = 1, \dots, L$, where $\operatorname{sign}(x_i) = 0$ if $x_i = 0$. In the rest of the proof, we set $v = v^{(L)}$. By construction, v is in the row space of A. The main idea of the golfing scheme is then to combine the results from the various Lemmas in Section 6.4 with an appropriate choice of L to show that the random vector v satisfies the assumptions of Lemma 2.20 with large probability. Using the shorthand notation $v_S^{(\ell)} = P_S v^{(\ell)}$, let us define

$$w^{(\ell)} = \operatorname{sign}(x) - v_S^{(\ell)}, \ \ell = 1, \dots, L,$$

where $x \in \mathbb{C}^n$ is the solution of Problem (2.1).

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

$$w^{(\ell)} = \left(P_S - \frac{m}{m_\ell} A_S^{(\ell)*} A_S^{(\ell)} \right) w^{(\ell-1)} = \prod_{j=1}^{\ell} \left(P_S - \frac{m}{m_j} A_S^{(j)*} A_S^{(j)} \right) \operatorname{sign}(x), \tag{2.24}$$

and

$$v = \sum_{\ell=1}^{L} \frac{m}{m_\ell} A^{(\ell)*} A_S^{(\ell)} w^{(\ell-1)}. \tag{2.25}$$

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

For all $1 \leq \ell \leq L$, assume that

$$\left\| w^{(\ell)} \right\|_2 \leq r_\ell \left\| w^{(\ell-1)} \right\|_2 \tag{C1-ℓ}$$

$$\left\| \frac{m}{m_\ell} \left(A_{S^c}^{(\ell)} \right)^* A_S^{(\ell)} w^{(\ell-1)} \right\|_\infty \leq t_\ell \| w^{(\ell-1)} \|_\infty \tag{C2-ℓ}$$

$$\left\| \left(\frac{m}{m_\ell} \left(A_S^{(\ell)} \right)^* A_S^{(\ell)} - P_S \right) w^{(\ell-1)} \right\|_\infty \leq t'_\ell \| w^{(\ell-1)} \|_\infty, \tag{C3-ℓ}$$

with

1. $L = 2 + \left\lceil \frac{\ln(s)}{2 \ln 2} \right\rceil$,
2. $r_\ell = \frac{1}{2}$, for $\ell = 1, \dots, L$,
3. $t_\ell = t'_\ell = \frac{1}{5}$ for $\ell = 1, \dots, L$.

Note that using (C1-ℓ), we can write that

$$\| \operatorname{sign}(x_S) - v_S \|_2 = \| w_S^{(L)} \|_2 \leq \| \operatorname{sign}(x_S) \|_2 \prod_{\ell=1}^{L} r_\ell \leq \sqrt{s} \prod_{\ell=1}^{L} r_\ell \leq \frac{\sqrt{s}}{2^L} \leq \frac{1}{4}, \tag{2.26}$$

where the last inequality follows from the previously specified choice on L.

Furthermore, Equation (C2-ℓ) implies that

$$\begin{aligned}\|v_{S^c}\|_\infty &= \left\|\sum_{\ell=1}^{L} \frac{m}{m_\ell}\left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)}\right\|_\infty \\ &\leq \sum_{\ell=1}^{L} \left\|\frac{m}{m_\ell}\left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)}\right\|_\infty \\ &\leq \sum_{\ell=1}^{L} t_\ell \left\|w^{(\ell-1)}\right\|_\infty \\ &\leq \sum_{\ell=1}^{L} t_\ell \prod_{j=1}^{\ell-1} t'_j \\ &= \left(\frac{1}{5}\right)\frac{1-(1/5)^L}{1-1/5} \leq \frac{1}{4}. \end{aligned} \tag{2.27}$$

Note that in Inequality (2.27), the control of the operator norms $\infty \to \infty$ avoids the apparition of $\sqrt{s}$ as in the usual golfing scheme of [20]. Indeed, in our proof strategy, we have used the fact that $\|w_0\|_\infty = \|\mathrm{sign}(x_S)\|_\infty = 1$, whereas in [20] $\|w_0\|_2 = \|\mathrm{sign}(x_S)\|_2 \leq \sqrt{s}$ is involved. This is a key step in the proof, since the absence of the degree of sparsity at this stage allows to derive results depending only on S and not on its cardinality $s = |S|$.

We denote by $p_1(\ell)$, $p_2(\ell)$ and $p_3(\ell)$ the probabilities that the upper bounds (C1-ℓ), (C2-ℓ) and (C3-ℓ) do not hold.

Let us call "failure C" the event in which one of the $3L$ inequalities (C1-ℓ), (C2-ℓ), (C3-ℓ) is not satisfied. Then,

$$\mathbb{P}\left(\text{failure C}\right) \leq \sum_{\ell=1}^{L} \mathbb{P}\left(\text{failure (C1-}\ell\text{)}\right) + \mathbb{P}\left(\text{failure (C2-}\ell\text{)}\right) + \mathbb{P}\left(\text{failure (C3-}\ell\text{)}\right).$$

Therefore a sufficient condition for $\mathbb{P}\left(\text{failure C}\right) \leq \varepsilon$ is $\sum_{\ell=1}^{L} p_1(\ell)+p_2(\ell)+p_3(\ell) \leq \varepsilon$ which holds provided that $p_1(\ell) \leq \varepsilon/3L$, $p_2(\ell) \leq \varepsilon/3L$ and $p_3(\ell) \leq \varepsilon/3L$ for every $\ell = 1, \ldots, L$. By Lemma 2.26, condition $p_1(\ell) \leq \varepsilon/3L$ is satisfied if

$$m_\ell \geq 32\Gamma(S,\pi)\left(\ln\left(\frac{3L}{\varepsilon}\right) + \frac{1}{4}\right).$$

By Lemma 2.27, condition $p_2(\ell) \leq \varepsilon/3L$ is satisfied if

$$m_\ell \geq 101\Gamma(S,\pi)\ln\left(\frac{12nL}{\varepsilon}\right).$$

By Lemma 2.28, condition $p_3(\ell) \leq \varepsilon/3L$ is satisfied if

$$m_\ell \geq 101\Gamma(S,\pi)\ln\left(\frac{12nL}{\varepsilon}\right).$$

Overall, condition

$$m_\ell \geq 101\Gamma(S,\pi)\ln\left(\frac{12nL}{\varepsilon}\right) \tag{2.28}$$

ensures that (2.26) and (2.27) are satisfied with probability $1-\varepsilon$. Condition

$$m = \sum_{\ell=1}^{L} m_\ell \geq 101\left(\frac{\ln(s)}{2\ln(2)} + 3\right)\Gamma(S,\pi)\ln\left(12nL\varepsilon^{-1}\right)$$

will imply (2.28). The latter condition can be simplified into

$$m \geq 73 \cdot \Gamma(S, \pi) \ln(64s) \left(\ln\left(\frac{9n}{\varepsilon}\right) + \ln\ln(55s) \right). \tag{2.29}$$

The latter condition ensures that the random vector v, defined by (2.25), satisfies Assumptions 2.20 of Lemma 2.20 with probability larger than $1 - \varepsilon$.

Hence, we have thus shown that if conditions (2.21), (2.22) and (2.29) are satisfied, then the Assumptions 2.19 and 2.20 of Lemma 2.20 simultaneously hold with probability larger than $1 - 3\varepsilon$. Note that bound (2.29) implies (2.21) and (2.22).

6.2 Proof of Theorem 2.8

To prove Theorem 2.8, one can show the following lemma.

Lemma 2.21 (Inexact duality [20, 48]). *Suppose that $x \in \mathbb{R}^n$ is supported on $S \subset \{1, \ldots, n\}$. Assume that A_S is full column rank and that*

$$\| (A_S^* A_S)^{-1} \|_{2 \to 2} \leq 2 \qquad \text{and} \qquad \max_{i \in S^c} \|A_S^* A e_i\|_2 \leq 1, \tag{2.30}$$

where $(A_S^* A_S)^{-1}$ only makes sense on the set span$\{e_i, i \in S\}$. Morever, suppose that there exists $v \in \mathbb{R}^n$ in the row space of A, i.e. $v = A^* h$, obeying

$$\|v_S - \operatorname{sign}(x_S)\|_2 \leq 1/4 \qquad \text{and} \qquad \|v_{S^c}\|_\infty \leq 1/4, \tag{2.31}$$

and

$$\|h\|_2 \leq \sqrt{2}\sqrt{s}. \tag{2.32}$$

Then, a minimizer $x^\sharp$ of $\|z\|_1$ subject to $\|Az - y\| \leq \eta$ satisfies

$$\|x - x^\sharp\|_2 \leq c_1 \sigma_s(x)_1 + (c_2 + c_3\sqrt{s})\eta,$$

for $c_1 = 24$, $c_2 = 16\sqrt{3/2}$, $c_3 = 24\sqrt{2}$ and $\sigma_s(x)_1$ is the error of the best s-term approximation given by $\sigma_s(x)_1 := \min_{\|x'\|_0 \leq s} \|x' - x\|_1$.

To prove assumptions (2.30) and (2.31), we can reuse what has been done in the previous section for the proof of Theorem 2.5. It remains to verify the last condition (2.32). By using the same notation as in the proof of Theorem 2.5, we can define

$$h^{(\ell)} = \frac{m}{m_\ell} A_S^{(\ell)} w^{(\ell-1)} \qquad \text{for} \quad \ell = 1, \ldots, L,$$

and then $h = (h^{(1)*}, h^{(2)*}, \ldots, h^{(L)*}, 0, \ldots, 0)^*$. We aim at controlling

$$\|h\|_2^2 = \sum_{\ell=1}^{L} \|h^{(\ell)}\|_2^2 = \sum_{\ell=1}^{L} \left\| \frac{m}{m_\ell} A_S^{(\ell)} w^{(\ell-1)} \right\|_2^2.$$

Note that

$$\begin{aligned}
\left\| \frac{m}{m_\ell} A_S^{(\ell)} w^{(\ell-1)} \right\|_2^2 &= \left\langle \frac{m}{m_\ell} \left(A_S^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)}, w^{(\ell-1)} \right\rangle \\
&= \left\langle \left(\frac{m}{m_\ell} \left(A_S^{(\ell)}\right)^* A_S^{(\ell)} - P_S \right) w^{(\ell-1)}, w^{(\ell-1)} + \|w^{(\ell-1)}\|_2^2 \right\rangle \\
&= \left\langle w^{(\ell)}, w^{(\ell-1)} \right\rangle + \|w^{(\ell-1)}\|_2^2 \\
&\leq \|w^{(\ell)}\|_2 \|w^{(\ell-1)}\|_2 + \|w^{(\ell-1)}\|_2^2,
\end{aligned}$$

using Cauchy-Schwarz inequality. Using (C1-ℓ) with $r_\ell = 1/2$ for every $\ell = 1, \dots, L$, we can conclude that

$$\left\| \frac{m}{m_\ell} A_S^{(\ell)} w^{(\ell-1)} \right\|_2^2 \leq \frac{3}{2} \|w^{(\ell-1)}\|_2^2 \leq \frac{3}{2} \prod_{j=1}^{\ell-1} \left(\frac{1}{2}\right)^2 \|w^{(0)}\|_2^2 \leq \frac{3}{2} \left(\frac{1}{4}\right)^{\ell-1} s.$$

Then, we obtain that

$$\|h\|_2^2 \leq \sum_{\ell=1}^{L} \frac{3}{2} \left(\frac{1}{4}\right)^{\ell-1} s \leq 2s,$$

which corresponds to Assumption (2.32) with a constant of $\sqrt{2}$.

Therefore, if $m \geq 73 \cdot \Gamma(S, \pi) \ln(64s) \left(\ln\left(\frac{27n}{\varepsilon}\right) + \ln\ln(55s) \right)$, then Assumptions (2.30), (2.31) and (2.32) holds with probability larger than $1 - \varepsilon$.

6.3 Bernstein's inequalities

Theorem 2.22 (Scalar Bernstein Inequality)**.** *Let $x_1, \dots, x_m$ be independent real-valued, zero-mean, 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 2.23 (Vector Bernstein Inequality (V1))**.** *[20, Theorem 2.6] Let $(y_k)_{1 \leq k \leq m}$ be a finite sequence of independent random complex vectors of dimension n. Suppose that $\mathbb{E}y_k = 0$ and $\|y_k\|_2 \leq K$ a.s. for some constant $K > 0$ and set $\sigma^2 \geq \sum_k \mathbb{E}\|y_k\|_2^2$. Let $Z = \|\sum_{k=1}^m y_k\|_2$. Then, for any $0 < t \leq \sigma^2/K$, we have that*

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

Theorem 2.24 (Bernstein Inequality for self-adjoint matrices)**.** *Let $(Z_k)_{1 \leq k \leq n}$ be a finite sequence of independent, random, self-adjoint matrices of dimension d, and let a_k be a sequence of fixed self-adjoint matrices. Suppose that Z_k is such that $\mathbb{E}Z_k = 0$ and $\|Z_k\|_{2\to 2} \leq K$ a.s. for some constant $K > 0$ that is independent of k. Moreover, assume that $\mathbb{E}Z_k^2 \preceq A_k^2$ for each $1 \leq k \leq n$. Define*

$$\sigma^2 = \left\| \sum_{k=1}^{n} A_k^2 \right\|_{2\to 2}$$

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

$$\mathbb{P}\left(\left\| \sum_{k=1}^{n} Z_k \right\|_{2\to 2} \geq t \right) \leq d \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3} \right).$$

Proof. This result is as an application of the techniques developed in [97] to obtain tail bounds for sum of random matrices. Our arguments follow those in the proof of Theorem 6.1 in [97]. We assume that $K = 1$ since the general result follows by a scaling argument. Using the assumption that $\mathbb{E}Z_k^2 \preceq A_k^2$, and by applying the arguments in the proof of Lemma 6.7 in [97], we obtain that

$$\mathbb{E}\exp(\theta Z_k) \preceq \exp\left(g(\theta)A_k^2\right),$$

for any real $\theta > 0$, where $g(\theta) = e^\theta - \theta - 1$, and the notation $\exp(A)$ denotes the matrix exponential of a self-adjoint matrice A (see [97] for further details). Therefore, by Corollary 3.7 in [97], it follows that

$$\mathbb{P}\left(\left\|\sum_{k=1}^n Z_k\right\|_{2\to 2} \geq t\right) \leq d \inf_{\theta>0}\left\{e^{-\theta t+\sigma^2 g(\theta)}\right\}, \tag{2.33}$$

where $\sigma^2 = \|\sum_{k=1}^n A_k^2\|_{2\to 2}$. To conclude, we follow the proof of Theorem 6.1 in [97]. The function $\theta \mapsto -\theta t + \sigma^2 g(\theta)$ attains its minimum for $\theta = \ln(1 + t/\sigma^2)$, which implies that the minimal value of the right-hand size of Inequality (2.33) is $d\exp\left(-\sigma^2 h(t/\sigma^2)\right)$ where $h(u) = (1+u)\ln(1+u) - u$ for $u \geq 0$. To complete the proof, it suffices to use the standard lower bound $h(u) \geq \frac{u^2/2}{1+u/3}$ for $u \geq 0$. □

6.4 Estimates: auxiliary results

Let S be the support of the signal to be reconstructed such that $|S| = s$. We set

$$\Lambda(S,\pi) := \max_{1\leq k\leq M} \frac{1}{\pi_k}\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2}.$$

Note that $\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2} \leq \left\|B_{k,S}^* B_{k,S}\right\|_{\infty\to\infty} \leq \|B_k^* B_{k,S}\|_{\infty\to\infty}$, therefore,

$$\Lambda(S,\pi) \leq \Theta(S,\pi).$$

To make the notation less cluttered, we will write Λ, Θ, Υ and Γ instead of $\Lambda(S,\pi)$, $\Theta(S,\pi)$, $\Upsilon(S,\pi)$ and $\Gamma(S,\pi)$.

Lemma 2.25. *Let $S \subset \{1,\dots,n\}$ be of cardinality of s. Suppose that $\Theta \geq 1$. Then, for any $\delta > 0$, one has that*

$$\mathbb{P}\left(\|A_S^* A_S - P_S\|_{2\to 2} \geq \delta\right) \leq 2s\exp\left(-\frac{m\delta^2/2}{\Theta(1+\delta/3)}\right). \tag{E1}$$

Proof. We decompose the matrix $A_S^* A_S - P_S$ as

$$A_S^* A_S - P_S = \frac{1}{m}\sum_{k=1}^m \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S = \frac{1}{m}\sum_{k=1}^m X_k,$$

where $X_k := \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right)$. It is clear that $\mathbb{E}X_k = 0$, and since for all $1 \leq k \leq M$, $\frac{\left\|B_{k,S}B_{k,S}^*\right\|_{2\to 2}}{\pi_k} \leq \Lambda \leq \Theta$, we have that

$$\|X_k\|_{2\to 2} \leq \max\left(\frac{\left\|B_{J_k,S}^* B_{J_k,S}\right\|_{2\to 2}}{\pi_{J_k}} - 1, 1\right) \leq \Theta.$$

Lastly, we remark that

$$\begin{aligned} 0 \quad \preceq \quad \mathbb{E}X_k^2 = \mathbb{E}\left[\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}}\right]^2 - P_S \quad &\preceq \quad \max_{1\le k\le M} \frac{\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2}}{\pi_k} \mathbb{E}\left[\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}}\right] \\ &\preceq \max_{1\le k\le M} \frac{\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2}}{\pi_k} P_S \preceq \Lambda P_S \\ &\preceq \Theta P_S. \end{aligned}$$

Therefore, using Theorem 2.24, we can set $\sigma^2 = \|\sum_{k=1}^m \mathbb{E}X_k^2\|_{2\to 2} \le m\Theta$. Hence, inequality (E1) immediately follows from Bernstein's inequality for random matrices (see Theorem 2.24). □

Lemma 2.26. *Let $S \subset \{1,\dots,n\}$, such that $|S| = s$. Let w be a vector in $\mathbb{C}^n$. Then, for any $0 \le t \le 1$, one has that*

$$\mathbb{P}\left(\left\|\left(A_S^* A_S - P_S\right) w\right\|_2 \ge t\|w\|_2\right) \le \exp\left(-\frac{mt^2}{8\Theta} + \frac{1}{4}\right). \tag{E2}$$

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

$$\left(A_S^* A_S - \mathrm{Id}_s\right) w_S = \frac{1}{m}\sum_{k=1}^m \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) w = \frac{1}{m}\sum_{k=1}^m y_k,$$

where $y_k = \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) w$ is a random vector with zero mean. Simple calculations yield that

$$\begin{aligned} \left\|\frac{1}{m} y_k\right\|_2^2 &= \frac{1}{m^2}\left(w^*\left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}}\right)^2 w - 2w^* \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w + w^* w\right) \\ &\le \frac{1}{m^2}\left(\Lambda w^* \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w - 2w^* \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w + 1\right) \\ &= \frac{1}{m^2}\left((\Lambda - 2)\, w^* \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w + 1\right) \\ &\le \frac{1}{m^2}\left((\Lambda - 2)\,\Lambda \|w\|_2^2 + 1\right) = \frac{1}{m^2}\left((\Lambda - 2)\,\Lambda + 1\right) \\ &\le \frac{1}{m^2}(\Lambda - 1)^2 \le \frac{1}{m^2}\Lambda^2 \le \frac{1}{m^2}\Theta^2. \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}\left[Z^2\right] &= \frac{1}{m}\mathbb{E}\left\|y_1\right\|_2^2 = \frac{1}{m}\mathbb{E}\left[\left\langle \frac{B_{J,S}^* B_{J,S}}{\pi_J} w, \frac{B_{J,S}^* B_{J,S}}{\pi_J} w\right\rangle - 2\left\langle \frac{B_{J,S}^* B_{J,S}}{\pi_J} w, w\right\rangle + \langle w, w\rangle\right] \\ &= \frac{1}{m}\mathbb{E}\left[\left\langle \left(\frac{B_{J,S}^* B_{J,S}}{\pi_J}\right)^2 w, w\right\rangle - 2\frac{\|B_{J,S} w\|_2^2}{\pi_J} + 1\right]. \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}\left[\left\langle \left(\frac{B_{J_1,S}^* B_{J_1,S}}{\pi_{J_1}}\right)^2 w, w\right\rangle\right] &= \left\langle \mathbb{E}\left[\left(\frac{B_{J_1,S}^* B_{J_1,S}}{\pi_{J_1}}\right)^2\right] w, w\right\rangle \\ &\le \Lambda\left\langle \mathbb{E}\left[\left(\frac{B_{J_1,S}^* B_{J_1,S}}{\pi_{J_1}}\right)\right] w, w\right\rangle \le \Lambda\|w\|_2^2 \le \Theta. \end{aligned}$$

One immediately has that $\mathbb{E}\frac{\left\|B_{J,S}w\right\|_2^2}{\pi_k} = \|w\|_2^2 = 1$. Therefore, one finally obtains that

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

Using the above upper bounds, namely $\left\|\frac{1}{m}y_k\right\|_2 \leq \frac{\Theta}{m}$ and $\mathbb{E}\left[Z^2\right] \leq \frac{\Theta}{m}$, the result of the lemma is thus a consequence of the Bernstein's inequality for random vectors (see Theorem 2.23), which completes the proof. □

Lemma 2.27. *Let $S \subset \{1, \ldots, n\}$, such that $|S| = s$. Let v be a vector of $\mathbb{C}^n$. Then we have*

$$\mathbb{P}\left(\|A_{S^c}^* A_S v\|_\infty \geq t\|v\|_\infty\right) \leq 4n \exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3}\right). \tag{E3}$$

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

$$\|A_{S^c}^* A_S v\|_\infty = \max_{i \in S^c} |\langle e_i, A^* A_S v\rangle| = \max_{i \in S^c} \frac{1}{m}\left|\sum_{k=1}^m \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right|.$$

Let us define $Z_k = \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle$. Note that $\mathbb{E}Z_k = 0$, since for $i \in S^c$, $\mathbb{E}\left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle = e_i^* \sum_{k=1}^M \pi_k \frac{B_k^* B_{k,S}}{\pi_k} v = e_i^* P_S v = 0$. From Holder's inequality, we get

$$\begin{aligned}
|Z_k| &= \left|\left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right| = \left|e_i^* \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right| \leq \max_{\substack{j \in S^c \\ 1 \leq k \leq M}} \frac{1}{\pi_k} \left\|B_{k,S}^* B_k e_j\right\|_1 \|v\|_\infty \\
&\leq \max_{1 \leq k \leq n} \max_{\substack{j \in S^c \\ 1 \leq k \leq M}} \frac{1}{\pi_k} \left\|e_j^* B_k^* B_{k,S}\right\|_1 = \Theta.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}|Z_k|^2 = \mathbb{E}\left|\left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right|^2 &= \sum_{\ell=1}^M \frac{|e_i^* B_\ell^* B_{\ell,S} v|^2}{\pi_\ell} \\
&\leq \Upsilon.
\end{aligned}$$

Therefore $\sum_{k=1}^m \mathbb{E}|Z_k|^2 \leq m\Upsilon$. Using real-valued Bernstein's inequality 2.22 in the case of complex random variables, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right| \geq t\right) \\
&\leq \mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \mathrm{Re}\left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right| \geq t/\sqrt{2}\right) \ldots \\
&\qquad + \mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \mathrm{Im}\left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right| \geq t/\sqrt{2}\right) \\
&\leq 4 \exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3}\right).
\end{aligned}$$

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

Lemma 2.28. *Let $S \subset \{1,\dots,n\}$, such that $|S| = s$. Suppose that $\Theta \geq 1$.Let v be a vector of $\mathbb{C}^n$. Then we have*

$$\mathbb{P}\left(\left\|\left(A_S^*A_S - P_S\right)v\right\|_\infty \geq t\|v\|_\infty\right) \leq 4s\exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3}\right). \tag{E4}$$

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

$$\left\|\left(A_S^*A_S - P_S\right)v\right\|_\infty = \max_{i\in S}\left|\left\langle e_i, \left(A_S^*A_S - P_S\right)v\right\rangle\right| = \max_{i\in S}\frac{1}{m}\left|\sum_{k=1}^m\left\langle e_i, \left(\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right)v\right\rangle\right|.$$

Let us define $Z_k = \left\langle e_i, \left(\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right)v\right\rangle$. Note that $\mathbb{E}Z_k = 0$. From Holder's inequality, we get

$$|Z_k| = \left|\left\langle e_i, \left(\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right)v\right\rangle\right| \leq \left\|\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right\|_{\infty\to\infty} \leq \max(\Theta - 1, 1) \leq \Theta,$$

since $\|B_{k,S}^*B_{k,S}\|_{\infty\to\infty} \leq \|B_k^*B_{k,S}\|_{\infty\to\infty}$, and using the same argument as in Lemma 2.27. Furthermore,

$$\begin{aligned}
\mathbb{E}|Z_k|^2 &= \mathbb{E}\left|\left\langle e_i, \left(\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right)v\right\rangle\right|^2 \\
&= \mathbb{E}\left|\left\langle e_i, \frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}}v\right\rangle\right|^2 - \langle e_i, v\rangle\,\mathbb{E}\left\langle e_i, \frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}}v\right\rangle - \langle e_i, v\rangle^*\,\mathbb{E}\left\langle e_i, \frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}}v\right\rangle \\
&\quad + |\langle e_i, v\rangle|^2 \\
&= \mathbb{E}\left|\left\langle e_i, \frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}}v\right\rangle\right|^2 - |\langle e_i, v\rangle|^2 \leq \mathbb{E}\left|\left\langle e_i, \frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}}v\right\rangle\right|^2 = \sum_{\ell=1}^M\frac{\left|e_i^*B_{\ell,S}^*B_{\ell,S}v\right|^2}{\pi_\ell} \\
&\leq \Upsilon.
\end{aligned}$$

Therefore, $\sum_{k=1}^m \mathbb{E}|Z_k|^2 \leq m\Upsilon$, and using real-valued Bernstein's inequality 2.22 in the case of complex random variables, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m\left\langle e_i, \left(\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right)v\right\rangle\right| \geq t\right) \\
&\leq \mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \mathrm{Re}\left\langle e_i, \left(\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right)v\right\rangle\right| \geq t/\sqrt{2}\right) \\
&+ \mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \mathrm{Im}\left\langle e_i, \left(\frac{B_{J_k,S}^*B_{J_k,S}}{\pi_{J_k}} - P_S\right)v\right\rangle\right| \geq t/\sqrt{2}\right) \\
&\leq 4\exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3}\right).
\end{aligned}$$

Taking the union bound over $i \in S$ completes the proof. □

Lemma 2.29. *Let S be a subset of $\{1,\dots,n\}$. Then, for any $0 \leq t \leq m$, one has that*

$$\mathbb{P}\left(\max_{i\in S^c}\|A_S^*Ae_i\|_2 \geq t\right) \leq n\exp\left(-\frac{\left(\sqrt{m}t/\sqrt{\Theta} - 1\right)^2}{4}\right). \tag{E5}$$

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

$$x_k := \frac{B^*_{J_k,S} B_{J_k}}{\pi_{J_k}} e_i.$$

Then, since $i \in S^c$ one easily gets $\mathbb{E}x_k = \sum_{\ell=1}^M B^*_{\ell,S} B_\ell e_i = \sum_{\ell=1}^M (B_\ell P_S)^* B_\ell e_i = P_S \sum_{\ell=1}^M B^*_\ell B_\ell e_i = P_S e_i = 0$ (note that P_S is self-adjoint). In addition, we can write

$$\|A^*_S A e_i\|_2 = \left\| \frac{1}{m} \sum_{k=1}^m \frac{B^*_{J_k,S} B_{J_k}}{\pi_{J_k}} e_i \right\|_2 = \left\| \frac{1}{m} \sum_{k=1}^M x_k \right\|_2.$$

Then,

$$\begin{aligned} \|x_k\|_2 = \left\| \frac{B^*_{J_k,S} B_{J_k}}{\pi_{J_k}} e_i \right\|_2 &\le \left\| \frac{B^*_{J_k,S} B_{J_k}}{\pi_{J_k}} e_i \right\|_1 = \left\| e_i^* \frac{B^*_{J_k} B_{J_k,S}}{\pi_{J_k}} \right\|_1 \\ &\le \frac{1}{\pi_{J_k}} \left\| B^*_{J_k} B_{J_k,S} \right\|_{\infty\to\infty} \le \Theta. \end{aligned}$$

Furthermore, one has that

$$\mathbb{E}\,\|x_k\|_2^2 = \mathbb{E} \left\| \frac{B^*_{J_k,S} B_{J_k}}{\pi_{J_k}} e_i \right\|_2^2 \le \mathbb{E} \left\| \frac{B_{J_k,S}}{\sqrt{\pi_{J_k}}} \right\|_{2\to 2}^2 \left\| \frac{B_{J_k}}{\sqrt{\pi_{J_k}}} e_i \right\|_2^2 \le \Lambda \mathbb{E} \left\| \frac{B_{J_k}}{\sqrt{\pi_{J_k}}} e_i \right\|_2^2 = \Lambda \|e_i\|_2^2 = \Lambda,$$

$$\sum_{k=1}^m \mathbb{E}\,\|x_k\|_2^2 \le m\Lambda \le m\Theta.$$

Hence, using the above upper bounds, it follows from Bernstein's inequality for random vectors (see Theorem 2.23) that

$$\mathbb{P}\left(\|A^*_S A e_i\|_2 \ge t\right) \le \exp\left(-\frac{\left(\sqrt{m}t/\sqrt{\Theta} - 1\right)^2}{4} \right),$$

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

6.5 Proof of results in Applications

6.5.1 Proof of Corollary 2.10

The proof relies on the evaluation of Θ and Υ in the case of isolated measurements. In this case, we have n blocks composed of isolated measurements. Then, each block corresponds to one of the rows $(a_k^*)_{1\le k\le n}$ of A_0. Recall that $\|a_k a^*_{k,S}\|_{\infty\to\infty} = \max_{1\le i\le n} \sup_{\|v\|_\infty \le 1} |e_i^* a_k a^*_{k,S} v|$, so the norm $\|a_k a^*_{k,S}\|_{\infty\to\infty}$ is the maximum ℓ_1-norm of the rows of the matrix $a_k a^*_{k,S}$.

Therefore, the quantities in Definition 2.3 can be rewritten as follows

$$\begin{aligned}\Theta(S,\pi) &:= \max_{1\le k\le n} \frac{\|a_k a_{k,S}^*\|_{\infty\to\infty}}{\pi_k} \\ &= \max_{1\le k\le n} \frac{\|a_k\|_\infty \|a_{k,S}\|_1}{\pi_k} \qquad (2.34)\\ &\le s\cdot \max_{1\le k\le n} \frac{\|a_k\|_\infty^2}{\pi_k},\end{aligned}$$

$$\begin{aligned}\Upsilon(S,\pi) &= \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 |a_{k,S}^* v|^2 \qquad (2.35)\\ &\le \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} \|a_k\|_\infty^2 |a_{k,S}^* v|^2 \\ &\le \sup_{\|v\|_\infty\le 1} \max_{1\le \ell\le n} \frac{\|a_\ell\|_\infty^2}{\pi_\ell} \sum_{k=1}^n |a_{k,S}^* v|^2 = \sup_{\|v\|_\infty\le 1} \|A_0 P_S v\|_2^2 \max_{1\le \ell\le n} \frac{\|a_\ell\|_\infty^2}{\pi_\ell} \\ &= \sup_{\|v\|_\infty\le 1} \|P_S v\|_2^2 \max_{1\le \ell\le n} \frac{\|a_\ell\|_\infty^2}{\pi_\ell} \\ &\le s\cdot \max_{1\le k\le n} \frac{\|a_k\|_\infty^2}{\pi_k}.\end{aligned}$$

Therefore we can choose $\Gamma(S,\pi) = s\cdot \max_{1\le k\le n} \frac{\|a_k\|_\infty^2}{\pi_k}$, and the result follows by Theorem 2.5.

6.5.2 Around Corollary 2.12

Proof of Corollary 2.12 Again, this is all about evaluating Θ and Υ in this specific case. Concerning the evaluation of Υ, we can use the expression (2.35) to conclude that

$$\Upsilon(S,\pi) = \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 |a_{k,S}^* v|^2.$$

To control Θ, using (2.34), it suffices to write:

$$\begin{aligned}\Theta(S,\pi) = \max_{1\le k\le n} \frac{\|a_k a_{k,S}^*\|_{\infty\to\infty}}{\pi_k} &\le \max_{1\le k\le n} \frac{\|a_k\|_\infty \|a_{k,S}\|_1}{\pi_k} \\ &\le \max_{1\le k\le n} \frac{\|a_k\|_\infty \sum_{\ell=1}^N \|a_{k,\Omega_\ell}\|_\infty s_\ell}{\pi_k}.\end{aligned}$$

By Theorem 2.5, the two conditions

$$\begin{aligned} m &\ge C\left(\max_{1\le k\le n} \frac{\sum_{\ell=1}^N s_\ell \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k}\right) \ln(s) \ln\left(\frac{n}{\varepsilon}\right), \\ m &\ge C\left(\max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2\right) \ln(s) \ln\left(\frac{n}{\varepsilon}\right),\end{aligned}$$

lead to the desired conclusion.

Comparison of Corollary 2.12 and the results in [2]. Note that the sampling in [2] is based on Bernoulli drawings structured by level. Their results are then easily transposable to the case of i.i.d. sampling with constant probability by level. The first condition on m in Corollary 2.12 is similar to condition (4.4) in Theorem 4.4 of [2], since we recognize the term $\frac{\|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k}$ as the (k,ℓ)-local coherence defined in [2]. Let us show that the second condition on m is similar to equation (4.5) in [2]. First, observe that

$$\begin{aligned}\max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2 &\le \max_{1\le \ell\le N} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty^2 \left|a_{k,S}^* v\right|^2 \\ &\le \max_{1\le \ell\le N} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty \left|a_{k,S}^* v\right|^2 .\end{aligned}$$

Let $\tilde{v}$ denote the maximizer in the last expression, and define $\widetilde{s_k} = \left|a_{k,S}^* \tilde{v}\right|^2$ for $1 \le k \le n$. It follows,

$$\max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2 \le \max_{1\le \ell\le N} \sum_{k=1}^n \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty \widetilde{s_k}, \tag{2.36}$$

and $\sum_{k=1}^n \widetilde{s_k} = \sum_{k=1}^n \left|a_{k,S}^* v\right|^2 = \|A_0 P_S \tilde{v}\|_2^2 = \|P_S \tilde{v}\|_2^2 \le \sum_{\ell=1}^N s_\ell$. The last inequality and Equation (2.36) for i.i.d sampling correspond to the condition (4.5) in Theorem 4.4 of [2] in the case of Bernoulli sampling. This completes the comparison between Corollary 2.12 and the results in [2].

6.5.3 Proof of Corollary 2.13

Recall that $(\Omega_j)_{0\le j\le J}$ the dyadic partition of the set of indexes $\{1,\dots,n\}$. Recall also the function $j : \{1,\dots,n\} \to \{0,\dots,J\}$ defined by $j(u) = j$ if $u \in \Omega_j$. In the interests of simplifying notation, in this section, the symbol '$\gtrsim$' will be equivalent to '$\ge C\cdot$', with C a universal constant. The following lemma will be useful to bound above the coefficients of A_0 in absolute value, and to derive Lemmas 2.31 and 2.32.

Lemma 2.30. *[3] The magnitude of the coefficients of matrix $A_0 = \mathcal{F}\phi^*$, where $\mathcal{F}$ is the 1D Fourier transform and ϕ is the 1D Haar transform, satisfies*

$$\|P_{\Omega_j} A_0 P_{\Omega_\ell}\|_{1\to\infty}^2 \lesssim 2^{-j} 2^{-|j-\ell|}, \quad \text{for} \quad 0 \le j,\ell \le J. \tag{2.37}$$

Lemma 2.31. *In the case of isolated measurements, with $A_0 = \mathcal{F}\phi^*$ with ϕ to be the inverse 1D Haar transform, suppose that the signal to reconstruct x is sparse by levels, meaning that $\|P_{\Omega_j} x\|_0 \le s_j$ for $0 \le j \le J$. Then,*

$$\Theta \lesssim \max_{1\le k\le n} \frac{2^{-j(k)}}{\pi_k} \left(s_{j(k)} + \sum_{\substack{\ell=0 \\ \ell\ne j(k)}}^J s_\ell 2^{-|j(k)-\ell|/2} \right). \tag{2.38}$$

Choosing π_k to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$, the last expression can be rewritten as follows

$$\Theta \lesssim \max_{0\le j\le J} \frac{2^{-j}}{\tilde{\pi}_j} \left(s_j + \sum_{\substack{\ell=0 \\ \ell\ne j}}^J s_\ell 2^{-|j-\ell|/2} \right). \tag{2.39}$$

Proof. Using (2.34), we can write

$$\begin{aligned}\Theta &= \max_{1\le k\le n} \frac{\|a_k\|_\infty \|a_{k,S}\|_1}{\pi_k} \le \max_{1\le k\le n} \frac{\|a_k\|_\infty \sum_{\ell=0}^J \|a_{k,\Omega_\ell}\|_\infty s_\ell}{\pi_k} \\ &\lesssim \max_{1\le k\le n} \frac{1}{\pi_k} 2^{-j(k)/2} \sum_{\ell=0}^J 2^{-j(k)/2} 2^{-|j(k)-\ell|/2} s_\ell \\ &\lesssim \max_{1\le k\le n} \frac{1}{\pi_k} 2^{-j(k)} \sum_{\ell=0}^J 2^{-|j(k)-\ell|/2} s_\ell,\end{aligned}$$

where we use (2.37) to bound above $\|a_{k,\Omega_\ell}\|_\infty$. □

Lemma 2.32. *In the case of isolated measurements, with $A_0 = \mathcal{F}\phi^*$ with ϕ to be the inverse Haar transform, suppose that the signal to reconstruct x is sparse by levels, meaning that $\|P_{\Omega_j} x\|_0 \le s_j$ for $0 \le j \le J$. Choosing π_k to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$, we have*

$$\Upsilon \lesssim \max_{0\le j\le J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^J 2^{-|j-p|/2} s_p. \tag{2.40}$$

Proof. Denoting $\tilde{v} = \tilde{v}(i)$ the argument of the supremum in the definition of Υ, we get

$$\begin{aligned}\Upsilon &:= \max_{1\le i\le n} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 |a_{k,S}\tilde{v}|^2 \le \max_{0\le \ell\le J} \sum_{k=1}^n \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty^2 |a_{k,S}\tilde{v}|^2 \\ &\lesssim \max_{0\le \ell\le J} \sum_{k=1}^n \frac{1}{\pi_k} 2^{-j(k)} 2^{-|j(k)-\ell|} |a_{k,S}\tilde{v}|^2 \lesssim \max_{0\le \ell\le J} \sum_{j=0}^J \frac{1}{\tilde{\pi}_j} 2^{-j} 2^{-|j-\ell|} \underbrace{\sum_{k\in\Omega_j} |a_{k,S}\tilde{v}|^2}_{=:K_j}\end{aligned}$$

We can rewrite K_j as follows $K_j = \|P_{\Omega_j} A_0 P_S \tilde{v}\|_2^2$. Therefore, since $\|\tilde{v}\|_\infty \le 1$,

$$\begin{aligned}\sqrt{K_j} &= \|P_{\Omega_j} A_0 P_S \tilde{v}\|_2 = \|P_{\Omega_j} A_0 \sum_{p=0}^J P_{\Omega_p} P_S \tilde{v}\|_2 \le \sum_{p=0}^J \|P_{\Omega_j} A_0 P_{\Omega_p} P_S \tilde{v}\|_2 \\ &\le \sum_{p=0}^J \|P_{\Omega_j} A_0 P_{\Omega_p}\|_{2\to 2} \|P_{\Omega_p} P_S \tilde{v}\|_2 \le \sum_{p=0}^J \|P_{\Omega_j} A_0 P_{\Omega_p}\|_{2\to 2} \sqrt{s_p}.\end{aligned}$$

Using Lemma 4.3 of [3], we have the following upper bound

$$\|P_{\Omega_j} A_0 P_{\Omega_p}\|_{2\to 2} \lesssim 2^{-|j-p|/2}, \qquad \text{for} \quad 0 \le j, p \le J.$$

Then, $\sqrt{K_j} \lesssim \sum_{p=0}^J 2^{-|j-p|/2} \sqrt{s_p}$, and thus

$$\begin{aligned}K_j &\lesssim \left(\sum_{p=0}^J 2^{-|j-p|/2} \sqrt{s_p} \right)^2 \lesssim \left(\sum_{p=0}^J 2^{-|j-p|/2} \right) \left(\sum_{p=0}^J 2^{-|j-p|/2} s_p \right) \\ &\lesssim \left(\sum_{p=0}^J 2^{-|j-p|/2} s_p \right)\end{aligned}$$

where in the second inequality we use Cauchy-Schwarz inequality. Therefore,

$$\begin{aligned}\Upsilon &\lesssim \max_{0\le\ell\le J}\sum_{j=0}^{J} 2^{-|j-\ell|}\frac{1}{\tilde{\pi}_j}2^{-j}\sum_{p=0}^{J}2^{-|j-p|/2}s_p \\ &\lesssim \left(\max_{0\le\ell\le J}\sum_{j=0}^{J}2^{-|j-\ell|}\right)\left(\max_{0\le j\le J}\frac{1}{\tilde{\pi}_j}2^{-j}\sum_{p=0}^{J}2^{-|j-p|/2}s_p\right) \\ &\lesssim \max_{0\le j\le J}\frac{1}{\tilde{\pi}_j}2^{-j}\sum_{p=0}^{J}2^{-|j-p|/2}s_p.\end{aligned}$$

□

Note that the upper bounds given in Lemmas 2.31 and 2.32 coincide. Therefore, we can apply Theorem 2.5 with the following upper bound for $\Gamma(S,\pi)$

$$\Gamma(S,\pi) \lesssim \max_{0\le j\le J}\frac{1}{\tilde{\pi}_j}2^{-j}\sum_{p=0}^{J}2^{-|j-p|/2}s_p,$$

and conclude the proof for Corollary 2.13.

6.5.4 Proof of Corollary 2.16

Recall that $A_0 = \phi\otimes\phi \in \mathbb{C}^{n\times n}$, where $\phi \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$ is a 1D orthogonal transform. Consider a blocks dictionary made of $\sqrt{n}$ horizontal lines, i.e. for $1 \le k \le \sqrt{n}$

$$B_k = \left(\phi_{k,1}\phi,\dots,\phi_{k,\sqrt{n}}\phi\right), \quad \text{and thus} \quad B_k^*B_k = \left(\phi_{k,i}^*\phi_{k,j}\mathrm{Id}_{\sqrt{n}}\right)_{1\le i,j\le\sqrt{n}}.$$

Now, let us fix that the signal support S is concentrated on q horizontal lines of the spatial plane. Formally,

$$S \subset \{(j-1)\sqrt{n} + \{1,\dots,\sqrt{n}\}, j\in J\} \tag{2.41}$$

where $J \subset \{1,\dots,\sqrt{n}\}$ and $|J| = q$. Therefore,

$$B_k^*B_{k,S} = \left(\delta_{j\in J}\phi_{k,i}^*\phi_{k,j}\mathrm{Id}_{\sqrt{n}}\right)_{1\le i,j\le\sqrt{n}},$$

where $\delta_{j\in J} = 1$ if $j\in J$ and 0 otherwise. In such a setting, the quantities in Definition 2.3 can be rewritten as follows:

$$\Theta(S,\pi) = \max_{1\le k\le M}\max_{1\le i\le n}\frac{\|e_i^*B_k^*B_{k,S}\|_1}{\pi_k} = \max_{1\le k\le\sqrt{n}}\max_{1\le\tilde{i}\le\sqrt{n}}\frac{|\phi_{k,\tilde{i}}|\sum_{j\in J}|\phi_{k,j}|}{\pi_k} \le \max_{1\le k\le\sqrt{n}} q\frac{\|\phi_{k,:}\|_\infty^2}{\pi_k}. \tag{2.42}$$

Recall that

$$\Upsilon(S,\pi) := \max_{1\le i\le n}\sup_{\|v\|_\infty\le 1}\sum_{k=1}^{M}\frac{1}{\pi_k}\left|e_i^*B_k^*B_{k,S}v\right|^2,$$

and call $(i^\star, v)$ the argument of the supremum over $\{1,\dots,n\}$ and $\{u, \|u\|_\infty \le 1\}$. Therefore,

$$\Upsilon(S,\pi) = \sum_{k=1}^{M}\frac{1}{\pi_k}\left|e_{i^\star}^*B_k^*B_{k,S}v\right|^2.$$

We can decompose $i^\star = (i_1 - 1)\sqrt{n} + i_2$ with i_1, i_2 integers of $\{1, \dots, \sqrt{n}\}$. We can write

$$\Upsilon(S, \pi) = \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left| \sum_{j=1}^{\sqrt{n}} \delta_{j \in J} \phi^*_{k,i_1} \phi_{k,j} e^*_{i_2} v[j] \right|^2 = \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} |\phi_{k,i_1}|^2 \left| \sum_{j=1}^{\sqrt{n}} \delta_{j \in J} \phi_{k,j} w_j \right|^2,$$

where $w \in \mathbb{C}^{\sqrt{n}}$ such that $w_j = e^*_{i_2} v[j]$ and $v[j] \in \mathbb{C}^{\sqrt{n}}$ is the restriction of v to the j-th horizontal line, i.e. to the components of v indexed by $\{(j-1)\sqrt{n}+1, \dots, j\sqrt{n}\}$. We can rewrite the last expression as follows

$$\begin{aligned}
\Upsilon(S, \pi) &= \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} |\phi_{k,i_1}|^2 \left| \langle e_k, \phi P_J w \rangle \right|^2 \leq \max_{1 \leq \ell \leq \sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \sum_{k=1}^{\sqrt{n}} \left| \langle e_k, \phi P_J w \rangle \right|^2 \\
&= \max_{1 \leq \ell \leq \sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \|\phi P_J w\|_2^2 = \max_{1 \leq \ell \leq \sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \|P_J w\|_2^2 \\
&\leq \max_{1 \leq \ell \leq \sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \cdot q,
\end{aligned}$$

where in the last expression we use that $\|w\|_\infty \leq 1$. Choosing ϕ as the 1D Fourier transform gives $\|\phi_{\ell,:}\|_\infty = \frac{1}{n^{1/4}}$ and choosing a uniform sampling among the $\sqrt{n}$ horizontal lines, i.e. $\pi^\star_\ell = 1/\sqrt{n}$ for $1 \leq \ell \leq \sqrt{n}$, leads to

$$\Gamma(S, \pi^\star) \leq q,$$

which ends the proof of Corollary 2.16.

6.5.5 Proof of Corollary 2.18

We recall that the sampling matrix is then constructed from the full sampling matrix $A_0 \in \mathbb{C}^{n \times n}$, in the 2D setting, where $A_0 = \mathcal{F}_{2D} \Psi^*$ with $\mathcal{F}_{2D} \in \mathbb{C}^{n \times n}$ the 2D Fourier transform and $\Psi^* \in \mathbb{C}^{n \times n}$ the 2D inverse wavelet transform. Since both transforms are separable, $\mathcal{F}_{2D} = \mathcal{F} \otimes \mathcal{F}$, $\Psi = \psi \otimes \psi$, with $\otimes$ the Kronecker product and $\mathcal{F}, \psi \in \mathbb{C}^{\sqrt{n} \times \sqrt{n}}$ the corresponding 1D transforms. Then A_0 can also be rewritten as $A_0 = \phi \otimes \phi$, the Kronecker product of the 1D transforms $\phi := \mathcal{F}\psi^* \in \mathbb{C}^{\sqrt{n} \times \sqrt{n}}$.

In this section, in order to avoid any confusion, we will denote by $\left(e_i^{(n)}\right)_{1 \leq i \leq n}$ the canonical basis in dimension n.

In Corollary 2.18, we focus on the case where $A_0 = \phi \otimes \phi \in \mathbb{C}^{n \times n}$ is the 2D Fourier-Shannon wavelet transform, then $\phi \in \mathbb{C}^{\sqrt{n} \times \sqrt{n}}$ is the 1D Fourier-Shannon wavelets transform. Therefore, ϕ and A_0 are block-diagonal orthogonal matrices. The sensing schemes are based on horizontal lines on the 2D plane, meaning that

$$B_k = \left(\phi_{k,1} \phi \dots \phi_{k,\sqrt{n}} \phi \right),$$

for $k = 1, \dots, \sqrt{n}$. By defintion of the Fourier-Shannon transform, we have that

$$B_k^* B_k = \left(\phi^*_{k,\ell} \phi_{k,m} \mathrm{Id}_{\sqrt{n}} \right)_{1 \leq \ell, m \leq \sqrt{n}} = \frac{1}{2^{j(k)}} \left(\delta_{\ell \in \tau_{j(k)}} \delta_{m \in \tau_{j(k)}} \mathrm{Id}_{\sqrt{n}} \right)_{1 \leq \ell, m \leq \sqrt{n}},$$

for $k = 1, \dots, \sqrt{n}$, where $\delta_{\ell \in \tau_j} = 1$ if $\ell \in \tau_j$, and 0 otherwise.

First let us start with the evaluation of Θ. By definition of $\| \cdot \|_{\infty \to \infty}$, we have

$$\|B_k^* B_{k,S}\|_{\infty \to \infty} = \max_{1 \leq \ell \leq n} \sup_{\substack{\|v\|_\infty \leq 1 \\ v \in \mathbb{C}^n}} \left| \left(e_\ell^{(n)}\right)^* B_k^* B_k P_S v \right|.$$

Setting $\tilde{v} = \tilde{v}(k)$ the argument of the supremum in the last expression, then

$$\Theta := \max_{1\le k\le\sqrt{n}} \max_{1\le\ell\le n} \frac{1}{\pi_k} \left| \left(e_\ell^{(n)}\right)^* B_k^* B_k P_S \tilde{v} \right|,$$

Note that $\|\tilde{v}\|_\infty \le 1$. The index ℓ can be rewritten as $\ell = (\ell_1 - 1)\sqrt{n} + \ell_2$, with $1 \le \ell_1, \ell_2 \le \sqrt{n}$.

$$\begin{aligned}\Theta &:= \max_{1\le k\le\sqrt{n}} \max_{1\le\ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \left(\phi_{k,m} \left(e_{\ell_2}^{(\sqrt{n})}\right)^* \right)_{1\le m\le\sqrt{n}} P_S \tilde{v} \right|, \\ &= \max_{1\le k\le\sqrt{n}} \max_{1\le\ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \sum_{m=1}^{\sqrt{n}} \phi_{k,m} \left(e_{\ell_2}^{(\sqrt{n})}\right)^* (P_S \tilde{v})[m] \right|,\end{aligned}$$

where $(v)[m] \in \mathbb{C}^{\sqrt{n}}$ is the restriction of the vector v to the m-th horizontal line, i.e. to the components indexed by $\{(m-1)\sqrt{n}+1, \dots, m\sqrt{n}\}$. Set $w^{|(m)} := (P_S\tilde{v})[m] \in \mathbb{C}^{\sqrt{n}}$, the restriction of $P_S\tilde{v}$ to the m-th horizontal line. Then the ℓ_2-th component of $w^{|(m)}$, written as $w_{\ell_2}^{|(m)}$ is equal to $\left(e_{\ell_2}^{(\sqrt{n})}\right)^* (P_S\tilde{v})[m]$. Note that $\left|w_{\ell_2}^{|(m)}\right| \le 1$ if $(m-1)\sqrt{n}+\ell_2 \in S$, and it is equal to 0 otherwise. Then,

$$\Theta \le \max_{1\le k\le\sqrt{n}} \max_{1\le\ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \sum_{m=1}^{\sqrt{n}} \phi_{k,m} w_{\ell_2}^{|(m)} \right|. \tag{2.43}$$

By the properties of block-diagonality of the Fourier-Shannon transform, we have

$$\begin{aligned}\Theta &\le \max_{1\le k\le\sqrt{n}} \max_{1\le\ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \sum_{m\in\tau_{j(k)}} \phi_{k,m} w_{\ell_2}^{|(m)} \right| \qquad (2.44) \\ &\le \max_{1\le k\le\sqrt{n}} \max_{1\le\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \|\phi_{k,:}\|_\infty^2 \left| \sum_{m\in\tau_{j(k)}} w_{\ell_2}^{|(m)} \right| \\ &\le \max_{1\le k\le\sqrt{n}} \max_{1\le\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \|\phi_{k,:}\|_\infty^2 \sum_{m\in\tau_{j(k)}} \left| w_{\ell_2}^{|(m)} \right| \\ &\lesssim \max_{1\le k\le\sqrt{n}} \frac{1}{\pi_k} \frac{1}{2^{j(k)}} s_{j(k)}^c. \qquad (2.45)\end{aligned}$$

Indeed, $\sum_{m\in\tau_{j(k)}} \left|w_{\ell_2}^{|(m)}\right|$ is bounded above by $\sum_{m\in\tau_{j(k)}} \delta_{(m-1)\sqrt{n}+\ell_2\in S}$, which counts the number of intersections between S, the ℓ_2th-column and the $j(k)$ (horizontal) level, see the blue line in Figure 2.3. Taking the maximum over $1 \le \ell_2 \le \sqrt{n}$ leads to $\sum_{m\in\tau_{j(k)}} \delta_{(m-1)\sqrt{n}+\ell_2\in S} \le s_{j(k)}^c$.

Secondly, let us evaluate Υ. We have that

$$\Upsilon := \max_{1\le\ell\le n} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left| \left(e_\ell^{(n)}\right)^* B_k^* B_k \tilde{v} \right|^2,$$

where $\tilde{v} = \tilde{v}(\ell)$ is the argument of the supremum on the ℓ_∞ unit-ball. Using (2.44), with $\ell = (\ell_1 - 1)\sqrt{n} + \ell_2$, we can rewrite

$$\Upsilon = \max_{1\le\ell_1,\ell_2\le\sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \sum_{m=1}^{\sqrt{n}} \phi_{k,m} w_{\ell_2}^{|(m)} \right|^2,$$

where $w_{\ell_2}^{|(m)} := \left(e_{\ell_2}^{(\sqrt{n})}\right)^* (P_S \tilde{v})\,[m]$. Note again that $\left|w_{\ell_2}^{|(m)}\right| \leq 1$ if $(m-1)\sqrt{n}+\ell_2 \in S$, and it is equal to 0 otherwise. By denoting $w^{|(:,\ell_2)}$ the vector with components

$$w^{|(:,\ell_2)} := \left(w_{\ell_2}^{|(1)}, \quad w_{\ell_2}^{|(2)}, \ldots, w_{\ell_2}^{|(\sqrt{n})}\right)^*, \tag{2.46}$$

we can rewrite the previous quantity as follows

$$\begin{aligned}\Upsilon &= \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left|\phi_{k,\ell_1}^* \left\langle \phi_{k,:}^*, w^{|(:,\ell_2)} \right\rangle\right|^2 \\ &= \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} |\phi_{k,\ell_1}|^2 \left|\left\langle \phi_{k,:}^*, w^{|(:,\ell_2)} \right\rangle\right|^2.\end{aligned} \tag{2.47}$$

Since ϕ is an orthogonal block-diagonal transform, we have

$$\Upsilon = \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \sum_{k\in\tau_{j(\ell_1)}} \frac{1}{\pi_k} |\phi_{k,\ell_1}|^2 \left|\left\langle \phi_{k,:}^*, w^{|(:,\ell_2)} \right\rangle\right|^2.$$

Choosing $\pi_k = \tilde{\pi}_j$ for $k \in \tau_j$ meaning that the probability of drawing lines is constant by levels, we can write that

$$\begin{aligned}\Upsilon &= \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \frac{1}{\tilde{\pi}_{j(\ell_1)}} \sum_{k\in\tau_{j(\ell_1)}} |\phi_{k,\ell_1}|^2 \left|\left\langle \phi_{k,:}^*, w^{|(:,\ell_2)} \right\rangle\right|^2, \\ &\leq \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \frac{1}{\tilde{\pi}_{j(\ell_1)}} \sum_{k\in\tau_{j(\ell_1)}} \|\phi_{k,:}\|_\infty^2 \left|\left\langle \phi_{k,:}^*, w^{|(:,\ell_2)} \right\rangle\right|^2, \\ &\lesssim \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} \sum_{k\in\tau_{j(\ell_1)}} \left|\left\langle \phi_{k,:}^*, w^{|(:,\ell_2)} \right\rangle\right|^2, \\ &= \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} \left\|P_{\tau_{j(\ell_1)}} \phi w^{|(:,\ell_2)}\right\|_2^2.\end{aligned}$$

Since ϕ is orthogonal and block diagonal we have $\left\|P_{\tau_{j(\ell_1)}} \phi w^{|(:,\ell_2)}\right\|_2^2 = \|P_{\tau_{j(\ell_1)}} w^{|(:,\ell_2)}\|_2^2$. Then,

$$\begin{aligned}\Upsilon &\lesssim \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} \left\|P_{\tau_{j(\ell_1)}} w^{|(:,\ell_2)}\right\|_2^2, \\ &\lesssim \max_{1\leq \ell_1 \leq \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} s_{j(\ell_1)}^c,\end{aligned} \tag{2.48}$$

where the last step invokes that $\|P_{\tau_{j(\ell_1)}} w^{|(:,\ell_2)}\|_2^2 \leq \sum_{m\in\tau_{j(\ell_1)}} \delta_{(m-1)\sqrt{n}+\ell_2 \in S} \leq s_{j(\ell_1)}^c$. Note that the upper bounds (2.45) and (2.48) on Υ and Θ coincide. They lead to the following choice for $1 \leq k \leq \sqrt{n}$,

$$\pi_k = \tilde{\pi}_{j(k)} = \frac{s_{j(k)}^c 2^{-j(k)}}{\sum_{\ell=1}^{\sqrt{n}} s_{j(\ell)}^c 2^{-j(\ell)}} = \frac{s_{j(k)}^c 2^{-j(k)}}{\sum_{j=0}^{J} \sum_{\ell\in\tau_j} s_j^c 2^{-j}} = \frac{s_{j(k)}^c 2^{-j(k)}}{\sum_{j=0}^{J} s_j^c}.$$

Then for this particular choice, we can rewrite

$$\max(\Theta, \Upsilon) \lesssim \sum_{j=0}^{J} s_j^c.$$

To conclude, by Theorem 2.5, a lower bound on the required number of horizontal lines to acquire is thus

$$m \gtrsim \sum_{j=0}^{J} s_j^c \ln(s) \ln(n/\varepsilon).$$

6.5.6 Proof of Corollary 2.19

In this part, using the formalism introduced in the last section, ψ is the 1D Haar transform, and ϕ is then the Fourier-Haar's wavelet transform. In such a case, we can reuse (2.43) in Section 6.5.5 to evaluate Θ:

$$\Theta \leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1} \sum_{m=1}^{\sqrt{n}} \phi_{k,m} w_m[\ell_2] \right|.$$

Using Lemma 2.30, we have for $1 \leq k, m \leq \sqrt{n}$,

$$|\phi_{k,m}| \lesssim 2^{-j(k)/2} 2^{-|j(k)-j(m)|/2}.$$

Therefore,

$$\begin{aligned}
\Theta &\leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right| \sum_{m=1}^{\sqrt{n}} \left|\phi_{k,m} w^{|(m)}_{\ell_2}\right| \\
&\leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right| \sum_{j=0}^{J} \sum_{m\in\tau_j} |\phi_{k,m}| \left|w^{|(m)}_{\ell_2}\right| \\
&\leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right| \sum_{j=0}^{J} \sum_{m\in\tau_j} |\phi_{k,m}| \left|w^{|(m)}_{\ell_2}\right| \\
&\lesssim \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_2\leq\sqrt{n}} \frac{1}{\pi_k} 2^{-j(k)} \sum_{j=0}^{J} 2^{-|j(k)-j|/2} \sum_{m\in\tau_j} \left|w^{|(m)}_{\ell_2}\right| \\
&\lesssim \max_{1\leq k\leq\sqrt{n}} \frac{1}{\pi_k} 2^{-j(k)} \sum_{j=0}^{J} 2^{-|j(k)-j|/2} s_j^c. \qquad (2.49)
\end{aligned}$$

Now let us study Υ. Recall the definition of $w^{|(:,\ell_2)}$ depending on ℓ_2 in (2.46), we can reuse (2.47) to have

$$\begin{aligned}
\Upsilon &= \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1} \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2 \\
&= \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right|^2 \left|\left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2, \\
&= \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \sum_{j=0}^{J} \frac{1}{\tilde{\pi}_j} \sum_{k\in\tau_j} \left|\phi^*_{k,\ell_1}\right|^2 \left|\left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2,
\end{aligned}$$

by choosing $\pi_k = \tilde{\pi}_j$ for $k \in \tau_j$, meaning that the drawing probability is constant by level. Since for $k \in \tau_j$, we have $\left|\phi^*_{k,\ell_1}\right|^2 \leq 2^{-j} 2^{-|j-j(\ell_1)|}$ by Lemma 2.30. Then,

$$\Upsilon = \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \sum_{j=0}^{J} \frac{1}{\tilde{\pi}_j} 2^{-j} 2^{-|j-j(\ell_1)|} \underbrace{\sum_{k\in\tau_j} \left|\left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2}_{=:K_j}.$$

Dealing with K_j, we can derive that

$$\sqrt{K_j} = \left\|P_{\tau_j}\phi^* w^{|(:,\ell_2)}\right\|_2 = \left\|P_{\tau_j}\phi^* \sum_{r=0}^{J} P_{\tau_r} w^{|(:,\ell_2)}\right\|_2 \leq \sum_{r=0}^{J} \left\|P_{\tau_j}\phi^* P_{\tau_r}\right\|_{2\to 2} \left\|P_{\tau_r} w^{|(:,\ell_2)}\right\|_2$$
$$\lesssim \sum_{r=0}^{J} 2^{-|j-r|/2}\sqrt{s_r^c},$$

where the upper bound $\left\|P_{\tau_j}\phi^* P_{\tau_r}\right\|_{2\to 2} \lesssim 2^{-|j-r|/2}$ can be found in [3, Lemma 4.3]. Then,

$$K_k \lesssim \left(\sum_{r=0}^{J} 2^{-|j-r|/2}\sqrt{s_r^c}\right)^2 \lesssim \left(\sum_{r=0}^{J} 2^{-|j-r|/2}\right)\left(\sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c\right)$$
$$\lesssim \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c.$$

Therefore,

$$\Upsilon \lesssim \max_{1\leq \ell_1 \leq \sqrt{n}} \sum_{j=0}^{J} \frac{1}{\tilde{\pi}_j} 2^{-j} 2^{-|j-j(\ell_1)|} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c$$
$$\lesssim \left(\max_{1\leq \ell_1 \leq \sqrt{n}} \sum_{j=0}^{J} 2^{-|j-j(\ell_1)|}\right)\left(\max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c\right)$$
$$\lesssim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c. \tag{2.50}$$

The upper bounds (2.49) and (2.50) give

$$\max(\Theta, \Upsilon) \lesssim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c.$$

Therefore, by Theorem 2.5, a lower bound on the required number of horizontal lines is

$$m \gtrsim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \ln(n/\varepsilon)\ln(s).$$

By choosing

$$\pi_k = \tilde{\pi}_{j(k)} = \frac{2^{-j(k)} \sum_{r=0}^{J} 2^{-|j(k)-r|/2} s_r^c}{\sum_{\ell=1}^{\sqrt{n}} 2^{-j(\ell)} \sum_{r=0}^{J} 2^{-|j(\ell)-r|/2} s_r^c},$$

for $1 \leq k \leq \sqrt{n}$, the lower bound on the required number of horizontal lines can be rewritten as

$$m \gtrsim \sum_{\ell=1}^{\sqrt{n}} 2^{-j(\ell)} \sum_{r=0}^{J} 2^{-|j(\ell)-r|/2} s_r^c \cdot \ln(n/\varepsilon)\ln(s)$$
$$\gtrsim \sum_{j=0}^{J} \sum_{\ell\in\tau_j} 2^{-j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \cdot \ln(n/\varepsilon)\ln(s)$$
$$\gtrsim \sum_{j=0}^{J} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \cdot \ln(n/\varepsilon)\ln(s)$$
$$\gtrsim \sum_{j=0}^{J} \left(s_j^c + \sum_{\substack{r=0 \\ r\neq j}}^{J} 2^{-|j-r|/2} s_r^c\right) \cdot \ln(n/\varepsilon)\ln(s),$$

which concludes the proof of Corollary 2.19.

Part II

Numerical analysis: on the generation of block-constrained sampling schemes

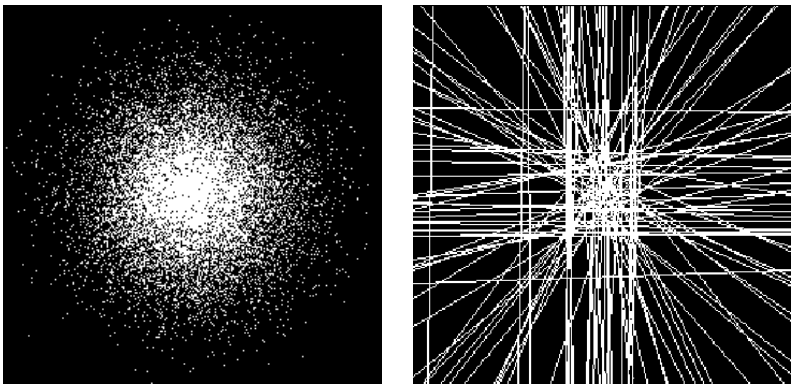
3 Finding a suitable distribution for drawing blocks of measurements

Contents

1 Introduction

Acquiring data by blocks of measurements raises the issue of designing appropriate sampling schemes. In this chapter, we propose to randomly extract blocks of measurements that are made of several rows from a full deterministic sensing matrix A_0. The main

This chapter has been accepted in a sligthly different form as [16], as joint work with Pierre Weiss and Jérémie Bigot.



(a) (b)

Figure 3.1: An example of MRI sampling schemes in the *k-space* (the 2D Fourier plane where low frequencies are centered) (a): Isolated measurements drawn from a probability measure p having a radial distribution. (b): Sampling scheme based on a dictionary of blocks of measurements: blocks consist of discrete lines of the same size.

question investigated is how to choose an appropriate probability distribution from which blocks of measurements will be drawn. A first step in this direction [11, 80] was recently proposed. Unfortunately, the probability distributions proposed in [11] and [80] are difficult to compute numerically and seem suboptimal in practice.

In this chapter, we propose an alternative strategy which is based on the numerical resolution of an optimization problem. The main idea is to construct a probability distribution π on a dictionary of blocks. The blocks are drawn independently at random according to this distribution. We propose to choose π in such a way that the resulting sampling patterns are similar to those based on isolated measurements, such as the ones proposed in the CS literature. For this purpose, we define a dissimilarity measure to compare a probability distribution π on a dictionary of blocks and a target probability distribution p defined on a set of isolated measurements. Then, we propose to choose an appropriate distribution $\pi[p]$ by minimizing its dissimilarity with a distribution p on isolated measurements that is known to lead to good sensing matrices.

This chapter is organized as follows. In Section 2, we describe the problem setting. Then, we construct a dissimilarity measure between probability distributions lying in different, but spatially related domains. We then formulate the problem of finding a probability distribution $\pi[p]$ on blocks of measurements as a convex optimization problem. In Section 3, we present an original and efficient way to solve this minimization problem via a dual formulation and an algorithm based on the accelerated gradient descents in metric spaces [77]. We study carefully how the theoretical rates of convergence are affected by the choice of norms and prox-functions on the primal and dual spaces. Finally, in Section 4, we propose a dictionary of blocks that is appropriate for MRI applications. Then, we compare the quality of MRI images reconstructions using the proposed sampling schemes and those currently used in the context of MRI acquisition, demonstrating the potential of the proposed approach on real scanners.

2 Variable density sampling with block constraints

2.1 Notation

We consider d-dimensional signals for any $d \in \mathbb{N}^*$, of size $n_1 \times n_2 \times \ldots \times n_d = n$. Let E and F denote finite-dimensional vector spaces endowed with their respective norms $\|.\|_E$ and $\|.\|_F$. In the paper, we identify E to $\mathbb{R}^m$ and F to $\mathbb{R}^n$. We denote by E^* and F^*, respectively the dual spaces of E and F. For $s \in E^*$ and $x \in E$ we denote by $\langle s, x \rangle_{E^* \times E}$ the value of s at x. The notation $\langle \cdot, \cdot \rangle$ will denote the usual inner product in a Euclidean space. The norm of the dual space E^* is defined by:

$$\|s\|_{E^*} = \max_{\substack{x \in E \\ \|x\|_E = 1}} \langle s, x \rangle_{E^* \times E} .$$

Let $M : E \to F^*$ denote some operator. When M is linear, we denote its adjoint operator by $M^* : F \to E^*$. The subordinate operator norm is defined by :

$$\|M\|_{E \to F^*} = \sup_{\|x\|_E \leq 1} \|Mx\|_{F^*}$$

When the spaces E^* and F are endowed with ℓ^q and ℓ^p norms respectively, we will use the following notation for the operator norm of M^*:

$$\|M^*\|_{F \to E^*} = \|M^*\|_{p \to q}.$$

We set $\Delta_m \subset E$ to be the simplex in $E = \mathbb{R}^m$, and $\Delta_n \subset F$ to be the simplex in $F = \mathbb{R}^n$. For $\pi \in \Delta_m$ and an index $j \in \{1, \ldots, m\}$ we denote by π_j the j-th component of π.

Let $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ denote a closed convex function. Its Fenchel conjugate is denoted g^*. The relative interior of a set $X \subseteq \mathbb{R}^n$ is denoted $\mathrm{ri}(X)$. Finally, the normal cone to X at a point x on the boundary of X is denoted $\mathcal{N}_X(x)$.

2.2 Problem setting

In this part, we assume that the acquisition system is capable of sensing a finite set $\{y_1, \ldots, y_n\}$ of linear measurements of a signal $x \in \mathbb{R}^{n_s}$ such that

$$y_i = \langle a_i^*, x \rangle, \qquad \forall i = 1, \ldots, n,$$

where a_i^* denotes the i-th row of the full sensing matrix A_0. Let us define a set $\mathcal{I} = \{I_1, \ldots, I_m\}$ where each $I_k \subseteq \{1, \ldots, n\}$ denotes a set of indexes. We assume that the acquisition system has physical constraints that impose sensing simultaneously the following sets of measurements

$$E_k = \{y_i, i \in I_k\}, \qquad \forall k = 1, \ldots, m.$$

In what follows, we refer to $\mathcal{I}$ as the blocks dictionary.

For example in MRI, $n = n_s$ is the number of pixels or voxels of a 2D or 3D image, and y_i represents the i-th discrete Fourier coefficient of this image. In this setting, the sets of indexes I_k may represent straight lines in the discrete Fourier domain as in Figure 3.1(b). In Section 4.1, we give further details on the construction of such a dictionary.

We propose to partially sense the signal using the following procedure:

1. Construct a discrete probability distribution $\pi \in \Delta_m$.

2. Draw i.i.d. indexes $k_1, \dots, k_b$ from the probability distribution π on the set $\{1, \dots, m\}$, with $1 \leq b \leq m$.

3. Sense randomly the signal x by considering the random set of blocks of measurements $\left(E_{k_j}\right)_{j \in \{1,\dots,b\}}$, which leads to the construction of the following sensing matrix

$$A = (a_i^*)_i \in \cup_{j=1}^{b} I_{k_j} \, .$$

The main objective of this work is to provide an algorithm to construct the discrete probability distribution π based on the knowledge of a target discrete probability distribution $p \in \Delta_n$ on the set $\{y_1, \dots, y_n\}$ of isolated measurements. The problem of choosing a distribution p leading to good image reconstruction is not addressed here, since there already exist various theoretical results and heuristic strategies in the CS literature on this topic [71, 30, 2, 63].

2.3 A variational formulation

In order to define π, we propose to minimize a dissimilarity measure between $\pi \in \Delta_m$ and $p \in \Delta_n$. The difficulty lies in the fact that these two probability distributions belong to different spaces. We propose to construct a dissimilarity measure $\mathcal{D}(\pi, p, \mathcal{I})$ that depends on the blocks dictionary $\mathcal{I}$. This dissimilarity measure will be minimized over $\pi \in \Delta_m$ using numerical algorithms with m being relatively large (typically $10^4 \leq m \leq 10^{10}$). Therefore, it must have appropriate properties such as convexity, for the problem to be solvable in an efficient way.

Mapping the m-dimensional simplex to the n-dimensional one

In order to define a reasonable dissimilarity measure, we propose to construct an operator M that maps a probability distribution $\pi \in \Delta_m$ to some $p' \in \Delta_n$:

$$\begin{aligned} M : \quad & E \longrightarrow F^* \\ & \pi \longmapsto p', \end{aligned}$$

where for $i \in \{1, \dots, n\}$,

$$p'_i = \frac{\sum_{k=1}^{m} \pi_k \mathbb{1}_{i \in I_k}}{\sum_{j=1}^{n} \sum_{k'=1}^{m} \pi_{k'} \mathbb{1}_{j \in I_{k'}}}, \tag{3.1}$$

where $\mathbb{1}_{i \in I_k}$ is equal to 1 if $i \in I_k$, 0 otherwise. The i-th element of p' represents the probability to draw the i-th measurement y_i by drawing blocks of measurements according to the probability distribution π. The operator M satisfies the following property by construction :

$$M\Delta_m \subseteq \Delta_n.$$

A sufficient condition for the mapping M to be a linear operator

Note that the operator M is generally non linear, due to the denominator in (3.1). This is usually an important drawback for the design of numerical algorithms involving the operator M. However, if the sets $(I_k)_{k \in \{1,\dots,m\}}$ all have the same cardinality (or length)

equal to ℓ, the denominator in (3.1) is equal to ℓ. In this case, M becomes a linear operator. In this chapter, we will focus on this setting, which is rich enough for many practical applications:

Assumption 3.1. *For $k \in \{1, \dots, m\}$, $\mathrm{Card}\,[I_k] = \ell$, where ℓ is some positive integer.*

Let us provide two important results for the sequel.

Proposition 3.2. *For $\ell > 1$, $M\Delta_m \subsetneq \Delta_n$, i.e. $M\Delta_m$ is a strict subset of Δ_n.*

Proof. By definition of the convex envelope, $M\Delta_m = \mathrm{conv}\,(\{M_{:,i}, i \in \{1, \dots, m\}\})$, where $M_{:,i}$ denotes the i-th column of M. For $\ell > 1$, $\{M_{:,i}, i \in \{1, \dots, m\}\}$ is a subset of Δ_n that does not contain the extreme points of the simplex. □

In practice, Proposition 3.2 means that it is impossible to reach exactly an arbitrary distribution $p \in \Delta_n$, except for the trivial case of isolated measurements.

Proposition 3.3. *Suppose that Assumption 3.1 holds, then for $p \in [1, \infty]$,*

$$\|M^*\|_{p\to\infty} = \ell^{-\frac{1}{p}}.$$

Proof. Under Assumpiton 3.1, all the columns of M have only ℓ non-zero coefficients equal to $1/\ell$. With $\|\cdot\|_F = \|\cdot\|_{\ell^p}$, we can thus derive that

$$\begin{aligned}\|M^*\|_{p\to\infty} &= \max_{\|x\|_p=1} \|M^* x\|_{\ell^\infty} = \max_{1\le i\le m} \max_{\|x\|_p=1} \langle M_{:,i}, x\rangle \\ &= \max_{1\le i\le m} \|M_{:,i}\|_{F^*} = \max_{1\le i\le m} \|M_{:,i}\|_q \\ &= \ell^{-\frac{1}{p}},\end{aligned}$$

where $M_{:,i}$ denotes the i-th column of M, and q is the conjugate of p satisfying $1/p + 1/q = 1$. □

Measuring the dissimilarity between π and p through the operator M

Now that we have introduced the mapping M, we propose to define a dissimilarity measure between $\pi \in \Delta_m$ and $p \in \Delta_n$. To do so, we propose to compare $M\pi$ and p that are both vectors belonging to the simplex Δ_n. Owing to Proposition 3.2, it is hopeless to find some $\tilde{\pi} \in \Delta_m$ satisfying $M\tilde{\pi} = p$ for an arbitrary target density p. Therefore, we can only expect to get an approximate solution by minimizing a dissimilarity measure $\mathcal{D}(M\pi, p)$. For obvious numerical reasons, $\mathcal{D}$ should be convex in π. Among statistical distances, the most natural ones are the total variation distance, Kullback-Leibler of more generally f-divergences. Among this family, total variation presents the interest of having a dual of bounded support. We will exploit this property to design efficient numerical algorithms in Section 3. In the sequel, we will thus use $\mathcal{D}(M\pi, p) = \|M\pi - p\|_{\ell^1}$ to compare the distributions $M\pi$ and p.

Entropic regularization

In applications such as MRI, the number m of columns of M is larger than the number n of its rows. Therefore, $\mathrm{Ker}(M) \neq \emptyset$ and there exist multiple $\pi \in \Delta_m$ with the same

dissimilarity measure $\mathcal{D}(M\pi, p)$. In this case, we propose to take among all these solutions, the one minimizing the neg-entropy $\mathcal{E}$ defined by

$$\mathcal{E} : \pi \in \Delta_m \longmapsto \sum_{j=1}^{m} \pi_j \log(\pi_j), \tag{3.2}$$

with the convention that $0 \log(0) = 0$. We recall that the entropy $\mathcal{E}(\pi)$ is proportional to the Kullback-Leibler divergence between π and the uniform distribution π^c in Δ_m (i.e. such that $\pi_j^c = \frac{1}{m}$ for all j). Therefore, among all the solutions minimizing $\mathcal{D}(M\pi, p)$, choosing the distribution $\pi(p)$ minimizing $\mathcal{E}(\pi)$ gives priority to entropic solutions, i.e. probability distributions which maximize the covering of the sampling space if we proceed to several drawings of blocks of measurements. Therefore, we can finally write the following regularized problem defined by

$$\min_{\pi \in \Delta_m} F_\alpha(\pi), \tag{PP}$$

where

$$F_\alpha(\pi) = \|M\pi - p\|_{\ell^1} + \alpha \mathcal{E}(\pi),$$

for some regularization parameter $\alpha > 0$. Adding the neg-entropy has the effect of spreading out the probability distribution π, which is a desirable property. Moreover, the neg-entropy is strongly convex on the simplex Δ_m. This feature is of primary importance for the numerical resolution of the above optimization problem. Note that an appropriate choice of the regularization parameter α is also important, but this issue will not be addressed here.

A toy example

To illustrate the interest of Problem (PP), we design a simple example. Consider a 3×3 image. Define the target distribution p as a dirac on the central pixel (numbered 5 in Figure 3.2). Consider a blocks dictionary composed of horizontal and vertical lines. In that setting, the operator M is given by

$$M = \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

For such a matrix, there are various distributions minimizing $\|M\pi - p\|_{\ell^1}$. For example, one can choose $\pi_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}^*$ or $\pi_2 = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 & 0 \end{pmatrix}^*$. The solution maximizing the entropy is π_2. In the case of image processing, this solution is preferable since it leads to better covering of the acquisition space. Note that, among all the ℓ^p-norms ($1 \leq p < +\infty$), only the ℓ^1-norm is such that $\|M\pi_1 - p\|_{\ell^1} = \|M\pi_2 - p\|_{\ell^1}$. This property is once again desirable since we want the regularizing term (and not the fidelity term) to force choosing the proper solution.

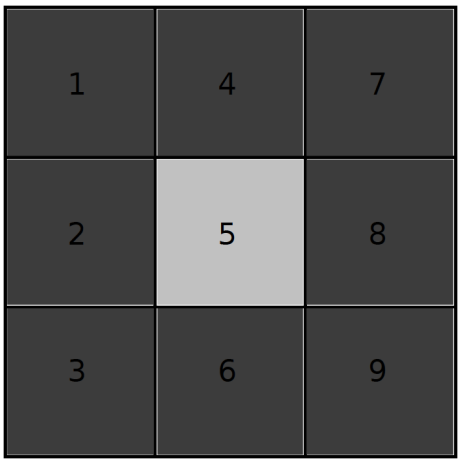


Figure 3.2: Illustration of a target distribution concentrated on the central pixel of a 3×3 images. The pixels are numbered, and this order is kept in the design of M and π.

3 Optimization

In this section, we propose a numerical algorithm to solve Problem (PP). Note that despite being convex, this optimization problem has some particularities that make it difficult to solve. Firstly, the parameter $\pi \in \Delta_m$ lies in a very high dimensional space. In our experiments, n varies between 10^4 and 10^7 while m varies between 10^4 and 10^{10}. Moreover, the function $\mathcal{E}$ is differentiable but its gradient is not Lipschitz, and the total variation distance $\|\cdot\|_{\ell^1}$ is non-differentiable.

The numerical resolution of Problem (PP) is thus a delicate issue. Below, we propose an efficient strategy based on the numerical optimization of the dual problem of (PP), and on the use of Nesterov's ideas [77]. Contrarily to most first order methods proposed recently in the literature [9, 78, 36] which are based on Hilbert space formalisms, Nesterov's algorithm is stated in a (finite dimensional) normed space. We thus perform the minimization of the dual problem on a metric space, and we carefully study the optimal choice of the norms in the primal and dual spaces. We show that depending on the blocks length ℓ, the optimal choice might well be different from the standard ℓ^2-norm. Such ideas stem back from (at least) [32], but were barely used in the domain of image processing.

3.1 Dualization of the problem

Our algorithm consists in solving the problem dual to (PP) in order to avoid the difficulties related to the non-differentiability of the ℓ^1-norm. Proposition 3.4 and 3.7 state that the dual of problem (PP) is differentiable. We will use this feature to design an efficient first-order algorithm and use the primal-dual relationships (Proposition 3.8) to retrieve the primal solution.

Proposition 3.4. *Let* $J_\alpha(q) := \langle p, q\rangle_{F^*\times F} - \alpha \log\left(\sum_{\ell=1}^m \exp\left(-\frac{(M^*q)_\ell}{\alpha}\right)\right)$, *for* $q \in F$. *The dual problem to* (PP) *is:*

$$-\min_{q\in B_\infty} J_\alpha(q), \tag{DP}$$

in the sense that $\min_{\pi\in\Delta_m} F_\alpha(\pi) = \max_{q\in B_\infty} -J_\alpha(q)$, *where* B_∞ *is the* ℓ^∞*-ball of unit radius in* F.

Proof. The proof is available in Appendix 5.1. □

In order to study the regularity properties of J_α, and so the solvability of (DP), we use the strong convexity of the neg-entropy $\mathcal{E}$ with respect to $\|\cdot\|_E$. First, let us recall one version of the definition of the strong convexity in Banach spaces.

Definition 3.5. *We say that $f : F \to \mathbb{R}$ is σ-strongly convex with respect to $\|\cdot\|_F$ on $F' \subset F$ if*

$$\forall x, y \in F', \quad \forall t \in [0,1], \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\sigma}{2}t(1-t)\|x-y\|_F^2. \tag{3.3}$$

We define the convexity modulus σ_f of f as the largest positive real σ satisfying Equation (3.3).

Proposition 3.6. *For $\|\cdot\|_E = \|\cdot\|_{\ell^p}$, $p \in [1, +\infty]$, the convexity modulus of the neg-entropy on the simplex Δ_m is $\sigma_\mathcal{E} = 1$.*

Proof. The proof is available in Appendix 5.2. □

Proposition 3.7. *The function J_α is convex and its gradient is Lipschitz continuous i.e.*

$$\|\nabla J_\alpha(q_1) - \nabla J_\alpha(q_2)\|_{F^*} \leq L_\alpha \|q_1 - q_2\|_F \qquad \forall (q_1, q_2) \in F^2.$$

with constant

$$L_\alpha = \frac{\|M^*\|_{F\to E^*}^2}{\alpha\sigma_\mathcal{E}}. \tag{3.4}$$

Moreover, ∇J_α is locally Lipschitz around $q \in F$ with constant

$$L_\alpha(q) = \frac{\|M^*\|_{F\to E^*}^2}{\alpha\sigma_\mathcal{E}(\pi(q))}, \tag{3.5}$$

where $\sigma_\mathcal{E}(\pi) := \inf_{\|h\|_E=1} \left\langle \mathcal{E}''(\pi)h, h \right\rangle$ is the local convexity modulus of $\mathcal{E}$ around π, and an explicit expression for $\pi(q)$ is given in (3.16).

Proof. The proof is available in Appendix 5.3. □

Note that a standard reasoning would rather lead to $L_\alpha = \frac{\|M^*\|_{2\to 2}^2}{\alpha\sigma_\mathcal{E}}$, which is usually much larger than bound (3.4). Proposition 3.7 implies that Problem (DP) is efficiently solvable by Nesterov's algorithm [77]. Therefore, we will first solve the dual problem (DP). Then, we use the relationships between the primal and dual solutions (as described in Proposition 3.8) to finally compute a primal solution $\pi^\star$ for Problem (PP).

Proposition 3.8. *The relationships between the primal and dual solutions*

$$\pi^\star = \arg\min_{\pi\in\Delta_m} F_\alpha(\pi) \qquad \text{and} \qquad q^\star \in \arg\min_{q\in B_\infty} J_\alpha(q)$$

are given by

$$\pi_j^\star = \frac{\exp\left(-\frac{(M^*q^\star)_j}{\alpha}\right)}{\sum_{k=1}^m \exp\left(-\frac{(M^*q^\star)_k}{\alpha}\right)}, \qquad \forall j \in \{1, \dots, m\}. \tag{3.6}$$

Furthermore,

$$\operatorname{sign}(M\pi^\star - p) = \operatorname{sign}(q^\star). \tag{3.7}$$

Proof. Equation (3.6) is a direct consequence of (3.16). To derive the second equation (3.7), it suffices to write the optimality conditions of the problem $\max_{q \in B_\infty} \langle M\pi^\star - p, q \rangle_{F^* \times F} + \alpha \mathcal{E}(\pi^\star)$. It leads to:

$$M\pi^\star - p \in \mathcal{N}_{B_\infty}(q^\star) \Leftrightarrow \operatorname{sign}(M\pi^\star - p) = \operatorname{sign}(q^\star).$$

□

3.2 Numerical optimization of the dual problem

Now that the dual problem (DP) is fully characterized, we propose to solve it using Nesterov's optimal accelerated projected gradient descent [77] for smooth convex optimization.

3.2.1 The algorithm

Nesterov's algorithm is based on the choice of a prox-function d of the set B_∞, i.e. a continuous function that is strongly convex on B_∞ w.r.t. $\|\cdot\|_F$. Let σ_d denote the convexity modulus of d, we further assume that $d(q_c) = 0$ so that

$$d(q) \geq \frac{\sigma_d}{2}\|q - q_c\|_F^2 \qquad \forall q \in B_\infty,$$

where $q_c = \arg\min_{q \in B_\infty} d(q)$. Nesterov's algorithm is described in Algorithm 1.

Algorithm 1 Resolution scheme for smooth optimization proposed by [77]

① Initialization: choose $q_0 \in B_\infty$.
② **for** $k = 0 \dots K$ **do**
③ Compute $J_\alpha(q_k)$ and $\nabla J_\alpha(q_k)$
④ Find $y_k \in \arg\min_{y \in B_\infty} \langle \nabla J_\alpha(q_k), y - q_k \rangle + \frac{1}{2} L_\alpha \|y - q_k\|_F^2$
⑤ Find $z_k \in \arg\min_{q \in B_\infty} \frac{L_\alpha}{\sigma_d} d(q) + \sum_{i=0}^{k} \frac{i+1}{2} \left[J_\alpha(q_i) + \langle \nabla J_\alpha(q_i), q - q_i \rangle \right]$
⑥ Set $q_{k+1} = \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$.
⑦ **end for**
⑧ Set the primal solution to $\pi_j = \dfrac{\exp\left(-\frac{(M^* y_K)_j}{\alpha}\right)}{\sum_{k=1}^{m} \exp\left(-\frac{(M^* y_K)_k}{\alpha}\right)}, \qquad \forall j \in \{1, \dots, m\}.$

Theorem (3.9) summarizes the theoretical guarantees of Algorithm 1.

Theorem 3.9. *[77, Theorem 2] Algorithm 1 ensures that*

$$\begin{aligned} J_\alpha(y_k) - J_\alpha(q^\star) &\leq \frac{4 L_\alpha d(q^\star)}{\sigma_d (k+1)(k+2)} \\ &\leq \frac{4 \|M^*\|_{F \to E^*}^2 d(q^\star)}{\alpha \sigma_{\mathcal{E}} \sigma_d (k+1)(k+2)}, \end{aligned} \tag{3.8}$$

where $q^\star$ is an optimal solution of Problem (DP).

Since $d(q^\star)$ is generally unknown, we can bound (3.8) by

$$\frac{4\|M^*\|^2_{F\to E^*} D}{\alpha\sigma_{\mathcal{E}}\sigma_d(k+1)(k+2)}. \tag{3.9}$$

where $D = \max_{q\in B_\infty} d(q)$. Note that until now, we got theoretical guarantees in the dual space but not in the primal. What matters to us is rather to obtain guarantees on the primal iterates, which can be summarized by the following theorem.

Theorem 3.10. *Denote*

$$\pi_k = \frac{\exp\left(-\frac{(M^* y_k)}{\alpha}\right)}{\left\|\exp\left(-\frac{(M^* y_k)}{\alpha}\right)\right\|_{\ell^1}}.$$

where y_k is defined in Algorithm 1. The following inequality holds:

$$\|\pi_k - \pi^\star\|_E^2 \le \frac{8\|M^*\|^2_{F\to E^*} D}{\alpha^2\sigma_{\mathcal{E}}^2\sigma_d(k+1)(k+2)}.$$

The proof is given in Appendix 5.4. It is a direct consequence of a more general result of independent interest.

3.2.2 Choosing the prox-function and the metrics

Algorithm 1 depends on the choice of $\|\cdot\|_E$, $\|\cdot\|_F$ and d. The usual accelerated projected gradient descents consist in setting $\|\cdot\|_E = \|\cdot\|_{\ell^2}$, $\|\cdot\|_F = \|\cdot\|_{\ell^2}$ and $d(\cdot) = \frac{1}{2}\|\cdot\|^2_{\ell^2}$. However, we will see that it is possible to change the algorithm's speed of convergence by making a different choice. To this end, we concentrate on the usual ℓ^p-norms, $p \in [1, +\infty]$.

Choosing a norm on E: The following proposition shows an optimal choice for $\|\cdot\|_{E^*}$.

Proposition 3.11. *The norm $\|\cdot\|_{E^*}$ that minimizes (3.9) among all ℓ^p-norms, $p \in [1, +\infty]$ is $\|\cdot\|_{\ell^\infty}$. Note however that the minimum local Lipschitz constant $L_\alpha(q)$ for $q \in F$ might be reached for another choice of $\|\cdot\|_{E^*}$.*

Proof. From Proposition 3.6, we get that $\sigma_{\mathcal{E}}$ remains unchanged no matter how $\|\cdot\|_E$ is chosen among ℓ^p-norms. The choice of $\|\cdot\|_E$ is thus driven by the minimization of $\|M^*\|_{F\to E^*}$. From the operator norm definition, it is clear that the best choice consists in setting $\|\cdot\|_{E^*} = \|\cdot\|_{\ell^\infty}$ since the ℓ^∞-norm is the smallest of all ℓ^p-norms. □

According to Proposition 3.11, choosing $\|\cdot\|_{E^*}$ to be $\|\cdot\|_{\ell^\infty}$ leads to consider $\|\cdot\|_E$ to be $\|\cdot\|_{\ell^1}$. As shown by Proposition 3.3, it is clear that the norm $\|M^*\|_{F\to E^*}$ may vary a lot with respect to $\|\cdot\|_F$ for the particular operator M considered here.

Choosing a norm on F and a prox-function d: by Proposition 3.11 the norm $\|\cdot\|_F$ and the prox function d should be chosen in order to minimize $\frac{\|M^*\|^2_{F\to\infty} D}{\sigma_d}$. We are unaware of a general theory to make an optimal choice despite recent progresses in that direction. The recent paper [40] proposes a systematic way of selecting $\|\cdot\|_F$ and d in order to make the algorithm complexity invariant to change of coordinates for a general optimization problem. The general idea in [40] is to choose $\|\cdot\|_F$ to be the Minkowski gauge of the constraints set (of the optimization problem), and d to be a strongly convex approximation of $\frac{1}{2}\|\cdot\|^2_F$. However, this strategy is not shown to be optimal. In our setting, since the

constraints set is B_∞, this would lead to choose $\|\cdot\|_F = \|\cdot\|_{\ell^\infty}$. Unfortunately, there is no good strongly convex approximation of $\frac{1}{2}\|\cdot\|^2_{\ell^\infty}$.

In this section, we thus study the influence of $\|\cdot\|_F$ and d both theoretically and experimentally, with $\|\cdot\|_F \in \{\|\cdot\|_{\ell^1}, \|\cdot\|_{\ell^2}, \|\cdot\|_{\ell^\infty}\}$. Propositions 3.12, 3.13 and 3.14 summarize the theoretical algorithm complexity in different regimes.

Proposition 3.12. *Let $p' \in]1, 2]$. Define $d_{p'}(x) = \frac{1}{2}\|x\|^2_{p'}$. Then*

- *For $p \in [p', \infty]$, $d_{p'}$ is $(p'-1)$-strongly convex w.r.t. $\|\cdot\|_p$.*
- *For $p \in [1, p']$, $d_{p'}$ is $(p'-1)n^{(1/p'-1/p)}$-strongly convex w.r.t. $\|\cdot\|_p$.*

Proof. The proof is a direct consequence of [60, Proposition 3.6] and of the fact that for $p' \geq p$,

$$\|x\|_{p'} \leq \|x\|_p \leq n^{(1/p-1/p')}\|x\|_{p'}.$$

□

Proposition 3.13. *Suppose that Assumption 3.1 holds. Set $\|\cdot\|_F = \|\cdot\|_p$ and $d = d_{p'}$ with $p \in [1, \infty]$ and $p' \in]1, 2]$. For all this family of norms and prox-functions, the one minimizing the complexity bound (3.9) is*

- *$p' = 2$ and $p \in [1, 2]$, if $\ell^2 = n$. For this choice, we get*

$$J_\alpha(y_k) - J_\alpha(q^\star) \leq \frac{2\sqrt{n}}{\alpha(k+1)(k+2)}. \tag{3.10}$$

- *$p = p' = 2$, if $\ell^2 < n$. For this choice, we get*

$$J_\alpha(y_k) - J_\alpha(q^\star) \leq \frac{2n}{\alpha\ell(k+1)(k+2)}. \tag{3.11}$$

- *$p = 1$ and $p' = 2$, if $\ell^2 > n$. For this choice, we get*

$$J_\alpha(y_k) - J_\alpha(q^\star) \leq \frac{2n^{3/2}}{\alpha\ell^2(k+1)(k+2)}. \tag{3.12}$$

Proof. The result is a direct consequence of Proposition 3.12. □

Unfortunately, the bounds in (3.10), (3.11) and (3.12) are dimension dependent. Moreover, the optimal choice suggested by Proposition 3.13 is different from the Minkowski gauge approach suggested in [40]. Indeed, in all the cases described in Proposition 3.9, the optimal choice $\|\cdot\|_F$ differs from $\|\cdot\|_{\ell^\infty}$. The difficulty to apply this approach is to find a function $d \simeq 1/2\|\cdot\|^2_{\ell^\infty}$ strongly convex w.r.t. $\|\cdot\|_{\ell^\infty}$. A simple choice consists in setting $d_\varepsilon = \frac{1}{2}\|\cdot\|^2_{\ell^\infty} + \frac{\varepsilon}{2}\|\cdot\|^2_{\ell^2}$. This function is ε-strongly convex w.r.t. $\|\cdot\|_{\ell^\infty}$. We thus get the following proposition:

Proposition 3.14. *Suppose that Assumption 3.1 holds, with $\ell = \sqrt{n}$. Set $\|\cdot\|_F = \|\cdot\|_{\ell^\infty}$, $d_\varepsilon(\cdot) = \frac{1}{2}\|\cdot\|^2_{\ell^\infty} + \frac{\varepsilon}{2}\|\cdot\|^2_{\ell^2}$.*

$$J_\alpha(y_k) - J_\alpha(q^\star) \leq \frac{2\,(1/\varepsilon + n)}{\alpha(k+1)(k+2)}.$$

In particular, for $\varepsilon \propto \frac{1}{n}$, $J_\alpha(y_k) - J_\alpha(q^\star) = O\left(\frac{n}{\alpha k^2}\right)$.

Note that this complexity bound is worse than that of Proposition 3.13 in the case where $\ell = \sqrt{n}$. In the next section, we intend to illustrate and to confirm in practice the different rates of convergence, predicted by the theoretical results in Proposition 3.13.

3.3 Numerical experiments on convergence

In this section, we are willing to emphasize the improvement achieved by appropriately choosing the norms $\|.\|_E$, $\|.\|_F$, and the prox-function d. To do so, we run experiments on a dictionary of blocks of measurements having all the same size $\ell = 256$, described in Section 4.1, for 2D images of size 256×256. At first, we choose $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^2}$ and $d = \frac{1}{2}\|.\|_{\ell^2}^2$ and we perform Algorithm 1 for this dictionary. In fact, this first case (the norm on E differs from the usual $\|\cdot\|_{\ell^2}$) nearly corresponds to a standard accelerated gradient descent [76]. In a second time, we set $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^\infty}$ $d = \frac{1}{2}\|.\|_{\ell^2}^2$. In Figure 3.3, we display the decrease of the objective function in both settings. Figure 3.3 points out that a judicious selection of norms on E and F can significantly speed up the convergence: for 29 000 iterations, the standard accelerated projected gradient descent reaches a precision of 10^{-5} whereas Algorithm 1 with $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^\infty}$, i.e. a "modified" gradient descent, reaches a precision of 10^{-3}. The conclusions for this numerical experiment appear to be faithful to what was predicted by the theory, see Proposition 3.9. For the sake of completeness, we add in Figure 3.3 (in green) the case where $\|.\|_E = \|.\|_{\ell^2}$, $\|.\|_F = \|.\|_{\ell^2}$ and $d = \frac{1}{2}\|.\|_{\ell^2}^2$, which is an usual choice in practice. Clearly, this is the slowest rate of convergence observed.

Finally, we perform the algorithm for $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^2}$, and $d = \frac{1}{2}\|.\|_{\ell^2}^2$ by changing the value of L_α. The value of L_α provided by Proposition 3.7 is tight uniformly on B_∞. However, the local Lipschitz constant of ∇J_α varies rapidly inside the domain. In practice, the Lipschitz constant around the minimizer may be much smaller than L_α (note that $\pi^\star \in \mathrm{ri}(\Delta_m)$ for all $\alpha > 0$). In this last heuristic approach, we will decrease L_α by substantial factors without losing practical convergence. This result is presented in Figure 3.3 where the black curve denotes convergence result when the Lipschitz constant L_α has been divided by 100. We can observe that in this case, it suffices 1500 iterations to reach the precision obtained by the case $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^2}$ and $d = \frac{1}{2}\|.\|_{\ell^2}^2$ (in red) after 29000 iterations. Let us give an intuitive explanation to this positive behaviour. To simplify the reasoning, let us assume that $\pi^\star$ is the uniform probability distribution. First notice that the choice of $\|\cdot\|_E$ only influences the Lipschitz constant of ∇J_α but does not change the algorithm, so that we can play with the norm on E to decrease the local Lipschitz constant. Furthermore, the choice of $\|\cdot\|_E$ minimizing the global Lipschitz constant may be different from the one minimizing the local Lipschitz constant. Considering that $\|\cdot\|_E = \|\cdot\|_{\ell^2}$, from Equation (3.5), we get that $L_\alpha(q^\star) = \frac{\|M^*\|_{2\to 2}^2}{\alpha\sigma_{\mathcal{E}}(\pi^\star)}$. Using Perron-Frobenius theorem, it can be shown that $\|M^*\|_{2\to 2}^2 = O(1)$ for our choice of dictionary, and $\sigma_{\mathcal{E}}(\pi^\star) = m$ for $\|\cdot\|_E = \|\cdot\|_{\ell^2}$. From this simple reasoning, we can conclude that the local Lipschitz constant around $\pi^\star$ is no greater than $O(1/m)$. This means that if the minimizer is sufficiently far away from the simplex boundary, we can decrease L_α by a significant factor without loosing convergence.

4 Numerical results

In this section, we assess the reconstruction performance of the sampling patterns using the approach described in Section 3.2 with $\alpha = 10^{-2}$. We compare it to standard approaches used in the context of MRI. We call $\pi\,[p]$ the probability distribution $\pi^\star$ resulting from the minimization problem (PP) for a given target distribution p on isolated measurements.

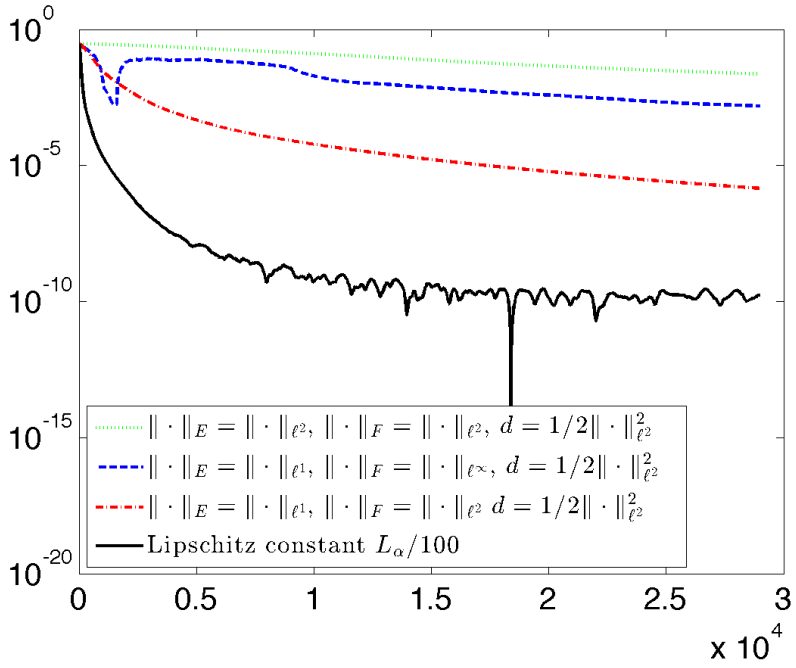


Figure 3.3: Convergence curves in a semi-logarithmic scale for Algorithm 1 ($\alpha = 10^{-2}$) (number of iterations on the x-axis) in green the case where $\|.\|_E = \|.\|_{\ell^2}$, $\|.\|_F = \|.\|_{\ell^2}$, $d = \frac{1}{2}\|.\|^2_{\ell^2}$, in red the case where $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^2}$, $d = \frac{1}{2}\|.\|^2_{\ell^2}$, in blue the case where $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^\infty}$ $d = \frac{1}{2}\|.\|^2_{\ell^2}$, and in black the case where $\|.\|_E = \|.\|_{\ell^1}$, $\|.\|_F = \|.\|_{\ell^2}$, $d = \frac{1}{2}\|.\|^2_{\ell^2}$ with a restricted Lipschitz constant $L'_\alpha = L_\alpha/100$.

4.1 The choice of a particular dictionary of blocks

From a numerical point of view, we study a particular system of blocks of measurements. The dictionary used in all numerical experiments of this article is composed of discrete lines of length ℓ, joining any pixel on the edge of the image to any pixel on the opposite edge, as in Figure 3.1(b). Note that the number of blocks in this dictionary is $n_1^2 + n_2^2$ for an image of size $n_1 \times n_2$. The choice of such a dictionary is particularly relevant in MRI, since the gradient waveforms that define the acquisition paths is subject to bounded-gradient and slew-rate constraints, see e.g. [71]. Moreover the practical implementation on the scanner of straight lines is straightforward since it is already in use in standard echo-planar imaging strategies.

Remark that, in such a setting, the mapping M, defined in (3.1), is a linear mapping that can be represented by a matrix of size $n \times m$ with $M_{i,j} = 1/\ell$ when the i-th pixel belongs to the j-th block, for $i = 1, \dots, n$ and $j = 1, \dots, m$.

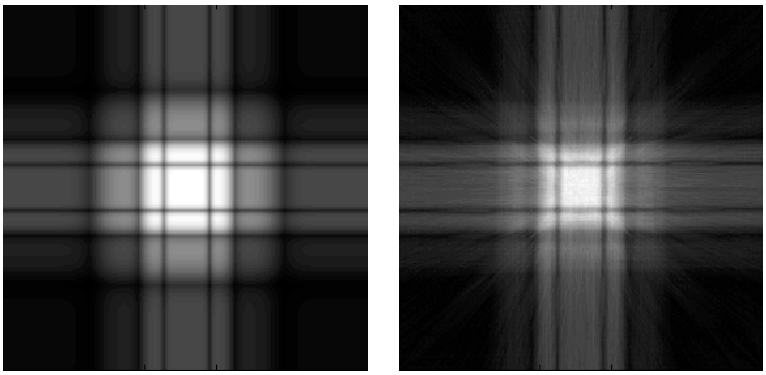
One may argue that in MRI, dealing with samples lying on continuous lines (and not discrete grids) is more realistic in the design of the MR sequences. To deal with this issue, one could resort to the use of the Non-Uniform Fast Fourier Transform. This technique is however much more computationally intensive. In this study we thus stick to values of the Fourier transform located on the Euclidean grid. This is commonly used in MRI with regridding techniques.

4.2 The reconstructed probability distribution

We are willing to illustrate the fidelity of $\pi[p]$, the solution of Problem (PP), to a given target p. In the setting of 2D MR sensing, with the dictionary of lines in dimension $n_1 = n_2 = 256$ described in the previous subsection. We set the target probability distribution $p = p_{\text{opt}}$ the one suggested by current CS theories on the set of isolated measurements. It is proportional to $\|a_k^*\|_{\ell^\infty}^2$, see [84, 30, 11]. To give an idea of what the resulting probability distribution $\pi[p_{\text{opt}}]$ looks like, we draw 100 000 independent blocks of measurements according to $\pi[p_{\text{opt}}]$ and count the number of measurement for each discrete Fourier coefficient. The result is displayed on Figure 3.4. This experiment underlines that our strategy manages to catch the overall structure of the target probability distibution.

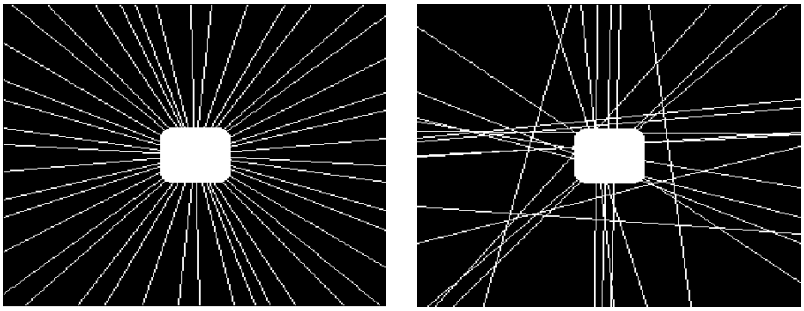
4.3 Reconstruction results

In this section, we compare the reconstruction quality of MR images for different acquisition schemes. The comparison is always performed for schemes with an equivalent number of isolated measurements. We recall that in the case of MR images, the acquisition is done in the Fourier domain, and MR images are supposed to be sparse in the wavelet domain. Therefore, the full sensing matrix $A_0 = (a_1|a_2|\dots|a_n)^*$, which models the acquisition process, is the composition of a Fourier transform with an inverse Wavelet transform. The reconstruction is done via ℓ^1-minimization, using Douglas-Rachford algorithm [38]. It was proven in various papers [30, 27, 2] that MRI image quality can be strongly improved by fully acquiring the center of the Fourier domain via a mask defined by the support of the mother wavelet, see Figure 3.5. Therefore, for every type of schemes used in our reconstruction test, we first fully acquire this mask.



(a) (b)

Figure 3.4: Illustration of the fidelity of $\pi[p_{\text{opt}}]$ to p_{opt}. (a): on the left hand side, we present the target probability distribution p_{opt} (b): on the right hand side, we perform 100000 i.i.d. drawings according to $\pi[p_{\text{opt}}]$ of blocks from the blocks dictionary and count the number of times that a point is sampled at each location.

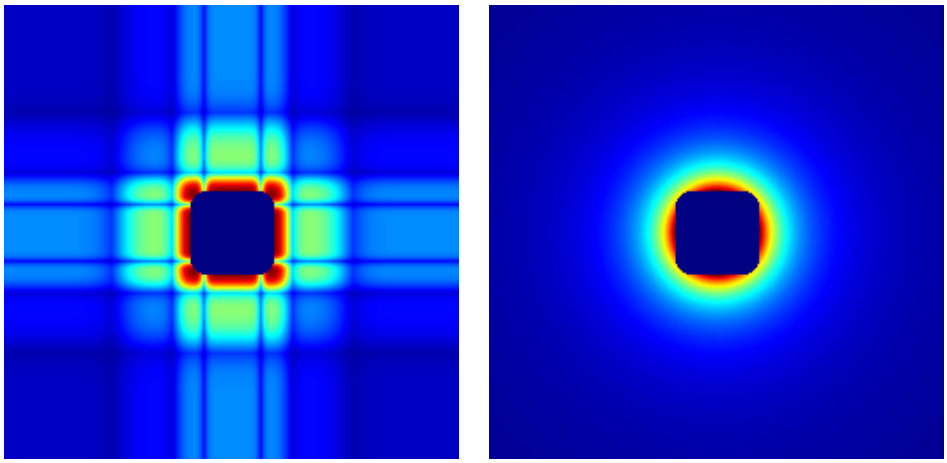


(a) (b)

Figure 3.5: Different schemes based (a) on the golden angle pattern, and (b) on the dictionary proposed in Section 4.1. Both schemes are combined with a mask which fully samples the center of the Fourier domain. In both cases, the proportion of total measurements represents 10% of the full image, while the mask defined by the support of the mother wavelet represents 3% of the full image.

The various schemes considered in this chapter are based on blocks of measurements and on heuristic schemes that are widely used in the context of MRI. They will consist in:

- Equiangularly distributed radial lines: the scheme is made of lines always intersecting the center of the acquisition domain, and that are distributed uniformly [70].
- Golden angle scheme: the scheme is made of radial lines separated from the golden angle, i.e 111.246°. This technique is used often in MRI sequences, and it gives good reconstruction results in practice [104].
- Random radial scheme: radial lines are drawn uniformly at random [26].
- Scheme based on the dictionary described in Section 4.1



(a) (b)

Figure 3.6: Target probabilities on pixels (in red, high values, and in dark blue, values close to 0). (a) displays the distribution proportional to $\|a_i^*\|_{\ell^\infty}^2$ defined in [30], (b) displays a radial distribution as presented in [63]. The center has been set to zero, since it will be sampled by the mask in a deterministic way.

- Blocks are drawn according to $\pi\left[p_{\text{rad}}\right]$ which is the resulting probability distribution obtained by minimizing Problem (PP) for $p = p_{\text{rad}}$. The distribution p_{rad} a radial distribution that decreases as $\mathcal{O}\left(\frac{1}{k_x^2+k_y^2}\right)$. This choice was justified recently in [63] and used extensively in practice. Note that p_{rad} is set to 0 on the k-space center since it is already sampled deterministically, see Figure 3.6(b).
- Blocks are drawn according to $\pi\left[p_{\text{opt}}\right]$, which is the resulting probability distribution obtained by minimizing Problem (PP) for $p = p_{\text{opt}}$ defined in [30, 11]. Once again, p_{opt} is set to 0 on the k-space center, see Figure 3.6(a).

Setting 256×256

The numerical experiment is run for images of size $n_0 \times n_0$ with $n_0 = 256$. The full dictionary described in Section 4.1 contains lines of length $\ell = n_0$ pixels connecting every point on the edge of the image to every point on the opposite side. For each proportion of measurements $(10\%, 15\%, 20\%, 25\%, 30\%, 40\%, 50\%)$, we proceed to 100 drawings of schemes when the considered scheme is random. Reconstruction results, for the reference images showed in Figure 3.7 and for various sampling schemes, are displayed in terms of means of PSNR in Figure 3.8(a)(c).

Figure 3.8 shows that the schemes based on the approach presented in this article give better results than random radial schemes, for any proportion of measurements. The improvement in terms of PSNR is generally between 1 and 2 dB. The schemes based on $\pi\left[p_{\text{opt}}\right]$ and $\pi\left[p_{\text{rad}}\right]$ are competitive with those based on the golden angle or equiangularly distributed schemes in the case where the proportion of measurements is low (less than 20% of measurements). We observe that for 10% measurements, schemes based on our dictionary and drawn according to $\pi\left[p_{\text{rad}}\right]$ outperform by more than 1 dB the standard sampling strategies. Increasing the PSNR of 1dB is significant and can be qualitatively

(a) Brain (b) Baboon

Figure 3.7: Reference images to reconstruct for the settings 256×256 and 512×512 .

observed in the reconstructed image.

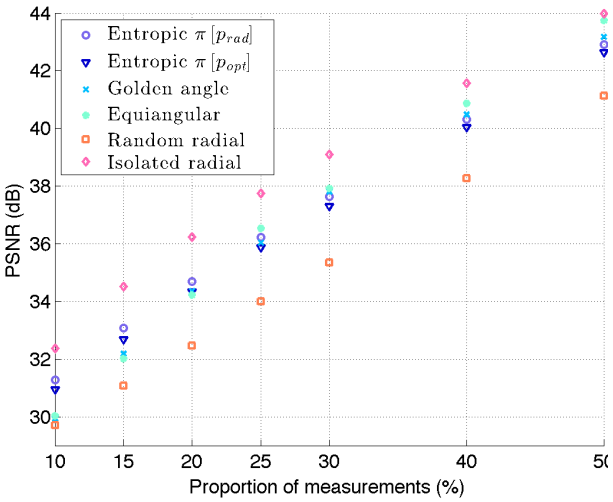
In Figures 3.8(a)(c), it can be seen that block-constrained acquisition never outperforms acquisitions based on independent measurements. This was to be expected since adding constraints reduces the space of possible sampling patterns. Once again, note that independent drawings are however not conceivable in many contexts such as MRI. In a nutshell, we can say that the proposed sampling approach always produces results comparable to the standard sampling schemes and tend to produce better results for low sampling ratios.

Finally, in Figure 3.9, we illustrate that block-constrained acquisition does not allow to reach an arbitrary target distribution by showing the difference between p_{rad} and the probability distribution $M(\pi[p_{\text{rad}}])$ which is defined on the set of isolated measurements. This confirms Proposition 3.2 experimentally.

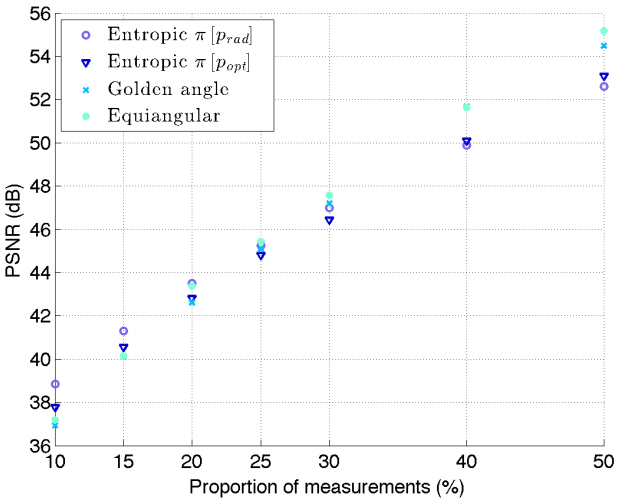
Setting 512×512

Given that in CS the quality of the reconstruction can be resolution dependent, as described in [2], we have decided to run the same numerical experiment on 512×512 images. The numerical experiment is run for images of size $n_0 \times n_0$ with $n_0 = 512$. The full dictionary described in Section 4.1 contains lines of length $\ell = n_0$ connecting every point on the edge of the image to every point on the opposite side. For each proportion of measurements $(10\%, 15\%, 20\%, 25\%, 30\%, 40\%, 50\%)$, we proceed to 10 drawings of sampling schemes when the considered scheme is random. The images of reference to reconstruct are the same as in the setting 256×256, see Figure 3.7.

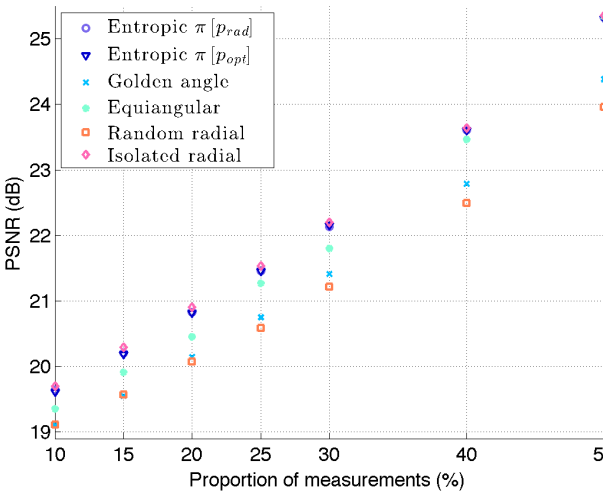
Quality of reconstructions are compared in Figure 3.8(b) and (d) for the golden or equiangularly distributed lines and our proposed method based on $\pi(p_{\text{opt}})$ and $\pi(p_{\text{rad}})$. This experiment shows that the PSNR of the reconstructed images is significantly improved by using the proposed method until 30% of measurements for the brain image and until 40% of measurements for the baboon one. We can remark that for both images, for a same proportion of measurements, the PSNR of the reconstructed images increases while the resolution increases. This numerical experiment suggests that the proposed sampling approach might be significantly better than traditional ones in the MRI context for high resolution images. In Figure 3.10(a), we present the reconstructed image of the



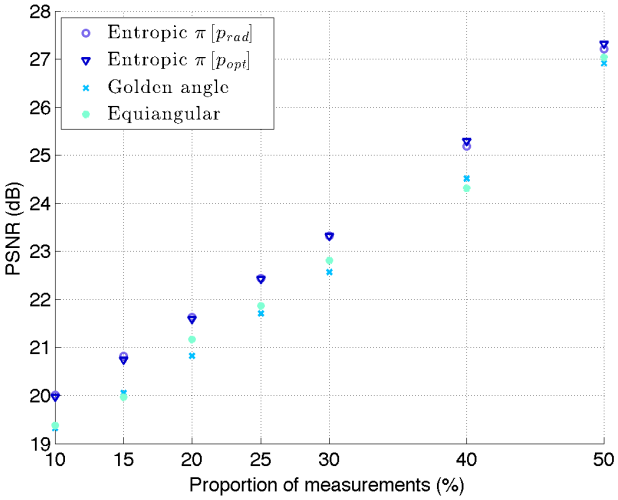
(a) Brain in 256×256



(b) Brain in 512×512

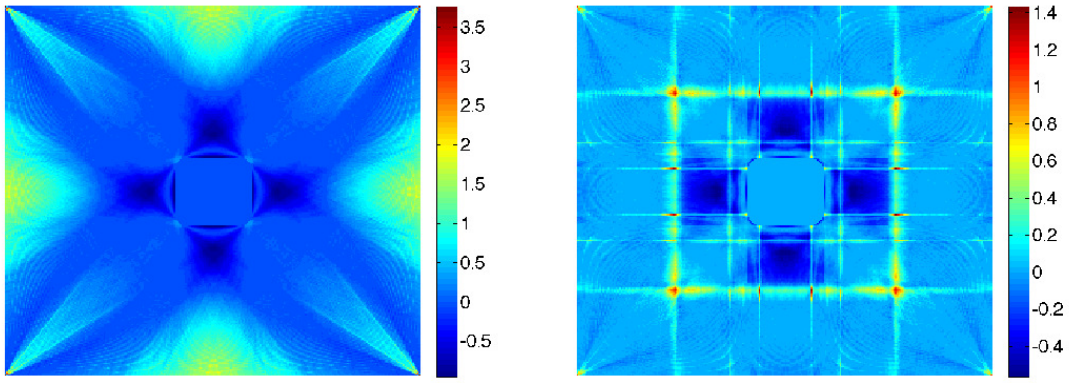


(c) Baboon in 256×256



(d) Baboon in 512×512

Figure 3.8: PSNR means of the reconstructed images (brain, baboon) with respect to the proportion of measurements chosen in the scheme $(10\%, 15\%, 20\%, 25\%, 30\%, 40\%, 50\%)$. The undersampling ratio for all schemes is the ratio between the number of sampled distinct frequencies and the total number of possible measurements. This means that duplicated frequencies are accounted for only once.



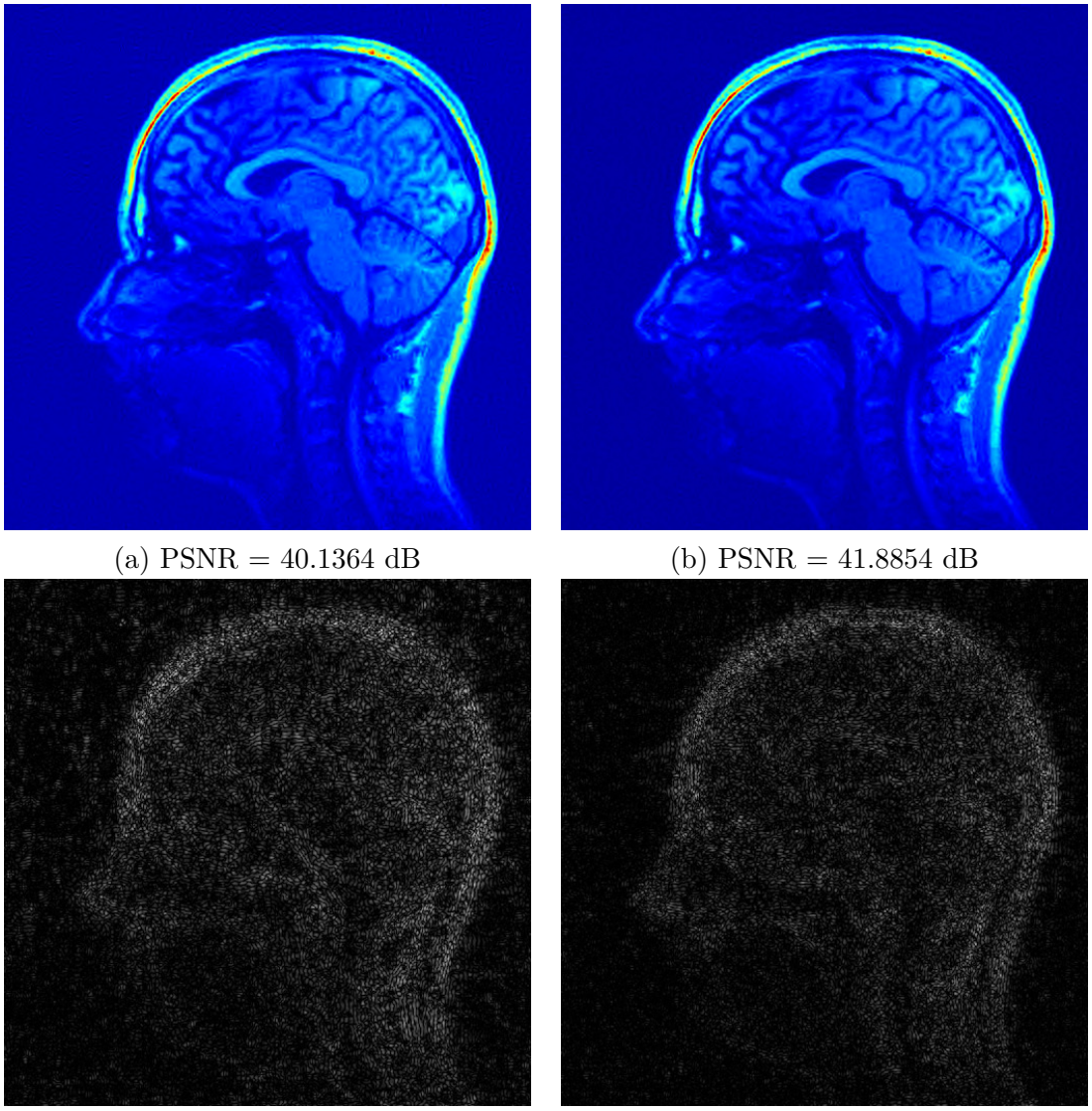
(a) (b)

Figure 3.9: Difference between the target probabilities p and $M\pi(p)$ relatively to the magnitude of p, i.e. we show the following quantity $\frac{(M(\pi(p)))_i - p_i}{p_i}$ for the i-th sampling location, (a) for the radial distribution p_{rad}, we see that we "sub-draw" by a factor of 50 % around the mask, and we "over-draw" by a factor of 150 % at the center of the edges. (b) for the CS optimal distribution p_{opt}, we see that we "sub-draw" by a factor of 40 % around the mask. Note that the sub-drawing effect cannot be avoided: indeed, we cannot reach any target probability distribution via M, see Proposition 3.2.

brain from 15% of measurements in the case of a golden angle scheme. In Figure 3.10(b), we present the reconstructed image of the brain from 15% of measurements in the case of a realization of schemes based on $\pi(p_{\text{rad}})$. The latter's PSNR is 41.88 dB whereas in the golden scheme case, the PSNR only reaches 40.13 dB. In Figure 3.10(c) and (d), we display the corresponding difference images to the reference image, which underlines the improvement of 1.7 dB in our method.

As a side remark, let us mention that in MRI, sampling diagonal or horizontal lines actually takes the same scanning time (even though the diagonals are longer), since gradient coils work independently in each direction. In the MRI example, the length of the path is thus less meaningful that the number of scanned lines. In Figure 3.11, we show different sampling schemes based on the golden angle pattern or on our method with the same number of lines, and we show the corresponding reconstructions of brain images.

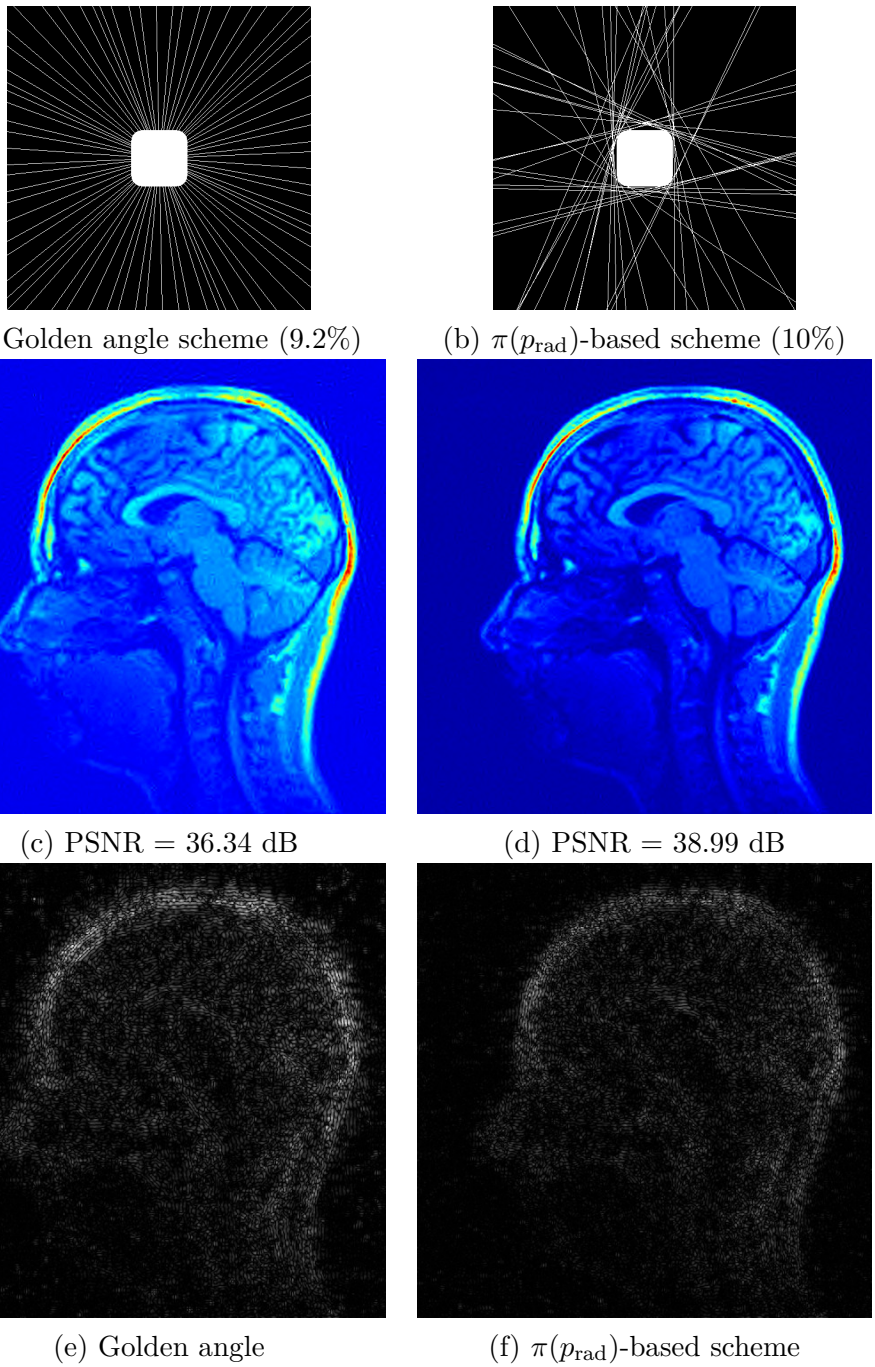
Remark. In both settings, for the brain image, schemes based on $\pi\left[p_{\text{rad}}\right]$ lead to better reconstruction results in terms of PSNR than schemes from $\pi\left[p_{\text{opt}}\right]$. This can be explained by the fact that p_{opt} is the probability density given by CS theory which provides guarantees for any s-sparse image to reconstruct. However, brain images or natural images have a structured sparsity in the wavelet domain: indeed, their wavelet transform is not uniformly s-sparse, the approximation part contains more non-zero coefficients than the rest of the details parts. We can infer that p_{rad} manages to catch the sparsity structure of the wavelet coefficients of the considered images.



(a) PSNR = 40.1364 dB (b) PSNR = 41.8854 dB

(c) (d)

Figure 3.10: Comparison of the reconstructed images from 15% of measurements for a 512×512 image for a golden angle scheme (a), and a scheme based on our dictionary and $\pi(p_{\text{rad}})$ (b). We respectively plot the absolute difference to the reference image for the reconstruction using a golden angle scheme in (c) and for the reconstruction using a scheme based on $\pi(p_{\text{rad}})$ in (d). Note that in (c) and (d), the gray levels are in the same scale.



(a) Golden angle scheme (9.2%) (b) $\pi(p_{\text{rad}})$-based scheme (10%)

(c) PSNR = 36.34 dB (d) PSNR = 38.99 dB

(e) Golden angle (f) $\pi(p_{\text{rad}})$-based scheme

Figure 3.11: Reconstruction examples. We plot schemes made of 37 lines based on the golden angle pattern (a), or based on our method with $\pi(p_{\text{rad}})$ (b). Drawing 37 lines in both cases leads to a cover of the sampling space by 9.2% in the case of the golden angle scheme, and by 10% for the $\pi(p_{\text{rad}})$-based scheme. Note that despite a difference of 0.8% in the covering of the k-space, the scanning time is the same for both schemes. In (c) and (d) we display the corresponding reconstructions via ℓ^1-minimization. We can see that we improve the reconstruction of more than 2 dB with our method. At the bottom, we show the corresponding absolute difference with the reference image. Note that the gray levels have the same scaling in (e) and (f).

5 Proof of the main results

5.1 Proof of Proposition 3.4

First, we express the Fenchel-Rockafellar dual problem [88]:

$$\begin{aligned}
&\min_{\pi\in\Delta_m} \|p - M\pi\|_{\ell^1} + \alpha\mathcal{E}(\pi) \\
&\quad = \min_{\pi\in\Delta_m} \max_{q\in B_\infty} \langle M\pi - p, q\rangle_{F^*\times F} + \alpha\mathcal{E}(\pi) \\
&\quad = \max_{q\in B_\infty} \min_{\pi\in\Delta_m} \langle M^*q, \pi\rangle_{E^*\times E} - \langle p, q\rangle_{F^*\times F} + \alpha\mathcal{E}(\pi) \\
&\quad = \max_{q\in B_\infty} -J_\alpha(q)
\end{aligned}$$

where B_∞ stands for the ℓ^∞-ball of unit radius and

$$J_\alpha(q) = -\min_{\pi\in\Delta_m} \langle M^*q, \pi\rangle_{E^*\times E} - \langle p, q\rangle_{F^*\times F} + \alpha\mathcal{E}(\pi). \tag{3.13}$$

The solution $\pi(q)$ of the minimization problem (3.13) satisfies

$$M^*q + \alpha\left(\log(\pi(q)) + 1\right) \in -\mathcal{N}_{\Delta_m}(\pi(q)) \quad \text{if} \quad \pi(q) \in \operatorname{ri}(\Delta_m), \tag{3.14}$$

where $\mathcal{N}_{\Delta_m}(\pi(q))$ denotes the normal cone to the set Δ_m at the point $\pi(q)$, and $\operatorname{ri}(\Delta_m)$ denotes the relative interior of Δ_m. Equation (3.14) can be rewritten in the following way

$$M^*q + \alpha\log(\pi(q)) = (-\lambda - \alpha)\mathbb{1}, \text{with} \quad \lambda \in \mathbb{R}^+ \quad \text{and} \quad \pi(q) \in \Delta_m. \tag{3.15}$$

By choosing $\lambda = \alpha\log\left(\sum_{k=1}^m \exp\left(-\frac{(M^*q)_k}{\alpha}\right)\right) - \alpha$ and plugging it into (3.15) we get that

$$\left(\pi(q)\right)_j = \frac{\exp\left(-\frac{(M^*q)_j}{\alpha}\right)}{\sum_{k=1}^m \exp\left(-\frac{(M^*q)_k}{\alpha}\right)}, \qquad \forall j \in \{1, \dots, m\}. \tag{3.16}$$

It remains to plug (3.16) in (3.13) to obtain (DP) with

$$J_\alpha(q) = \langle p, q\rangle_{F^*\times F} - \alpha\log\left(\sum_{k=1}^m \exp\left(-\frac{(M^*q)_k}{\alpha}\right)\right).$$

5.2 Proof of Proposition 3.6

The neg-entropy is continuous, and twice continuously differentiable on $\operatorname{ri}(\Delta_m)$. Then, in order to prove its strong convexity, it is sufficient to bound from below its positive diagonal Hessian with respect to $\|\cdot\|_E$. We have

$$\left\langle \mathcal{E}''(\pi)h, h\right\rangle = \sum_{i=1}^m \frac{(h_i)^2}{\pi_i}, \qquad \text{for} \quad \pi \in \operatorname{ri}(\Delta_m), \quad \text{and} \quad h \in \mathbb{R}^m. \tag{3.17}$$

Using Cauchy-Schwartz's inequality,

$$\begin{aligned}
\|h\|_{\ell^1}^2 &= \left(\sum_{i=1}^m \frac{|h_i|}{\sqrt{\pi_i}}\sqrt{\pi_i}\right)^2 \le \left(\sum_{i=1}^m \frac{h_i^2}{\pi_i}\right)\left(\sum_{i=1}^m \pi_i\right) \\
&\le \underbrace{\|\pi\|_{\ell^1}}_{=1} \left\langle \mathcal{E}''(\pi)h, h\right\rangle.
\end{aligned}$$

Therefore, $\mathcal{E}$ is 1-strongly convex on the simplex with respect to $\|.\|_{\ell^1}$. Since for all $p \in [1, \infty]$, $\|.\|_{\ell^1} \geq \|.\|_p$, we get:

$$\|h\|_{\ell^p}^2 \leq \left\langle \mathcal{E}''(\pi)h, h \right\rangle, \quad \forall h \in \mathbb{R}^m, \pi \in \mathrm{ri}\,(\Delta_m)\,.$$

Moreover if $(\pi_n)_{n\in\mathbb{N}}$ denotes a sequence of $\mathrm{ri}(\Delta_m)$ pointwise converging to the first element of the canonical basis e_1 and $h = e_1$, then

$$\lim_{n\to+\infty} \langle \mathcal{E}''(\pi_n)h, h\rangle = \|h\|_{\ell^p}^2 = 1$$

so that the inequality is tight.

5.3 Proof of Proposition 3.7

The proof is based on similar arguments as [77, Theorem 1]. First, notice that

$$\begin{aligned}
&\langle \nabla\mathcal{E}\left(\pi(q_2)\right) - \nabla\mathcal{E}\left(\pi(q_1)\right), \pi(q_2) - \pi(q_1)\rangle \\
&= \left\langle \int_{t=0}^{1} \mathcal{E}''(\pi_1 + t(\pi_2 - \pi_1))(\pi_2 - \pi_1)dt, \pi(q_2) - \pi(q_1) \right\rangle \\
&\geq \sigma_{\mathcal{E}}\left[\pi_1, \pi_2\right] \|\pi_2 - \pi_1\|_E^2, \qquad (3.18)
\end{aligned}$$

where

$$\sigma_{\mathcal{E}}\left[\pi_1, \pi_2\right] = \inf_{t\in[0,1]} \sigma_{\mathcal{E}}(t\pi_1 + (1-t)\pi_2)$$

is the local convexity modulus of $\mathcal{E}$ on the segment $[\pi(q_1), \pi(q_2)]$.

Next, recall that

$$J_\alpha(q) = \max_{\pi\in\Delta_m} \langle -M^*q, \pi\rangle_{E^*\times E} + \langle p, q\rangle_{F^*\times F} - \alpha\mathcal{E}(\pi).$$

The optimality conditions of the previous maximization problem for $J_\alpha(q_1)$ and $J_\alpha(q_1)$, $q_1, q_2 \in F$, lead to

$$\begin{aligned}
\langle -M^*q_1 - \alpha\nabla\mathcal{E}\left(\pi(q_1)\right), \pi(q_2) - \pi(q_1)\rangle \leq 0, \\
\langle -M^*q_2 - \alpha\nabla\mathcal{E}\left(\pi(q_2)\right), \pi(q_1) - \pi(q_2)\rangle \leq 0.
\end{aligned}$$

Combining the two previous inequalities, we can write that for $q_1, q_2 \in F$:

$$\begin{aligned}
&\langle M^*(q_1 - q_2), \pi(q_1) - \pi(q_2)\rangle \geq \alpha \left\langle \nabla\mathcal{E}\left(\pi(q_2)\right) - \nabla\mathcal{E}\left(\pi(q_1)\right), \pi(q_2) - \pi(q_1)\right\rangle, \\
&\implies \|M^*(q_1 - q_2)\|_{E^*} \|\pi(q_1) - \pi(q_2)\|_E \overset{(3.18)}{\geq} \alpha\sigma_{\mathcal{E}}\left[\pi(q_1), \pi(q_2)\right] \|\pi(q_1) - \pi(q_2)\|_E^2, \\
&\implies \|M^*\|_{F\to E^*} \|q_1 - q_2\|_F \|\pi(q_1) - \pi(q_2)\|_E \geq \alpha\sigma_{\mathcal{E}}\left[\pi(q_1), \pi(q_2)\right] \|\pi(q_1) - \pi(q_2)\|_E^2.
\end{aligned}$$

Therefore, we can write that

$$\|\pi(q_1) - \pi(q_2)\|_E \leq \frac{\|M^*\|_{F\to E^*} \|q_1 - q_2\|_F}{\alpha\sigma_{\mathcal{E}}\left[\pi(q_1), \pi(q_2)\right]}.$$

Noting that $\nabla J_\alpha(q) = -M\pi(q) + p$, we can conclude that

$$\begin{aligned}
\|\nabla J_\alpha(q_1) - \nabla J_\alpha(q_2)\|_{F^*} &= \|M\left(\pi(q_1) - \pi(q_2)\right)\|_{F^*} \\
&\leq \|M^*\|_{F\to E^*} \|\pi(q_1) - \pi(q_2)\|_E \\
&\leq \frac{\|M^*\|_{F\to E^*}^2}{\alpha\sigma_{\mathcal{E}}\left[\pi(q_1), \pi(q_2)\right]} \|q_1 - q_2\|_F.
\end{aligned}$$

Let us consider a sequence $(q_n)_{n\in\mathbb{N}}$ converging uniformly towards $q \in B_\infty \subset F$. Since $\pi : q \in B_\infty \longmapsto \pi(q) \in \Delta_m$ is a continuous mapping, $\pi_n := \pi(q_n)$ converges uniformly towards $\pi(q)$. Thus,

$$\sigma_{\mathcal{E}}\left[\pi(q_n), \pi(q)\right] \underset{n\to+\infty}{\longrightarrow} \sigma_{\mathcal{E}}(\pi(q)).$$

5.4 Proof of Theorem 3.10

Theorem 3.10 is a direct consequence of Lemma (3.15) below. A similar proof was proposed in [45], but not extended to a general setting.

Lemma 3.15. *Let $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ denote closed convex functions such that $A \cdot ri(dom(f)) \cap ri(dom(g)) \neq \emptyset$. Assume further that g is σ-strongly convex w.r.t. an arbitrary norm $\|\cdot\|$. Then*

1. *Function g^* satisfies $dom(g^*) = \mathbb{R}^n$ and it is differentiable on $\mathbb{R}^n$.*
2. *Denote*

$$p(x) = f(Ax) + g(x) \tag{3.19}$$

$$d(y) = -g^*(-A^*y) - f^*(y) \tag{3.20}$$

and

$$x(y) = \nabla g^*(-A^*y).$$

Let x^* denote the minimizer of (3.19) and y^* denote any minimizer of (3.20). Then for any $y \in \mathbb{R}^m$ we have

$$\|x(y) - x^*\|^2 \leq \frac{2}{\sigma}\left(d(y) - d(y^*)\right). \tag{3.21}$$

Proof. Point (i) is a standard result in convex analysis. See e.g. [57]. We did not find the result (ii) in standard textbooks and to our knowledge it is new. We assume for simplicity that g, g^*, f and f^* are differentiable. This hypothesis is not necessary and can be avoided at the expense of longer proofs. First note that

$$\inf_{x\in\mathbb{R}^n} p(x) = \sup_{y\in\mathbb{R}^m} -g^*(-A^*y) - f^*(y)$$

by Fenchel-Rockafellar duality. Since g is strongly convex ∇g is a one-to-one mapping and

$$\nabla g(\nabla g^*(x)) = x, \ \forall x \in \mathbb{R}^n. \tag{3.22}$$

The primal-dual relationships read

$$\begin{cases} A^*y^* + \nabla g(x^*) &= 0 \\ Ax^* - \nabla f^*(y^*) &= 0 \end{cases}$$

So that

$$\begin{aligned} x^* &= (\nabla g)^{-1}(-A^*y^*) \\ &= (\nabla g^*)(-A^*y^*). \end{aligned}$$

Let us define the following Bregman divergences quantities:

$$D_1(y) := f^*(y) - f^*(y^*) - \langle A\nabla g^*(-A^*y^*), y - y^*\rangle$$
$$D_2(y) := g^*(-A^*y) - g^*(-A^*y^*) + \langle A\nabla g^*(-A^*y^*), y - y^*\rangle.$$

By construction

$$D_1(y) + D_2(y) = d(y) - d(y^*).$$

Moreover since y^* is the minimizer of d it satisfies $A\nabla g^*(-A^*y^*) = \nabla f^*(y^*)$. By replacing this expression in D_1 and using the fact that f^* is convex we get that

$$D_1(y) \geq 0, \ \forall y \in \mathbb{R}^n.$$

Using identity (3.22) we get:

$$D_2(y) = g^*(\nabla g(x(y))) - g^*(\nabla g(x^*)) + \langle x^*, \nabla g(x^*) - \nabla g(x(y))\rangle. \tag{3.23}$$

Moroever, since (see e.g. [57])

$$g(x) + g^*(x^*) = \langle x, x^*\rangle \Leftrightarrow x^* = \nabla g(x),$$

we get that

$$g^*(\nabla g(x(y))) = \langle \nabla g(x(y)), x(y)\rangle - g(x(y)),$$

and

$$g^*(\nabla g(x^*)) = \langle \nabla g(x^*), x^*\rangle - g(x^*).$$

Replacing these expressions in (3.23) we obtain

$$\begin{aligned} D_2(y) &= g(x^*) - g(x(y)) + \langle \nabla g(x(y)), x(y) - x^*\rangle \\ &\geq \frac{\sigma}{2}\|x(y) - x^*\|^2 \end{aligned}$$

since g is σ strongly convex w.r.t $\|\cdot\|$. To sum up we have:

$$\begin{aligned} d(y) - d(y^*) &= D_1(y) + D_2(y) \\ &\geq D_2(y) \\ &\geq \frac{\sigma}{2}\|x(y) - x^*\|^2 \end{aligned}$$

which is the desired inequality. □

We now have all the ingredients to prove Theorem 3.10.

Proof of Theorem 3.10. The proof is a direct consequence of Lemma 3.15. It can be obtained by setting $A \equiv M$, $f(y) \equiv \|y - p\|_1$ and $g(x) \equiv \alpha\mathcal{E}(x) + \chi_{\Delta_m}(\pi)$, with χ_{Δ_m} the indicator function of the set Δ_m. Thus $p(x) = f(Ax) + g(x) = F_\alpha(x)$ and $d(y) = J_\alpha(y)$. Then remark that π_k defined in Theorem 3.10 satisfies $\pi_k = \nabla g^*(-A^*y_k)$. By Proposition 3.6, we get that g is $\alpha\sigma_{\mathcal{E}}$-strongly convex w.r.t. $\|\cdot\|_{\ell^p}$, for all $p \in [1;\infty]$. It then suffices to use bound (3.9) together with Lemma 3.15 to conclude. □

6 Conclusion

In this chapter, we have focused on constrained acquisition by blocks of measurements. Sampling schemes are constructed by drawing blocks of measurements from a given dictionary of blocks according to a probability distribution π that needs to be chosen in an appropriate way. We have presented a novel approach to compute π in order to imitate existing sampling schemes in CS that are based on the drawing of isolated measurements. For this purpose, we have defined a notion of dissimilarity measure between a probability distribution on a dictionary of blocks and a probability distribution on a set of isolated measurements. This setting leads to a convex minimization problem in high dimension. In order to compute a solution to this optimization problem, we proposed an efficient numerical approach based on the work of [77]. Our numerical study highlights the fact that performing minimization on a metric space rather than a Hilbert space might lead to significant acceleration. Finally, we have presented reconstruction results using this new approach in the case of MRI acquisition. Our method seems to provide better reconstruction results than standard strategies used in this domain. We believe that this last point brings interesting perspectives for 3D MRI reconstruction.

As an outlook, we plan to extend the proposed numerical method to a wider setting and to provide better theoretical guarantees for cases where the Lipschitz constant of the gradient may vary across the domain. A first step in this direction was proposed recently in [50]. We also plan to accelerate the matrix-vectors product involving M by using fast Radon transforms. This step is unavoidable to apply our algorithm in 3D or 3D+t problems for which we expect important benefits compared to the small images we tested until now. Finally we are currently collaborating with physicists at Neurospin, CEA and plan to implement the proposed sampling schemes on real MRI scanners.

4 Generating block-constrained sampling schemes via measure projection

Contents

In this chapter, we propose a new way to generate block-constrained sampling schemes, via measure projection. In the previous chapter, the strategy is to construct a projected distribution probability $\pi^{(\text{old})}$ of a target density p, with $\pi^{(\text{old})}$ supported on a blocks dictionary; then the sampling scheme is generated by drawing N i.i.d. blocks according to $\pi^{(\text{old})}$. In this chapter, we propose to directly construct a measure π supported on N blocks to best approximate the target density p. Then, we give an answer to the question: how to choose these N elements to best fit the target measure? The proposed approach is an heuristic presented in [29] inspired by halftoning techniques. Halftoning (or stippling) is the process that aims at approximating a gray-scale image by N black dots: the dots should be more concentrated in the dark regions of the image, see Figure 4.1(b). In this chapter, we will mainly focus on the application to MRI sampling.

This chapter is an introduction to an ongoing (and inchoate for now) work.

1 Introduction

In [28], the authors study the efficiency of a variable density sampler. They exhibit three main properties that characterize an efficient variable density sampler: (i) it must rapidly cover the space (this is related to the notion of *mixing time*), (ii) it must have a given limit empirical measure, (iii) it must belong to a class of admissible samplers, e.g. straight lines in MRI. In this chapter, we propose a new way to generate such efficient sampling schemes.

2 The principle

2.1 A measure projection formulation

Let p denote a target continuous measure defined on the set Ω. For the sake of clarity, fix $\Omega = [0,1]^d$. Let $(\mathcal{M}_N)_{N\in\mathbb{N}}$ denote a sequence of admissible measures sets. In the sequel, N may represent the number of admissible shapes in Ω supporting the admissible measures in $\mathcal{M}_N$. In order to project p onto $\mathcal{M}_N$, we need to define a distance between p and any measure μ_N lying in $\mathcal{M}_N$. The authors in [29] have proposed the following distance:

$$D_h(\mu_N, p) = \|h \star (\mu_N - p)\|_2 ,$$

with h a convolution kernel on Ω. The convolution product between h and μ is defined for all $x \in \Omega$ by:

$$\mu \star h(x) = \int_\Omega h(x-y) d\mu(y) = \mu\left(h(x-\cdot)\right).$$

The function D_h is a distance if the Fourier series of h is real and positive. Moreover, in that case, it metrizes the weak convergence: $\mu_N \underset{N\to\infty}{\rightharpoonup} p$ if and only if $D_h(\mu_N, p) \to 0$.

Then, the main idea proposed in this chapter is to define the sampling pattern as a solution of the following variational problem:

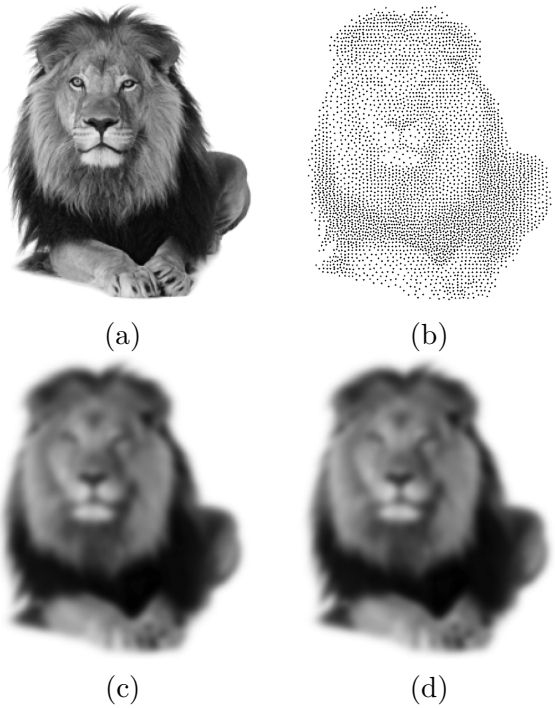
$$\min_{\mu\in\mathcal{M}_N} \frac{1}{2} \|h \star (\mu - p)\|_2^2 . \tag{4.1}$$

This problem yields the best approximation of p lying in the admissible space $\mathcal{M}_N$.

2.2 Why could this work?

In order to understand why the previous definition of D_h is founded, let us take a practical example. The human eye is able to recognize the image in Figure 4.1[1](a) from the image in Figure 4.1(b) even though the differences between both pixel values are large. This phenomenon can be explained by the multiresolution feature of the human visual system. Namely, the human eye associates Figure 4.1(a) to Figure 4.1(b), because blurred versions of both, Figures 4.1(c)(d), are similar. This explanation suggests a way to place dots in order to approximate a measure: it suffices to solve Problem (4.1) and to set $\mathcal{M}_N := \mathcal{M}(\Omega^N) = \left\{ \frac{1}{N}\sum_{i=1}^N \delta_{\omega_i}, \omega_i \in \Omega \right\}$, with δ_{ω_i} the Dirac mass supported on ω_i.

[1] the courtesy of Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn and Pierre Weiss.



(a) (b)

(c) (d)

Figure 4.1: In (a), an image of a lion corresponding to the target measure p. In (b), the computed dotted image by halftoning. Figures (c) and (d) are obtained by convolving (a) and (b) with a Gaussian of variance equal to 3 pixels. After convolution, the images cannot be distinguished.

2.3 Some theoretical guarantees on Problem 4.1

Let γ be the normalized Lebesgue measure on $[0,T]$. For a function $f : [0,T] \to \mathbb{R}^d$, the pushforward measure associated to f is denoted by $f_*\gamma$ and it is defined by

$$f_*\gamma(B) = \gamma(f^{-1}(B))$$

where B is any measurable set. Let $\mathcal{F}$ be a set of parametrizations $f : [0,T] \to \mathbb{R}^d$ and let $\mathcal{M}(\mathcal{F}) = \{\mu = f_*\gamma, f \in \mathcal{F}\}$ denote the set of pushforward measures associated to the set of parametrizations $\mathcal{F}$. In [29], the authors show that if $\mathcal{F}$ is compact for the topology of pointwise convergence, then there exists a minimizer to Problem 4.1 with $\mathcal{M}_N = \mathcal{M}(\mathcal{F})$. From this, they deduce a necessary and sufficient condition on $(\mathcal{M}_N)$ with $N \in \mathbb{N}$ to ensure consistency, i.e. $\mu_N^\star \underset{N\to\infty}{\rightharpoonup} p$. They also provide a result on the rate of convergence of $\mu_N^\star$ to p.

2.4 Choosing $\mathcal{M}_N$

2.4.1 MRI constraints

In this chapter we will focus on the application to MRI. In MRI, the acquisition consists of sensing some values of the Fourier transform $\widehat{u}$ of an image $u : \Omega \to \mathbb{R}$. These sampling locations are acquired along parametrized curves

$$\lambda : [0,T] \to \mathbb{R}^d,$$

that should verify the following feasibility conditions [31]

$$\Lambda_T = \{\lambda : [0,T] \to \mathbb{R}^d, \|\dot{\lambda}\|_\infty \leq \alpha, \|\ddot{\lambda}\|_\infty \leq \beta\} \tag{4.2}$$

where α and β depend on the imaging device. This condition ensures that the sampling trajectory is smooth and regular. Therefore, an MR acquisition gives access to the set of measurements

$$E = \left\{ \hat{u}\left(\lambda(i\Delta t)\right), 1 \leq i \leq \left\lfloor \frac{T}{\Delta t} \right\rfloor \right\},$$

where $\Delta t > 0$ denotes the sampling rate of the curve λ. Note that it is also possible to consider Cartesian products of the set Λ when multiple trajectories are used.

2.4.2 Considered projection sets for MR sampling

Let p be a given continuous target sampling density. It can be obtained by heuristics [69] or derived from CS theories [2, 28, 69]. Given the conditions (4.2) of the admissible pattern for MR sampling, we will consider various sets of constraints in the numerical experiments:

① The sums of N Dirac masses:

$$\mathcal{M}(\Omega^N) := \left\{ \mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{\omega_i}, \omega_i \in \Omega \right\}.$$

This case is not feasible on a real MR scan, but it is commonly used in simulations. Moreover, solving the projection problem is similar to the so-called Poisson Disc sampling [17, 68].

② The empirical measures of parametrized curves described in (4.2):

$$\mathcal{M}(\Lambda_T) := \{\mu = \lambda_* \gamma, \lambda \in \Lambda_T\}.$$

This case is feasible for MR acquisition and has been proposed in [29].

③ The empirical measures lying on a set of N segments with arbitrary length. Let

$$\mathcal{L} = \{\lambda : [0,1] \to \Omega, \exists (x_1, x_2) \in \Omega^2,\ \lambda(t) = (1-t)x_1 + t x_2 \quad \forall t \in [0,1]\}.$$

The admissible set is defined as:

$$\mathcal{M}(\mathcal{L}^N) := \left\{ \mu = \frac{1}{N} \sum_{i=1}^{N} (\lambda_i)_* \gamma, \lambda_i \in \mathcal{L} \right\}.$$

This setting is of particular interest in MRI, since a straight line is really easy to implement on real scanners.

3 Numerical resolution

3.1 Repulsion-Attraction form

In order to numerically solve Problem 4.1, we need to discretize it. It was shown in [29] that a measure set $\mathcal{M}_N$ can be approximated by a subset of n-point measures $\mathcal{N}_n \subseteq \mathcal{M}(\Omega^n)$ with an arbitrary precision ϵ in the sense that the Hausdorff distance:

$$\mathcal{H}_{\text{dist}}(\mathcal{N}_n, \mathcal{M}_N) = \max\left(\sup_{\pi \in \mathcal{N}_n} \inf_{\mu \in \mathcal{M}_N} \text{dist}(\mu, \pi), \sup_{\mu \in \mathcal{M}_N} \inf_{\pi \in \mathcal{N}_n} \text{dist}(\mu, \pi) \right) \leq \epsilon.$$

The set $\mathcal{N}_n$ may always be written as $\mathcal{N}_n = \{\mu = \frac{1}{n}\sum_{i=1}^n \delta_{q_i}, \quad \text{for } q = (q_i)_{1\leq i\leq n} \in \mathcal{Q}_n^\epsilon\}$, where the parameterization set $\mathcal{Q}_n^\epsilon$ depends on $\mathcal{M}_N$. The process of finding $\mathcal{Q}_n^\epsilon$ is non constructive, but one can show that it can be constructed explicitly for the parameterization sets defined in the previous paragraph [29].

This remark allows discretizing problem (4.1) by

$$\min_{\mu\in\mathcal{N}_n} \frac{1}{2}\|h\star(\mu-p)\|_2^2\,, \tag{4.3}$$

where $\mathcal{N}_n \subseteq \mathcal{M}(\Omega^n)$ is a suitable approximation of $\mathcal{M}_N$.

Then, by developing the L^2-norm in the definition of D_h, we may rewrite Problem 4.3 as follows

$$\min_{q\in\mathcal{Q}_n^\epsilon} \underbrace{\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n H(q_i-q_j)}_{\text{Repulsion potential}} - \underbrace{\sum_{i=1}^n\int_\Omega H(x-q_i)dp(x)}_{\text{Attraction potential}}, \tag{4.4}$$

where H is is defined in the Fourier domain by $\widehat{H}(\xi) = |\widehat{h}|^2(\xi)$ for all $\xi\in\mathbb{Z}^d$. It is a simple exercise to show that if H is real and even (meaning that $\widehat{H}(\xi)$ is real and positive for all $\xi\in\mathbb{Z}^d$), then Problem 4.1 and Problem 4.4 admit the same solution set. Define the objective J by

$$J(q) = \underbrace{\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n H(q_i-q_j)}_{J_1(q)} - \underbrace{\sum_{i=1}^n\int_\Omega H(x-q_i)dp(x)}_{J_2(q)}\,.$$

We can distinguish two terms in the expression of J:

- the first one J_1 is a repulsion potential, tending to maximize the distance between all pairs of points. It then ensures no cluster and so, a good cover of the space. This is in the same flavor as the Poisson disk sampling [101], known to give good practical results.
- the second one J_2 is the attraction potential, attracting q in the regions of high density of p, and ensuring that the solution of Problem (4.4) will be an approximation of p.

In practice, H is chosen to be $\|\cdot\|_{L^2(\Omega)}$. This choice ensures rotation and scale invariance.

3.2 The algorithm

Problem (4.4) can be shown to be a generalization of Thompson's problem (optimal repartition of electrons on a sphere) [96]. It belongs to Smale's list of problems to solve for the XXIst century. A global optimization of J is therefore out of reach for high-dimensional problems (in our experiments, n varies from 1000 to 200000). Therefore, we only present an algorithm converging to a critical point of Problem 4.3. Surprisingly, the critical points of J seem to always provide satisfactory practical results. We denote $P_{\mathcal{Q}_n^\epsilon}$ the projector onto the set $\mathcal{Q}_n^\epsilon$. Since $\mathcal{Q}_n^\epsilon$ is nonconvex, this projector may be a point-to-set mapping. In Algorithm 2, the symbol ∂J denotes the Clarke-subdifferential of J. This algorithm is proposed in [29] and it is based on the ideas of [5]. It is remarkable to see that

Algorithm 2 Resolution scheme for Problem 4.3[29]

Input :

- p : a target measure
- n : the number of discretization points
- L : the number of segments
- $q^{(0)}$: an initial segments-shaped measure
- nit : number of iterations

Output :

- $q^{(\text{nit})}$: an approximation of a curve
- $\mu^{(\text{nit})} = q_*^{(\text{nit})}\gamma$: an approximation of measure in $\widetilde{\mathcal{M}}_N$

for $k = 0 \dots \text{nit}$ **do**

  Compute $\eta^{(k)} = \eta$, $\eta \in \partial J(q^{(k)})$

  Set $q^{(k+1)} \in P_{\mathcal{Q}_n^\epsilon}\left(q^{(k)} - \tau\eta^{(k)}\right)$.

end for

very mild conditions on $\mathcal{Q}_n^\epsilon$ (e.g. semi-algebraicity) are sufficient to ensure convergence to critical points.

Algorithm 2 presents two critical steps: (i) the computation of an element of the sub-differential of J necessitates $O(n^2)$ operations by direct calculations, with n the number of discrete points used to represent q and p. In practice, we use fast summation methods [82, 61] (relying on nfft) introduced for particles systems. See [95] for an efficient computation of the gradient of J_1; (ii) the step of projection on the set $\mathcal{Q}_n^\epsilon$. The latter can have an explicit expression or can be solved with iterative methods depending on the geometrical structure of $\mathcal{Q}_n^\epsilon$.

An important note. The resolution of Problem (4.4) is very involved in terms of numerical analysis and computer science. Many subtle tricks have to be used and traps have to be avoided. We also used parallel programming (Open-MP) quite a lot to manage performing large scale computations.

3.3 Discretization

In this paragraph, we briefly describe the discretization schemes used for the specific sets defined in paragraph 2.4.2.

Case (1) The discretization of this projection problem is straightforward.

Case (2) We adapt the previous problem for discrete parameterized curves q [29]. For a discrete curve $q \in \mathbb{R}^{2\times n}$ of the 2D space, let q_i denote its position at time $i\Delta t$, with $\Delta t = T/n$. Let D_1 denote the discretized version of the first-order differential operator such that

$$(D_1 q)_i = \begin{cases} 0 & \text{if} \quad i = 1 \\ q_i - q_{i-1} & \text{otherwise.} \end{cases}$$

Similarly we let D_j denote the j-th order differential operator. In the numerical experiment we set $D_2 = -D_1^* D_1$. The set of discretized curves can be thus written as follows [29]

$$P_n^{2,\infty} = \{q = (q_i)_{1\leq i\leq n}, \quad \text{s.t.} \quad q_i \in \Omega, \forall 1 \leq i \leq n, \text{ and } \|D_1 q\|_\infty \leq \alpha, \text{ and } \|D_2 q\|_\infty \leq \beta\} \tag{4.5}$$

In order to have degree of freedom for the values of q_1 and $(D_1 q)_1$ corresponding to the position and speed at time 0 in the acquisition space, we add an affine constraints set $\mathcal{A}$ such that

$$\mathcal{A} = \{q \in \mathbb{R}^{nd}, A(q) = a\},$$

where A is a linear mapping from $\mathbb{R}^{nd}$ to $\mathbb{R}^c$, with c the number of constraints, and $a \in \mathbb{R}^c$ is a parameter vector. To compute the projection on $P_n^{2,\infty} \cap \mathcal{A}$, we use accelerated proximal gradient descents on the dual problem. Indeed, the dual problem of the projection problem

$$\min_{q\in\mathcal{A}\cap P_n^{2,\infty}} \frac{1}{2}\|q - q_0\|_2^2, \tag{4.6}$$

can be written as follows

$$\sup_{\rho_1,\rho_2\in\mathbb{R}^{nd}} F(\rho_1, \rho_2) - \alpha\|\rho_1\|_* - \beta\|\rho_2\|_*, \tag{4.7}$$

with the dual norm $\|\rho\|_* = \sup_{\|q\|_\infty\leq 1} \langle q, \rho\rangle$ and F is defined as follows

$$F(\rho_1, \rho_2) = \min_{q\in\mathcal{A}} \langle D_1 q, \rho_1\rangle + \langle D_2 q, \rho_2\rangle + \frac{1}{2}\|q - q_0\|_2^2. \tag{4.8}$$

The proof of this result can be found in [31]. It can be derived that the minimizer

$$q^\star(\rho_1^\star, \rho_2^\star) = \arg\min_{q\in\mathcal{A}} \langle D_1 q, \rho_1^\star\rangle + \langle D_2 q, \rho_2^\star\rangle + \frac{1}{2}\|q - q_0\|_2^2$$

is given by

$$q^\star(\rho_1^\star, \rho_2^\star) = z + A^+(v - Az).$$

with $z = q_0 - D_1^*\rho_1 - D_2^*\rho_2$ and $A^+ = A^*(A^*A)^{-1}$ the pseudo-inverse of A. Therefore, the dual problem (4.7) consists in a sum of the differentiable convex function $(-F)$ and G, such that $G(\rho_1, \rho_2) = \alpha\|\rho_1\|_* + \beta\|\rho_2\|_*$. Moreover, the gradient of $(-F)$ is Lipschitz continuous of constant $L = |||D_1^*D_1 + D_2^*D_2|||$, with $|||\cdot|||$ the spectral norm. This sum can be minimized by using accelerated proximal gradient descent [75]. Algorithm 3 presents the numerical scheme for the optimization of the dual problem.

Case ③ In this case, the admissible set is $\mathcal{M}\left(\mathcal{L}^N\right)$ meaning that the resulting measure is supported on N segments of variable length. Supposing that each segment is discretized in k points, the discretized version of $\mathcal{M}\left(\mathcal{L}^N\right)$ can be written as follows

$$P_{L_N}^k = \left\{q \in \Omega^{kN},\ q_j = q_i + \frac{j-i-1}{k-1}\left(q_{i+k} - q_i\right), \text{ for } i = 1 : k : kN \text{ and } i \leq j \leq i+k\right\},$$

where $1 : k : kN$ denotes the set $\{1, k+1, 2k+1, \ldots, (N-1)k+1\}$.

The projection on $P_{L_N}^k$ can be explicitly computed, as presented in Algorithm 4. For the sake of clarity, we present the projection on $P_{L_1}^k$ in Algorithm 4. Note that the projection onto $P_{L_N}^k \cap \Omega^{kN}$ can be computed iteratively using the alternating projection method (POCS) [8, 79].

Algorithm 3 Projection on $P_n^{2,\infty} \cap \mathcal{A}$ [31]

Input :

- q_0 : a target discretized curve
- α, β : the parameters for $P_n^{2,\infty}$
- nit : number of iterations
- $\rho^{(0)} = (\rho_1^{(0)}, \rho_2^{(0)})$: initialization

Output :

- $q^{(\text{nit})}$: an approximation of the projection of q_0 on $P_n^{2,\infty} \cap \mathcal{A}$

for $k = 1 \dots \text{nit}$ **do**

  Compute $\rho^{(k)} = \text{prox}_{\frac{1}{L}G}\left(y^{(k-1)} - \frac{1}{L}\nabla(-F)(y^{(k-1)})\right)$

  Compute $y^{(k)} = \rho^{(k)} + \frac{k-1}{k+2}(\rho^{(k)} - \rho^{(k-1)})$

end for

Return : $q^{(\text{nit})} = q^\star(\rho_1^{(\text{nit})}, \rho_1^{(\text{nit})})$

Algorithm 4 Projection on $P_{L_1}^k$

Input :

- u : a vector of k points

Output :

- q : a vector of $P_{L_1}^k$

Compute $C = k(k^2 - 3k + 2)/(6(k-1)^2)$

Compute $D = k(2k^2 - 3k + 1)/(6(k-1)^2)$

Compute $x_i^{(1)} = (k-i)u_i$ for $1 \le i \le k$

Compute $x_i^{(2)} = (i-1)u_i$ for $1 \le i \le k$

Compute $s^{(1)} = \sum_{i=1}^k x_i^{(1)}$

Compute $s^{(2)} = \sum_{i=1}^k x_i^{(2)}$

Evaluation of the end points

- $q_k = C/(C^2 - D^2)\left(s^{(1)} - D/Cs^{(2)}\right)$
- $q_1 = 1/C(s^{(2)} - Dq_k)$

Place $(q_i)_{2 \le i \le k-1}$ uniformly distributed on $[q_1, q_k]$

4 Results for MRI

To evaluate the relevancy of this approach, we test it on the reconstruction of a MR phantom[2] of size 1024×1024.

4.1 Resulting sampling schemes

The target density p is chosen to have a radial decay in the frequency domain, as [64] suggests. The considered projection sets are those described in Section 2.4.2 ①, ②, ③, the last one with a number of segments equal to 200 segments. The generated sampling schemes by Algorithm 2 are computed using a discretization of 50000 samples, meaning that the subsampling ratio for these sampling schemes is of 4.7% for an image 1024×1024. In Figure 4.2, we present the resulting schemes for the projection sets ①, ② and ③. We also add a typical CS sampling scheme obtained by drawing 50000 i.i.d. points according to p in Figure 4.2(a'), in order to make the comparison with Figure 4.2(a) obtained with the projection set ①.

It should be noted that the sampling schemes in Figure 4.2(a)(a') are not feasible on real scanners. However, they can be relevant to quantify the reconstruction quality loss when the acquisition is structured by blocks of measurements.

4.2 Reconstructions

In Figure 4.3, we present image reconstructions based on the sampling schemes displayed in Figure 4.2. In Figure 4.3(a'), we show the reconstruction obtained by the sampling scheme of Figure 4.2(a') with a SNR= 19.6 dB. It will be used as a reference for the reconstruction even if the corresponding sampling scheme is not implementable in practice.

The sampling based a repulsion of isolated measurements in Figure 4.2(a) slightly improves the reconstruction in Figure 4.3(a) compared to Figure 4.3(a'): the increase of 0.1 dB in the SNR does not represent a visual significant improvement.

The sampling scheme based on the projection set ③ in Figure 4.2(c) has the advantage to be implementable on real scanners, and leads to a reconstruction in Figure 4.3(c) with a SNR= 18.9 dB. The structure in the acquisition implies a SNR loss of 0.7 dB, which is reasonable.

However, the sampling scheme based on the projection set ② in Figure 4.2(b) leads to a considerable loss of SNR, even if it remains implementable on real MR scanners. This scheme is completely realistic in the sense that the parameters α and β in Equation (4.2) correspond to the parameters of scanners available in the Neurospin project at CEA, Saclay. The reason for failure is due to the fact that α is low imposing consecutive samples to be close to each other. We are currently testing the methods with longer scanning times. Those schemes are being implemented on the world largest pre-clinical scanners (Bruker) in Neurospin at 7T and 17T.

4.3 Concluding remarks

From the previous numerical experiments, the following conclusions can be drawn:

[2] available online at http://bigwww.epfl.ch/algorithms/mriphantom/

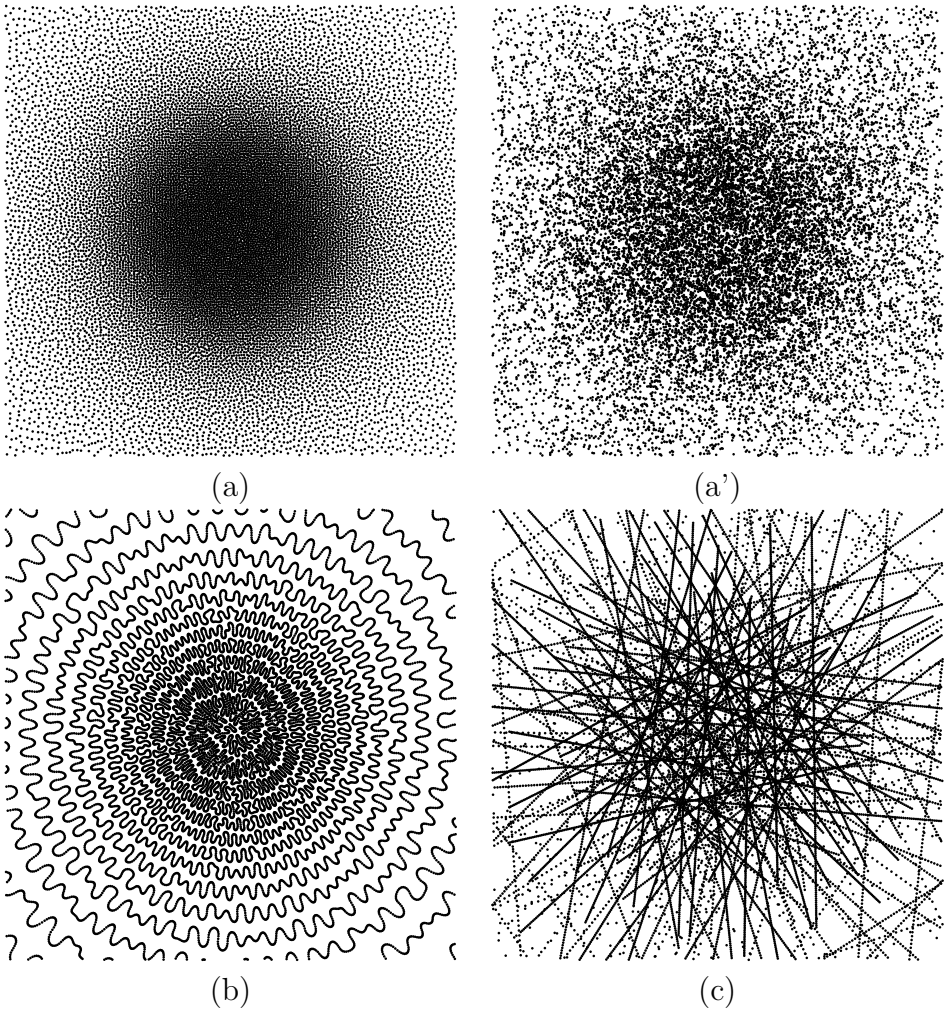


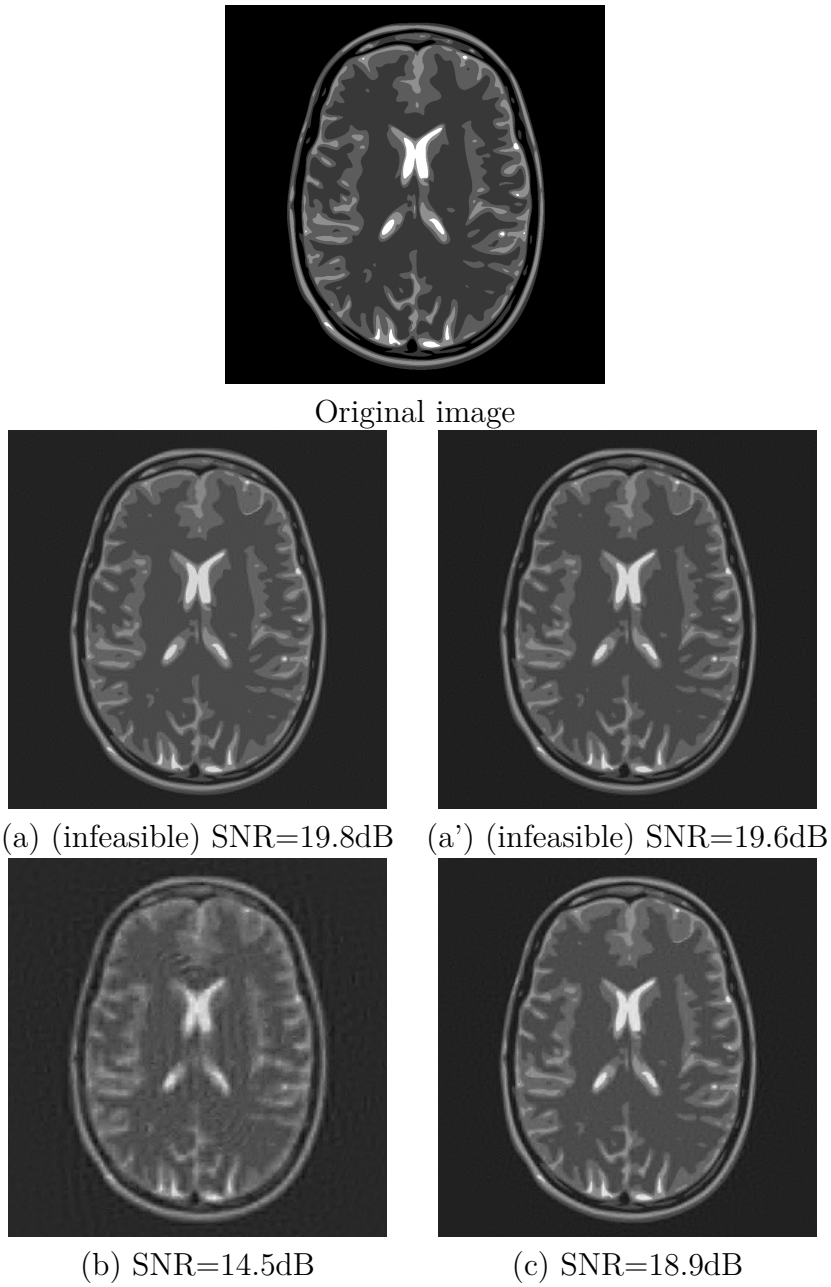
Figure 4.2: Resulting sampling schemes based on ($n =$)50000 samples (4.7% of the total number of measurements). (a) projection on sum of Dirac masses ① (infeasible). (a') sampling scheme based on i.i.d. drawing for comparison. (b) projection on admissible curves for MRI -corresponding to the set $P_n^{2,\infty}$ ②. (c) projection on 200 segments of arbitrary length ③.

(i) Adding repulsion between isolated points such as in Figure 4.2(a) marginally improves practical recovery compared to i.i.d. drawings such as in Figure 4.2(a'). This result questions the practical efficiency of sampling techniques based on Poisson Disc sampling [17].

(ii) The computed sampling schemes obtained by Algorithm 2 with a target radial density are very close to the sampling schemes used on real machine. Indeed, the projection set ② leads to a noisy spiral which has already been proposed in MRI setting by [72].

(iii) The novel technique of generating sampling scheme presented in this chapter also allows to discover original patterns such as those in Figure 4.2(c) based on the projection set ③.

(iv) Adding structure in the acquisition seems to reduce reconstruction quality, but the loss can be reasonable (~ 1 dB).



Original image

(a) (infeasible) SNR=19.8dB (a') (infeasible) SNR=19.6dB

(b) SNR=14.5dB (c) SNR=18.9dB

Figure 4.3: Reconstructed images from respectively the sampling schemes presented in Figure 4.2. At the top, we present the original image of size 1024×1024. In (a)(a')(b)(c), we respectively present the reconstructions obtained by the sampling schemes in Figure 4.2(a)(a')(b)(c)

.

Résumés

A Résumé des chapitres 1 et 2

L'échantillonnage compressé (CS) a eu un fort impact dans des domaines variés. Cependant, les schémas d'échantillonnage proposés par le CS consistent en des mesures isolées, et sont donc souvent incompatibles avec la physique d'acquisition. Afin de prendre en compte les contraintes d'acquisition, nous introduisons la notion de bloc de mesures, qui est un ensemble de mesures pouvant représenter une forme arbitraire, par exemple une ligne droite. L'acquisition structurée par blocs de mesures est utilisée en pratique et donne de bons résultats de reconstruction. Afin de combler le fossé entre les théories CS et les acquisitions faites en pratiques, nous donnons des résultats théoriques de reconstructions CS avec acquisition de blocs de mesures pour un vecteur $x \in \mathbb{C}^n$ de support S fixé - et pas seulement s-parcimonieux. Nous proposons ainsi une nouvelle classe de matrices d'échantillonnage, compatibles avec de nombreux domaines d'applications.

1 Introduction

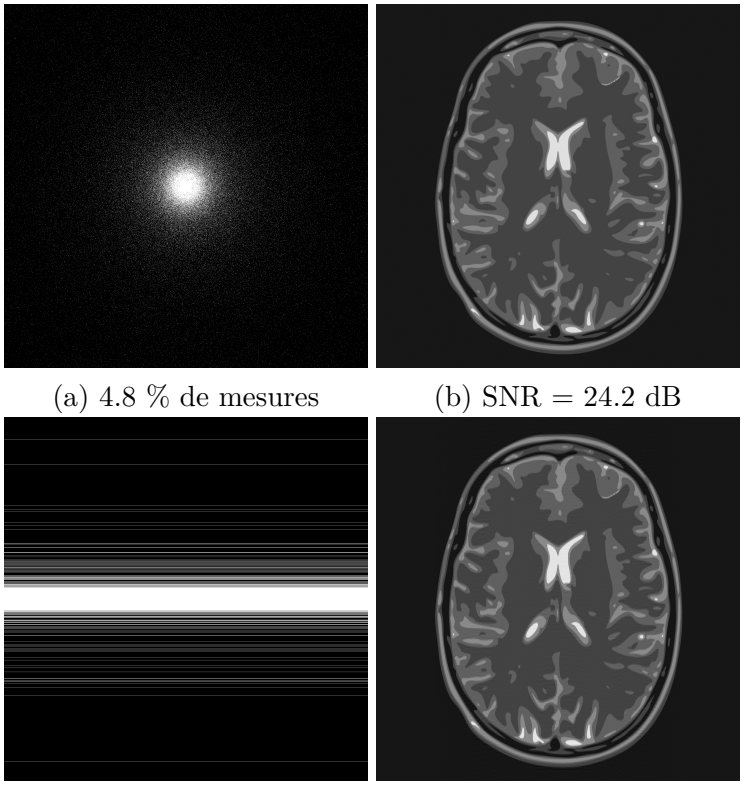
Les résultats typiques d'échantillonnage compressé (CS) garantissent que si $x \in \mathbb{C}^n$ est un vecteur s-parcimonieux, si la matrice d'échantillonnage $A \in \mathbb{C}^{m \times n}$ est incohérente, alors x peut être reconstruit exactement via

$$\min_{x \in \mathbb{C}^n, Ax=y} \|x\|_1. \tag{A.1}$$

avec seulement $m \gtrsim s \ln(n)$ mesures [20]. Malgré des résultats récents, les théories existantes peinent à justifier le succès du CS dans les applications pratiques. Dans ce travail, nous cherchons à étendre l'applicabilité du CS en combinant deux nouveaux ingrédients : la parcimonie structurée et l'acquisition structurée.

Les matrices d'échantillonnage principalement étudiées par le CS sont (i) les matrices aléatoires à entrées i.i.d. (ii) les transformées orthogonales sous-échantillonnées aléatoirement. Les premières ont l'avantage d'être quasi-optimales [24], mais sont rares dans les domaines d'application. Les deuxièmes ont déjà plus d'intérêt d'un point de vue applicatif, puisque une transformée orthogonale (déterministe) peut modéliser la physique d'acquisition du système d'imagerie. Et même si elles ne vérifient pas toujours la propriété de RIP, elles peuvent apporter de bonnes garanties de reconstructions si leur cohérence est faible [20]. Cependant, ces catégories de matrices restent peu utilisables en pratique. Prenons l'exemple de l'Imagerie par Résonance Magnétique (IRM). En IRM, les données sont acquises dans l'espace de Fourier, et les images à reconstruire sont considérées parcimonieuses dans le domaine des ondelettes. Des théories standards de CS, peuvent

justifier des stratégies d'échantillonnage à densité variable, voir Figure A.1(a). Les points blancs dans la Figure A.1(a) représentent les mesures prises dans le domaine de Fourier. Seulement 4,6% des échantillons permettent de reconstruire une image de bonne qualité. Cependant, l'acquisition de mesures isolées tirées aléatoirement en Fourier n'est pas envisageable en IRM : les mesures doivent rester le long de trajectoires continues et régulières. Par exemple, une stratégie très usitée en IRM est d'acquérir des lignes entières de coefficients de Fourier, voir Figure A.1(c), et les reconstructions obtenues sont de bonne qualité. A notre connaissance, aucune théorie n'est capable d'expliquer ces résultats. Les seuls travaux considérant des mesures très structurées tels que ceux présentés dans le chapitre 1 manquent d'expliquer leurs bons résultats pratiques car aucune hypothèse n'est faite sur la structure du support. Nous montrons dans le chapitre 2 que la structure de parcimonie joue un rôle crucial pour obtenir des garanties de reconstructions quand l'acquisition est structurée.



(a) 4.8 % de mesures (b) SNR = 24.2 dB

(c) 13 % de mesures (d) SNR = 24.1 dB

Figure A.1: Exemple de reconstruction d'une image 2048×2048 à partir de (a) mesures isolées et (c) de blocs de mesures. Les reconstructions correspondantes via minimisation ℓ_1 sont présentées en (b) et (d).

2 Principaux résultats

2.1 Cadre général

Soit $A_0 \in \mathbb{C}^{n \times n}$ une matrice orthogonale représentant l'ensemble des projections linéaires possibles. Les vecteurs-lignes $(a_i^*)_{1 \leq i \leq n} \in \mathbb{C}^n$ de A_0 sont partitionnés en blocs $(B_k)_{1 \leq k \leq M}$,

tels que

$$B_k = (a_i^*)_{i\in\mathcal{I}_k} \in \mathbb{C}^{|\mathcal{I}_k|\times n} \quad \text{avec} \quad \sqcup_{k=1}^M \mathcal{I}_k = \{1,\dots,n\}.$$

La matrice d'échantillonnage A est alors construite en tirant des blocs aléatoirement

$$A = \frac{1}{\sqrt{m}} \left(\frac{1}{\sqrt{\pi_{K_\ell}}} B_{K_\ell} \right)_{1\le\ell\le m}, \tag{A.2}$$

où $(K_\ell)_{1\le\ell\le m}$ sont des copies i.i.d. de la variable aléatoire K telle que

$$\mathbb{P}(K=k) = \pi_k \quad \text{pour tout } 1\le k\le M$$

On notera que $\mathbb{E}\left(\frac{B_K^* B_K}{\pi_K}\right) = \mathrm{Id}$.

2.2 De nouvelles quantités d'intérêt

La très grande majorité des papiers de CS utilisent la notion de cohérence ou les propriétés de Restricted Isometry Property (RIP) pour obtenir des résultats de reconstruction exacte. Ces quantités sont le bon outil lorsqu'on s'intéresse à la reconstruction d'un vecteur s-parcimonieux de support arbitraire. Un résultat important du chapitre 2 est de proposer de nouvelles quantités utiles lorsqu'on s'intéresse à la reconstruction d'un vecteur S-parcimonieux dont le support est S.

Définition A.1. *Soient le dictionnaire de blocs $(B_k)_{1\le k\le M}$, $S\subset\{1,\dots,n\}$ et π une loi de probabilité sur $\{1,\dots,M\}$. On pose*

$$\Theta(S,\pi) := \max_{1\le k\le M} \frac{1}{\pi_k} \|B_k^* B_{k,S}\|_{\infty\to\infty} \tag{A.3}$$

$$\Upsilon(S,\pi) := \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^M \frac{1}{\pi_k} \left|e_i^* B_k^* B_{k,S} v\right|^2, \tag{A.4}$$

$$\Gamma(S,\pi) := \max\left(\Upsilon(S,\pi), \Theta(S,\pi)\right), \tag{A.5}$$

où $\|B\|_{\infty\to\infty} = \sup_{\|v\|_\infty\le 1} \|Bv\|_\infty$, et B_S est la matrice extraite de B dont les colonnes sont indéxées par S.

2.3 Garanties de reconstruction exacte

Le théorème suivant apporte des garanties de reconstruction quand l'acquisition est structurée par blocs.

Théorème 1. *Soit $S\subset\{1,\dots,n\}$ un ensemble d'indices $s\ge 16$ et supposons que $x\in\mathbb{C}^n$ est un vecteur s-parcimonieux de support S. Fixons $\varepsilon\in(0,1)$. Soit A la matrice d'échantillonnage définie en (A.2). Si le nombre de blocs*

$$m \ge 73\cdot\Gamma(S,\pi)\ln(64s)\left(\ln\left(\frac{27n}{\varepsilon}\right) + \ln\ln(55s)\right), \tag{A.6}$$

alors x est l'unique solution du problème de minimisation (A.1) avec probabilité supérieure à $1-\varepsilon$.

On notera que la borne sur m donnée dans le Théorème 1 dépend explicitement du support S et de la probabilité de tirage π. Illustrons dès à présent l'intérêt du Théorème 1 dans des cas concrets.

3 Applications

3.1 Mesures isolées et signaux s-parcimonieux

Dans un premier temps, nous pouvons supposer que les blocs ne sont constitués que d'un seul vecteur ligne, ce qui revient aux acquisitions CS standards de mesures isolées. Ceci revient à choisir $B_k = a_k^*$ pour $1 \leq k \leq n$ avec $M = n$, où a_k^* sont les lignes de la matrice orthogonale A_0. La matrice de mesure peut alors s'écrire comme suit

$$A = \frac{1}{\sqrt{m}} \left(\frac{1}{\sqrt{\pi_{K_\ell}}} a_{K_\ell}^* \right)_{1 \leq \ell \leq m}, \tag{A.7}$$

où $(K_\ell)_{1 \leq \ell \leq m}$ sont des copies i.i.d. de K vérifiant $\mathbb{P}(K = k) = \pi_k$, pour $1 \leq k \leq n$.

Appliquons le Theorème 1 quand seule la parcimonie s du signal est connue.

Corollaire A.2. *Soit $x \in \mathbb{C}^n$ un vecteur s-parcimonieux. Supposons que A est la matrice d'échantillonnage définie en (A.7). Si le nombre de mesures vérifie*

$$m \geq C \cdot s \cdot \max_{1 \leq k \leq n} \frac{\|a_k\|_\infty^2}{\pi_k} \ln(s) \ln\left(\frac{n}{\varepsilon}\right), \tag{A.8}$$

alors x est l'unique solution du problème de minimisation (A.1) avec probabilité supérieure à $1 - \varepsilon$.

La borne (A.8) est similaire à celle de [20], à une constante multiplicative et un terme logarithmique près.

3.2 Mesures isolées et parcimonie structurée

Dans les applications, la matrice de mesure peut être partiellement cohérente. Le corollaire précédent devient insuffisant à expliquer le succès du CS. Il est alors nécessaire de prendre en compte des informations supplémentaires sur le support S du signal.

Illustrons le Théorème 1 sur un exemple pratique. Soit $A_0 = \mathcal{F}\phi^*$, où $\mathcal{F} \in \mathbb{C}^{n \times n}$ est la tranformée de Fourier 1d, et $\phi^* \in \mathbb{C}^{n \times n}$ est la transformée inverse en ondelettes 1d. On décompose le signal en ondelettes au niveau maximal $J = \log_2(n) - 1$. On définit la partition dyadique $(\Omega_j)_{0 \leq j \leq J}$ de l'ensemble $\{1, \ldots, n\}$, i.e. $\Omega_0 = \{1\}$, $\Omega_1 = \{2\}$, $\Omega_3 = \{3, 4\}$, $\ldots$, $\Omega_J = \{n/2 + 1, \ldots, n\}$. On définit également :
$j : \{1, \ldots, n\} \to \{0, \ldots, J\}$ par $j(u) = j$ si $u \in \Omega_j$.

Corollaire A.3. *Soit $S \subset \{1, \ldots, n\}$ tel que $|S \cap \Omega_j| = s_j$ pour $0 \leq j \leq J$. Soit $x \in \mathbb{C}^n$ un vecteur supporté par S. La matrice de mesure A est définie en (A.7) avec A_0 la transformée Fourier-Haar. On choisit π_k constant par niveau, i.e. $\pi_k = \tilde{\pi}_{j(k)}$. Si*

$$m \geq C \max_{0 \leq j \leq J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p \ln(s) \ln\left(\frac{n}{\varepsilon}\right) \tag{A.9}$$

alors x est l'unique solution du problème de minimisation (A.1) avec probabilité supérieure à $1 - \varepsilon$.

En particulier, la distribution minimisant (A.9) est

$$\tilde{\pi}_j = \frac{2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p}{\sum_{\ell=1}^{n} 2^{-j(\ell)} \sum_{p=0}^{J} 2^{-|j(\ell)-p|/2} s_p},$$

ce qui mène à

$$m \geq \sum_{j=0}^{J} \left(s_j + \sum_{\substack{p=0 \\ p \neq j}}^{J} 2^{-|j-p|/2} s_p \right) \ln(s) \ln\left(\frac{n}{\varepsilon}\right). \tag{A.10}$$

Notons que ce résultat est très similaire aux travaux récents de [2]. Ce corollaire peut être interprété de la façon suivante : le nombre de mesures au niveau j doit dépendre essentiellement du degré de parcimonie restreinte à ce niveau ; la parcimonie des autres niveaux vient également interférer mais son influence est d'autant plus atténuée que la distance entre les niveaux croît. Notons aussi l'intérêt d'un théorème dépendant du support S : si on a un a priori sur le signal (ici une parcimonie structurée par sous-bandes), on peut exploiter cette information pour i) dériver des densités de tirage optimales et ii) étendre l'applicabilité du CS dans des cas non-couverts par la théorie actuelle.

3.3 Acquisition structurée et parcimonie structurée

Dans cette section, nous nous intéressons au cas où les blocs ne sont plus de simples vecteurs lignes de A_0 mais un ensemble de lignes de A_0.

3.3.1 Les limites de l'acquisition structurée

Dans le chapitre 1, nous donnons des résultats théoriques de CS lorsque l'acquisition est structurée par blocs. En particulier, nous mettons en exergue une des principales limites du CS lorsque l'on échantillonne des lignes horizontales de transformées tensorielles. Ce cadre est notamment très utilisé en imagerie, cf [69].

Proposition A.4 ([11]). *Soit $A_0 = \phi \otimes \phi \in \mathbb{C}^{n \times n}$ une transformée 2D séparable, où $\phi \in \mathbb{C}^{\sqrt{n} \times \sqrt{n}}$ est une transformée orthogonale. Les blocs de mesures consistent en $\sqrt{n}$ lignes horizontales de l'espace d'acquisition 2D, i.e. pour $1 \leq k \leq \sqrt{n}$*

$$B_k = \left(\phi_{k,1}\phi, \ldots, \phi_{k,\sqrt{n}}\phi \right).$$

Si le nombre m de blocs mesurés est inférieur à $\min(2s, \sqrt{n})$, alors il n'existe pas de décodeur Δ tel que $\Delta(Ax) = x$ pour tout vecteur $x \in \mathbb{C}^n$ s-parcimonieux. Autrement dit, le nombre minimal m de blocs distincts requis pour identifier tout vecteur s-parcimonieux est nécessairement plus grand que $\min(2s, \sqrt{n})$.

Ce théorème semble indiquer que pour reconstruire un vecteur s-parcimonieux, il faut tirer s blocs de $\sqrt{n}$ mesures, soit un total de $s\sqrt{n}$ mesures isolées. A première vue, cette borne théorique semble entrer en contradiction avec les résultats observés en pratique comme ceux de la Figure A.1(d) ou comme la stratégie proposée dans [69]. La subtilité est que si on ne peut pas reconstruire *tous* les vecteur s-parcimonieux, on peut en reconstruire une sous-classe importante pour les applications. Par la suite, nous montrons comment le Théorème 1 peut nous affranchir de telles limites.

3.3.2 Passer outre ces limites

Dans ce paragraphe, nous montrons l'importance d'hypothèses supplémentaires telles que la parcimonie structurée pour expliquer le succès du CS dans les applications.

Corollaire A.5. *Soit $A_0 \in \mathbb{C}^{n\times n}$ la transformée de Fourier 2D. Soit x un signal qui peut être représenté en 2D, et tel que son support S est concentré sur seulement q lignes horizontales de l'espace. En choisissant une stratégie de tirage uniforme entre les lignes, i.e. $\pi_k^\star = 1/\sqrt{n}$ pour $1 \leq k \leq \sqrt{n}$, le nombre m de lignes horizontales suffisant pour reconstruire x avec probabilité $1-\varepsilon$ est*

$$m \geq C \cdot q \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

Le nombre m de lignes suffisant pour la reconstruction exacte de x n'est plus de l'ordre de s comme le prédisait la Proposition A.4, mais de l'ordre du nombre de lignes horizontales constituant le support !

3.3.3 Conséquences pour l'échantillonnage en IRM

Nous nous intéressons maintenant à un cas pratique, celui de l'échantillonnage en IRM. On considère une matrice $A_0 \in \mathbb{C}^{n\times n}$, produit de la transformée de Fourier 2D $\mathcal{F}_{2D}$ avec la transformée 2D inverse en ondelettes Φ^*. On souhaite reconstruire un vecteur $x \in \mathbb{C}^n$ (qui peut être considéré comme une transformée en ondelettes de taille $\sqrt{n}\times\sqrt{n}$). Soit $J = \log_2(\sqrt{n}) - 1$ et posons $(\tau_j)_{0\leq j\leq J}$ la partition dyadique de l'ensemble $\{1,\dots,\sqrt{n}\}$, i.e. $\tau_0 = \{1\}$, $\tau_1 = \{2\}$, $\tau_2 = \{3,4\}, \dots, \tau_J = \{\sqrt{n}/2+1,\dots,\sqrt{n}\}$. Soit $j : \{1,\dots,\sqrt{n}\} \to \{0,\dots,J\}$ définie par $j(u) = j$ si $u \in \tau_j$. Enfin, définissons $\Omega_{\ell,\ell'} = \tau_\ell \times \tau_{\ell'}$, pour $0 \leq \ell,\ell' \leq J$.

Définition A.6. *Soit $S = supp(x)$, on définit*

$$s_\ell^c := \max_{0\leq l'\leq J} \max_{k\in\tau_{\ell'}} |S \cap \Omega_{\ell,\ell'} \cap C_k|, \tag{A.11}$$

où C_k représente les indices correspondant à la k-ème ligne verticale.

La quantité s_ℓ^c représente le degré de parcimonie maximal de x restreint à ses colonnes et aux niveaux $(\Omega_{\ell,\ell'})_{1\leq l'\leq J}$.

Corollaire A.7. *Soit $x \in \mathbb{C}^n$ un vecteur supporté par S. Fixons A_0 comme la transformée 2D Fourier-Haar. Considérons que les blocs de mesures sont les $\sqrt{n}$ lignes horizontales de l'espace 2D. On choisit $(\pi_k)_{1\leq k\leq\sqrt{n}}$ constant par niveau, i.e. $\pi_k = \tilde{\pi}_{j(k)}$. Si le nombre m de lignes horizontales acquises vérifie*

$$m \gtrsim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \ln(s) \ln\left(\frac{n}{\varepsilon}\right) \tag{A.12}$$

alors x est l'unique solution du problème de minimisation (A.1) avec probabilité supérieure à $1-\varepsilon$.

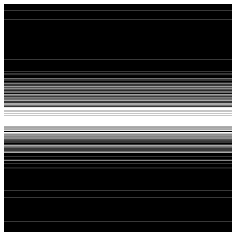
En particulier, la distribution minimisant (A.12) est

$$\pi_k = \frac{2^{-j(k)} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c}{\sum_{\ell=1}^{\sqrt{n}} 2^{-j(\ell)} \sum_{r=0}^{J} 2^{-|j(\ell)-r|/2} s_r^c},$$

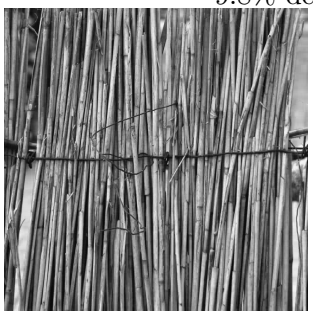
menant à

$$m \gtrsim \sum_{j=0}^{J} \left(s_j^c + \sum_{\substack{r=0 \\ r\neq j}}^{J} 2^{-|j-r|/2} s_r^c \right) \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

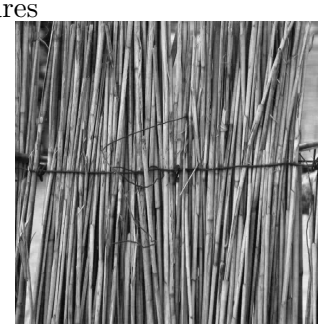
Ce résultat indique que le nombre de lignes à acquérir dans le niveau "horizontal" j dépend essentiellement de la quantité s_j^c. De plus, le Corollaire A.7 s'affranchit totalement de la condition $m \geq min\,(2s, \sqrt{n})$ requise par la Proposition A.4, ceci au prix d'une meilleure connaissance de la structure du support. Finalement, le Corollaire A.7 révèle que plus l'acquisition est structurée, plus le support doit être structuré afin d'assurer des résultats de reconstruction exacte. On notera que cette structure dans le support est explicitement donnée. Nous illustrons le Corollaire A.7 dans la Figure A.2.



9.8% de mesures

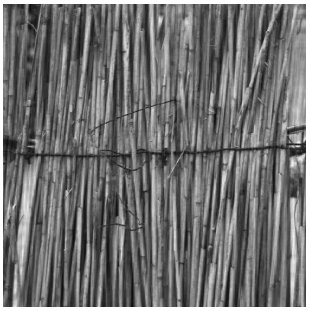


(a) Image originale
$s^c = (16, 16, 32, 59, 81, 75, 48)$



(b) SNR = 27.8 dB

(c) Image originale
$s^c = (16, 16, 32, 64, 124, 240, 411)$



(d) SNR = 14.7 dB

Figure A.2: Dans cet exemple, on cherche à reconstruire une image, en mesurant tous ses coefficients de Fourier, le long de lignes horizontales. Cette stratégie est similaire à ce qui se fait en IRM. L'image est de taille 2048×2048. (a) et (c) sont les images à reconstruire. Notons que l'image (c) est la même que celle en (a) mais tournée de 90°. On évalue les quantités $s^c = \left(s_j^c\right)_{1 \leq j \leq 7}$ pour (a) et (c). Nous utilisons pour les 2 reconstructions le schéma d'échantillonnage en haut de la figure. Les images en (b) et (d) sont les reconstructions correspondantes via minimisation ℓ_1. L'image (d) a été tournée de 90° pour faciliter la comparaison entre (b) et (d). On voit que la reconstruction de (b) est de très bonne qualité alors que la reconstruction de (c) est de mauvaise qualité. Ce résultat est prédit par le Corollaire A.7.

B Résumé du chapitre 3

Dans ce chapitre, nous proposons une méthode numérique de définition de densité variable d'échantillonnage supportée par des blocs de mesures. La première contribution est d'apporter une nouvelle façon de tirer les blocs afin d'imiter au mieux les stratégies issues du CS basées sur des mesures isolées. Pour ce faire, l'idée est de minimiser une mesure de dissimilarité entre une loi de probabilité définie sur un ensemble de mesures isolées et une loi de probabilité définie sur un dictionnaire de blocs de mesures. Ce problème s'avère être convexe et soluble, même en grande dimension. La deuxième contribution est de définir un algorithme rapide de minimisation basé sur la méthode de descente de gradient accélérée proposée par Nesterov dans les espaces métriques. Nous étudierons le choix des métriques et de la fonction proximale. Et l'on montrera que le choix de la métrique peut dépendre du type de bloc considéré. Enfin, nous utiliserons les distributions ainsi construites pour la définition de schémas d'échantillonnage en IRM : les résultats de reconstruction obtenus se montreront meilleurs que des schémas d'IRM standards tels que les schémas radiaux.

1 Notation

Nous considérons des signaux d-dimensionnels pour $d \in \mathbb{N}^*$, de taille $n_1 \times n_2 \times \ldots \times n_d = n$. Soient E et F des espaces vectoriels de dimension finie avec leur norme respective $\|.\|_E$ et $\|.\|_F$. Dans ce chapitre, on identifie E à $\mathbb{R}^m$ et F à $\mathbb{R}^n$. On note E^* et F^*, les espaces duaux de E and F. Pour $s \in E^*$ et $x \in E$ on définit $\langle s, x \rangle_{E^* \times E}$ la valeur de s au point x. La notation $\langle \cdot, \cdot \rangle$ sera utilisée pour le produit scalaire usuel dans un espace euclidien. La norme de E^* est définie par:

$$\|s\|_{E^*} = \max_{\substack{x \in E \\ \|x\|_E = 1}} \langle s, x \rangle_{E^* \times E} .$$

Soit $M : E \to F^*$ un opérateur. On définit la norme opérateur suivante

$$\|M\|_{E \to F^*} = \sup_{\|x\|_E \leq 1} \|Mx\|_{F^*}$$

Quand les espaces E^* et F sont respectivement dotés des normes ℓ^q et ℓ^p, on notera:

$$\|M^*\|_{F \to E^*} = \|M^*\|_{p \to q}.$$

On pose $\Delta_m \subset E$ le simplexe dans $E = \mathbb{R}^m$, et $\Delta_n \subset F$ le simplexe dans $F = \mathbb{R}^n$. Pour $\pi \in \Delta_m$ et un indice $j \in \{1, \ldots, m\}$ on désigne par π_j la j-ème composante de π.

2 Echantillonnage à densité variable avec contraintes d'acquisition de blocs

2.1 Cadre

On suppose que le système d'acquisition peut nous donner accès, dans l'idéal, à un ensemble fini $\{y_1, \dots, y_n\}$ de mesures d'un signal $x \in \mathbb{R}^{n_s}$ tels que

$$y_i = \langle a_i^*, x \rangle, \qquad \forall i = 1, \dots, n,$$

où a_i^* est la i-ème ligne de la matrice pleine A_0.

Soit $\mathcal{I} = \{I_1, \dots, I_m\}$ où chaque $I_k \subseteq \{1, \dots, n\}$ est un sous-ensemble d'indices. Nous supposons que le système d'acquisition est régi par des contraintes physiques qui oblige à mesurer simultanément les ensembles de mesures suivants

$$E_k = \{y_i, i \in I_k\}, \qquad \forall k = 1, \dots, m.$$

Dans le reste du chapitre, par un abus de notation, $\mathcal{I}$ désignera le dictionnaire de blocs. Par exemple en IRM, $n = n_s$ est le nombre de pixels d'une image 2D et y_i représente le i-ème coefficient de Fourier de l'image. On pourrait alors choisir I_k de sorte à décrire les lignes horizontales de l'espace de Fourier 2D.

Nous proposons d'acquérir partiellement le signal suivant la procédure suivante:

1. Construire une loi de probabilité $\pi \in \Delta_m$ imitant une densité cible $p \in \Delta_n$.
2. Tirer des indices i.i.d. $k_1, \dots, k_b$ à valeurs dans $\{1, \dots, m\}$, selon la loi π, avec $1 \leq b \leq m$.
3. Mesurer aléatoirement le signal x en utilisant l'ensemble aléatoire de blocs $\left(E_{k_j}\right)_{j \in \{1,\dots,b\}}$. La matrice d'échantillonnage s'écrira alors

$$A = (a_i^*)_{i \,\in\, \cup_{j=1}^b I_{k_j}} .$$

2.2 Une formulation variationnelle

Afin de comparer $\pi \in \Delta_m$ et la probabilité cible $p \in \Delta_n$, nous introduisons l'opérateur M suivant

$$\begin{aligned} M : \quad & E \longrightarrow F^* \\ & \pi \longmapsto p', \end{aligned}$$

où pour $i \in \{1, \dots, n\}$,

$$p'_i = \frac{\sum_{k=1}^m \pi_k \mathbb{1}_{i \in I_k}}{\sum_{j=1}^n \sum_{k'=1}^m \pi_{k'} \mathbb{1}_{j \in I_{k'}}}, \tag{B.1}$$

avec $\mathbb{1}_{i \in I_k}$ égal à 1 si $i \in I_k$, et 0 sinon.

Nous proposons de résoudre le problème suivant afin de calculer π:

$$\min_{\pi \in \Delta_m} F_\alpha(\pi), \tag{PP}$$

où

$$F_\alpha(\pi) = \|M\pi - p\|_{\ell^1} + \alpha\mathcal{E}(\pi),$$

avec $\alpha \in \mathbb{R}_+$ et où $\mathcal{E}$ représente la fonction entropie. Le premier terme de F représente un terme d'attache aux données garantissant que π sera bien une approximation de p. Par ailleurs, la fonction entropie $\mathcal{E}$ assure d'étaler au maximum la probabilité π sur l'ensemble du dictionnaire, et ainsi permet une meilleure couverture de l'espace d'acquisition. De plus, on notera que $\mathcal{E}$ est une fonciton fortement convexe sur Δ_m de paramètre $\sigma_\mathcal{E}$ (défini par rapport à la norme $\|\cdot\|_E$).

3 Optimisation

Dans cette section, nous présentons un algorithme afin de résoudre le problème PP. Il convient de noter que même si le problème PP est convexe, il reste délicat de le résoudre. Premièrement, π peut vivre dans un espace de très grande dimension. Deuxièmement, la fonction $\mathcal{E}$ est différentiable mais son gradient n'est pas Lipschitz continu, et la distance en variation totale $\|\cdot\|_{\ell^1}$ n'est pas differentiable.

Nous optons alors pour la résolution du problème dual de (PP) en utilisant les idées de Nesterov [77], dans un formalisme d'espaces métriques. Nous nous intéressons au choix optimal des normes dans l'espace primal et dual. Nous montrons que suivant la longueur ℓ des blocs, le choix optimal peut différer de la norme ℓ_2 très usitée en pratique.

3.1 Problème dual

En premier lieu, nous dérivons le problème dual de (PP).

Proposition B.1. *Soit* $J_\alpha(q) := \langle p, q\rangle_{F^*\times F} - \alpha \log\left(\sum_{\ell=1}^m \exp\left(-\frac{(M^*q)_\ell}{\alpha}\right)\right)$, *pour* $q \in F$. *Le problème dual de* (PP) *est:*

$$-\min_{q\in B_\infty} J_\alpha(q), \tag{DP}$$

au sens où $\min_{\pi\in\Delta_m} F_\alpha(\pi) = \max_{q\in B_\infty} -J_\alpha(q)$, *où* B_∞ *est la boule unité* ℓ^∞ *de* F.

Nous pouvons exhiber les propriétés suivantes pour le problème dual.

Proposition B.2. *La fonction* J_α *est convexe et son gradient est Lipschitz continu i.e.*

$$\|\nabla J_\alpha(q_1) - \nabla J_\alpha(q_2)\|_{F^*} \leq L_\alpha\|q_1 - q_2\|_F \qquad \forall (q_1, q_2) \in F^2,$$

avec pour constante

$$L_\alpha = \frac{\|M^*\|^2_{F\to E^*}}{\alpha\sigma_\mathcal{E}}. \tag{B.2}$$

De plus, ∇J_α *est localement Lipschitz en* $q \in F$ *de constante*

$$L_\alpha(q) = \frac{\|M^*\|^2_{F\to E^*}}{\alpha\sigma_\mathcal{E}(\pi(q))}, \tag{B.3}$$

où $\sigma_\mathcal{E}(\pi) := \inf_{\|h\|_E=1} \left\langle \mathcal{E}''(\pi)h, h\right\rangle$ *est le paramètre de convexité locale* $\mathcal{E}$ *autour de* π, *une expression explicite de* $\pi(q)$ *est donnée en* (B.4).

Nous allons donc résoudre le problème dual (B.1). Afin de revenir à la solution primale, nous explicitons les relations primales-duales.

Proposition B.3. *Les relations primales-duales*

$$\pi^\star = \arg\min_{\pi\in\Delta_m} F_\alpha(\pi) \quad et \quad q^\star = \arg\min_{q\in B_\infty} J_\alpha(q)$$

peuvent se réécrire

$$\pi_j^\star = \frac{\exp\left(-\frac{(M^*q^\star)_j}{\alpha}\right)}{\sum_{k=1}^m \exp\left(-\frac{(M^*q^\star)_k}{\alpha}\right)}, \qquad \forall j \in \{1,\dots,m\}\,. \tag{B.4}$$

De plus,

$$\operatorname{sign}\left(M\pi^\star - p\right) = \operatorname{sign}\left(q^\star\right). \tag{B.5}$$

3.2 Algorithme de résolution

L'algorithme de Nesterov est basé sur le choix d'un fonction prox d de l'ensemble B_∞, i.e. une fonction continue fortement convexe sur B_∞ par rapport à $\|\cdot\|_F$. On appelle σ_d le paramètre de convexité de d. Le schéma de résolution est donné dans l'algorithme 5. On

Algorithm 5 Algorithme de résolution proposé par Nesterov [77]

① Initialisation: fixer $q_0 \in B_\infty$.
② **for** $k = 0 \dots K$ **do**
③ Calculer $J_\alpha(q_k)$ et $\nabla J_\alpha(q_k)$
④ Trouver $y_k \in \arg\min_{y\in B_\infty} \langle \nabla J_\alpha(q_k), y - q_k\rangle + \frac{1}{2}L_\alpha\|y-q_k\|_F^2$
⑤ Trouver $z_k \in \arg\min_{q\in B_\infty} \frac{L_\alpha}{\sigma_d}d(q) + \sum_{i=0}^k \frac{i+1}{2}\left[J_\alpha(q_i) + \langle\nabla J_\alpha(q_i), q - q_i\rangle\right]$
⑥ Evaluer $q_{k+1} = \frac{2}{k+3}z_k + \frac{k+1}{k+3}y_k$.
⑦ **end for**
⑧ Calcul de la solution primale $\pi_j = \dfrac{\exp\left(-\frac{(M^*y_K)_j}{\alpha}\right)}{\sum_{k=1}^m \exp\left(-\frac{(M^*y_K)_k}{\alpha}\right)}, \qquad \forall j \in \{1,\dots,m\}\,.$

peut montrer que l'algorithm 5 assure que

$$\begin{aligned} J_\alpha(y_k) - J_\alpha(q^\star) &\leq \frac{4L_\alpha d(q^\star)}{\sigma_d(k+1)(k+2)} \\ &\leq \frac{4\|M^*\|^2_{F\to E^*}D}{\alpha\sigma_\mathcal{E}\sigma_d(k+1)(k+2)}, \end{aligned} \tag{B.6}$$

avec $q^\star$ une solution optimale du problème (DP), et $D = \max_{q\in B_\infty} d(q)$. Nous pouvons alors donner des garanties sur la distance au primal.

Théorème 2. *Soit*

$$\pi_k = \frac{\exp\left(-\frac{(M^* y_k)}{\alpha}\right)}{\left\|\exp\left(-\frac{(M^* y_k)}{\alpha}\right)\right\|_{\ell^1}}.$$

où y_k est défini dans l'algorithme 5. Nous avons

$$\|\pi_k - \pi^\star\|_E^2 \leq \frac{8\|M^*\|_{F\to E^*}^2 D}{\alpha^2 \sigma_{\mathcal{E}}^2 \sigma_d (k+1)(k+2)}.$$

3.3 Choix des normes sur E et F

Nous rappelons ici les principaux résultats sur les choix de normes.

Proposition B.4. *La norme $\|\cdot\|_{E^*}$ minimisant (B.6) parmi toutes les normes ℓ^p, $p \in [1, +\infty]$ est $\|\cdot\|_{\ell^\infty}$.*

Proposition B.5. *Soient $\|\cdot\|_F = \|\cdot\|_p$ et $d = d_{p'}$ avec $p \in [1, \infty]$ et $p' \in]1, 2]$. Pour ces familles de normes et de fonctions prox, celles qui minimisent la borne de complexité (B.6) sont*

- *$p' = 2$ et $p \in [1, 2]$, if $\ell^2 = n$. Ainsi*

$$J_\alpha(y_k) - J_\alpha(q^\star) \leq \frac{2\sqrt{n}}{\alpha(k+1)(k+2)}. \tag{B.7}$$

- *$p = p' = 2$, si $\ell^2 < n$. Ainsi,*

$$J_\alpha(y_k) - J_\alpha(q^\star) \leq \frac{2n}{\alpha\ell(k+1)(k+2)}. \tag{B.8}$$

- *$p = 1$ et $p' = 2$, si $\ell^2 > n$. Ainsi,*

$$J_\alpha(y_k) - J_\alpha(q^\star) \leq \frac{2n^{3/2}}{\alpha\ell^2(k+1)(k+2)}. \tag{B.9}$$

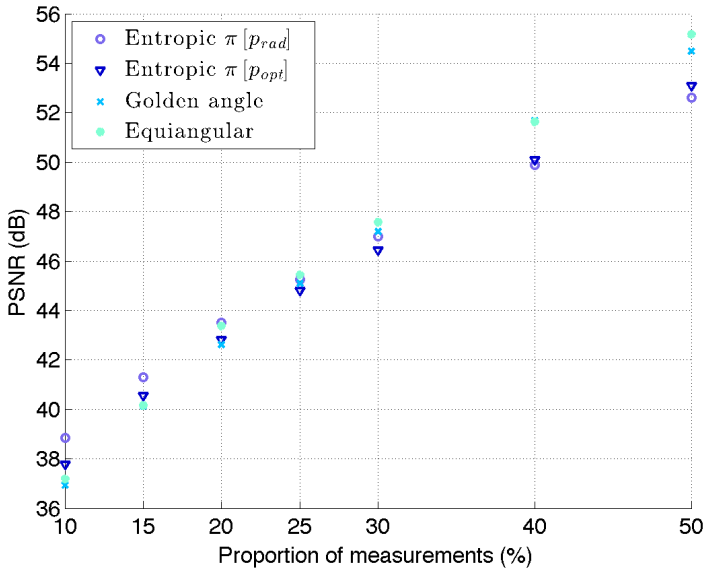
4 Application à la reconstruction d'images IRM

Dans cette partie, nous comparons les qualités de reconstruction d'images IRM obtenus à partir de différents schémas d'acquisition. Les schémas considérés dans ce chapitre sont basés sur des blocs de mesures et sur des schémas heuristiques utilisés en pratique sur des vrais scanners IRM, consistant en:

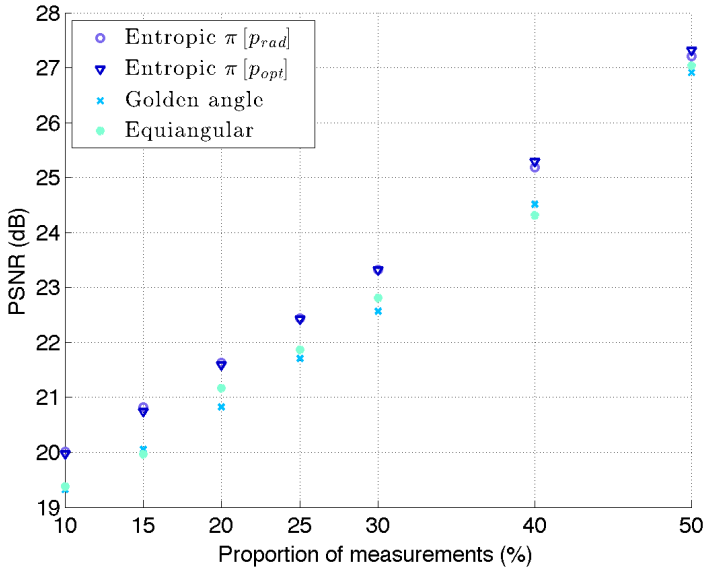
- lignes radiales uniformément distribuées [70].
- schéma basé sur l'angle d'or [104].
- lignes radiales tirées aléatoirement [26].
- schéma basé sur un dictionnaire de lignes droites rejoignant les bords opposés du cadre de l'espace d'acquisition

- les blocs sont tirés selon $\pi\left[p_{\text{rad}}\right]$ qui est la probabilité résultante de la minimisation du problème (PP) pour $p = p_{\text{rad}}$ une densité cible radiale.
- les blocs sont tirés selon $\pi\left[p_{\text{opt}}\right]$ qui est la probabilité résultante de la minimisation du problème (PP) pour $p = p_{\text{opt}}$ une densité définie dans [30, 11].

On reconstruit, en utilisant les schémas précédemment décrits, une image IRM et une image de babouin. Pour les 2 images, on considère que le cadre est celui de l'IRM : l'image est parcimonieuse en ondelettes et les données sont acquises dans l'espace de Fourier. Nous comparons dans la Figure B.1 l'efficacité des schémas d'échantillonnage précédemment décrits. Nous pouvons donc conclure que pour de faibles proportions de mesures ($\sim 10\%$) les schémas d'échantillonnage générés par la nouvelle méthode proposée dans ce chapitre permettent d'obtenir de meilleures reconstructions que les schémas standards contraints par blocs de mesures et utilisés en IRM.



(a) Image IRM de taille 512×512



(d) Image d'un babouin de taille 512×512

Figure B.1: Moyennes des PSNR obtenus pour les images reconstruites (du cerveau et du babouin) en fonction de la proportion de mesures $10\%, 15\%, 20\%, 25\%, 30\%, 40\%, 50\%$ dans les différents types de schémas d'échantillonnage.

C

Résumé du chapitre 4

Ce chapitre a deux finalités. Premièrement, nous proposons un état de l'art des théories d'échantillonnage compressé pour l'Imagerie par résonance magnétique (IRM). Ceci nous permet de dégager quelques grands principes à suivre pour générer des schémas performants en termes de temps d'acquisition et de qualité de reconstruction. Dans une deuxième partie, nous proposons une méthodologie originale de conception de schémas qui repose sur des algorithmes de projection de densités sur des espaces de mesures. Nous proposons finalement des comparaisons avec des stratégies actuelles d'échantillonnage sur des simulations et montrons ainsi le bien-fondé de notre approche.

1 Introduction

La génération de schémas d'échantillonnage compressé efficaces en Imagerie par Résonance Magnétique (IRM) est une question qui reste aujourd'hui largement ouverte d'un point de vue théorique et applicatif. A notre connaissance, la majorité des méthodes implémentées actuellement sur des imageurs repose sur le bon sens, mais n'a pas de fondements mathématiques solides. Ceci s'explique par la grande diversité des modalités d'IRM (anatomique, fonctionnelle, parallèle,...), la complexité des contraintes d'acquisition et des images, ainsi que les nombreuses difficultés mathématiques pour produire une théorie unifiée capable de prendre en compte ces différents aspects. Il nous semble qu'il est encore possible d'obtenir des gains significatifs en terme de temps d'échantillonnage ou de résolution spatio-temporelle. Ceci aurait des conséquences très importantes pour la médecine et la biologie.

Dans une série de travaux récents [16, 28, 31, 11], nous avons proposé de nouvelles idées pour combler une partie de ces lacunes. Bien que beaucoup de questions restent en suspens, ces travaux nous ont permis d'acquérir une bonne intuition des lignes à suivre pour générer des schémas performants. L'objectif de ce chapitre est (i) de résumer les conclusions principales de ces travaux et (ii) de proposer une technique de génération de schémas d'échantillonnage originale suivant ces principes.

2 Notations

Dans ce travail, nous considérons une image $u \in L^2(\Omega)$ à valeurs dans $\mathbb{C}$. Nous supposons pour simplifier la discussion que $\Omega = [0,1]^d$ avec $d = 2$ ou $d = 3$. La transformée de Fourier d'une image u est notée $\hat{u}$. Les images discrétisées sont notées en gras $u \in \mathbb{C}^n$, où

n représente le nombre de pixels ou de voxels. La matrice de la transformée de Fourier discrète est notée $F \in \mathbb{C}^{n\times n}$. On suppose de plus que les images sont compressibles avec une certaine transformée $\Psi \in \mathbb{C}^{n\times n}$. Dans les expériences numériques, la transformée Ψ est une transformée en ondelettes orthogonale. La compressibilité signifie que l'essentiel de l'énergie de u est concentré dans un petit nombre de coefficients de la transformée $\Psi^* u$. L'espace $\mathcal{M}$ désigne l'ensemble des mesures de probabilité sur Ω. L'espace Δ_n est le simplexe de $\mathbb{R}^n$.

3 Etat de l'art

3.1 Echantillonnage en IRM

L'IRM permet de mesurer un ensemble de valeurs de la transformée de Fourier d'une image u. [1] Les valeurs de la transformée de Fourier sont acquises le long d'une courbe paramétrée $s : [0, T] \to \mathbb{R}^d$ qui doit appartenir à un certain ensemble de contraintes de faisabilité $\mathcal{S}_T$ [31]. Par exemple, dans le cas où l'acquisition est obtenue à partir d'une seule excitation radio-fréquence, la courbe s doit appartenir à l'ensemble suivant :

$$\mathcal{S}_T = \{s : [0, T] \to \mathbb{R}^d, \|\dot{s}\|_\infty \leq \alpha, \|\ddot{s}\|_\infty \leq \beta\}. \tag{C.1}$$

Les paramètres α et β dépendent de l'imageur. Si l'acquisition se fait à partir de plusieurs excitations radio-fréquences, s appartient à des produits cartésiens de l'ensemble $\mathcal{S}_T$.

Après une acquisition en IRM, on a donc accès à l'ensemble E de mesures :

$$E = \left\{\hat{u}(s(i\Delta t)), 1 \leq i \leq \left\lfloor \frac{T}{\Delta t} \right\rfloor\right\} \tag{C.2}$$

où $\Delta t > 0$ décrit la cadence d'échantillonnage de la courbe s. Dans ce chapitre, nous négligeons les dégradations de mesures liées au bruit ou aux inhomogénéité des champs magnétiques.

Maintenant que les possibilités d'une IRM sont posées, nous pouvons présenter les questions typiques qui se posent en échantillonnage pour l'IRM. A T fixé comment trouver une courbe $s \in \mathcal{S}_T$ optimale, au sens où elle permet d'obtenir la reconstruction la plus fidèle ? Quel est le temps d'acquisition T minimal qui permette d'obtenir une précision de reconstruction donnée ? Comment reconstruire les images connaissant E ?

Ces questions semblent extrêmement complexes mathématiquement, ce qui explique la prédominance d'heuristiques [69] dans la littérature. Dans la suite, nous présentons quelques résultats théoriques qui permettent tout de même de guider la conception d'une trajectoire.

3.2 Echantillonnage compressé usuel

Dans cette partie, nous décrivons la théorie de l'échantillonnage compressé telle qu'elle a été posée dans les papiers fondateurs [23] et plus récemment [20]. Les auteurs considèrent une matrice orthogonale $A_0 = [a_1^*; \ldots; a_n^*]$ (le séparateur ; indique une concaténation verticale, comme en Matlab). Ils proposent de construire une matrice de mesure aléatoire

[1] Ceci est vrai de façon conventionnelle, cependant, on peut modifier la signification des mesures en utilisant des pulses radio-fréquence particuliers [55], l'IRM parallèle [65] ou encore en modifiant l'utilisation des correcteurs de champs magnétiques [83].

$A = [a_{J_1}^*; \ldots; a_{J_m}^*]$ où les quantités $J_k \in \{1, \ldots, n\}$ sont des variables aléatoires i.i.d. uniformes. Connaissant $y = Ax$, les auteurs proposent de reconstruire une estimée $\bar{x}$ de x en résolvant le problème ℓ_1 suivant :

$$\bar{x} \in \arg\min_{Ax=y} \|x\|_1$$

Le type de théorème qu'ils obtiennent dans [20] est de la forme suivante[2].

Théorème 3. *Supposons que x est s-parcimonieux, i.e. qu'il ne contient que s composantes non nulles parmi n. Si le nombre de mesures m satisfait :*

$$m \geq Cs \left(n \max_{1 \leq k \leq n} \|a_k\|_\infty^2 \right) \log\left(\frac{n}{\epsilon}\right),$$

où $C > 0$ est une constante universelle, alors $\bar{x} = x$ avec probabilité $1 - \epsilon$.

La cohérence $\kappa(A_0) = n \max_{1 \leq k \leq n} \|a_k\|_\infty^2$ est comprise entre 1 et n. En particulier, $\kappa(F) = 1$ et $\kappa(\mathrm{Id}) = n$. Dans le cas favorable de la transformée de Fourier, ce théorème indique donc qu'il suffit d'environ $s \log\left(\frac{n}{\epsilon}\right)$ mesures pour obtenir une reconstruction exacte d'un signal s-parcimonieux.

Bien que ce type de théorème ait eu un impact énorme dans la littérature, il n'est pas applicable en IRM. En effet, la transformée naturelle en IRM est $A_0 = F^*\Psi$, i.e. le produit d'une transformée de Fourier et d'une transformée en ondelettes. On peut montrer que $\kappa(F^*\Psi) = O(n)$. Le théorème n'a donc pas d'intérêt dans ce cas.

3.3 Les avancées récentes

Dans cette partie, nous indiquons les avancées récentes en échantillonnage compressé qui nous semblent les plus importantes. Des omissions sont possibles car une telle entreprise contient nécessairement une bonne dose de subjectivité.

3.3.1 Echantillonnage à densité variable

Dans la majorité des applications réelles, les transformées à utiliser sont cohérentes. C'est notamment le cas de l'IRM. Une technique assez simple pour briser la cohérence consiste à ne pas tirer les échantillons suivant une loi uniforme, mais à tirer les échantillons cohérents plus souvent. Précisons cette idée. Soit $\boldsymbol{\pi} \in \Delta_n$ la distribution des variables aléatoires J_k. Le théorème suivant [28] justifie l'utilisation d'échantillonnage à densité variable. Notons que cette technique est la base (non justifiée) de l'article fondateur [69].

Théorème 4. *Supposons que x est s-parcimonieux, i.e. qu'il ne contient que s composantes non nulles parmi n. Posons $\boldsymbol{\pi}_k = \frac{\|a_k\|_\infty^2}{\sum_{i=1}^n \|a_i\|_\infty^2}$. Si le nombre de mesures m satisfait :*

$$m \geq Cs \left(\sum_{i=1}^n \|a_i\|_\infty^2 \right) \log\left(\frac{n}{\epsilon}\right),$$

où $C > 0$ est une constante universelle, alors $\bar{x} = x$ avec probabilité $1 - \epsilon$.

[2] Notez que nous ne parlons à aucun moment de Restricted Isometry Property (RIP). Ce type de notion est très répandu car il permet la reconstruction de *tous* les signaux s-parcimonieux. Cette propriété - bien que très plaisante - est extrêmement exigeante. Il nous semble clair qu'elle doit être abandonnée à l'avenir pour obtenir des résultats théoriques convaincants en IRM.

On peut montrer que pour le cas de l'IRM (i.e. $A_0 = F^*\Psi$), $\sum_{i=1}^{n} \|a_i\|_\infty^2 \propto \log(n)$. Ainsi, il suffit de $O(s\log(n)^2)$ coefficients de Fourier pour pouvoir reconstruire exactement une image s-parcimonieuse.

3.3.2 Parcimonie structurée

Bien que le théorème (4) soit plutôt attrayant théoriquement, des expériences numériques [2] montrent qu'il est encore insuffisant pour une application en IRM : les constantes apparaissant dans les O sont en effet grandes, et même le terme $\log(n)^2$ n'est pas négligeable en pratique. Une avancée récente qui nous semble importante est le papier [2]. Les auteurs montrent qu'il est possible d'exploiter une parcimonie structurée (e.g. plus de coefficients d'ondelettes non nuls dans les sous-bandes correspondant aux basses fréquences) pour obtenir de meilleures garanties théoriques. Nous ne décrivons pas les théorèmes résultants par manque de place. Ils permettent cependant d'obtenir deux conclusions importantes. Premièrement, il faut utiliser des schémas d'échantillonnage à densité variable comme dans le paragraphe précédent. Les densités doivent dépendre à la fois de la cohérence locale $\|a_k\|_\infty$ et de l'a priori qu'on se donne sur le support de x. Deuxièmement, comme la parcimonie est asymptotique (elle n'apparait que dans les sous-bandes hautes-fréquences), l'acquisition comprimée en IRM peut difficilement être utilisée comme une technique pour gagner du temps. Elle permet par contre - à temps d'acquisition constant - d'obtenir de meilleures résolutions spatiales.

3.3.3 Acquisition structurée

Tous les résultats donnés jusqu'à présent ne contiennent aucune notion de structure dans l'acquisition. Ils ne sont donc pas applicables en pratique. A notre connaissance, les seuls travaux commençant à proposer des solutions d'échantillonnage plausibles et justifiées théoriquement sont [28, 11].

Dans [28] nous avons d'abord proposé une définition générique précise de ce qu'est un échantillonneur à densité variable : c'est un processus dont la mesure empirique convergence vers une densité cible quand le temps tend vers l'infini. Nous avons analysé l'utilisation de tels échantillonneurs théoriquement et en avons déduit deux propriétés essentielles pour qu'ils permettent des reconstruction précises :

1. La *mesure empirique* du processus doit correspondre à une mesure cible, i.e. il faut effectuer un échantillonnage à densité variable.

2. Le processus ne sera efficace que si son *temps de mélange* (i.e. sa rapidité à converger vers la densité cible) est faible.

Ces résultats indiquent qu'un bon échantillonneur doit i) couvrir l'espace rapidement (temps de mélange) ii) avoir une densité cible fixée et iii) appartenir à un ensemble admissible pour l'IRM. Ces conclusions sont très proches des travaux plus heuristiques menés par les leaders mondiaux en échantillonnage compressé pour l'IRM [101]. Dans la suite de ce travail, nous proposons un algorithme original permettant de générer de tels schémas automatiquement.

4 Génération automatique de schémas d'échantillonnage

Nous supposons connue une densité cible $\pi \in \mathcal{M}$. Celle-ci peut être obtenue par des considérations heuristiques [69] ou en utilisant des théories de l'échantillonnage compressé [2, 28, 11]. Nous fixons aussi le temps total d'acquisition T et cherchons à trouver une courbe d'échantillonnage $s \in \mathcal{S}_T$ efficace.

À une courbe $s \in \mathcal{S}_T$, nous pouvons associer une mesure image $\mu_s \in \mathcal{M}(\mathcal{S}_T)$ définie pour tout Borélien $B \subseteq \Omega$ par

$$\mu_s(B) = \gamma(s^{-1}(B))$$

où γ est la mesure de Lebesgue normalisée par T. Ainsi $\mu_s(B)$ mesure le temps relatif que s passe dans l'ensemble B. On souhaite que $\mu_s \simeq \pi$ et il faut donc définir une notion de distance entre mesures. Dans un travail récent [29], nous avons proposé de définir une distance entre mesures comme suit :

$$\mathrm{dist}(\mu_s, \pi) = \|h \star (\mu_s - \pi)\|_2,$$

où h est un noyau de convolution continu sur Ω. Si la transformée de Fourier de h est réelle et non nulle partout, on a montré que dist est effectivement une distance et qu'elle métrise la convergence faible entre mesures. Elle est de plus majorée par la distance de Wasserstein si h est Lipschitz continue.

L'idée principale de ce chapitre est de définir s comme la courbe paramétrisant la solution du problème variationnel suivant :

$$\min_{\mu_s \in \mathcal{M}(\mathcal{S}_T)} \frac{1}{2}\|h \star (\mu_s - \pi)\|_2^2. \tag{C.3}$$

Ce problème de projection consiste à chercher la courbe s admissible dont la mesure approche au mieux la densité cible π.

Nous avons montré dans [29] que des algorithmes de programmation non-linéaire permettent de trouver des point critiques de la fonctionnelle (C.3). De plus, ces algorithmes donnent des résultats de projection d'une bonne qualité quelque soit l'initialisation. Nous renvoyons le lecteur intéressé à ce papier pour plus de détails.

5 Résultats

Pour évaluer la pertinence, de cette approche, nous travaillons sur un fantôme couramment utilisé en IRM de taille 1024^2, voir Figure C.1(a). Nous projetons une densité cible avec décroissance radiale sur : des sommes de N Diracs, Figure C.2(b), les mesures empiriques de courbes dans l'ensemble $\mathcal{S}_T$ défini en (C.1), Figure C.2(c), la mesure empirique d'un ensemble de 200 segments de longueurs arbitraires, Figure C.2(d). L'exemple de la Figure C.2(a) correspond à des tirages i.i.d. Il sert uniquement de référence car il n'est pas implémentable, mais correspond aux approches standards utilisées en simulation. Nous indiquons uniquement les SNR des images reconstruites par souci de concision. On peut donner plusieurs conclusions. (i) Utiliser une répulsion entre points (telle que le Poisson disk sampling [17]) améliore très légèrement les résultats par rapport à des tirages i.i.d. (ii) Le schéma de projection permet de retrouver des schémas correspondant à l'état de l'art (e.g. spirales bruitées) automatiquement, Figure C.2(c). (iii) Le schéma de projection permet aussi de trouver de nouveaux schémas inédits (e.g. lignes de longueur variables,

Figure C.2(d)). iv) Ajouter des contraintes cinématiques sur les trajectoires semble réduire la qualité de reconstruction, mais la baisse est raisonnable (~1dB).

Les schémas proposés sont en cours d'implémentation sur les scanners pré-cliniques Bruker 7T et 17T du projet Neurospin.

(a) (b)

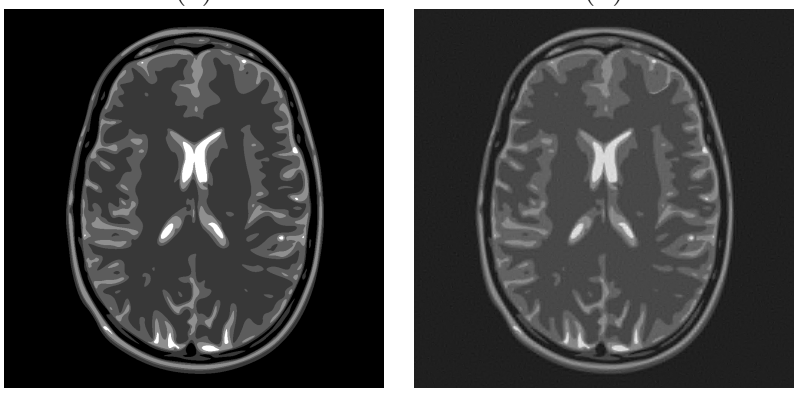


Figure C.1: (a) image à reconstruire de taille 1024x1024. (b) Image reconstruite avec le schéma de la Figure C.2(a), SNR=19.6dB.

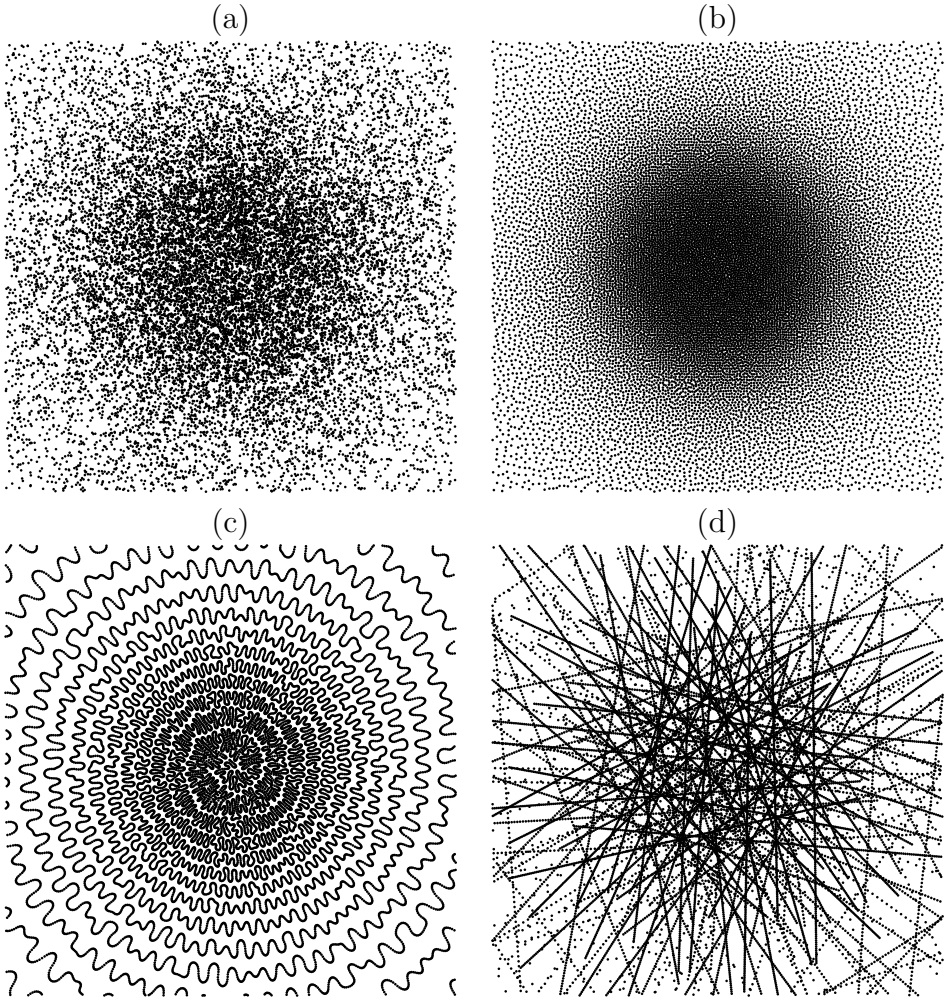


Figure C.2: Exemples de schémas d'échantillonnage générés avec 50000 échantillons, soit 4.7% des mesures. (a) Tirages i.i.d. (infaisable - SNR=19.6dB) (b) projection sur les mesures discrètes (infaisable - SNR=19.8dB) (c) projection sur les mesures admissibles pour une trajectoire d'IRM (SNR=18.6dB). (d) projection sur 200 segments de longueur variable (SNR=18.9dB).

Bibliography

[1] Ben Adcock and Anders C Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Foundations of Computational Mathematics, to appear*, 2015.

[2] Ben Adcock, Anders C. Hansen, Clarice Poon, and Bogdan Roman. Breaking the coherence barrier: A new theory for compressed sensing. *arXiv preprint arXiv:1302.0561*, 2013.

[3] Ben Adcock, Anders C Hansen, and Bogdan Roman. A note on compressed sensing of structured sparse wavelet coefficients from subsampled fourier measurements. *arXiv preprint arXiv:1403.6541*, 2014.

[4] Ben Adcock, Anders C. Hansen, and Bogdan Roman. *Compressed Sensing and its Applications*, chapter "The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing". 2016.

[5] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods. *Mathematical Programming*, 137(1-2):91–129, 2013.

[6] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[7] Alexander Bastounis and Anders C Hansen. On the absence of the rip in real-world applications of compressed sensing and the rip in levels. *arXiv preprint arXiv:1411.4449*, 2014.

[8] Heinz Bauschke and Jonathan Borwein. On the convergence of von neumann's alternating projection algorithm for two sets. *Set-Valued Analysis*, 1(2):185–212, 1993.

[9] Heinz Bauschke and Patrick Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011.

[10] Peter Bickel, Ya'acov Ritov, and Alexandre Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.

[11] Jérémie Bigot, Claire Boyer, and Pierre Weiss. An analysis of blocks sampling strategies in compressed sensing. *arXiv preprint arXiv:1310.4393*, 2014.

[12] Jérôme Bobin, Jean-Luc Starck, and Roland Ottensamer. Compressed sensing in astronomy. *Selected Topics in Signal Processing, IEEE Journal of*, 2(5):718–726, 2008.

[13] Stéphane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[14] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[15] Claire Boyer, Philippe Ciuciu, Pierre Weiss, and Sébastien Mériaux. Hyr 2 pics: Hybrid regularized reconstruction for combined parallel imaging and compressive sensing in mri. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 66–69. IEEE, 2012.

[16] Claire Boyer, Pierre Weiss, and Jérémie Bigot. An algorithm for variable density sampling with block-constrained acquisition. *SIAM Journal on Imaging Sciences*, 7(2):1080–1107, 2014.

[17] Robert Bridson. Fast poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH*, volume 2007, page 5, 2007.

[18] Peter Bühlmann and S van de Geer. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.

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

[20] Emmanuel Candes and Yaniv Plan. A probabilistic and ripless theory of compressed sensing. *Information Theory, IEEE Transactions on*, 57(11):7235–7254, 2011.

[21] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[22] Emmanuel Candes, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[23] Emmanuel Candes, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[24] Emmanuel Candes and Terence Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[25] Emmanuel Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006.

[26] Rachel Chan, Elizabeth Ramsay, Edward Cheung, and Donald Plewes. The influence of radial undersampling schemes on compressed sensing reconstruction in breast mri. *Magnetic Resonance in Medicine*, 67(2):363–377, 2012.

[27] Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, and Pierre. Weiss. Variable density sampling with continuous sampling trajectories. *preprint*, 2013.

[28] Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, and Pierre Weiss. Variable density sampling with continous sampling trajectories. *SIAM Journal on Imaging Sciences, in press*, 2014.

[29] Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, and Pierre Weiss. A projection method on measures sets. *submitted*, 2015.

[30] Nicolas Chauffert, Philippe Ciuciu, and Pierre Weiss. Variable density compressed sensing in MRI. theoretical vs heuristic sampling strategies. In *proceedings of IEEE ISBI*, 2013.

[31] Nicolas Chauffert, Pierre Weiss, Jonas Kahn, and Philippe Ciuciu. Gradient waveform design for variable density sampling in magnetic resonance imaging. *arXiv arXiv:1412.4621*, 2014.

[32] Gong Chen and Marc Teboulle. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

[33] Scott Shaobing Chen, David Donoho, and Michael Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.

[34] Kihwan Choi, Jing Wang, Lei Zhu, Tae-Suk Suh, Stephen Boyd, and Lei Xing. Compressed sensing based cone-beam computed tomography reconstruction with a first-order method. *Medical physics*, 37(9):5113–5125, 2010.

[35] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best k-term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.

[36] Patrick Combettes, Djinh Dũng, and Bang Công Vũ. Dualization of signal recovery problems. *Set-Valued and Variational Analysis*, 18(3-4):373–404, 2010.

[37] Patrick Combettes and Jean-Christophe Pesquet. A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):564–574, 2007.

[38] Patrick Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.

[39] George Bernard Dantzig. *Linear programming and extensions*. Princeton university press, 1998.

[40] Alexandre d'Aspremont and Martin Jaggi. An optimal affine invariant smooth minimization algorithm. *arXiv preprint arXiv:1301.0465*, 2013.

[41] Nicolas Dobigeon, Adrian Basarab, Denis Kouamé, and Jean-Yves Tourneret. Regularized bayesian compressed sensing in ultrasound imaging. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 2600–2604. IEEE, 2012.

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

[43] David Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 47(7):2845–2862, 2001.

[44] Marco Duarte and Yonina Eldar. Structured compressed sensing: From theory to applications. *Signal Processing, IEEE Transactions on*, 59(9):4053–4085, 2011.

[45] Jalal Fadili and Gabriel Peyré. Total variation projection with first order schemes. *Image Processing, IEEE Transactions on*, 20(3):657–669, 2011.

[46] Willliam Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 2008.

[47] Simon Foucart, Alain Pajor, Holger Rauhut, and Tino Ullrich. The Gelfand widths of l_p-balls for $0 < p \leq 1$. *J. Complexity*, 26:629–640, 2010. Best Paper Award 2010 of the Journal of Complexity.

[48] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[49] Andrej Garnaev and Efim Gluskin. The widths of a euclidean ball. In *Dokl. Akad. Nauk SSSR*, volume 277, pages 1048–1052, 1984.

[50] Clóvis Gonzaga and Elizabeth Karas. Fine tuning nesterov's steepest descent algorithm for differentiable convex programming. *Mathematical Programming*, pages 1–26, 2013.

[51] Rémi Gribonval and Morten Nielsen. Beyond sparsity: Recovering structured representations by {\ ell}^ 1 minimization and greedy algorithms. *Advances in computational mathematics*, 28(1):23–41, 2008.

[52] Karlheinz Gröchenig, José Luis Romero, Jayakrishnan Unnikrishnan, and Martin Vetterli. On minimal trajectories for mobile sampling of bandlimited fields. *Applied and Computational Harmonic Analysis*, 2014.

[53] David Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566, 2011.

[54] David Gross, Felix Krahmer, and Richard Kueng. A partial derandomization of phaselift using spherical designs. *Journal of Fourier Analysis and Applications*, pages 1–38.

[55] Justin P Haldar, Diego Hernando, and Zhi-Pei Liang. Compressed-sensing mri with random encoding. *Medical Imaging, IEEE Transactions on*, 30(4):893–903, 2011.

[56] Cédric Herzet, Charles Soussen, Jérôme Idier, and Rémi Gribonval. Exact recovery conditions for sparse representations with partial support information. *Information Theory, IEEE Transactions on*, 59(11):7509–7524, 2013.

[57] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals*, volume 1. Springer, 1996.

[58] Anatoli Juditsky, Fatma Kılınç Karzan, Arkadi Nemirovski, Boris Polyak, et al. Accuracy guaranties for l1 recovery of block-sparse signals. *The Annals of Statistics*, 40(6):3077–3107, 2012.

[59] Anatoli Juditsky and Arkadi Nemirovski. On verifiable sufficient conditions for sparse signal recovery via l1 minimization. *Mathematical programming*, 127(1):57–88, 2011.

[60] Anatoli Juditsky and Arkadii S Nemirovski. Large deviations of vector-valued martingales in 2-smooth normed spaces. *arXiv preprint arXiv:0809.0813*, 2008.

[61] Jens Keiner, Stefan Kunis, and Daniel Potts. Using nfft 3—a software library for various nonequispaced fast fourier transforms. *ACM Transactions on Mathematical Software (TOMS)*, 36(4):19, 2009.

[62] Felix Krahmer, Shahar Mendelson, and Holger Rauhut. Suprema of chaos processes and the restricted isometry property. *Communications on Pure and Applied Mathematics*, 67(11):1877–1904, 2014.

[63] Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Trans. Image Proc.*, 23(2):612–622, 2014.

[64] Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *Image Processing, IEEE Transactions on*, 23(2):612–622, 2014.

[65] David J Larkman and Rita G Nunes. Parallel magnetic resonance imaging. *Physics in medicine and biology*, 52(7):R15, 2007.

[66] Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electron. J. Probab.*, 15:no. 41, 1319–1343, 2010.

[67] Li Liu, Yuntao He, Jianguo Zhang, Huayu Jia, and Jun Ma. Optimum linear array for aperture synthesis imaging based on redundant spacing calibration. *Optical Engineering*, 53(5):053109, 2014.

[68] Michael Lustig, Marcus Alley, Shreyas Vasanawala, David Donoho, and John Pauly. L1 spir-it: autocalibrating parallel imaging compressed sensing. In *Proc Intl Soc Mag Reson Med*, volume 17, page 379, 2009.

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

[70] Michael Lustig, David Donoho, Juan M. Santos, and John M. Pauly. Compressed sensing mri. *Signal Processing Magazine, IEEE*, 25(2):72–82, 2008.

[71] Michael Lustig, Seung-Jean Kim, and John M. Pauly. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. *Medical Imaging, IEEE Transactions on*, 27(6):866–873, 2008.

[72] Michael Lustig, Jin Hyung Lee, David L Donoho, and John M Pauly. Faster imaging with randomly perturbed, under-sampled spirals and l1 reconstruction. In *Proceedings of the 13th annual meeting of ISMRM, Miami Beach*, page 685, 2005.

[73] Marcio Marim, Michael Atlan, Elsa Angelini, and Jean-Christophe Olivo-Marin. Compressed sensing with off-axis frequency-shifting holography. *Optics letters*, 35(6):871–873, 2010.

[74] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Uniform uncertainty principle for bernoulli and subgaussian ensembles. *Constructive Approximation*, 28(3):277–289, 2008.

[75] Yurii Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). In *Soviet Mathematics Doklady*, volume 27, pages 372–376, 1983.

[76] Yurii Nesterov. *Introductory lectures on convex optimization*, volume 87. Springer Science & Business Media, 2004.

[77] Yurii Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[78] Yurii Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.

[79] John von Neumann. Functional operators. In *The Geometry of Orthogonal Spaces, Ann. Math. Studies No. 22*, volume 2. Princeton Univ. Press Princeton, N. J, 1950.

[80] Adam Polak, Marco Duarte, and Dennis Goeckel. Performance bounds for grouped incoherent measurements in compressive sensing. *arXiv preprint arXiv:1205.2118*, 2014.

[81] Adam Polak, Marco Duarte, and Dennis Goeckel. Performance bounds for grouped incoherent measurements in compressive sensing. *Signal Processing, IEEE Transactions on*, PP(99):1–1, 2015.

[82] Daniel Potts and Gabriele Steidl. Fast summation at nonequispaced knots by nfft. *SIAM Journal on Scientific Computing*, 24(6):2013–2037, 2003.

[83] Gilles Puy, Jose Marques, Rolf Gruetter, J. Thiran, Dimitri Van De Ville, Pierre Vandergheynst, and Yves Wiaux. Spread spectrum magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 31(3):586–598, 2012.

[84] Gilles Puy, Pierre Vandergheynst, and Yves Wiaux. On variable density compressive sampling. *Signal Processing Letters, IEEE*, 18(10):595–598, 2011.

[85] Céline Quinsac, Adrian Basarab, Jean-Marc Girault, and Denis Kouame. Compressed sensing of ultrasound images: sampling of spatial and frequency domains. In *Signal Processing Systems (SIPS), 2010 IEEE Workshop on*, pages 231–236. IEEE, 2010.

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

[87] Yair Rivenson, Adrian Stern, and Bahram Javidi. Compressive fresnel holography. *Journal of Display Technology*, 6(10):506–509, 2010.

[88] R Tyrell Rockafellar. *Convex analysis*, volume 28. Princeton university press, 1997.

[89] Bogdan Roman, Anders Hansen, and Ben Adcock. On asymptotic structure in compressed sensing. *arXiv preprint arXiv:1406.4178*, 2014.

[90] Justin Romberg. Compressive sensing by random convolution. *SIAM Journal on Imaging Sciences*, 2(4):1098–1128, 2009.

[91] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.

[92] Christian Schmaltz, Pascal Gwosdek, Andrés Bruhn, and Joachim Weickert. Electrostatic halftoning. In *Computer Graphics Forum*, volume 29, pages 2313–2327. Wiley Online Library, 2010.

[93] Emil Y Sidky, Chien-Min Kao, and Xiaochuan Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct. *Journal of X-ray Science and Technology*, 14(2):119–139, 2006.

[94] Steven Smith. *Digital Signal Processing: A Practical Guide for Engineers and Scientists: A Practical Guide for Engineers and Scientists*. Newnes, 2013.

[95] Tanja Teuber, Gabriele Steidl, Pascal Gwosdek, Christian Schmaltz, and Joachim Weickert. Dithering by differences of convex functions. *SIAM Journal on Imaging Sciences*, 4(1):79–108, 2011.

[96] Joseph Thompson, Hans Geiger, Ernest Marsden, Ernest Rutherford, Henry Moseley, and James Chadwick. On the structure of the atom. *Phylos Mag*, 7:237–265, 1904.

[97] Joel Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[98] Jayakrishnan Unnikrishnan and Martin Vetterli. Sampling and reconstructing spatial fields using mobile sensors. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 3789–3792. Ieee, 2012.

[99] Jayakrishnan Unnikrishnan and Martin Vetterli. Sampling and reconstruction of spatial fields using mobile sensors. *Signal Processing, IEEE Transactions on*, 61(9):2328–2340, 2013.

[100] Jayakrishnan Unnikrishnan and Martin Vetterli. Sampling high-dimensional bandlimited fields on low-dimensional manifolds. *Information Theory, IEEE Transactions on*, 59(4):2103–2127, 2013.

[101] Shreyas Vasanawala, Mark Murphy, Marcus Alley, P Lai, Kurt Keutzer, John Pauly, and Michael Lustig. Practical parallel imaging compressed sensing mri: Summary of two years of experience in accelerating body mri of pediatric patients. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1039–1043. IEEE, 2011.

[102] Yves Wiaux, Laurent Jacques, Gilles Puy, Anna Scaife, and Pierre Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 395(3):1733–1742, 2009.

[103] Yves Wiaux, Gilles Puy, Yannick Boursier, and Pierre Vandergheynst. Spread spectrum for imaging techniques in radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 400(2):1029–1038, 2009.

[104] Stefanie Winkelmann, Tobias Schaeffter, Thomas Koehler, Holger Eggers, and Olaf Doessel. An optimal radial profile order based on the golden ratio for time-resolved mri. *Medical Imaging, IEEE Transactions on*, 26(1):68–76, 2007.