HAL
open science

On the generation of sampling schemes for Magnetic Resonance Imaging

Claire Boyer, Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, Pierre Weiss

To cite this version:

Claire Boyer, Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, Pierre Weiss. On the generation of sampling schemes for Magnetic Resonance Imaging. SIAM Journal on Imaging Sciences, 2016, 9 (4), pp.2039-20972. hal-01373758v2

HAL Id: hal-01373758

https://hal.science/hal-01373758v2

Submitted on 12 Jan 2017

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

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

On the generation of sampling schemes for Magnetic Resonance Imaging.

Claire Boyer [*] Nicolas Chauffert[†] Philippe Ciuciu [‡]
Jonas Kahn[§] Pierre Weiss[¶]

September 28, 2016

Abstract

Magnetic resonance imaging (MRI) is probably one of the most successful application fields of compressed sensing. Despite recent advances, there is still a large discrepancy between theories and most actual implementations. Overall, many important questions related to sampling theory remain open. In this paper, we attack one of them: given a set of sampling constraints (e.g. measuring Fourier coefficients along physically plausible trajectories), how to optimally design a sampling pattern? We first outline three aspects that should be carefully designed by inspecting the literature, namely *admissibility*, *limit of the empirical measure* and *coverage speed*. To address them jointly, we then propose an original approach which consists of projecting a probability distribution onto a set of admissible measures. The proposed algorithm permits to handle arbitrary constraints and automatically generates efficient sampling patterns for MRI as shown on realistic simulations. We achieve a 20-fold undersampling factor at very high 2D resolution (100 μm isotropic) on physically plausible sampling trajectories with a gain in SNR of 2-3 dB on reconstructed MR images as compared to more standard sampling patterns (e.g. radial, spiral).

[*]Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, CNRS (claire.boyer11@gmail.com).

[†]Inria Saclay, Parietal team. CEA/NeusoSpin, 91191 Gif-sur-Yvette, France (nicolas.chauffert@cea.fr).

[‡]Inria Saclay, Parietal team. CEA/NeusoSpin, 91191 Gif-sur-Yvette, France (philippe.ciuciu@cea.fr).

[§]Laboratoire Painlevé, UMR8524 Université de Lille 1, CNRS. Cité Scientifique Bât. M2, 59655 Villeneuve d'Asq Cedex, France (jonas.kahn@math.univ-lille1.fr).

[¶]ITAV, USR 3505. PRIMO Team, Université de Toulouse, CNRS and Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, CNRS (pierre.armand.weiss@gmail.com).

Contents

1 Introduction **2**

2 MRI acquisition and reconstruction **6**
2.1 MRI acquisition 7
2.2 MRI reconstruction 8
2.2.1 Modeling the observation operator 9
2.2.2 ℓ^1-norm reconstruction 10

3 Theoretical foundations of variable density sampling **10**
3.1 The first compressed sensing results 11
3.2 The emergence of variable density sampling 12
3.3 Variable density sampling with structured sparsity 12
3.4 Variable density sampling with structured acquisition 13

4 Generation of sampling schemes by projection **15**
4.1 Pushforward measures 15
4.2 Measure sets in the MRI context 17
4.3 Measuring distances between measures 18
4.4 Design of sampling scheme as a projection problem 19

5 Numerical implementation **19**
5.1 The attraction-repulsion formulation 19
5.2 Projected gradient descents 20
5.3 Discretization of the parameterization sets 21
5.4 Implementation details 24

6 Results **24**
6.1 Constraints used in our experiments 25
6.2 Empirical choice of the target density π 25
6.3 New sampling patterns 26
6.3.1 Standard resolution imaging 27
6.3.2 Very high resolution imaging 30

7 Conclusion **34**

1 Introduction

Magnetic resonance imaging (MRI) is one of the flagship applications of compressed sensing (CS). The combination of CS and MRI initially appeared

in [34], very shortly after the seminal CS papers [11, 10, 19]. However, the way CS was originally implemented on real scanners strongly departed from theory. Despite having limited theoretical foundations, empirical implementations turned out to be useful in practice and triggered a massive interest both in the MRI and mathematics communities. Since then, many researchers have tried improving the way CS-MRI is implemented. These attempts can be divided into two distinct tracks:

- The first one consists of reducing the coherence of the sensing basis by using techniques termed phase scrambling [26] or spread spectrum [46]. This can be implemented using specific radio-frequency pulses [26] or shim-coils [46]. A few available theories support these techniques [52, 47].

- The second one consists of keeping the sensing basis unchanged: images are acquired by collecting Fourier samples and assuming sparsity in a wavelet basis. The problem then reformulates as the design of new sampling patterns either in 2D or in 3D. Examples ranging in this second category include patterns made of parallel lines [34], radial lines [65], spirals [43], noisy spirals [36], Poisson disc sampling [64], ...

The second approach is adopted more widely, probably due to its ease of implementation since collecting Fourier coefficients along lines for instance is practically feasible for the magnetic field gradients. In addition, it is unclear that using totally incoherent bases is better for structured signals, as illustrated by [51]. In this paper, we will therefore focus on the second track too, especially in the context of 2D sampling even though our approach can extend to 3D imaging.

Contributions. The sampling patterns proposed in the literature for 2D imaging[1] may seem somewhat arbitrary (horizontal parallel lines, radial lines, spirals, ...) but they actually match what the magnetic field gradients can easily play while satisfying the hardware constraints. For instance, although many existing theories recommend using completely random sampling patterns, it is not clear that adding random perturbations to a spiral will improve its practical efficiency. In 3D, the use of parallel lines in the orthogonal direction (i.e., readout direction) to the slices of interest permits to easily implement a 2D variable density sampling (VDS) [34] within each slice and in this regard to stick to the VDS theory [48, 1, 16, 31]. However,

[1] Slice-by-slice acquisition.

in a 3D perspective, this 2D-VDS is likely suboptimal since high frequencies along the readout direction are sampled too densely, hence increasing the scanning time uselessly as compared to a pure 3D VDS. Moreover, the recent CS theory with block-structured acquisition [5] predicts that the above mentioned parallel line strategy will produce some artifacts in the readout direction.

The first contribution of this paper is to provide a review of existing theoretical CS results in Section 3. This review permits to establish general principles for designing efficient sampling patterns.

The second and most significant contribution is to show that our recent projection algorithm [15] can be used to generate feasible sampling patterns complying with the proposed principles. The main idea is to project a probability distribution onto a space of admissible measures. The reader can look at the result on Figure 1 to get an idea of what the algorithm does: given an initial distribution (here a piece of text), the algorithm finds a sampling pattern complying with physical constraints that best fits the distribution. The core of this algorithm was proposed by a subset of the authors [15]. It is based on the use of fast projections algorithms on the set of admissible curves for MRI proposed in [17]. We also analyze new constraint sets relevant for applications, based on unions of segments of variable length.

The third and last contribution consists of a series of numerical experiments led on standard and high resolution 2D MR images. The results suggest that the proposed sampling patterns significantly outperform more traditional approaches (radial and spiral trajectories).

Related works. A few works in the literature address the problem of optimizing the acquisition space coverage using computational techniques.

The contributions [39, 58] propose an algorithm to cover the whole k-space as fast as possible by relying on techniques used for missile guidance. This idea departs from the proposed one since the objective of these authors was to satisfy Shannon's sampling theorem, meaning that the samples should cover the space uniformly.

In [32, 18], the authors have proposed to synthesize random feasible trajectories using optimization techniques. Their idea was to generate random control points uniformly distributed over the surface of a sphere. They then searched for a feasible trajectory that passed close to them using second order cone programming. Multiple random trajectories were then generated this way and a genetic algorithm was involved to select the most relevant

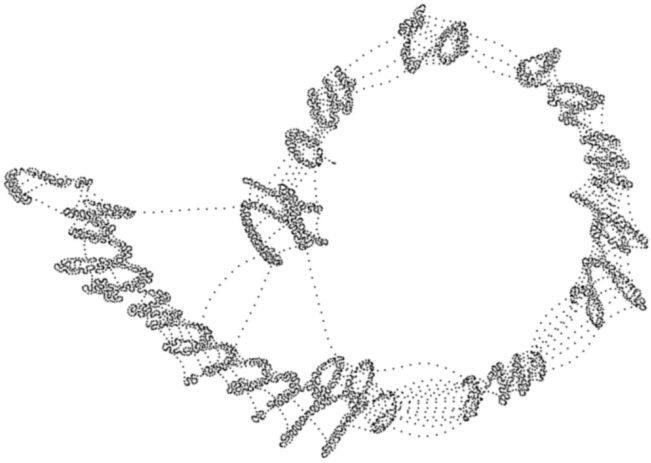

Figure 1: A glance at our contribution: our algorithm generates a sampling pattern complying with the MRI scanner constraints in which sampling locations consist of a piece of text, namely *How to sample me efficiently?*.

ones so as to ensure a uniform k-space coverage. This idea does not stem from a clear sampling theory and is based on randomness in contrast to the proposed approach.

In [7], two of the authors of this paper proposed to generate sampling schemes with ideas quite similar to the ones exposed here. Given a set of blocks of measurements (e.g. segments), an efficient drawing distribution was constructed by solving an original convex program. Drawing i.i.d. blocks usually leads to suboptimal image reconstruction results since neighbouring blocks can be sampled multiple times, hence local clusters of samples can emerge at the expense of a complete coverage of the k-space.

Finally, a few authors [54, 49, 20] have borrowed ideas from statistical design for generating efficient sampling trajectories. In [54], the key point is to fix a set of feasible trajectories (e.g. pieces of spirals) and to select them iteratively by picking the one that brings the largest amount of information at each step. Hence, finding the most meaningful trajectory becomes computationally intensive and hardly compatible with a real-time acquisition. The main contribution of [49, 20] is to propose alternative approaches to reduce the computational burden, by working on training images. These adaptive approaches suffer from a few drawbacks. First, the whole versatility of MRI scanners is not exploited since fixed trajectories are imposed. Our formalism does not impose such a restriction. Second, even though adaptivity to the sampled image may seem appealing at the first glance, it still seems unclear whether this learning step is really helpful [2]. Finally, these approaches strongly depart from existing sampling theories, whereas our contribution is still motivated by solid and recently established theories.

Outline of the paper. We first recall in Section 2 how data are collected in MRI and then how MR images are reconstructed. We then propose a short review of theoretical compressed sensing results in Section 3. Section 4 describes the main idea of the paper: we explain how the design of sampling patterns can be formulated as a measure projection problem. We then develop a numerical algorithm to solve this projection problem in Section 5. Finally, numerical experiments in a retrospective CS framework are conducted in Section 6 and conclusions are drawn in Section 7.

2 MRI acquisition and reconstruction

In this section, we start by presenting how MRI data are collected in a concise manner. This summary is motivated by the fact that in many papers

dealing with retrospective compressed sensing for 2D MRI, the authors assume that the data are collected point-by-point. This strategy is feasible in practice but it dramatically slows down the acquisition. In order to really accelerate acquisition, the data should be acquired along continuous (eg, lines), piecewise continuous (eg, spokes) or more regular trajectories (eg, spirals). We then describe the ℓ^1 reconstruction framework we adopt in this paper to get MR images.

2.1 MRI acquisition

In MRI, images are usually sampled in the so-called k-space, which corresponds to the 2D or 3D Fourier domain [60]. The acquisition domain can be slightly different (i) in the parallel MRI context, where spatial sensitivity encoding associated with the multiple channel coil introduces a convolution in k-space [56, 45] or (ii) when shim coils (e.g. phase scrambling/spread spectrum) are involved [38, 26, 46]. In this paper, we focus on the Fourier domain, but the proposed ideas could be extended to these other settings.

The samples lie along parameterized curves $s : [0, T] \mapsto \mathbb{R}^d$, where $d \in \{2, 3\}$ denotes the image dimensions. The i-th coordinate of s is denoted s_i. Let $u : \mathbb{R}^d \to \mathbb{C}$ denote a d dimensional image and $\hat{u}$ be its Fourier transform. Given an image u, a curve $s : [0, T] \to \mathbb{R}^d$ and a sampling period Δt (also termed dwell time in MRI), the image u shall be reconstructed from the following dataset:

$$\mathcal{E} = \left\{ \hat{u}(s(j\Delta t)), 0 \leqslant j \leqslant \left\lfloor \frac{T}{\Delta t} \right\rfloor \right\}. \tag{1}$$

In what follows, the scalar $m = \left\lfloor \frac{T}{\Delta t} \right\rfloor + 1$ denotes the total number of collected samples. Hence, vector $\mathbf{y} \in \mathcal{C}^m$ with components $y_j = \hat{u}(s(j\Delta t))$ denotes the vector of measurements. In this paper, we neglect typical distortions occurring in MRI such as noise, geometric distortions, signal loss at tissue/air interfaces or off-resonance effects which would affect the dataset $\mathcal{E}$ in Equation (1). We also neglect imprecisions in the trajectory due to Eddy currents that induce gradient errors [9]. These are very important features that we plan to consider in forthcoming works (see [62, 63] for details).

The gradient waveform associated with a curve s is defined by $g(t) = \gamma^{-1}\dot{s}(t)$, where γ denotes the gyro-magnetic ratio [27]. The gradient waveform is obtained by supplying electric power to gradient coils. This electric current has a bounded amplitude and cannot vary too rapidly (slew rate). Mathematically, these constraints read:

$$\|g\| \leqslant G_{\max} \qquad \text{and} \qquad \|\dot{g}\| \leqslant S_{\max}$$

where $\|\cdot\|$ denotes either the ℓ^∞-norm defined by

$$\|f\|_\infty := \max_{1\leq i\leq d} \sup_{t\in[0,T]} |f_i(t)|,$$

or the $\ell^{\infty,2}$-norm defined by

$$\|f\|_{\infty,2} := \sup_{t\in[0,T]} \left(\sum_{i=1}^{d} |f_i(t)|^2\right)^{\frac{1}{2}}.$$

Additional affine constraints (eg, k-space position at the echo time) could be added depending on the targeted application (e.g. structural or functional imaging) and the chronogram of the sequence (i.e. the interplay between the orthogonal gradients). For instance, s usually starts from the k-space center, i.e. $s(0) = 0$. Multiple sampling trajectories (or interleaves) starting from the origin can be used to improve the signal-to-noise ratio: this typically leads to additional linear constraints of type $s(k \cdot TR) = 0$ for all $k \in \mathbb{N}$, where TR is the time of repetition, ie the time that separates the delivery of two successive radio-frequency pulses. Overall these additional constraints can be summarized under the compact form $A(s) = b$ where A is a linear mapping and b is a fixed vector. We refer to [27, 17] for a more thorough discussion on these issues.

A sampling trajectory $s : [0, T] \to \mathbb{R}^d$ will be said *admissible* if it belongs to the convex set:

$$\mathcal{S}_T := \left\{ s \in \left(\mathcal{C}^2([0,T])\right)^d, \|\dot{s}\| \leqslant \alpha, \|\ddot{s}\| \leqslant \beta, A(s) = b \right\}, \qquad (2)$$
$$\text{with} \quad \alpha = \gamma G_{\max} \quad \text{and} \quad \beta = \gamma S_{\max}.$$

In addition to the above mentioned kinematics constraints, important considerations regarding the MR signal acquisition have to be taken into account. The first one is the exponential signal intensity decay due to transverse relaxation of spins. In this paper, we will assume that the MR signal is available for 200 ms and therefore limit the trajectories to that sampling time. The second supplementary constraint is the maximal number of samples that can be stored in the buffer of the analog-to-digital converter. This buffer length may depend on the imaging device but here we set this constraint to 8192 samples per readout.

2.2 MRI reconstruction

Reconstruction of MRI images from k-space measurements $\mathcal{E}$ is an involved problem that has been studied thoroughly. The main technical difficulties

to solve it are (i) the fact that k-space locations $s(j\Delta t)$ do not lie on a Cartesian grid, (ii) the ill-posedness of the problem, (iii) the large image dimensions and (iv) an inaccurate knowledge of the acquisition operator owing to magnetic field inhomogeneities, subject movements,... In this paragraph, we describe the methodology adopted to solve issues (i), (ii) and (iii). Although of primary importance, problem (iv) is beyond the scope of this paper.

2.2.1 Modeling the observation operator

In order to define an inverse problem, we first need to model the sampling operator S that maps u to its sampling set $(\hat{u}(s(j\Delta t)))_{0\leq j\leq m-1}$. The mapping S should map a continuous space like $L^2(\mathbb{R}^d)$ to $\mathbb{C}^m$. For the purpose of practical implementation, we consider instead a mapping $\mathbf{S} : \mathbb{C}^n \to \mathbb{C}^m$ between two finite dimensional spaces. By assuming that $u = h \star v$, where h is an interpolation kernel and

$$v = \sum_{0\leq i_1,\ldots,i_d\leq n^{1/d}-1} v_{i_1,\ldots,i_d} \cdot \delta_{\frac{(i_1,\ldots,i_d)}{n^{1/d}}}, \tag{3}$$

the analytical expression of $\hat{u}$ is given by:

$$\hat{u}(\xi) = \hat{h}(\xi) \cdot \left(\sum_{0\leq i_1,\ldots,i_d\leq n^{1/d}-1} v_i \cdot \exp\left(-\frac{2\imath\pi}{n^{1/d}} \langle \xi, i \rangle \right) \right), \tag{4}$$

where $i = (i_1, \ldots, i_d)$. Computing the sums

$$\sum_{0\leq i_1,\ldots,i_d\leq n^{1/d}-1} v_i \cdot \exp\left(-\frac{2\imath\pi}{n^{1/d}} \langle \xi, i \rangle \right) \tag{5}$$

for all $\xi \in \{s(j\Delta t),\ 0 \leq j \leq m-1\}$ using the above expression directly is an $O(nm)$ algorithm. This complexity is prohibitive for large m and n. If the samples lie on a Cartesian grid, the complexity can be decreased to $O(n\log(n))$ using fast Fourier transforms (FFT). In this paper, we depart from this simplifying assumption by considering non-Cartesian sampling schemes. We therefore need to resort to more advanced techniques called Non-Uniform Fast Fourier transforms (NUFFT) [30]. They come with good parallel implementations on multi-core or GPU architectures [29, 22]. All the numerical experiments of this paper are based on the NUFFT3 package on multi-core architecture delivered by Chemnitz university [29]. In all the numerical experiments, we simply set $h = \delta$. This is a reasonable choice, given that we only work on simulations with discretized images.

2.2.2 ℓ^1-norm reconstruction

A large set of reconstruction procedures have been developed over the past years. The simplest techniques are based on regridding [28, 44]. Lately, techniques based on ℓ^1-regularization of wavelet, frame, or learned dictionary coefficients were proven more efficient for large undersampling ratios [37, 24, 6, 50]. In this paper, we resort to ℓ^1 regularization using an orthogonal wavelet transform. This setting comes with strong theoretical guarantees of reconstruction, as will be seen in Section 3.

The idea is to decompose the image $\mathbf{u}$ in an orthogonal wavelet basis. Let $\mathbf{\Psi} \in \mathbb{C}^{n\times n}$, denote the wavelet synthesis operator. Here, we simply assume that $\mathbf{\Psi}$ decomposes the real and complex part of $\mathbf{u}$ separately using the same orthogonal basis. The wavelet coefficients of a discrete image $\mathbf{u}$ are denoted $\mathbf{x} = \mathbf{\Psi}^*\mathbf{u}$. If $\mathbf{u}$ describes a piecewise smooth image, it is well known that its wavelet coefficients $\mathbf{x}$ are compressible. This observation motivated the introduction of the basis pursuit algorithm that consists of solving:

$$\min_{\mathbf{x}\in\mathbb{R}^p, \mathbf{S\Psi x}=\mathbf{y}} \|\mathbf{x}\|_1. \tag{6}$$

The use of the ℓ^1-norm is often justified as a convex relaxation of the ℓ^0-counting function, that counts the number of nonzero components in $\mathbf{x}$. When the data $\mathbf{y}$ is degraded by noise, the exact constraint $\mathbf{S\Psi x} = \mathbf{y}$ is relaxed and transformed into a penalized data consistency term. Then, the following quadratic programming problem has to be solved instead:

$$\min_{\mathbf{x}\in\mathbb{R}^p} \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{S\Psi x} - \mathbf{y}\|_2^2. \tag{7}$$

Scalar $\lambda > 0$ is a parameter that balances the quadratic data consistency term and the regularization term. In all the paper, $\mathbf{\Psi}$ is defined as an orthogonal wavelet transform with Symlet wavelets and 3 vanishing moments.

Problem (7) can be solved by using various well documented techniques. In this paper we will use an accelerated proximal gradient descent algorithm (aka FISTA) [41, 42, 3].

3 Theoretical foundations of variable density sampling

In this section, we briefly review the existing theoretical CS results. The conclusions of this section motivate the main contribution of this work: the design of undersampling patterns by projecting i.i.d. drawings on measure sets.

3.1 The first compressed sensing results

Let us first describe the compressed sensing theory as it appeared in the seminal paper [10] and more recently in [12]. The authors consider an orthogonal matrix

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_n^* \end{pmatrix}.$$

They propose to construct a random sensing matrix as:

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

where the integers $J_k \in \{1, \ldots, n\}$ are i.i.d. uniform random variables. Knowing that $\mathbf{y} = \mathbf{A}\mathbf{x}$ in the noise-free case, the authors propose to recover $\mathbf{x}$ by solving Problem (6). Let

$$\bar{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^p, \mathbf{A}\mathbf{x} = \mathbf{y}}{\arg\min} \|\mathbf{x}\|_1. \tag{8}$$

In this context, their main result reads as follows:

Theorem 1. *Assume that* $\mathbf{x}$ *is* s*-sparse, i.e. that it contains at most* s *nonzero components. If the number of measurements* m *satisfies:*

$$m \geq Cs \left(n \max_{1 \leq k \leq n} \|\mathbf{a}_k\|_\infty^2 \right) \log \left(\frac{n}{\epsilon} \right),$$

where C *is a universal constant, then* $\bar{\mathbf{x}} = \mathbf{x}$ *with probability* $1 - \epsilon$.

Moreover, the authors show that if the measurements are noisy, i.e. $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{b}$ is a random perturbation, then the solution of the relaxed Problem (7) also provides stable reconstruction results.

The *coherence* $\kappa(\mathbf{A}_0) = n \max_{1 \leq k \leq n} \|\mathbf{a}_k\|_\infty^2$ belongs to the interval $[1, n]$. In particular, $\kappa(\mathbf{F}) = 1$ and $\kappa(\mathbf{Id}) = n$. In the favorable case of a Fourier transform, this theorem indicates that only $s \log\left(\frac{n}{\epsilon}\right)$ measurements are enough to perfectly recover an arbitrary s-sparse signal.

Even though this type of theorem got a huge impact in the literature, it is not applicable to MRI. The natural transform $\mathbf{A}_0$ in MRI reads $\mathbf{A}_0 = \mathbf{F}^* \mathbf{\Psi}$, i.e. the product of Fourier and wavelet transforms. In that case, one can show that $\kappa(\mathbf{A}_0) = O(n)$. Theorem 1 is thus irrelevant in such a setting.

3.2 The emergence of variable density sampling

In most practical applications, the transforms $\mathbf{A}_0$ are coherent. This is the case in MRI and more generally in Fourier or space imaging [1]. A simple technique to break down the so-called "coherence barrier" consists of drawing the coherent samples more often than incoherent ones [48, 31, 14]. Let us clarify this idea. Let $\boldsymbol{\pi} \in \Delta_n$ denote the distribution of the i.i.d. random variables J_k, i.e. $P(J_k = i) = \boldsymbol{\pi}_i$. The following theorem [14] justifies the use of variable density sampling.

Theorem 2. *Assume that* $\mathbf{x}$ *is* s*-sparse. Set*

$$\boldsymbol{\pi}_k = \frac{\|\mathbf{a}_k\|_\infty^2}{\sum_{j=1}^n \|\mathbf{a}_j\|_\infty^2}.$$

If the number of measurements satisfies

$$m \geq Cs \left(\sum_{j=1}^n \|\mathbf{a}_j\|_\infty^2 \right) \log\left(\frac{n}{\epsilon}\right),$$

where C *is a universal constant, then* $\bar{\mathbf{x}} = \mathbf{x}$ *with probability* $1 - \epsilon$.

In the MRI case, one can show that $\sum_{j=1}^n \|\mathbf{a}_j\|_\infty^2 = O(\log(n))$. Hence, it becomes possible to perfectly reconstruct an s-sparse image with $O(s \log(n)^2)$ measurements. Let us mention that variable density sampling was the basis for the seminal paper on compressed sensing MRI [34]. Theorem 2 is a first argument that supports that type of technique.

3.3 Variable density sampling with structured sparsity

Theorem 2 is quite attractive from a theoretical point of view. A simple analysis however suggests that it is still insufficient to justify the use of compressed sensing in MRI. First, the constant appearing in the O is large. This may only be an artifact of the proofs, but it is currently unknown how much it can be lowered. More importantly, the term $\log(n)^2$ that appears when using the Fourier-Wavelet pair cannot be improved by using only variable density sampling arguments. Most often, the logarithmic terms are disregarded and considered as negligible. It seems however important to look at them carefully since for instance $\log(1024 \times 1024)^2 = 192$. A method that needs $192s$ samples to reconstruct a 1024×1024 image is actually of limited practical interest.

A recent breakthrough that has been proposed in [1], consists of exploiting *structured sparsity* to derive better reconstruction guarantees. In the case of imaging, structured sparsity may mean that the wavelet subbands become sparser as the scale increases. Let us provide a typical result from this active field of research, coming from our recent work [5].

Let $(\Omega_j)_{0\leq j\leq J}$ denote the wavelet subbands with J the number of decomposition levels. Assume that $\mathbf{x}$ is supported on $S \subset \{1,\ldots,n\}$ with $|S\cap\Omega_j| = s_j$. This means that $\mathbf{x}$ restricted to the subband Ω_j is s_j-sparse. This model is called sparsity by levels in [1]. In such a setting, the following theorem holds.

Theorem 3. *Assume that matrix $\mathbf{A}_0$ is the product of the Fourier and Haar wavelet matrices. Let $j(k)$ denote the scale of index k, i.e. $j(k) = j$ if $k \in \Omega_j$. Set*

$$\boldsymbol{\pi}_k = \frac{2^{-j(k)}\sum_{p=0}^{J} 2^{-|j(k)-p|/2} s_p}{\gamma} \quad \textit{with} \quad \gamma = \sum_{j=0}^{J}\sum_{p=0}^{J} 2^{-|j-p|/2} s_p.$$

Set

$$m \geq C\gamma \log(s) \log\left(\frac{n}{\epsilon}\right) \tag{9}$$

where C is a universal constant. Under the previous sparsity-by-level hypothesis $\bar{\mathbf{x}} = \mathbf{x}$ with probability $1-\epsilon$.

Note that contrarily to previous results, the drawing probability $\boldsymbol{\pi}$ in Theorem 3 explicitly depends on the sparsity structure. The number of measurements in Theorem 3 is always lower than that in Theorem 2, but the gain once again depends on the signal support. Using the oversampling trick proposed in [1], the term $\log(s)$ in Equation (9) can also be discarded.

3.4 Variable density sampling with structured acquisition

Another feature that was not considered in the seminal works on compressed sensing is *structured acquisition*. In practice, sampling isolated measurements takes too much time to be appealing in practice. In MRI, radio-interferometry, X-ray tomography and many other systems, the samples are collected either line by line or along more complex trajectories (eg, spirals). In some cases (X-ray, PET imaging), the readout shape is imposed by the physics of acquisition. The vast majority of compressed sampling schemes are based on heuristic sampling patterns such as radial lines [33, 65], spirals

[57], noisy spirals [64] or other exotic shapes. Even though they often perform well, until very recently, theoretical results that allow to justify their use in practice were missing.

In the spirit of traditional Shannon's sampling theorem, the papers [61, 23] propose theoretical guarantees for the reconstruction of bandlimited functions from sets of measurements along lines or curves. These results usually lead to sampling patterns that span the acquisition space uniformly.

Concomitantly to these developments, we have proposed a few results in [4, 14, 5] to explain the success of structured acquisitions by using sparsity assumptions on the signal to be reconstructed. These results promoted variable density sampling strategies. In [4, 5], theoretical guarantees were derived for block sampling strategies: instead of probing isolated measurements, fixed groups of measurements are acquired, irrespective of the structured sparsity assumptions. Still in these references, it is shown that only specific sparsity patterns that depend on the acquisition constraints can be recovered.

In [14], we have proposed to sample signals using generic stochastic processes. The conclusions of this work actually define the starting point of the present paper. We first gave a mathematical definition of variable density samplers as sequences of stochastic processes with a prescribed limit empirical measure, termed density. We have also shown through mathematical arguments and experimental validation that the key features characterizing the efficiency of a variable density sampler are:

i) The *density*: the stochastic processes should cover the space non-uniformly according to a certain density.

ii) The *coverage speed*: a sampler will be efficient only if it covers the space quickly enough. More precisely, we proved that the mixing time should be as low as possible. The mixing time characterizes the speed at which the empirical measure converges to its limit.

Since most readers may not be familiar with these concepts, we illustrate them in Figure 2. In this Figure, we constructed three different variable density samplers with a density $\boldsymbol{\pi}$ illustrated on Figure 2 (a). This density was defined as suggested by Theorem 2 by setting $\boldsymbol{\pi}_k \propto \|\mathbf{a}_k\|_\infty$, where $\mathbf{a}_k$ is the k-th row of the Fourier-wavelet matrix $\mathbf{A} = \mathbf{F}^*\boldsymbol{\Psi}$. The wavelet transforms was defined using Symlet filters. The sampling schemes in Figure 2(b-d) all cover the 256×256 grid non uniformly with 20% measurements. For the sampling patterns (b) and (d), the samples density in a given region of space looks like $\boldsymbol{\pi}$. The same property holds for (c) even though this

does not seem obvious at the first glance. This property of non uniform coverage is captured by the sampler's density (more precisely, the limit of the empirical measure, see Section 4), i.e. feature i).

It is pretty intuitive when looking at Figure 2(b-d) that they are likely to have different efficiencies. The samples in Figure 2(b) cover the space quite uniformly locally, while the samplers in Figure 2(c)-(d) leave large portions of the space unexplored. Clearly, this lack of information might result in poor reconstruction results. This feature is captured by the notion of coverage speed, i.e. feature ii). Let us mention that the so-called Poisson disc sampling [8, 40], which is quite popular in compressed sensing MRI, is also based on the idea of covering the k-space as fast as possible.

4 Generation of sampling schemes by projection

In this section, we describe the main idea of this paper. We propose a general principle to construct samplers that comply with the three following guidelines:

- *Admissibility*: the sampler should be feasible. For instance in the MRI case, the samples should belong to a set of segments or curves defined in Equation (2).
- *Density*: as mentioned earlier, a sampler should approximate a given density π.
- *Coverage speed*: the sampler should cover the space as fast as possible.

This problem is probably more complex than it looks at first sight. Here, we first recall the notion of pushforward measure that is crucial to establish our algorithm. We then present its overall principle. Let us mention that this idea, the associated algorithm and some of its theoretical guarantees were presented in more detail in [15] for a completely different purpose, namely image stippling or continuous line drawing.

4.1 Pushforward measures

As shown in Figure 2, the density (a) is somehow similar to the sampling schemes (b-d). To make this statement more accurate, we resort to measure theory. Let us introduce a few definitions. Here, we work on the space $\Omega = [0, 1]^d$ where $d = 2$ denotes the space dimension. Extensions to other dimensions are straightforward. We equip Ω with the Borel algebra $\mathcal{B}$. Let

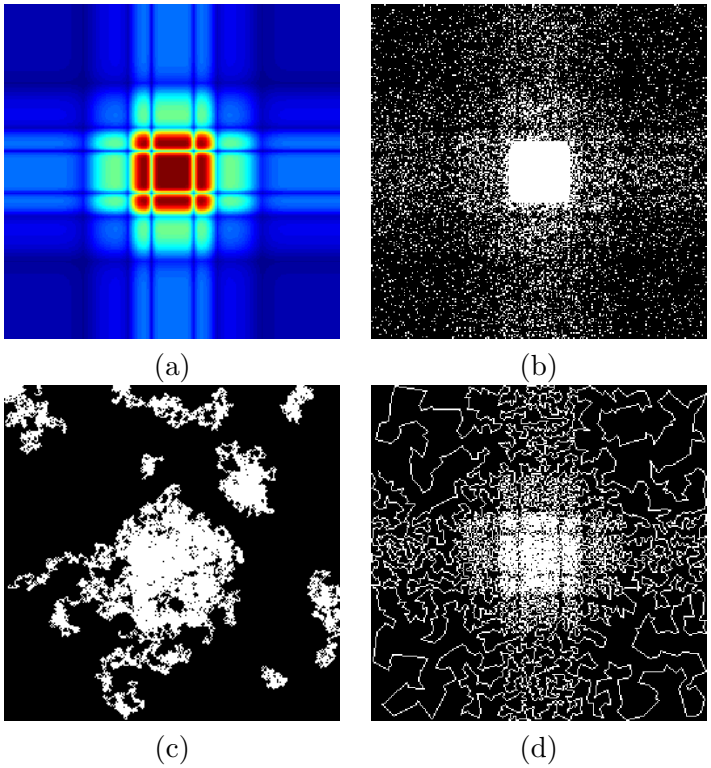

Figure 2: A few variable density samplers. (a) density π. (b) π-variable density sampler with i.i.d. drawings. (c) π-variable density sampler constructed using a Markov chain. (d) π-variable density sampler with a traveling salesman problem solution.

(X, Σ) be a measurable space, $f : X \to \Omega$ denote a measurable mapping and $\mu : X \to [0, +\infty]$ denote a measure. The *pushforward* measure of μ is denoted $\nu : \mathcal{B} \to \mathbb{R}$ and defined by

$$\nu(B) = f_*\mu(B) = \mu\left(f^{-1}(B)\right), \quad \forall B \in \mathcal{B}.$$

The function f is called *parameterization* of ν. Note that if μ is a probability measure, then ν is also a probability measure. Let us now illustrate this concept with two concrete examples.

Example 1 (Atomic measures)**.** *The set of m points in Figure 2 (b) can be ordered and parameterized as a function $f : \{1, \ldots, m\} \to \Omega$, where $f(i) = p_i$ denotes the i-th point. Set μ as the normalized counting measure defined for any set $I \subseteq \{1, \ldots, m\}$ by $\mu(I) = \frac{|I|}{m}$. Let $B \in \mathcal{B}$, then $f^{-1}(B)$ is the set of indices of points in B. The pushforward of μ is therefore an atomic measure defined by*

$$\nu = f_*\mu = \frac{1}{m}\sum_{i=1}^{m} \delta_{p_i}.$$

Example 2 (Measures supported on curves)**.** *Let $s : [0, T] \to \Omega$ denote a parameterized curve. Set μ as the normalized Lebesgue measure on $[0, T]$ defined for any interval $I \subseteq [0, T]$ by $\mu(I) = \frac{|I|}{T}$. Then $\nu(B) = s_*\mu(B)$ measures the relative time spent by the curve s in the set B.*

4.2 Measure sets in the MRI context

Now, let $\mathcal{P}$ denote a set of admissible parameterizations. Let $\mathcal{M}(\mathcal{P})$ be the set of pushforward measures associated with elements of $\mathcal{P}$:

$$\mathcal{M}(\mathcal{P}) = \{\nu = f_*\mu,\ f \in \mathcal{P}\}.$$

Depending on the context, μ will be either the normalized counting measure or the normalized Lebesgue measure. Hereafter, we will be particularly interested in exploring 3 different sets $\mathcal{P}$ which are particularly relevant in MRI.

Isolated points. The set of sums of m Dirac delta functions is:

$$\mathcal{M}(\Omega^m) = \left\{\nu = \frac{1}{m}\sum_{i=1}^{m} \delta_{p_i},\ p_i \in \Omega\right\}. \tag{10}$$

This case is time-consuming and thus inefficient in MRI, but it is commonly used in retrospective CS simulations. We will therefore use it as reference.

Segments of variable length. A more promising parameterization is the set of N segments with variable lengths (or crossed with variable speed at constant time). To this end, let

$$\mathcal{L} = \{\lambda : [0,1] \to \Omega,\ \exists (x_1, x_2) \in \Omega^2,\ \lambda(t) = (1-t)x_1 + tx_2,\ \forall t \in [0,1]\}.$$

The associated set of measures is defined by:

$$\mathcal{M}(\mathcal{L}^N) = \left\{ \nu = \frac{1}{N} \sum_{i=1}^{N} (\lambda_i)_* \mu,\ \lambda_i \in \mathcal{L} \right\}, \tag{11}$$

where μ is the Lebesgue measure on $[0,1]$. In this description, we implicitly assume that segments of different lengths are traversed at different speeds since the traversal time is fixed to 1.

Admissible curves for MRI. It corresponds to $\mathcal{M}(\mathcal{S}_T)$, where $\mathcal{S}_T$ is defined by Equation (2). This case allows us to exploit the full sampling potential in MRI.

4.3 Measuring distances between measures

Pushforward measures allow us to map a sampling pattern to the space of probability measures $\mathcal{M}_\Delta$ on Ω. The target distribution π also belongs to $\mathcal{M}_\Delta$. This mapping therefore permits to perform quantitative comparisons by defining distances on $\mathcal{M}_\Delta$. Various distances exist to compare probability measures (e.g. total variation, Wasserstein distance, ...). In this work, motivated by our previous results in [15], we propose to construct a distance as follows. Let $h : \Omega \to \mathbb{R}$ denote a continuous function with a Fourier series that does not vanish. The following mapping:

$$\mathrm{dist}(\pi, \nu) = \|h \star (\pi - \nu)\|_2^2 \tag{12}$$

defines a distance (or metric) on $\mathcal{M}_\Delta$. Moreover, we showed in [15] that it metrizes the weak convergence. Therefore, if π and ν are sufficiently weakly close, their distance will be small.

This measure is interesting numerically for at least two reasons. First, it has a simple direct expression compared to more standard tools such as the Wasserstein distance. Second, it is quadratic and this property will be exploited intensively in the numerical algorithms.

4.4 Design of sampling scheme as a projection problem

The distance on $\mathcal{M}_\Delta$ being defined, we can construct a sampler by solving the following variational problem:

$$\min_{\nu \in \mathcal{M}(\mathcal{P})} \operatorname{dist}(\pi, \nu) \tag{13}$$

where $\mathcal{P}$ is the set of admissible parameterizations. In other words, we are looking for the admissible measure ν^* that is the closest to the target measure π. This is therefore a projection problem onto $\mathcal{M}(\mathcal{P})$.

Let us mention that the mapping $\nu \mapsto \operatorname{dist}(\pi, \nu)$ is a nice convex and smooth function. However, for most parameterization sets $\mathcal{P}$, the associated measure set $\mathcal{M}(\mathcal{P})$ is highly nonconvex. This makes the resolution of Problem (13) very involved. In fact, in the "simple case" $\mathcal{P} = \Omega^m$, Problem (13) corresponds to Smale's 7th problem to solve for the XXIst century [55].

5 Numerical implementation

In this section, we propose a numerical algorithm to solve Problem (13).

5.1 The attraction-repulsion formulation

To numerically solve the infinite dimensional Problem (13), we need to discretize it. It was shown in [15] that any measure set $\mathcal{M}(\mathcal{P})$ can be approximated by a subset of p-point measures $\mathcal{N}_p \subseteq \mathcal{M}(\Omega^p)$ with an arbitrary precision. More precisely, it is possible to control the Hausdorff distance, defined by:

$$\mathcal{H}_{\text{dist}}(\mathcal{N}_p, \mathcal{M}(\mathcal{P})) = \max\left(\sup_{\pi \in \mathcal{N}_p} \inf_{\nu \in \mathcal{M}(\mathcal{P})} \operatorname{dist}(\nu, \pi), \sup_{\nu \in \mathcal{M}(\mathcal{P})} \inf_{\pi \in \mathcal{N}_p} \operatorname{dist}(\nu, \pi) \right).$$

Moreover, the set $\mathcal{N}_p$ can always be written as

$$\mathcal{N}_p = \left\{ \nu = \frac{1}{p} \sum_{i=1}^{p} \delta_{q_i}, \quad \text{for } q = (q_i)_{1 \le i \le p} \in \mathcal{Q}_p \right\},$$

where the parameterization set $\mathcal{Q}_p$ depends on $\mathcal{P}$. The abstract definition of $\mathcal{Q}_p$ proposed in [15] is not constructive. Explicit constructions for the parameterizations given in section 4.2 are provided in the next part. Notice that the discretization step is very different from what was done in many

papers [7, 54, 49, 20] where the authors propose to select samples among fixed blocks of measurements.

Once an approximate space of parameterizations $\mathcal{Q}_p$ has been constructed, Problem (13) can be replaced by its discrete approximation:

$$\min_{\nu \in \mathcal{N}_p} \frac{1}{2} \|h \star (\nu - \pi)\|_2^2 , \tag{14}$$

where $\mathcal{N}_p = \mathcal{M}(\mathcal{Q}_p)$ is a suitable approximation of $\mathcal{M}(\mathcal{P})$. Then, by expanding the L^2-norm, we may rewrite Problem (14) as follows:

$$\min_{q \in \mathcal{Q}_p} \underbrace{\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} H(q_i - q_j)}_{J_1(q)} - \underbrace{\sum_{i=1}^{p} \int_{\Omega} H(x - q_i) d\pi(x)}_{J_2(q)}, \tag{15}$$

where H is defined in the Fourier domain by $\widehat{H}(\xi) = |\widehat{h}|^2(\xi)$ for all $\xi \in \mathbb{Z}^d$. In this paper, we consider a kernel H defined by $H(x) = -\|x\|_2$. This choice ensures rotation and translation invariance with respect to the input measure π. In addition, it is nonlocal: the forces between particles is *independent* of the distance. This choice was initially introduced in [53].
Then, functional (15) can be decomposed in two terms:

- The first one J_1 is a *repulsion potential*: it tends to maximize the distance between all point pairs. It will guarantee that no cluster of points emerges and therefore ensures a good space coverage.
- The second one J_2 is an *attraction potential*: it attracts the particles q_i in the high density regions of π. This term ensures that the solution of Problem (15) will match the target density π.

Let us point out that the attraction-repulsion functional (15) was initially proposed in [53, 59] as an alternative to Poisson disk sampling [8, 64]. The proposed idea can therefore be considered as a generalization of Poisson disk sampling, allowing to handle arbitrary additional constraints.

5.2 Projected gradient descents

The attraction-repulsion formulation (15) of the projection problem (13) is amenable to a numerical resolution. Similarly to [15], we propose to use a projected gradient descent. We only describe it briefly hereafter and refer to [15] for its theoretical guarantees and more details. The algorithm reads

as follows:

Input:
An initial parameterization $q^{(0)} \in \mathcal{Q}_p$
A number of iterations n_{it}.
Output:
An approximation $\tilde{q}$ of the solution q^* of (15)

for $k = 1$ *to* n_{it} **do**

$$q^{(k+1)} \in \Pi_{\mathcal{Q}_p}\left(q^{(k)} - \tau\nabla(J_1 - J_2)(q^{(k)})\right) \tag{16}$$

end

Algorithm 1: Projected gradient descent to solve the projection Problem (15).

The step-size τ should be selected depending on the regularity of the kernel h. The projector $\Pi_{\mathcal{Q}_p}$ can be expressed as an optimization problem and we will provide algorithms adapted to specific choices of $\mathcal{Q}_p$ in the next sections. Note that $\mathcal{Q}_p$ has no reason to be convex in general and the projection on $\mathcal{Q}_p$ (ie, $\Pi_{\mathcal{Q}_p}$) might therefore not be unique. This explains the sign $\in$ instead of $=$ in Equation (16). If τ is well chosen, this algorithm is shown to converge to critical points of (15) in [15]. Let us finally mention that computing the gradients ∇J_1 and ∇J_2 is also a challenging issue that requires the use of tools developed for particle simulations such as fast multipole methods. In this work, we used the parallelized non-uniform fast Fourier transform [29, 59].

5.3 Discretization of the parameterization sets

In this section, we explicitly give the expressions of $\mathcal{Q}_p$ and $\Pi_{\mathcal{Q}_p}$ for the measures sets given in paragraph 4.1.

Isolated points. In the context of isolated points, $\mathcal{Q}_p = \Omega^p$, hence the projection $\Pi_{\mathcal{Q}_p}$ is the identity on Ω^p. The updating step 16 in Algorithm 1 is then $q^{(k+1)} = q^{(k)} - \tau\nabla(J_1 - J_2)(q^{(k)})$.

Segments of variable length. In this case, the measures are supported by N segments. Assuming that each segment is discretized into k points,

the total number of discretization points is $p = kN$ and the set $\mathcal{Q}_p$ reads:

$$\mathcal{Q}_p(\mathcal{L}^N) = \left\{ \begin{array}{ll} q \in \Omega^{p\times d} & , q_j = q_i + \frac{j-i-1}{k-1}\left(q_{i+k-1} - q_i\right), \\ & \text{for } i = 1:k:kN \text{ and } i \leq j < i+k \end{array} \right\},$$

where $1:k:kN$ denotes the set $\{1, k+1, 2k+1, \ldots, (N-1)k+1\}$.
The projection onto this set can be computed via Algorithm 2. For the sake of clarity, Algorithm 2 describes the projection onto the set of measures supported by only one segment ($N = 1$) in two dimensions.

Input:
u : a vector of k points
Output:
q : a vector of $\mathcal{Q}_p(\mathcal{L})$

- 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 $1 \leq i \leq k$
- Compute $x_i^{(2)} = (i-1)u_i$ for $1 \leq i \leq k$
- Compute $s^{(1)} = \frac{1}{k-1}\sum_{i=1}^{k} x_i^{(1)}$
- Compute $s^{(2)} = \frac{1}{k-1}\sum_{i=1}^{k} x_i^{(2)}$
- 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 $(q_i)_{2\leq i\leq k-1}$ uniformly spaced on $[q_1, q_k]$

Algorithm 2: Projection on $\mathcal{Q}_p(\mathcal{L}^1)$.

Proof. The set $\mathcal{Q}_p\left(\mathcal{L}^1\right)$ can be rewritten as follows

$$\mathcal{Q}_p\left(\mathcal{L}^1\right) = \left\{ p = (p_i)_{1\leq i\leq k}, \quad p_i = \frac{i-1}{k-1}p_k + \frac{k-i}{k-1}p_1 \right\}.$$

To define a projector on $\mathcal{Q}_p\left(\mathcal{L}^1\right)$, one should solve the following optimization problem

$$\Pi_{\mathcal{Q}_p(\mathcal{L}^1)}(q) = \underset{p\in\mathcal{Q}_p(\mathcal{L}^1)}{\arg\min}\, \frac{1}{2}\|p-q\|_2^2,$$

for some fixed $q = (q_i)_{1\leq i\leq k}$, where $q_i \in \mathbb{R}^d$ for all $1 \leq i \leq k$

This problem can be reformulated by optimizing only the end points of the projected segment, as follows

$$\min_{(p_1,p_k)\in\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}\sum_{i=1}^{k}\left\|\frac{i-1}{k-1}p_k + \frac{k-i}{k-1}p_1 - q_i\right\|_2^2. \tag{17}$$

The optimality conditions of Problem 17 read

$$\begin{cases} \sum_{i=1}^{k} \frac{k-i}{k-1}\left(\frac{i-1}{k-1}p_k + \frac{k-i}{k-1}p_1 - q_i\right)=0 \\ \sum_{i=1}^{k} \frac{i-1}{k-1}\left(\frac{i-1}{k-1}p_k + \frac{k-i}{k-1}p_1 - q_i\right)=0 \end{cases}$$

Set

$$C := \sum_{i=1}^{k}\frac{(k-i)(i-1)}{(k-1)^2} = \frac{k^3-3k^2+2k}{6(k-1)^2} \quad \text{and} \quad D := \sum_{i=1}^{k}\frac{(i-1)^2}{(k-1)^2} = \frac{2k^3-3k^2+k}{6(k-1)^2}.$$

The system can be rewritten as follows

$$\begin{cases} Cp_k + Dp_1 - \sum_{i=1}^{k}\frac{k-i}{k-1}q_i=0 \\ Dp_k + Cp_1 - \sum_{i=1}^{k}\frac{i-1}{k-1}q_i=0 \end{cases}$$

This 2×2 system can be easily inverted leading to Algorithm 2. □

Admissible curves for MRI. The projection onto $\mathcal{M}(\mathcal{S}_T)$ is the topic of [17]. The discretization of an element of $\mathcal{S}_T$ is a vector of $\mathbb{R}^{p\cdot d}$ where d is the space dimension and p is the number of points. Let $s(i)$ denote the curve location at time $(i-1)\delta t$ with $\delta t = \dfrac{T}{p-1}$. We define the first-order derivative by:

$$\dot{s}(i) = \begin{cases} 0 & \text{if } i = 1, \\ (s(i)-s(i-1))/\delta t & \text{if } i \in \{2,\ldots,p\}. \end{cases}$$

In the discrete setting, the first-order differential operator can be represented by a matrix $\dot{\mathbf{M}} \in \mathbb{R}^{p\cdot d\times p\cdot d}$, i.e. $\dot{s} = \dot{\mathbf{M}}s$. We define the discrete second-order

differential operator by $\ddot{\mathbf{M}} = -\dot{\mathbf{M}}^*\dot{\mathbf{M}} \in \mathbb{R}^{p\cdot d \times p\cdot d}$. In a discrete setting, the projection problem reads:

$$\Pi_{\mathcal{Q}_p}(c) = \underset{\substack{\|\dot{\mathbf{M}}s\| \leqslant \alpha \\ \|\ddot{\mathbf{M}}s\| \leqslant \beta}}{\arg\min} \|s - c\|_2^2$$

This problem can be solved using an accelerated proximal gradient descent algorithm by resorting to the dual formulation of the problem [17].

5.4 Implementation details

Solving the projection problem (15) is computationally demanding. Hopefully, the design of sampling patterns is performed offline and large computing times are therefore acceptable. In practice, we used a workstation with 192 Gb of RAM and 32 Cores at 2.4 GHz and all codes were multithreaded.

The computing times varied from 2 hours to generate the sampling schemes for low resolution images proposed in Figure 5 up to 48 hours for the schemes adapted to very high resolutions images in Figure 12. In practice, we used 4,000 iterations to generate the sampling schemes with isolated measurements. For the sampling schemes composed of lines or curves, we used a multi-resolution strategy: we first optimize an undersampled curve and progressively interpolate it, thus reducing the number of iterations as the resolution increases. We observed that this strategy provides improved results and speeds up convergence. As detailed in the next section, our trajectories based on lines or curves are not made by a connected path but instead by several disconnected pieces. In that case, the optimization of (15) is performed over multiple curves simultaneously. The set of independent curves with kinematic constraints is still a convex set and projections onto this set can be performed efficiently using specific convex programming approaches [17].

6 Results

In this section, we test the proposed ideas for reconstructing a 2D image (i.e. a slice) of a brain phantom at 2 different resolutions on a field of view of 20 cm. In all experiments, we used the analytical phantoms provided in [25].

The first image is of size 256×256, which approximately corresponds to an isotropic resolution of $780 \times 780\,\mu$m. This is a pretty standard resolution for actual MRI scanners (eg, 3 Tesla machine). The second image size is 2048×2048, which corresponds to an isotropic resolution of $98 \times 98\,\mu$m.

The latter is really uncommon in the literature and is actually an important challenge since it might permit to uncover the meso-scale brain architecture at ultra-high magnetic field (7 T and above). For instance, [21] reported ex-vivo experiments on brains at a resolution of $78 \times 78 \times 500\,\mu m$ allowing to much better understand the cytoarchitecture of the human cortex. However, such spatial resolution cannot be achieved during in-vivo experiments owing to the very long scanning times. For instance, the images used in [21] took more than 14 scanning hours. Compressed sensing may therefore play a key role in the future to push forward such resolutions, especially with the emergence of ultra-high field MRI at 7T or even 11.7T in the near future. Moreover, recent theoretical results [51] suggest that compressed sensing should be used as a resolution enhancer rather than a time saver.

6.1 Constraints used in our experiments

To apply our projection algorithm, the kinematic constraints have to be specified. To this end, we used typical constraints met on real MRI scanners, namely the same as the ones specified in [35]. The kinematic constraints imposed by MRI acquisition are the gradient magnitude and slew-rate: here, we set $G_{\max} = 40\,\mathrm{mT.m^{-1}}$ and $S_{\max} = 150\,\mathrm{mT.m^{-1}.ms^{-1}}$. For proton imaging, $\gamma = 42.576\,\mathrm{MHz.T^{-1}}$, which allows to compute $\alpha = \gamma G_{\max}$ and $\beta = \gamma S_{\max}$ in Equation (2). In addition to those constraints, we imposed our trajectories to last less than 200 ms[2] to keep a sufficient amount of signal.

6.2 Empirical choice of the target density π

The theorems in Section 3 provide some general guidelines to design a reasonable density. However, finding the *best* target density π is still an open issue depending on the number of measurements, the sparsity basis and the signal structure.

In this paper, we therefore used an empirical method. The basic idea was to optimize π experimentally in the family of polynomially decaying densities of type $1/(|k|+1)^{\eta}$. Those simple parametric densities have been used a lot in recent articles [1, 31] and have proved their efficiency in practice. Note however that they increase rapidly at the origin, leading to high samples concentrations. For Cartesian sampling, it was proved in references [1, 31] that the density should not exceed some threshold. Here, we are considering non-Cartesian sampling and there is no formal proof of this fact. We still observed that high concentrations were deleterious. The basic reason is that

[2] Beyond this limit, the T_2^* relaxation decay makes the noise predominant.

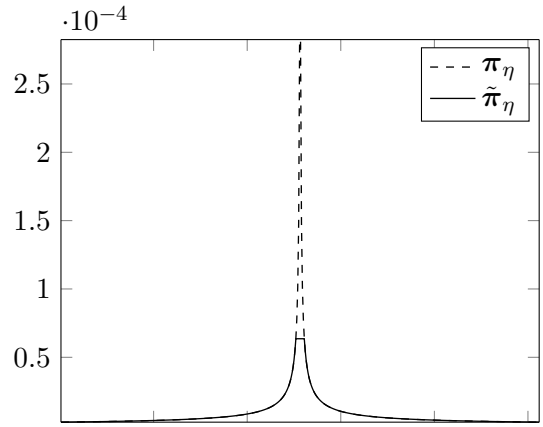

Figure 3: Action of the thresholding algorithm. The initial density $\boldsymbol{\pi}_\eta$ in dashed line and its thresholded version $\tilde{\boldsymbol{\pi}}_\eta$ defined in (18) in solid line.

they bring more information than necessary for low frequencies, which in turn, reduces the number of samples available for higher frequencies.

Given an initial discrete distribution $\boldsymbol{\pi}_\eta$ with a profile proportional to $1/(|k|+1)^\eta$, we therefore constructed a *truncated* version $\tilde{\boldsymbol{\pi}}_\eta$ of $\boldsymbol{\pi}_\eta$ defined by

$$\tilde{\boldsymbol{\pi}}_\eta = \min(\lambda \boldsymbol{\pi}_\eta, \tau) \tag{18}$$

where λ is chosen in such a way that $\|\tilde{\boldsymbol{\pi}}_\eta\|_1 = 1$. The distribution $\tilde{\boldsymbol{\pi}}_\eta$ has all components less than τ, and approximates $\boldsymbol{\pi}_\eta$.

In all our experiments, the threshold τ was chosen in such a way that the expectation of the number of samples in each pixel does not exceed 4 with an i.i.d. drawing. Assuming that $\boldsymbol{\pi}_\eta \in \mathbb{R}^n$ where n is the number of pixels in the image, this means that $m\tau = 4$, where m is the number of drawn samples. An illustration of density (18) is given in Figure 3.

6.3 New sampling patterns

We designed sampling schemes with the proposed algorithm and compared them to the state-of-the-art on the reconstructed brain phantom images. We compared 6 sampling patterns identified by letters:

- Standard patterns:

- (a) **Independent and identically distributed drawings** according to a prescribed density $\boldsymbol{\pi}_\eta$. This is the pattern considered in most compressed sensing theories. This pattern is not feasible in 2D in reasonable acquisition times, but serves as a reference.
- (b) **Equispaced radial lines**. This is another commonly used sampling pattern in MRI [65]. We assume that a spoke is a segment composed of $\sqrt{n}/2$ samples. Samples are equispaced along a line and the distance between two samples depends on the segment length.
- (c) **Spiral sampling**. We consider a spiral with the chosen target density $\boldsymbol{\pi}_\eta$ (see [13]), and reparameterize it to be admissible [35]. We replicate and rotate it a few times, to obtain a pattern made of interleaved spirals.

• Measure projection patterns:

- (d) **Projection of $\boldsymbol{\pi}_\eta$ on the set of isolated measurements** defined in (10). The initial parameterization $q^{(0)}$ in Algorithm 1 is defined as an independent point process with distribution π.
- (e) **Projection of $\boldsymbol{\pi}_\eta$ on the set of segments with varying lengths**. It is denoted $\mathcal{M}(\mathcal{L}^N)$ and defined in (11). Each segment contains the same number of samples $\sqrt{n}/2$ as a radial spoke. The initial parameterization $q^{(0)}$ in Algorithm 1 is defined as a set of equispaced radial segments.
- (f) **Projection of $\boldsymbol{\pi}_\eta$ on the set of admissible curves $\mathcal{S}_T$**, defined in Equation (2). The initial parameterization $q^{(0)}$ in Algorithm 1 is defined as a set of equispaced radial segments.

6.3.1 Standard resolution imaging

In this section, we focus on the reconstruction of 256×256 images.

Parameters setting. In this experiment, 25% of the k-space is sampled, corresponding to $m = 16,384$ samples. The sampling period (or dwell time) was fixed to 20 μs, which would correspond to a high SNR in a clinical setting. The total sampling time for each pattern is therefore $16384 * 20\mu$s = 327.68 ms. The minimal amount of time for a full acquisition (i.e., time necessary to probe each of the 256×256 discrete Fourier coefficients) is $256 \times 256 \times 20$ μs = 1.31s. This time is too long given the scanning constraints (MR signal decay, ...), but will serve as reference to measure the acceleration

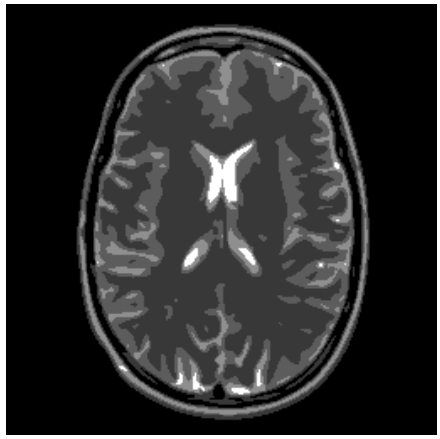

Figure 4: Axial slice of the phantom used in the experiments of size 256×256. The left brain hemisphere is shown on the right: left is right.

provided by each undersampling strategy. For this resolution, we found out that the best decay η defined in Section 6.2 was $\eta = 1.5$. This number was optimized by reconstructing images from i.i.d. drawings and keeping the decay corresponding to the best reconstruction.

To collect m Fourier coefficients, we used specific parameters to each sampling scheme as detailed below:

(b) 128 equispaced radial segments made of 128 samples;

(c) 2 spirals made of 8192 samples each; More details are given in Section 6.3.

(e) 128 segments made of 128 samples;

(f) 2 curves made of $8,192$ samples each; This corresponds to a typical buffer size.

Image and reconstruction. Data were simulated using the phantom depicted in Figure 4. The inverse problem used to reconstruct an image from simulated k-space data is Problem (7). Parameter λ was selected by hand once for all ($\lambda = 10^{-5}$) so as to nearly reach the equality constraint $\mathbf{S\Psi x} = \mathbf{y}$ and to provide a visually satisfactory solution in less than 1,000 iterations.

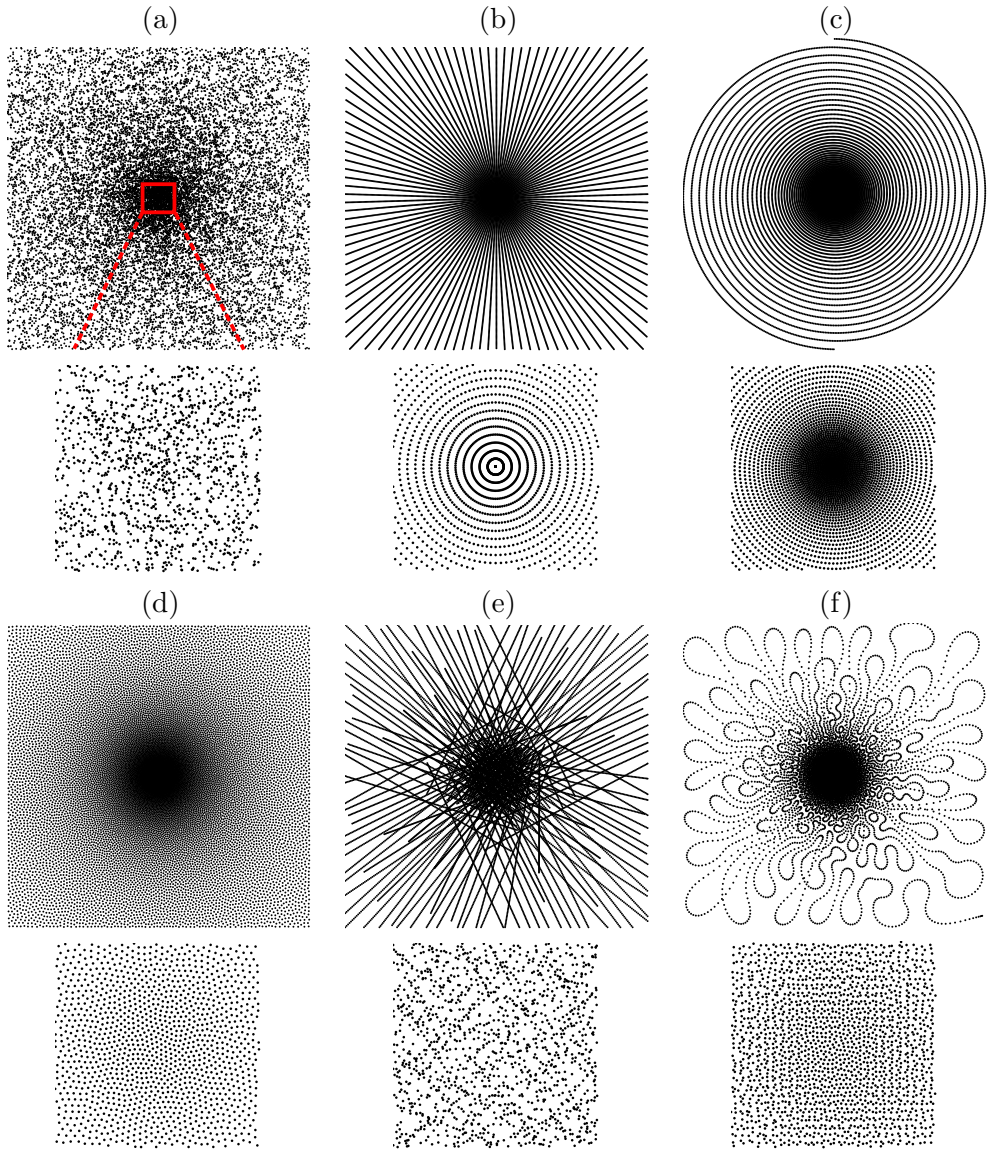

Figure 5: Classical sampling schemes (a-c) and sampling schemes obtained with the proposed projection algorithm (d-f). Top row: (a): independent drawing; (b): radial lines ; (c): spiral trajectory. Second row: zooms in the k-space centers. Third row: (d): isolated points; (e): segments of variable length; (f): admissible curves for MRI. Bottom row: zooms in the k-space center. Corresponding reconstruction results are provided in Figure 6.

Results. In Figure 6, we show the reconstruction results for the different sampling schemes depicted in Figure 5. Hereafter, we summarize our main findings.

- First, we noticed that the two schemes composed of isolated measurements provided rather satisfactory reconstruction results despite a few artifacts (17.7 and 18.3 dB in (a) and (d), respectively) with one fourth of the measurements. This is an appealing result, but unfortunately the schemes cannot be implemented on a scanner, at least in a time efficient manner.

- The repulsion between isolated samples in (d) improved the reconstruction result slightly by 0.6 dB. This result tends to validate the interest for this strategy as it provides improved coverage of the sampling space.

- Classical sampling patterns were feasible and yield a 4-fold acceleration of scanning time but delivered images that cannot be considered as good enough by clinicians (15.4 dB for radial in (b) and 13.2 dB for spirals in (c)). The reconstruction based on radial lines induced many small artifacts whereas the reconstruction based on spirals suffered from ringing effects.

- In this experiment, the new sampling patterns generated by our algorithm yielded improved reconstruction results as compared to i.i.d. drawings. This may be surprising since our sampling schemes are constrained to satisfy additional kinematic constraints. The basic reason for this phenomenon is that i.i.d. sampling tends to produce clusters in some regions of space, while the repulsion term J_2 in equation (15) avoids this deleterious effect.

  This result shows that adding complicated but realistic sampling constraints can still permit to get competitive reconstruction results. In particular, the sampling pattern in Figure 6(f) took only one fourth of the reference scanning time and yielded satisfactory reconstructed images.

6.3.2 Very high resolution imaging

Here, we focused on the reconstruction of very high resolution ($2,048 \times 2,048$) images.

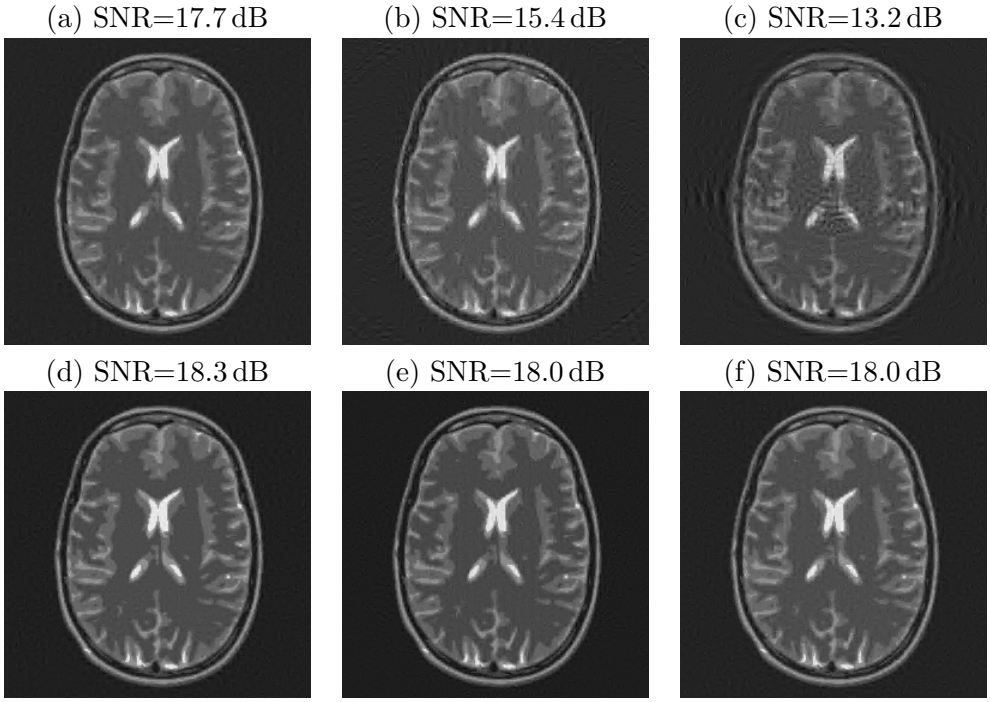

Figure 6: Reconstruction results for the sampling patterns proposed in Figure 5 on the phantom (Figure 4).

Parameters setting. We used the same constraints as before including the maximum sampling time $t_s = 200\,\text{ms}$ per trajectory. Hence, we decreased the sampling period down to its minimal value for a clinical scanner: $\Delta t = 8\,\mu\text{s}$. We no longer managed the buffer size constraint and performed experiments with 100,000 and 200,000 measurements. This corresponds to 2.4% and 4.8% of the total number of pixels in the image respectively. This also corresponds to a total acquisition duration of 0.8 s or 1.6 s, respectively.

The parameters specific to each sampling scheme are provided below:

(b) For the radial lines, we used 98 equispaced radial segments made of $1,024$ samples each for the experiment with $100,000$ samples experiment. We used 176 segments made of $1,024$ samples for the experiment with $200,000$ samples.

(c) For the spirals, we used 4 (resp. 8) rotated versions of spirals made of $25,000$ samples each for the $100,000$ (resp. $200,000$) experiment.

(e) For the repulsed segments, we used 196 (resp. 391) segments made of 512 samples for the $100,000$ (resp. $200,000$) experiment.

(f) For the projected curves, we used 4 (resp. 8) curves made of $25,000$ samples each.

Similarly to the previous section, the sampling density was optimized experimentally in the family of truncated, polynomially decaying densities of type $1/(|k|+1)^\eta$. For this resolution, the best decay was achieved for $\eta = 2$.

Image and reconstruction. We aimed at reconstructing the very high resolution phantom depicted in [25]. We modified it slightly by superimposing the high resolution text `COGITO ERGO SUM` to white matter in the left frontal region (see Figure 7).

Results. The resulting patterns are shown at different resolutions in Figures 8–9 for 100,000 measurements and Figures 11–12 for 200,000 measurements. For each scheme we reconstructed a $2,048 \times 2,048$ image by solving Problem (7). Hereafter, we summarize our main observations.

- The use of $200,000$ measurements yielded significantly better reconstruction results than $100,000$ samples. However, the relative differences between the sampling schemes did not vary between the two sampling ratios. In what follows, we therefore draw conclusions that are valid for both.

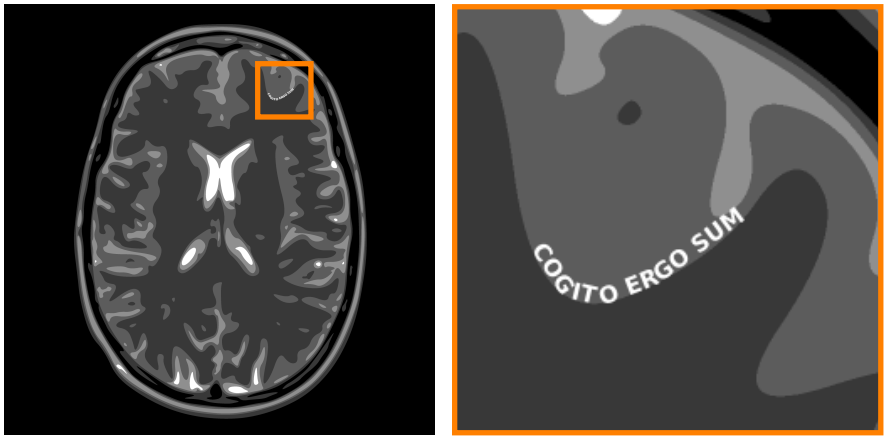

Figure 7: Axial slice of the brain phantom used in our $2,048 \times 2,048$ images (left) with a magnification on the left frontal area where the text has been superimposed (right).

- Similarly to the standard resolution experiment, sampling schemes made of i.i.d. drawings significantly outperformed radial lines and spirals sampling.
- Radial lines performed particularly poorly. This was probably due to the fact that for this resolution, the best sampling decay was $\eta = 2$, whereas we found $\eta = 1.5$ for the standard resolution experiment. Note that radial lines have a slow decay of order $1/|k|$, which might explain the observed discrepancy. Also note that the embedded text for radial reconstruction was readable, whereas it was not for spiral sampling. Once again, this is very likely a consequence of the slower decay for the sampling density. In contrast, the cortex was not correctly recovered by radial lines, whereas the reconstruction was acceptable for spirals. This experiment thus suggests that the sampling density should depend on the relative importance of low and high resolution details.
- The repulsed isolated measurements scheme performed slightly better than i.i.d. drawings, but not significantly so.
- Similarly to the previous section, the sampling schemes generated by our algorithm performed significantly better than spiral and radial patterns. The gain ranged from 1.7 dB to 3.6 dB which is significant, since they require the same scanning time.

- In contrast to the previous section, we observed that the feasible sampling schemes performed significantly worse than i.i.d. drawings in terms of SNR. A reason that might explain this behavior was that $\Delta t = 8\,\mu$s for this resolution while we used $\Delta t = 20\,\mu$s in the previous experiment. This means that the distance between consecutive samples was more than twice smaller (harder constraint). It is also important to realize that, although the differences between reconstructions were strong in terms of SNR, the visual perceptual differences mainly rely on small artifacts which do not severely degrade image analysis.

- The results obtained with $200,000$ samples were of a high quality, despite the realistic sampling constraints added. This very positive result suggests that obtaining $2,048 \times 2,048$ images might be feasible in 1.6 s by using a segmented acquisition (8 segments) scheme. This should be definitely deemed as a major advance for MRI. Of course, these results were preliminary since we did not manage all degradations appearing on actual scanners such as noise, Eddy currents, off-resonance effects ...

- Last, it is possible to infer the gain in terms of scanning times using the proposed approach by comparing Figure 10 and Figure 13. The SNR of the reconstructed image with 4 admissible curves and 0.8 s is 20.7 dB (see Figure 10 (f)). To reach the same quality, radial lines and spirals need roughly twice longer acquisition times, ie 1.6 s (see Figure 13(b)-(c)). This result shows that the proposed ideas may reduce the actual scanning times by a factor 2 compared to existing compressed sensing approaches.

7 Conclusion

This paper has provided an overview of existing compressed sensing results for MRI, both from theoretical and practical points of view. We also proposed an original approach to design efficient sampling schemes complying with physical constraints of MRI scanners. Even though we focused on standard anatomical MRI, the proposed ideas could be used, with some adaptations, in nearly all MRI fields (functional imaging, diffusion-weighted imaging, perfusion imaging, ...) and might have applications well beyond.

The numerical procedure we proposed for generating sampling schemes was based on a projection of sampling distributions onto a set of admissible

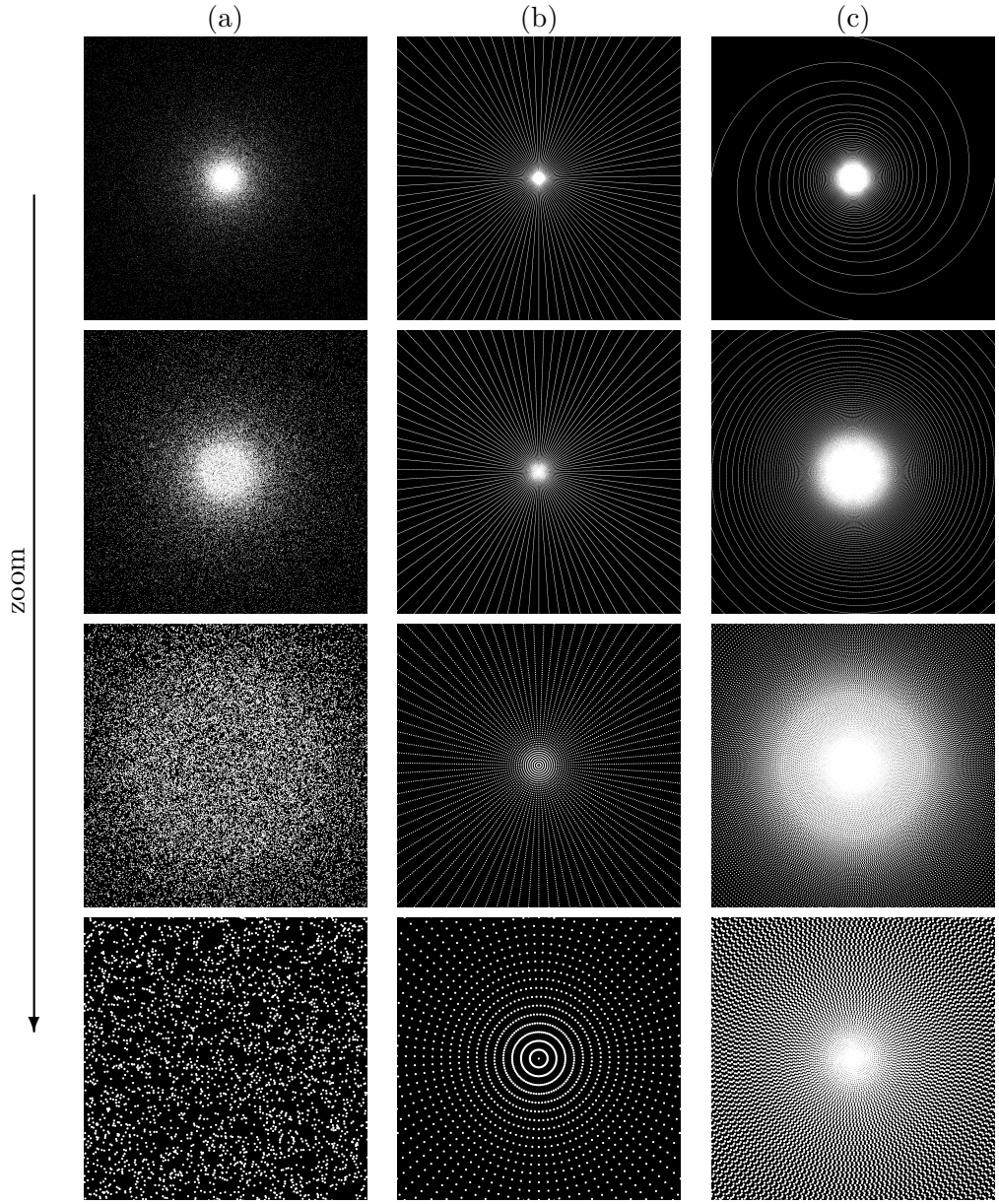

Figure 8: Standard sampling schemes composed of $100,000$ samples. (a): i.i.d. drawings. (b): Radial lines. (c): 4 interleaved spirals.

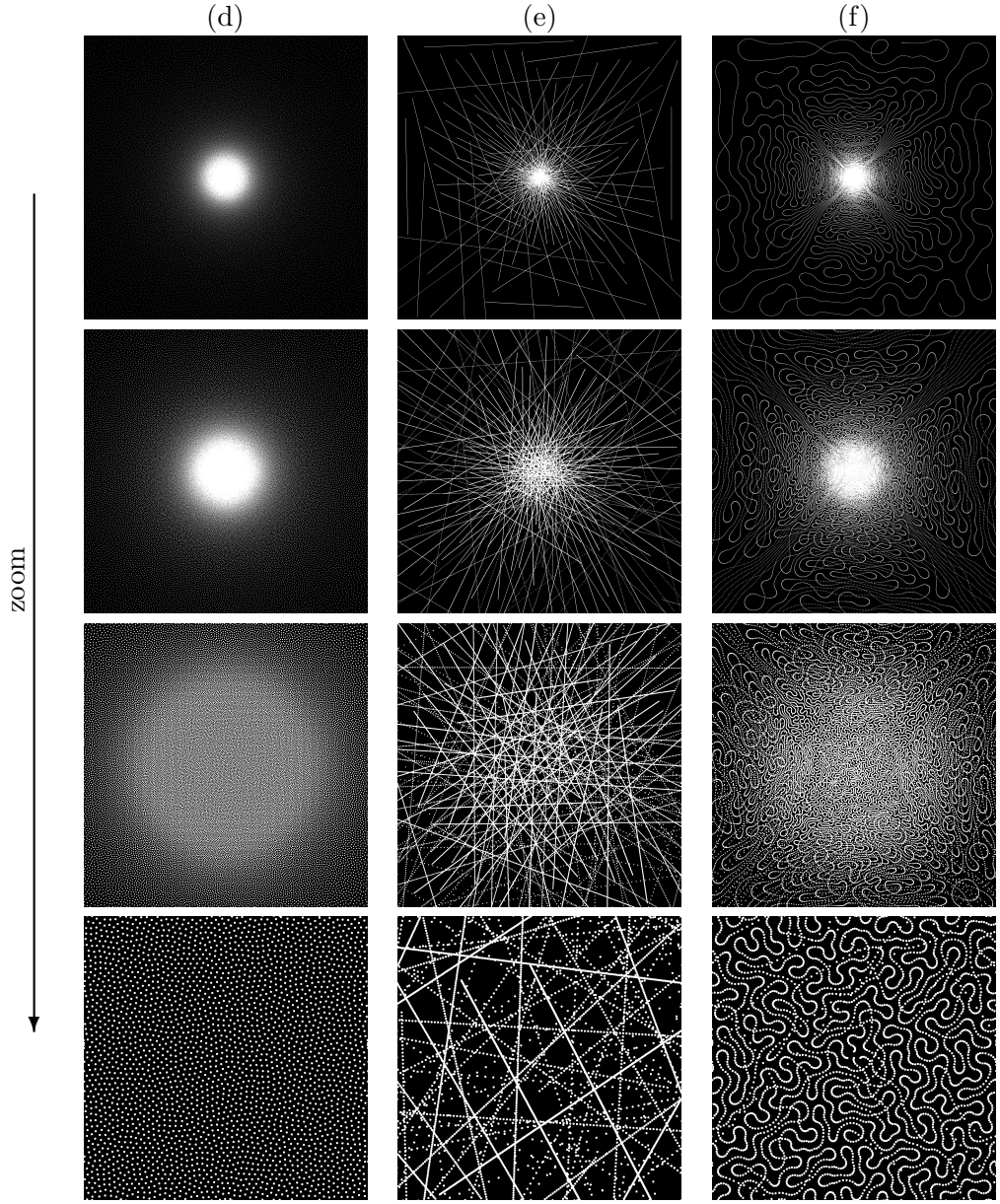

Figure 9: Sampling schemes yielded by our algorithm and composed of $100,000$ samples. (d): Isolated measurements. (e): Segments of variable length. (f): 4 feasible curves in MRI.

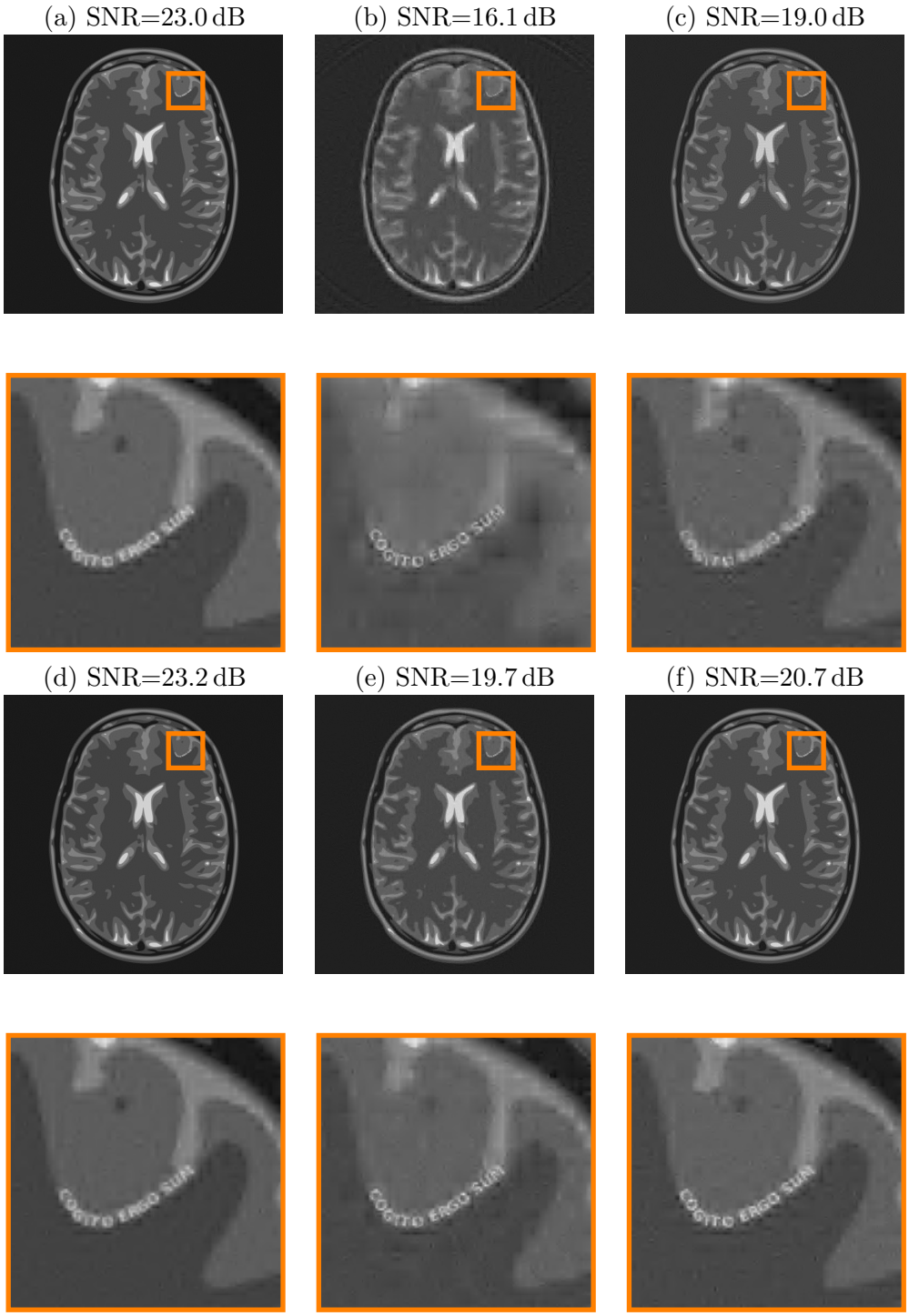

Figure 10: Very high resolution reconstructions using 100,000 samples (2.4% of the number of pixels) and different sampling schemes. Letters correspond to Figures 8–9.

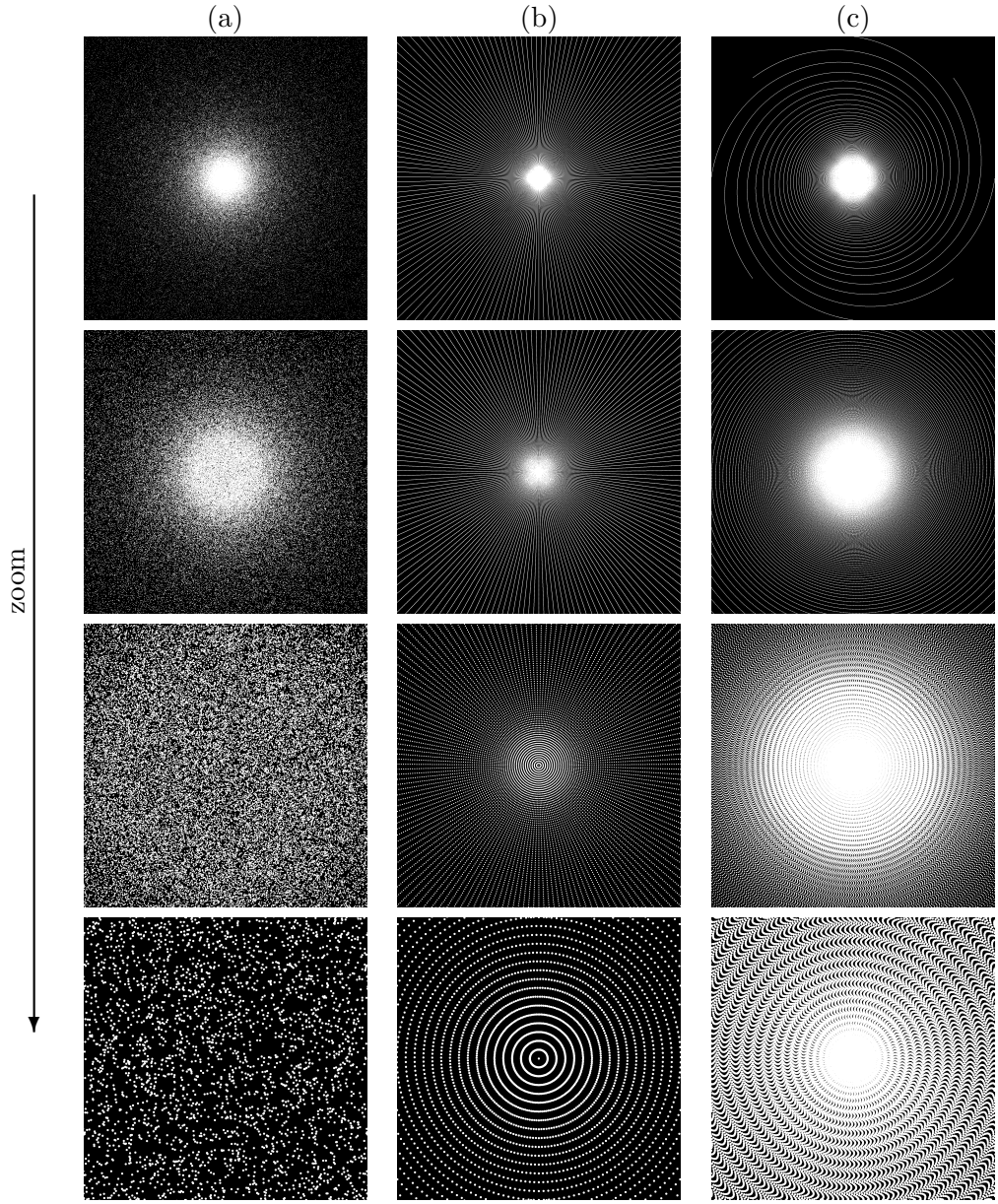

Figure 11: Standard sampling schemes composed of $200,000$ samples. (a): i.i.d. drawings. (b): Radial lines. (c): 8 interleaved spirals.

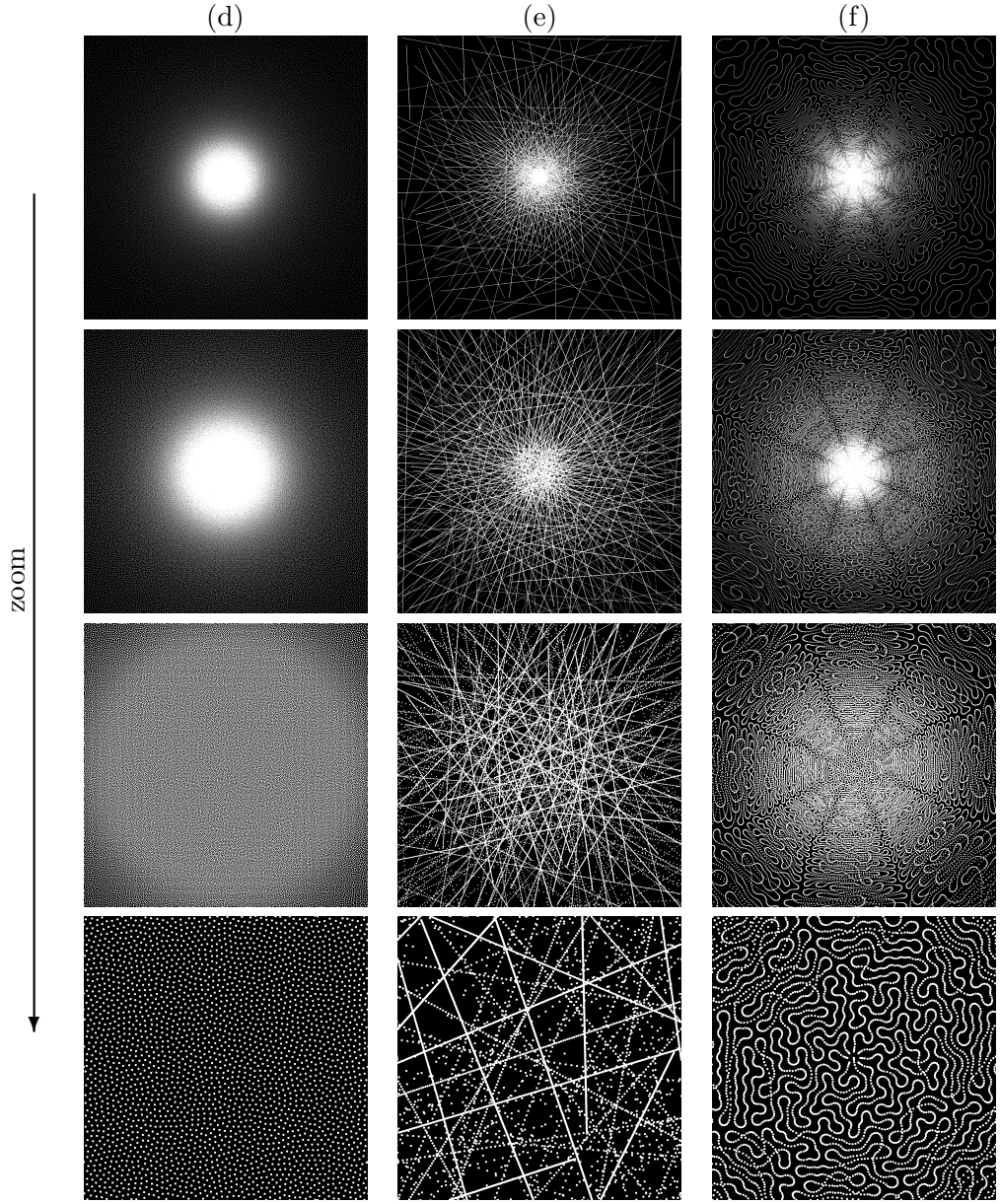

Figure 12: Sampling schemes yielded by our algorithm and composed of $200,000$ samples. (d): Isolated measurements. (e): Segments of variable length. (f): 8 feasible curves in MRI.

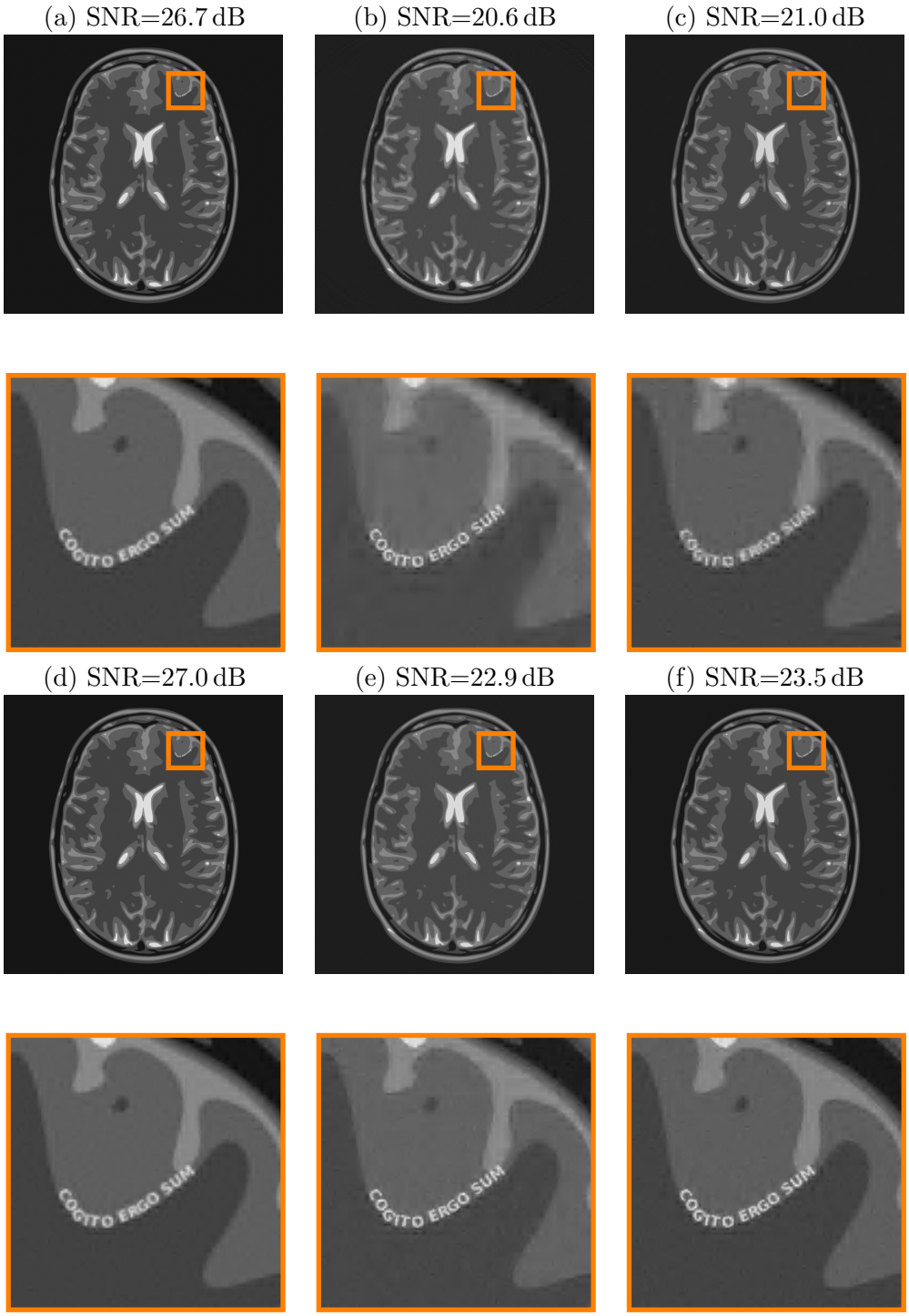

Figure 13: Very high resolution reconstructions using 200,000 samples (4.8% of the pixels number) and different sampling schemes. Letters correspond to Figures 11–12.

measures using a tailored dissimilarity measure. Even though computationally intensive, this algorithm was able to solve very large scale problems and could be extended to 3D quite easily. Probably the most promising result of this paper is practical: we showed through simulations that 1.6 s using a multi-shot acquisition (8 segments) might be enough to reconstruct a very high resolution slice of size 2048×2048. The validity of this result will be tested quite soon on the 7T scanner of NeuroSpin to check whether this constitutes a major improvement over existing sampling strategies which currently need a dozen of hours to reconstruct a hundred slices at this spatial resolution.

Acknowledgments

The authors wish to thank Daniel Potts, Toni Volkmer, Gabriele Steidl for their support and help to run the excellent NFFT library [29]. They also thank Anders Hansen for his review and insights on a preliminary version of this paper and Alexandre Vignaud for his numerous explanations on MR physics. They thank Laurent Jacques and Gabriel Peyré for reviewing a first draft of this work for Nicolas Chauffert's PhD defense. This work was partially supported by the CIMI (Centre International de Mathématiques et d'Informatique) Excellence program of Toulouse university.

References

[1] B. Adcock, A. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing. 2013.

[2] E. Arias-Castro, E. J. Candes, M. Davenport, et al. On the fundamental limits of adaptive sensing. *IEEE Transactions on Information Theory*, 59(1):472–481, 2013.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[4] J. Bigot, C. Boyer, and P. Weiss. An analysis of block sampling strategies in compressed sensing. *IEEE Transactions on Information Theory*, 62(4):2125–2139, 2016.

[5] C. Boyer, J. Bigot, and P. Weiss. Compressed sensing with structured sparsity and structured acquisition. *arXiv preprint arXiv:1505.01619*, 2015.

[6] C. Boyer, P. Ciuciu, P. Weiss, and S. Mériaux. HYR2PICS: Hybrid Regularized Reconstruction for Combined Parallel Imaging and Compressive Sensing in MRI. In *Proc. of 9th IEEE ISBI conference*, pages 66–69, Barcelona, Spain, May 2012.

[7] C. Boyer, P. Weiss, and J. Bigot. An algorithm for variable density sampling with block-constrained acquisition. *SIAM Journal on Imaging Sciences*, 7(2):1080–1107, 2014.

[8] R. Bridson. Fast poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH*, volume 2007, page 5, 2007.

[9] E. K. Brodsky, A. A. Samsonov, and W. F. Block. Characterizing and correcting gradient errors in non-Cartesian imaging: Are gradient errors linear time-invariant (LTI)? *Magnetic Resonance in Medicine*, 62(6):1466–1476, 2009.

[10] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[11] E. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies. *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[12] E. J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.

[13] N. Chauffert. *Compressed sensing along physicaly plausible sampling trajectories*. PhD thesis, Université Paris-Sud, Orsay, France, Sept. 2015.

[14] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss. Variable density sampling with continuous trajectories. Application to MRI. *SIAM Journal on Imaging Science*, 7(4):1962–1992, 2014.

[15] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss. A projection method on measure sets. *Constructive Approximation*, Jan. 2016.

[16] N. Chauffert, P. Ciuciu, and P. Weiss. Variable density compressed sensing in MRI. Theoretical vs. heuristic sampling strategies. In *Proc. of 10th IEEE ISBI conference*, pages 298–301, San Francisco, USA, Apr. 2013.

[17] N. Chauffert, P. Weiss, J. Kahn, and **P. Ciuciu**. A projection algorithm for gradient waveforms design in Magnetic Resonance Imaging. *IEEE Transaction on Medical Imaging*, 35(9):2026–2039, Sept. 2016.

[18] A. T. Curtis and C. K. Anand. Random volumetric MRI trajectories via genetic algorithms. *Journal of Biomedical Imaging*, 2008:6, 2008.

[19] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, Apr. 2006.

[20] D. duan Liu, D. Liang, X. Liu, and Y.-T. Zhang. Under-sampling trajectory design for compressed sensing MRI. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 73–76, Aug 2012.

[21] G. M. Fatterpekar, T. P. Naidich, B. N. Delman, J. G. Aguinaldo, S. H. Gultekin, C. C. Sherwood, P. R. Hof, B. P. Drayer, and Z. A. Fayad. Cytoarchitecture of the human cerebral cortex: MR microscopy of excised specimens at 9.4 Tesla. *American Journal of Neuroradiology*, 23(8):1313–1321, 2002.

[22] M. Freiberger, F. Knoll, K. Bredies, H. Scharfetter, and R. Stollberger. The agile library for biomedical image reconstruction using GPU acceleration. *Computing in Science & Engineering*, 15(1):34–44, 2013.

[23] K. Gröchenig, J. L. Romero, J. Unnikrishnan, and M. Vetterli. On minimal trajectories for mobile sampling of bandlimited fields. *Applied and Computational Harmonic Analysis*, 2014.

[24] M. Guerquin-Kern, M. Häberlin, K. P. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Transaction on Medical Imaging*, 30(9):1649–1660, 2011.

[25] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *IEEE Transaction on Medical Imaging*, 31(3):626–636, 2012.

[26] J. P. Haldar, D. Hernando, and Z.-P. Liang. Compressed-sensing MRI with random encoding. *IEEE Transactions on Medical Imaging*, 30(4):893–903, 2011.

[27] B. A. Hargreaves, D. G. Nishimura, and S. M. Conolly. Time-optimal multidimensional gradient waveform design for rapid imaging. *Magnetic Resonance in Medicine*, 51(1):81–92, 2004.

[28] J. I. Jackson, C. H. Meyer, D. G. Nishimura, and A. Macovski. Selection of a convolution function for Fourier inversion using gridding [computerised tomography application]. *IEEE Transactions on Medical Imaging*, 10(3):473–478, 1991.

[29] J. Keiner, S. Kunis, and D. Potts. Using NFFT 3—a software library for various nonequispaced fast Fourier transforms. *ACM Transactions on Mathematical Software (TOMS)*, 36(4):19, 2009.

[30] T. Knopp, S. Kunis, and D. Potts. A note on the iterative MRI reconstruction from nonuniform k-space data. *International journal of biomedical imaging*, 2007, 2007.

[31] F. Krahmer and R. Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Transactions on Image Processing*, 23(2):612–622, 2014.

[32] C. Kumar Anand, A. Thomas Curtis, and R. Kumar. Durga: A heuristically-optimized data collection strategy for volumetric magnetic resonance imaging. *Engineering Optimization*, 40(2):117–136, 2008.

[33] P. C. Lauterbur et al. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191, 1973.

[34] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, Dec. 2007.

[35] M. Lustig, S. J. Kim, and J. M. Pauly. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. *IEEE Transactions on Medical Imaging*, 27(6):866–873, 2008.

[36] M. Lustig, J. H. Lee, D. L. Donoho, and J. M. Pauly. Faster imaging with randomly perturbed, under-sampled spirals and ℓ_1 reconstruction.

In *Proceedings of the 13th annual meeting of ISMRM, Miami Beach*, page 685, 2005.

[37] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty. An efficient algorithm for compressed MR imaging using total variation and wavelets. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[38] A. Maudsley. Dynamic range improvement in NMR imaging using phase scrambling. *Journal of Magnetic Resonance (1969)*, 76(2):287–305, 1988.

[39] R. Mir, A. Guesalaga, J. Spiniak, M. Guarini, and P. Irarrazaval. Fast three-dimensional k-space trajectory design using missile guidance ideas. *Magnetic Resonance in Medicine*, 52(2):329–336, 2004.

[40] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig. Fast-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime. *IEEE Transactions on Medical Imaging*, 31(6):1250–1262, 2012.

[41] Y. Nesterov. A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. In *Soviet Mathematics Doklady*, volume 27, pages 372–376. 1983.

[42] Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.

[43] D. G. Nishimura, P. Irarrazabal, and C. H. Meyer. A velocity k-space analysis of flow effects in echo-planar and spiral imaging. *Magnetic Resonance in Medicine*, 33(4):549–556, 1995.

[44] J. O'sullivan. A fast sinc function gridding algorithm for Fourier inversion in computer tomography. *IEEE Transactions on Medical Imaging*, 4(4):200–207, 1985.

[45] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger, et al. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[46] G. Puy, J. P. Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux. Spread spectrum magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 31(3):586–598, 2012.

[47] G. Puy, P. Vandergheynst, R. Gribonval, and Y. Wiaux. Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–13, 2012.

[48] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *IEEE Transactions on Signal Processing letters*, 18(10):595–598, 2011.

[49] S. Ravishankar and Y. Bresler. Adaptive sampling design for compressed sensing MRI. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 3751–3755, Aug 2011.

[50] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041, 2011.

[51] B. Roman, A. Hansen, and B. Adcock. On asymptotic structure in compressed sensing. *arXiv preprint arXiv:1406.4178*, 2014.

[52] J. Romberg. Compressive sensing by random convolution. *SIAM Journal on Imaging Sciences*, 2(4):1098–1128, 2009.

[53] C. Schmaltz, P. Gwosdek, A. Bruhn, and J. Weickert. Electrostatic halftoning. In *Computer Graphics Forum*, volume 29, pages 2313–2327. Wiley Online Library, 2010.

[54] M. Seeger, H. Nickisch, R. Pohmann, and B. Schölkopf. Optimization of k-space trajectories for compressed sensing by Bayesian experimental design. *Magnetic Resonance in Medicine*, 63(1):116–126, 2010.

[55] S. Smale. Mathematical problems for the next century. *The Mathematical Intelligencer*, 20(2):7–15, 1998.

[56] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603, Jun. 1997.

[57] D. M. Spielman, J. M. Pauly, and C. H. Meyer. Magnetic resonance fluoroscopy using spirals with variable sampling densities. *Magnetic Resonance in Medicine*, 34(3):388–394, 1995.

[58] J. Spiniak, A. Guesalaga, R. Mir, M. Guarini, and P. Irarrazaval. Undersampling k-space using fast progressive 3D trajectories. *Magnetic Resonance in Medicine*, 54(4):886–892, 2005.

[59] T. Teuber, G. Steidl, P. Gwosdek, C. Schmaltz, and J. Weickert. Dithering by differences of convex functions. *SIAM Journal on Imaging Sciences*, 4(1):79–108, 2011.

[60] D. B. Twieg. The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Medical Physics*, 10(5):610–621, 1983.

[61] J. Unnikrishnan and M. Vetterli. Sampling high-dimensional bandlimited fields on low-dimensional manifolds. *IEEE Transactions on Information Theory*, 59(4):2103–2127, 2013.

[62] S. J. Vannesjo, N. N. Graedel, L. Kasper, S. Gross, J. Busch, M. Haeberlin, C. Barmet, and K. P. Pruessmann. Image reconstruction using a gradient impulse response model for trajectory prediction. *Magnetic Resonance in Medicine*, 2015.

[63] S. J. Vannesjo, B. J. Wilm, Y. Duerst, S. Gross, D. O. Brunner, B. E. Dietrich, T. Schmid, C. Barmet, and K. P. Pruessmann. Retrospective correction of physiological field fluctuations in high-field brain mri using concurrent field monitoring. *Magnetic Resonance in Medicine*, 73(5):1833–1843, 2015.

[64] S. Vasanawala, M. Murphy, M. T. Alley, P. Lai, K. Keutzer, J. M. Pauly, and M. Lustig. Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients. In *Proc. ISBI*, pages 1039–1043, 2011.

[65] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel. An optimal radial profile order based on the golden ratio for time-resolved MRI. *IEEE Transaction on Medical Imaging*, 26(1):68–76, 2007.