

# VARIABLE DENSITY SAMPLING WITH CONTINUOUS TRAJECTORIES. APPLICATION TO MRI.

NICOLAS CHAUFFERT\*, PHILIPPE CIUCIU†, JONAS KAHN‡ AND PIERRE WEISS§

**Abstract.** Reducing acquisition time is a crucial challenge for many imaging techniques. Compressed Sensing (CS) theory offers an appealing framework to address this issue since it provides theoretical guarantees on the reconstruction of sparse signals by projection on a low dimensional linear subspace. In this paper, we focus on a setting where the imaging device allows to sense a fixed set of measurements. We first discuss the choice of an optimal sampling subspace allowing perfect reconstruction of sparse signals. Its design relies on the random drawing of independent measurements. We discuss how to select the drawing distribution and show that a mixed strategy involving partial deterministic sampling and independent drawings can help breaking the so-called coherence barrier. Unfortunately, independent random sampling is irrelevant for many acquisition devices owing to acquisition constraints. To overcome this limitation, the notion of Variable Density Samplers (VDS) is introduced and defined as a stochastic process with a prescribed limit empirical measure. It encompasses samplers based on independent measurements or continuous curves. The latter are crucial to extend CS results to actual applications. We propose two original approaches to design continuous VDS, one based on random walks over the acquisition space, and one based on Traveling Salesman Problem. Following theoretical considerations and retrospective CS simulations in magnetic resonance imaging, we intend to highlight the key properties of a VDS to ensure accurate sparse reconstructions, namely its limit empirical measure and its mixing time<sup>1</sup>.

**Key words.** Variable density sampling, compressed sensing, CS-MRI, stochastic processes, empirical measure, TSP, Markov Chains,  $l^1$  reconstruction.

**AMS subject classifications.** 94A20, 60G20, 15A52, 94A08

**1. Introduction.** Variable density sampling is a technique that is extensively used in various sensing devices such as magnetic resonance imaging (MRI), in order to shorten scanning time. It consists in measuring only a small number of random projections of a signal/image on elements of a basis drawn according to a given density. For instance, in MRI where measurements consist of Fourier (or more generally  $k$ -space) coefficients, it is common to sample the Fourier plane center more densely than the high frequencies. The image is then reconstructed from this incomplete information by dedicated signal processing methods. To the best of our knowledge, *variable density sampling* has been proposed first in the MRI context by [45] where spiral trajectories were pushed forward. Hereafter, it has been used in this application (see e.g. [49, 27, 35] to quote a few), but also in other applications such as holography [43, 34]. This technique can hardly be avoided in specific imaging techniques such as radio interferometry or tomographic modalities (e.g., X-ray) where sensing is made along fixed sets of measurements [51, 44].

In the early days of its development, variable density sampling was merely an efficient heuristic to shorten acquisition time. It has recently found a partial justification in the Compressed Sensing (CS) literature. Even though this theory is not yet

---

\*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, Toulouse, France.(pierre.armand.weiss@gmail.com).

<sup>1</sup>Part of this work is based on the conference proceedings: [12, 11, 13].

mature enough to fully explain the practical success of variable density sampling, CS provides good hints on how to choose the measurements (i.e., the density), how the signal/image should be reconstructed and why it works. Let us now recall a typical result emanating from the CS literature for orthogonal systems. A vector  $x \in \mathbb{C}^n$  is said  $s$ -sparse if it contains at most  $s$  non-zero entries. Denote by  $a_i$ ,  $i \in \{1, \dots, n\}$  the sensing vectors and by  $y_i = \langle a_i, x \rangle$  the possible measurements. Typical CS results

state that if the signal (or image)  $x$  is  $s$ -sparse and if  $\mathbf{A} = \begin{pmatrix} a_1^* \\ \vdots \\ a_n^* \end{pmatrix}$  satisfies an incoherence property (defined in the sequel), then  $m = O(s \log(n)^\alpha)$  measurements chosen randomly among the elements of  $y = \mathbf{A}x$  are enough to ensure perfect reconstruction of  $x$ . The constant  $\alpha > 0$  depends on additional properties on  $x$  and  $\mathbf{A}$ . The set of actual measurements is denoted  $\Omega \subseteq \{1, \dots, n\}$  and  $\mathbf{A}_\Omega$  is the matrix formed by selecting a subset of rows of  $\mathbf{A}$  in  $\Omega$ . The reconstruction of  $x$  knowing  $y_\Omega = \mathbf{A}_\Omega x$  is guaranteed if it results from solving the following  $\ell_1$  minimization problem:

$$(1.1) \quad \min_{z \in \mathbb{C}^n} \|z\|_1 \quad \text{subject to} \quad \mathbf{A}_\Omega z = y_\Omega.$$

Until recent works [42, 24, 9], no general theory for selecting the rows was available. In the latter, the authors have proposed to construct  $\mathbf{A}_\Omega$  by drawing  $m$  rows of  $\mathbf{A}$  at random according to a discrete probability distribution or density  $p = (p_1, \dots, p_n)$ . The choice of an optimal distribution  $p$  is an active field of research (see e.g. [12, 29, 1]) that remains open in many regards.

Drawing independent rows of  $\mathbf{A}$  is interesting from a theoretical perspective, however it has little practical relevance since standard acquisition devices come with acquisition constraints. For instance, in MRI, the coefficients are acquired along *piece-wise continuous* curves on the  $k$ -space. The first paper performing variable density sampling in MRI [45] has fulfilled this constraint by considering spiral sampling trajectories. The standard reference about CS-MRI [32] has proposed to sample the MRI signal along parallel lines in the 3D  $k$ -space. Though spirals and lines can be implemented easily on a scanner, it is likely that more general trajectories could provide better reconstruction results, or save more scanning time.

The main objective of this paper is to propose new strategies to sample a signal along more general continuous curves. Although continuity is often not sufficient for practical implementation on actual scanner, we believe that it is a first important step towards more physically plausible compressed sampling paradigms. As far as we know, this research avenue is relatively new. The problem was first discussed in [52] and some heuristics were proposed. The recent contributions [38, 4] have provided theoretical guarantees when sampling is performed along fixed sets of measurements (e.g. straight lines in the Fourier plane), but have not addressed generic continuous sampling curves yet.

The contributions of this paper are threefold. First, we bring a well mathematically grounded definition of *variable density samplers* and provide various examples. Second, we discuss how the sampling density should be chosen in practice. This discussion mostly relies on variations around the theorems provided in [42, 9]. In particular, we justify the deterministic sampling of a set of highly coherent vectors to overcome the so-called ‘‘coherence barrier’’. In the MRI case, this amounts to deterministically sampling the  $k$ -space center. Our third and maybe most impacting contribution is to provide practical examples of variable density samplers along continuous curves and to derive some of their theoretical properties. These samplers are

defined as parametrized random curves that asymptotically fit a target distribution (e.g. the one shown in Fig. 1.1 (a)). More specifically, we first propose a *local* sampler based on random walks over the acquisition space (see Fig. 1.1 (b)). Second, we introduce a *global* sampler based on the solution of a Travelling Salesman Problem amongst randomly drawn “cities” (see Fig. 1.1 (c)). In both cases, we investigate the resulting density. To finish with, we illustrate the proposed sampling schemes on 2D and 3D MRI simulations. The reconstruction results provided by the proposed techniques show that the PSNR can be substantially improved compared to existing strategies proposed e.g. in [32]. Our theoretical results and numerical experiments on retrospective CS show that two key features of variable density samplers are the **limit of their empirical measure** and their **mixing properties**.

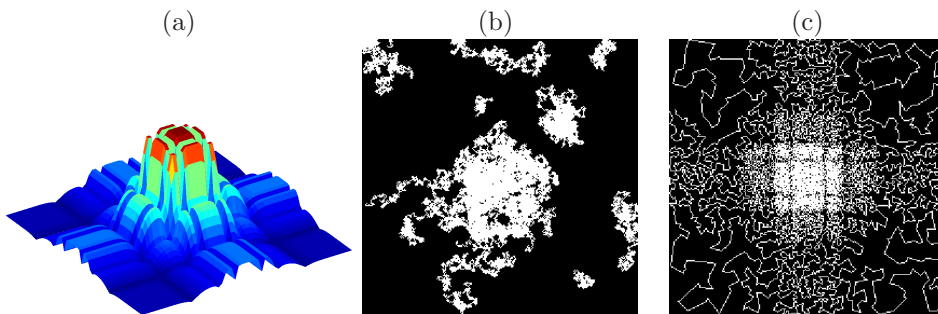


FIG. 1.1. (a): Target distribution  $\pi$ . Continuous random trajectories reaching distribution  $\pi$  based on Markov chains (b) and on a TSP solution (c).

The rest of this paper is organized as follows. First, we introduce a precise definition of a variable density sampler (VDS) and recall CS results in the special case of independent drawings. Then, we give a closed form expression for the optimal distribution depending on the sensing matrix  $\mathbf{A}$ , and justify that a partial deterministic sampling may provide better reconstruction guarantees. Hereafter, in Sections 3 and 4, we introduce two strategies to design continuous trajectories over the acquisition space. We show that the corresponding sampling distributions converge to a target distribution when the curve length tends to infinity. Finally, we demonstrate on simulation results that our TSP-based approach is promising in the MRI context (Section 5) since it outperforms its competing alternatives either in terms of PSNR at fixed sampling rate, or in terms of acceleration factor at fixed PSNR.

**Notation.** The main definitions used throughout the paper are defined in Tab. 1.1.

**2. Variable density sampling and its theoretical foundations.** To the best of our knowledge, there is currently no rigorous definition of variable density sampling. Hence, to fill this gap, we provide a precise definition below.

DEFINITION 2.1. Let  $p$  be a probability measure defined on a measurable space  $\Xi$ . A stochastic process  $X = \{X_i\}_{i \in \mathbb{N}}$  or  $X = \{X_t\}_{t \in \mathbb{R}_+}$  on state space  $\Xi$  is called a  $p$ -variable density sampler if its empirical measure (or occupation measure) weakly

TABLE 1.1  
General notation used in the paper.

	Notation	Definition	Domain
Compressed Sensing	$n$	Acquisition and signal space dimensions	$\mathbb{N}$
	$m$	Number of measurements	$\mathbb{N}$
	$R = n/m$	Sampling ratio	$\mathbb{Q}$
	$\mathbf{A}$	Full orthogonal acquisition matrix	$\mathbb{C}^{n \times n}$
	$\Omega$	Set of measurements	$\{1, \dots, n\}^m$
	$\mathbf{A}_\Omega$	Matrix formed with the rows of $\mathbf{A}$ corresponding to indexes belonging to $\Omega$	$\mathbb{C}^{m \times n}$
	$x$	Sparse signal	$\mathbb{C}^n$
	$s$	Number of non zero coefficients of $x$	$\mathbb{N}$
	$\Delta_n$	$\left\{ p = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}, 0 \leq p_i \leq 1, \sum_{i=1}^n p_i = 1 \right\}$	$\mathbb{R}^n$
		$\  \cdot \ _1$	$\ell_1$ norm defined for $z \in \mathbb{C}^n$ by $\ z\ _1 = \sum_{i=1}^n  z_i $
	$\  \cdot \ _\infty$	$\ell_\infty$ norm defined for $z \in \mathbb{C}^n$ by $\ z\ _\infty = \max_{1 \leq i \leq n}  z_i $	
MRI application	$k = \begin{pmatrix} k_x \\ k_y \end{pmatrix}$ or $\begin{pmatrix} k_x \\ k_z \end{pmatrix}$	Fourier frequencies	$\mathbb{R}^2$ or $\mathbb{R}^3$
	$\mathbf{F}_n^*$	$d$ -dimensional discrete Fourier transform on an $n$ pixels image	$\mathbb{C}^{n \times n}$
	$\mathbf{\Psi}_n$	$d$ -dimensional inverse discrete Wavelet transform on an image of $n$ pixels	$\mathbb{C}^{n \times n}$
		$\mathbf{F}_n^*$ and $\mathbf{\Psi}_n$ are denoted $\mathbf{F}^*$ and $\mathbf{\Psi}$ if no ambiguity	
VDS	$\Xi$	A measurable space which is typically $\{1, \dots, n\}$ or $[0, 1]^d$	
	$\mathcal{H}$	The unit cube $[0, 1]^d$	
	$p$	A probability measure defined on $\Xi$	
	$p(f)$	$= \int_{x \in \Xi} f(x) dp(x)$ , for $f$ continuous and bounded	$\mathbb{R}$
	$\lambda_{[0,1]}$	The Lebesgue measure on the interval $[0, 1]$	
	$X = (X_n)_{n \in \mathbb{N}^*}$	A time-homogeneous Markov chain on the state space $\{1, \dots, n\}$	$\{1, \dots, n\}^{\mathbb{N}^*}$
	$\mathbf{P}$	$:= (P_{ij})_{1 \leq i, j \leq n}$ the transition matrix: $P_{ij} := \mathbb{P}(X_k = j   X_{k-1} = i), \forall k > 1$	$\mathbb{R}^{n \times n}$
	$\lambda_i(\mathbf{P})$	The ordered eigenvalues of $\mathbf{P}$ : $1 = \lambda_1(\mathbf{P}) \geq \dots \geq \lambda_n(\mathbf{P}) \geq -1$	$[-1, 1]$
	$\epsilon(\mathbf{P})$	$= 1 - \lambda_2(\mathbf{P})$ , the spectral gap of $\mathbf{P}$	$[-1, 1]$
	$F$	A set of points $\subset \mathcal{H}$	$\mathcal{H}^N$
$C(F)$	The shortest Hamiltonian path (TSP) amongst points of set $F$	$\subset \mathcal{H}$	
$T(F, \mathcal{H})$	The length of $C(F)$	$\mathbb{R}_+$	
$T(F, R)$	For any set $R \subseteq \mathcal{H}$ , $T(F, R) := T(F \cap R, \mathcal{H})$	$\mathbb{R}_+$	

converges to  $p$  almost surely, that is:

$$\frac{1}{N} \sum_{i=1}^N f(X_i) \rightarrow p(f) \quad a.s.$$

or

$$\frac{1}{T} \int_{t=0}^T f(X_t) dt \rightarrow p(f) \quad a.s.$$

for all continuous bounded  $f$ .

EXAMPLE 1. In the case where  $X = (X_i)_{i \in \mathbb{N}}$  is a discrete time stochastic process with discrete state space  $\Xi = \{1, \dots, n\}$ , definition 2.1 can be slightly simplified. Let

us set  $Z_j^N = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{X_i=j}$ . The random variable  $Z_j^N$  represents the proportion of

points that fall on position  $j$ . Let  $p$  denote a discrete probability distribution function. Using these notations,  $X$  is a  $p$ -variable density sampler if:

$$\lim_{N \rightarrow +\infty} Z_j^N = p_j \quad a.s.$$

In particular, if  $(X_i)_{i \in \mathbb{N}}$  are i.i.d. samples drawn from  $p$ , then  $X$  is a  $p$ -variable density sampler. This simple example is the most commonly encountered in the compressed sensing literature and we will review its properties in paragraph 2.1.

EXAMPLE 2. More generally, drawing independent random variables according to distribution  $p$  is a VDS if the space  $\Xi$  is second countable, owing to the strong law of large numbers.

EXAMPLE 3. An irreducible aperiodic Markov chain on a finite sample space is a VDS for its stationary distribution (or invariant measure); see Section 3.3.

EXAMPLE 4. In the deterministic case, for a dynamical system, definition 2.1 closely corresponds to the ergodic hypothesis, that is time averages are equal to expectations over space. We discuss an example that makes use of the TSP solution in section 4.

The following proposition directly relates the VDS concept to the time spent by the process in a part of the space, as an immediate consequence of the porte-manteau lemma (see e.g. [5]).

PROPOSITION 2.2. Let  $p$  denote a Borel measure defined on a set  $\Xi$ . Let  $B \subseteq \Xi$  be a measurable set. Let  $X : \mathbb{R}_+ \rightarrow \Xi$  (resp.  $X : \mathbb{N} \rightarrow \Xi$ ) be a stochastic process. Let  $\mu$  denote the Lebesgue measure on  $\mathbb{R}$ . Define  $\mu_X^t(B) = \frac{1}{t} \mu(\{s \in [0, t], X(s) \in B\})$  (resp.  $\mu_X^n(B) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X(i) \in B}$ ). Then, the following two propositions are equivalent:

- (i)  $X$  is a  $p$ -VDS
- (ii) Almost surely,  $\forall B \subseteq \Xi$  a Borel set with  $p(\partial B) = 0$ ,
 
$$\lim_{t \rightarrow +\infty} \mu_X^t(B) = p(B) \quad a.s.$$
 (resp.) 
$$\lim_{n \rightarrow +\infty} \mu_X^n(B) = p(B) \quad a.s.$$

REMARK 1. Definition 2.1 is a generic definition that encompasses both discrete and continuous time and discrete and continuous state space since  $\Xi$  can be any measurable space. In particular, the recent CS framework on orthogonal systems [42, 9] falls within this definition.

Definition 2.1 does not encompass some useful sampling strategies. We propose a definition of a generalized VDS, which encompasses stochastic processes indexed over a bounded time set.

DEFINITION 2.3. A sequence  $\{\{X_t^{(n)}\}_{0 \leq t \leq T_n}\}_{n \in \mathbb{N}}$  is a generalized  $p$ -VDS if the sequence of occupation measures converges to  $p$  almost surely, that is:

$$\frac{1}{T_n} \int_{t=0}^{T_n} f(X_t^{(n)}) dt \rightarrow p(f) \quad a.s.$$

REMARK 2. Let  $(X_t)_{t \in \mathbb{R}}$  be a VDS, and  $(T_n)_{n \in \mathbb{N}}$  be any positive sequence such that  $T_n \rightarrow \infty$ . Then the sequence defined by  $X_t^{(n)} = X_t$  for  $0 \leq t \leq T_n$  is a generalized VDS.

EXAMPLE 5. Let  $\Xi = \mathbb{R}^2$ , and consider  $r : [0, 1] \mapsto \mathbb{R}^+$  a strictly increasing smooth function. We denote by  $r^{-1} : [r(0), r(1)] \rightarrow \mathbb{R}$  its inverse function and by  $r^{-1}$  the

derivative of  $r^{-1}$ . Consider a sequence of spiral trajectories  $s_N : [0, N] \rightarrow \mathbb{R}^2$  defined by  $s_N(t) = r\left(\frac{t}{N}\right) \begin{pmatrix} \cos(2\pi t) \\ \sin(2\pi t) \end{pmatrix}$ . Then  $s_N$  is a generalized VDS for the distribution  $p$  defined by:

$$p(x, y) = \begin{cases} \frac{r^{-1}(\sqrt{x^2+y^2})}{2\pi \int_{\rho=r(0)}^{r(1)} r^{-1}(\rho) \rho d\rho} & \text{if } r(0) \leq \sqrt{x^2+y^2} \leq r(1) \\ 0 & \text{otherwise} \end{cases}$$

A simple justification is that the time spent by the spiral in the infinitesimal ring  $\{(x, y) \in \mathbb{R}^2, \rho \leq \sqrt{x^2+y^2} \leq \rho + d\rho\}$  is  $\int_{r^{-1}(\rho)}^{r^{-1}(\rho+d\rho)} dt \propto r^{-1}(\rho)$ .

**2.1. Theoretical foundations - Independent VDS.** CS theories provide strong theoretical foundations of VDS based on independent drawings. In this paragraph, we recall a typical result that motivates independent drawing in the  $\ell_1$  recovery context [42, 17, 9, 29, 12, 4, 1]. Using the notation defined in the introduction, let us give a slightly modified version of [42, Theorem 4.2].

**THEOREM 2.4.** *Let  $p = (p_1, \dots, p_n)$  denote a probability distribution on  $\{1, \dots, n\}$  and  $\Omega \subset \{1, \dots, n\}$  denote a random set obtained by  $m$  independent drawings with respect to distribution  $p$ . Let  $S \in \{1, \dots, n\}$  be an arbitrary set of cardinality  $s$ . Let  $x$  be an  $s$ -sparse vector with support  $S$  such that the signs of its non-zero entries is a Rademacher or Steinhaus sequence<sup>2</sup>. Define:*

$$(2.1) \quad K(\mathbf{A}, p) := \max_{k \in \{1 \dots n\}} \frac{\|a_k\|_\infty^2}{p_k}$$

Assume that:

$$(2.2) \quad m \geq CK(\mathbf{A}, p) s \ln^2 \left( \frac{6n}{\eta} \right)$$

where  $C \approx 26.25$  is a constant. Then, with probability  $1 - \eta$ , vector  $x$  is the unique solution of the  $\ell_1$  minimization problem (1.1).

**REMARK 3.** *Candès and Plan have stated stronger results in the case of real matrices in [9]. Namely, the number of necessary measurements was decreased to  $O(s \log(n))$ , with lower constants and without any assumption on the vector signs. Their results have been derived using the so-called “golfing scheme” proposed in [19]. It is likely that these results could be extended to the complex case, however it would not change the optimal distribution which is the main point of this paper. We thus decided to stick to Theorem 2.4.*

The choice of an accurate distribution  $p$  is crucial since it directly impacts the number of measurements required. In the MRI community, a lot of heuristics have been proposed so far to identify the *best* sampling distribution. In the seminal paper on CS-MRI [32], Lustig *et al* have proposed to sample the  $k$ -space using a density that polynomially decays towards high frequencies. More recently, Knoll *et al* have generalized this approach by inferring the best exponent from MRI image databases [28]. It is actually easy to derive the theoretically *optimal distribution*, i.e. the one that minimizes the right hand-side in (2.2) as shown in Proposition 2.5, introduced in [12].

**PROPOSITION 2.5.** *Denote by  $K^*(\mathbf{A}) := \min_{p \in \Delta_n} K(\mathbf{A}, p)$ .*

<sup>2</sup>A Rademacher (resp. Steinhaus) random variable is uniformly distributed on  $\{-1; 1\}$  (resp. on the torus  $\{z \in \mathbb{C}; |z| = 1\}$ ).

(i) the optimal distribution  $\pi \in \Delta_n$  that minimizes  $K(\mathbf{A}, p)$  is:

$$(2.3) \quad \pi_i = \frac{\|a_i\|_\infty^2}{\sum_{i=1}^n \|a_i\|_\infty^2}$$

(ii)  $K^*(\mathbf{A}) = K(\mathbf{A}, \pi) = \sum_{i=1}^n \|a_i\|_\infty^2$ .

*Proof.*

- (i) Taking  $p = \pi$ , we get  $K(\mathbf{A}, \pi) = \sum_{i=1}^n \|a_i\|_\infty^2$ . Now assume that  $q \neq \pi$ , since  $\sum_{k=1}^n q_k = \sum_{k=1}^n \pi_k = 1$ ,  $\exists j \in \{1, \dots, n\}$  such that  $q_j < \pi_j$ . Then  $K(\mathbf{A}, q) \geq \|a_j\|_\infty^2 / q_j > \|a_j\|_\infty^2 / \pi_j = \sum_{i=1}^n \|a_i\|_\infty^2 = K(\mathbf{A}, \pi)$ . So,  $\pi$  is the distribution that minimizes  $K(\mathbf{A}, p)$ .
- (ii) This equality is a consequence of  $\pi$ 's definition.

□

The theoretical optimal distribution only depends on the acquisition matrix, i.e. on the acquisition and sparsifying bases. For instance, if we measure some Fourier frequencies of a sparse signal in the time domain (a sum of diracs), we should sample the frequencies according to a uniform distribution, since  $\|a_i\|_\infty = 1/\sqrt{n}$  for all  $1 \leq i \leq n$ . In this case,  $K^*(\mathbf{F}) = 1$  and the number of measurements  $m$  is proportional to  $s$ , which is in accordance with the seminal paper by Candès *et al.* [10].

**Independent drawings in MRI.** In the MRI case, the images are usually assumed sparse (or at least compressible) in a wavelet basis, while the acquisition is performed in the Fourier space. In this setting, the acquisition matrix can be written as  $\mathbf{A} = \mathbf{F}^* \Psi$ . In that case, the *optimal distribution* only depends on the choice of the wavelet basis. The optimal distributions in 2D and 3D are depicted in Fig. 2.1(a)-(b), respectively if we assume that the MR images are sparse in the Symmlet basis with 3 decomposition levels in the wavelet transform.

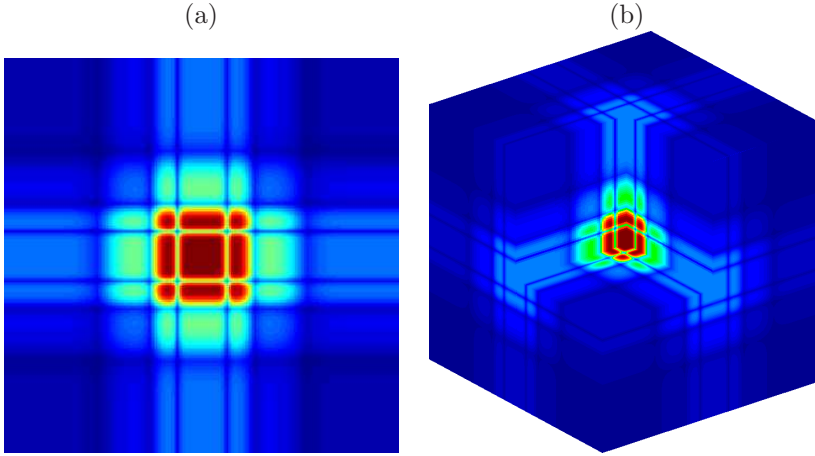


FIG. 2.1. *Optimal distribution  $\pi$  for a Symmlet-10 transform in 2D (a) and a maximal projection of the optimal distribution in 3D (b).*

Let us mention that similar distributions have been proposed in the literature. First, an alternative to independent drawing was proposed by Puy *et al.* [41]. Their approach consists in selecting or not a frequency by drawing a Bernoulli random variable. Its parameter is determined by minimizing a quantity that slightly differs



from  $K(\mathbf{A}, p)$ . Second, Krahmer and Ward [29] tried to unify theoretical results and empirical observations in the MRI framework. For Haar wavelets, they have shown that a polynomial distribution on the 2D  $k$ -space which varies as  $1/(k_x^2 + k_y^2)$  is close to the optimal solution since it verifies  $K(\mathbf{A}, p) = O(\log(n))$ . Our numerical experiments have confirmed that a decay as a power of 2 is near optimal in 2D.

In the next section, we improve the existing theories by showing that a deterministic sampling of highly coherent vectors (i.e. those satisfying  $\|a_i\|_\infty^2 \gg \frac{1}{n}$ ) may decrease the total number of required measurements. In MRI, this amounts to fully sampling the low frequencies, which exactly matches what has been done heuristically hitherto.

**2.2. Mixing deterministic and independent samplings.** In a recent work [12], we observed and partially justified the fact that a deterministic sampling of the low frequencies in MRI could drastically improve reconstruction quality. The following theorem proven in Appendix 1 provides a theoretical justification to this approach.

**THEOREM 2.6.** *Let  $S \in \{1, \dots, n\}$  be a set of cardinality  $s$ . Let  $x$  be an  $s$ -sparse vector with support  $S$  such that the signs of its non-zero entries is a Rademacher or Steinhaus sequence. Define the acquisition set  $\Omega \subseteq \{1, \dots, n\}$  as the union of:*

- (i) a deterministic set  $\Omega_1$  of cardinality  $m_1$ .
- (ii) a random set  $\Omega_2$  obtained by  $m_2$  independent drawing according to distribution  $p$  defined on  $\{1 \dots n\} \setminus \Omega_1$ .

Denote  $m = m_1 + m_2$ ,  $\Omega_1^c = \{1, \dots, n\} \setminus \Omega_1$  and let  $\Omega = \Omega_1 \cup \Omega_2$ . Assume that:

$$(2.4) \quad m \geq m_1 + CK(\mathbf{A}_{\Omega_1^c}, p)s \ln^2 \left( \frac{6n}{\eta} \right)$$

where  $C = 7/3$  is a constant, and  $K(\mathbf{A}_{\Omega_1^c}, p) = \max_{i \in \{1, \dots, n\} \setminus \Omega_1} \frac{\|a_i\|_\infty^2}{p_i}$ . Then, with probability  $1 - \eta$ , vector  $x$  is the unique solution of the  $\ell_1$  minimization problem (1.1).

This result implies that there exists an optimal partition between deterministically and randomly selected samples, which is moreover easy to compute. For example, consider the optimal distribution  $p_i \propto \|a_i\|_\infty^2$ , then  $K^*(\mathbf{A}_{\Omega_1^c}) = \sum_{i \in \{1, \dots, n\} \setminus \Omega_1} \|a_i\|_\infty^2$ .

If the measurement matrix contains rows with large values of  $\|a_i\|_\infty$ , we notice from inequality (2.4) that these frequencies should be sampled deterministically, whereas the rest of the measurements should be obtained from independent drawings. This simple idea is another way of overcoming the so-called coherence barrier [29, 1].

A striking example raised in [4] is the following. Assume that  $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{F}_{n-1}^* \end{pmatrix}$ . The assumed optimal independent sampling strategy would consist in independently drawing the rows with distribution  $p_1 = 1/2$  and  $p_k = 1/\sqrt{n-1}$  for  $k \geq 2$ . According to Theorem 2.4, the number of required measurements is  $2Cs \ln^2 \left( \frac{6n}{\eta} \right)$ . The alternative approach proposed in Theorem 2.6 basically performs a deterministic drawing of the first row combined with an independent uniform drawing over the remaining rows. In total, this scheme requires  $1 + Cs \ln^2 \left( \frac{6n}{\eta} \right)$  measurements and thus reduces the number of measurements by almost a factor 2. Note that the same gain would be obtained by using independent drawings with rejection.



**Mixed deterministic and independent sampling in MRI.** In our experiments, we will consider wavelet transforms with three decomposition levels and the Symmlet basis with 10 vanishing moments. Fig. 2.2(a)-(b) shows the modulus of  $\mathbf{A}$ 's entries with a specific reordering in (b) according to decaying values of  $\|a_i\|_\infty$ . This decay is illustrated in Fig. 2.2(c). We observe that a typical acquisition matrix in MRI shows large differences between its  $\|a_i\|_\infty$  values. More Precisely, there is a small number of rows with a large infinite norm, sticking perfectly to the framework of Theorem 2.6. This observation justifies the use of a partial deterministic  $k$ -space sampling, which had already been used in [32, 12]. In Fig. 2.2(d), the set  $\Omega_1$  is depicted for a fixed number of deterministic samples  $m_1$ , by selecting the rows with the largest infinite norms.

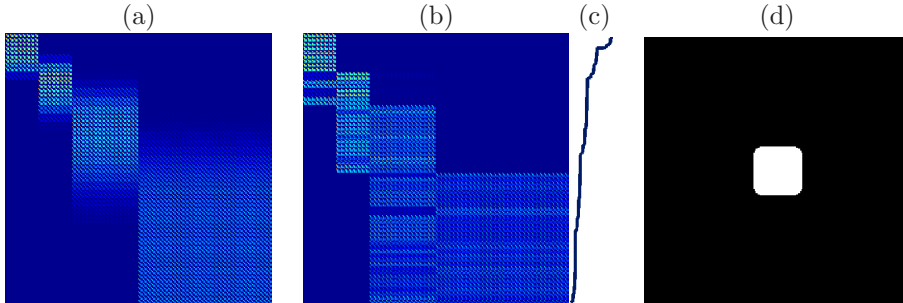


FIG. 2.2. (a): Absolute magnitudes of  $\mathbf{A}$  for a 2D Symmlet basis with 10 vanishing moments and 3 levels of decomposition. (b): same quantities as in (a) but sorted by decaying  $\|a_i\|_\infty$  (i.e. by decreasing order). (c): decay of  $\|a_i\|_\infty$ . (d): Set  $\Omega_1$  depicted in the 2D  $k$ -space.

Hereafter, the strategy we adopt is driven by the previous remarks. All our sampling schemes are performed according to Theorem. 2.6: a deterministic part is sampled, and a VDS is performed on the rest of the acquisition space (e.g. the high frequencies in MRI).

### 3. Variable density samplers along continuous curves.

**3.1. Why independent drawing can be irrelevant.** In many imaging applications, the number of samples is of secondary importance compared to the time spent to collect the samples. A typical example is MRI, where the important variable to control is the scanning time. It depends on the total length of the pathway used to visit the  $k$ -space rather than the number of collected samples. MRI is not an exception and many other acquisition devices have to meet such physical constraints amongst which are scanning probe microscopes, ultrasound imaging, ecosystem monitoring, radio-interferometry or sampling using vehicles subject to kinematic constraints [52]. In these conditions, measuring isolated points is not relevant and existing practical CS approaches consist in designing parametrized curves performing a variable density sampling. In what follows, we first review existing variable density sampling approaches based on continuous curves. Then, we propose two original contributions and analyze some of their theoretical properties. We mostly concentrate on continuity of the trajectory which is not sufficient for implementability in many applications. For instance, in MRI the actual requirement for a trajectory to be implementable is piecewise smoothness. More realistic constraints are discussed in Section 6.

**3.2. A short review of samplers along continuous trajectories.** The prototypical variable density samplers in MRI were based on spiral trajectories [45].

Similar works investigating different shapes and densities from a heuristic point of view were proposed in [49, 27, 35]. The first reference to compressed sensing appeared in the seminal paper [32]. In this work, Lustig *et al* have proposed to perform independent drawings in a 2D plane (defined by the partition and phase encoding directions) and sample continuously along the orthogonal direction to design piecewise continuous schemes in the 3D  $k$ -space (see Fig. 3.1). These authors have also suggested to make use of randomly perturbed spirals. The main advantage of these schemes lies in their simplicity of practical implementation since they only require minor modifications of classical MRI acquisition sequences.

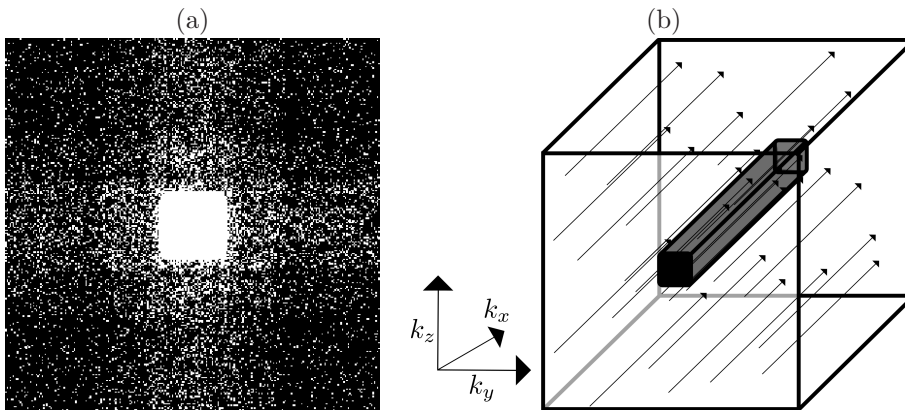


FIG. 3.1. Classical CS-MRI strategy. (a): 2D independent sampling according to a distribution  $\pi$ . (b): measurements performed in the orthogonal readout direction.

Recent papers [37, 4, 7] have generalized CS results from independent drawing of isolated measurements to independent drawings of blocks of measurements. In these contributions, the blocks can be chosen arbitrarily and may thus represent continuous trajectories. Interestingly, these authors have provided closed form expressions for the optimal distribution on the block set. Nevertheless, this distribution is very challenging to compute in large scale problems. Moreover, the restriction to sets of admissible blocks reduces the versatility of many devices such as MRI and can therefore impact the image reconstruction quality.

In many applications the length of the sampling trajectory is more critical than the number of acquired samples, therefore, finding the shortest pathway amongst random points drawn independently has been studied as a way of designing continuous trajectories [52, 50]. Since this problem is NP-hard, one usually resorts to a TSP solver to get a reasonable suboptimal trajectory. To the best of our knowledge, the only practical results obtained using the TSP were given by Wang *et al* [50]. In this work, the authors did not investigate the relationship between the initial sample locations and the empirical measure of the TSP curve. In Section 4, it is shown that this relationship is crucial to make efficient TSP-based sampling schemes.

In what follows, we first introduce an original sampler based on random walks on the acquisition space and then analyse its asymptotic properties. Our theoretical investigations together with practical experiments allows us to show that the VDS mixing properties play a central role to control its efficiency. This then motivates the need for more global VDS schemes.

**3.3. Random walks on the acquisition space.** Perhaps the simplest way to transform independent random drawings into continuous random curves consists in performing random walks on the acquisition space. Here, we discuss this approach and provide a brief analysis of its practical performance in the discrete setting. Through both experimental and theoretical results, we show that this technique is doomed to fail. However, we believe that this theoretical analysis provides a deep insight on what VDS properties characterize its performance.

Let us consider a time-homogeneous Markov chain  $X = (X_n)_{n \in \mathbb{N}}$  on the set  $\{1, \dots, n\}$  and its transition matrix denoted  $\mathbf{P} \in \mathbb{R}^{n \times n}$ . If  $X$  possesses a stationary distribution, i.e. a row vector  $p \in \mathbb{R}^n$  such that  $p = p\mathbf{P}$  then, by definition,  $X$  is a  $p$ -variable density sampler.

**3.3.1. Construction of the transition matrix  $\mathbf{P}$ .** A classical way to design a transition kernel ensuring that (i)  $p$  is the stationary distribution of the chain and (ii) the trajectory defined by the chain is continuous, is the Metropolis algorithm [21]. For a pixel/voxel position  $i$  in the 2D/3D acquisition space, let us define by  $\mathcal{N}(i) \subseteq \{1, \dots, n\}$  its neighbourhood, i.e. the set of possible measurement locations allowed when staying on position  $i$ . Let  $|\mathcal{N}(i)|$  denote the cardinal of  $\mathcal{N}(i)$  and define the *proposal kernel*  $\mathbf{P}^*$  as  $\mathbf{P}_{i,j}^* = |\mathcal{N}(i)|^{-1} \delta_{j \in \mathcal{N}(i)}$ . The Metropolis algorithm proceeds as follows:

1. from state  $i$ , draw a state  $i^*$  with respect to the distribution  $\mathbf{P}_{i,\cdot}^*$ .
2. accept the new state  $i^*$  with probability:

$$(3.1) \quad q(i, i^*) = \min \left( 1, \frac{p(i^*)\mathbf{P}_{i^*,i}^*}{p(i)\mathbf{P}_{i,i^*}^*} \right).$$

Otherwise stay in state  $i$ .

The transition matrix  $\mathbf{P}$  can then be defined by  $\mathbf{P}_{i,j} = q(i, j)\mathbf{P}_{i,j}^*$  for  $i \neq j$ . The diagonal is defined in a such a way that  $\mathbf{P}$  is a stochastic matrix. It is easy to check that  $p$  is an invariant distribution for this chain<sup>3</sup>. It is worth noticing that if the chain is irreducible positive recurrent (which is fulfilled if the graph is connected and the distribution  $p$  positive), the ergodic theorem ensures that  $X$  is a  $p$ -VDS.

Unfortunately, trajectories designed by this technique leave huge parts of the acquisition space unexplored (see Fig. 3.2 (a)). To circumvent this problem, we may allow the chain to *jump* to independent locations over the acquisition space. Let  $\tilde{\mathbf{P}}$  be the Markov kernel corresponding to independent drawing with respect to  $p$ , i.e.  $\tilde{\mathbf{P}}_{i,j} = p_j$  for all  $1 \leq i, j \leq n$ . Define:

$$(3.2) \quad \mathbf{P}^{(\alpha)} = (1 - \alpha)\mathbf{P} + \alpha\tilde{\mathbf{P}} \quad \forall 0 \leq \alpha \leq 1.$$

Then the Markov chain associated with  $\mathbf{P}^{(0)}$  corresponds to a continuous random walk, while the Markov chain associated with  $\mathbf{P}^{(\alpha)}$ ,  $\alpha > 0$  has a nonzero *jump* probability. This means that the trajectory is composed of continuous parts of average length  $1/\alpha$ .

**3.3.2. Example.** In Fig. 3.2, we show illustrations in the 2D MRI context where the discrete  $k$ -space is of size  $64 \times 64$ . On this domain, we set a distribution  $p$  which matches distribution  $\pi$  in Fig. 2.1 (a). We perform a random walk on the acquisition space until 10% of the coefficients are selected. In Fig. 3.2(a), we set  $\alpha = 0$  whereas

<sup>3</sup>If the neighboring system is such that the corresponding graph is connected, then the invariant distribution is unique.

$\alpha = 0.1$  in Fig. 3.2(b). As expected,  $\alpha = 0$  leads to a sampling pattern where large parts of the  $k$ -space are left unvisited. The phenomenon is partially corrected using a nonzero value of  $\alpha$ .

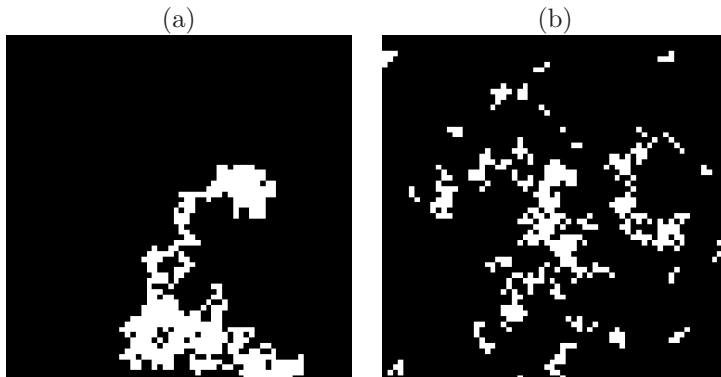


FIG. 3.2. Example of sampling trajectories in 2D MRI. (a) (resp. (b)): 2D sampling scheme of the  $k$ -space with  $\alpha = 0$  (resp.  $\alpha = 0.1$ ). Drawings are performed until 10% of the coefficients are selected ( $m = 0.1n$ ).

REMARK 4. Performing  $N$  iterations of the Metropolis algorithm requires  $O(N)$  computations leading to a fast sampling scheme design procedure. In our experiments, we iterate the algorithm until  $m$  different measurements are probed. Therefore, the number of iterations  $N$  required increases non linearly with respect to  $m$ , and can be time consuming especially when  $R = m/n$  is close to 1. This is not a tough limitation of the method since the sampling scheme is computed off-line.

**3.3.3. Compressed sensing results.** Let us assume<sup>4</sup> that  $\mathbb{P}(X_1 = i) = p_i$  and that  $X_i$  is drawn using  $\mathbf{P}$  as a transition matrix. The following result provides theoretical guarantees about the performance of the VDS  $X$ :

PROPOSITION 3.1 (see [13]). Let  $\Omega := X_1, \dots, X_m \subset \{1, \dots, n\}$  denote a set of  $m$  indexes selected using the Markov chain  $X$ .

Then, with probability  $1 - \eta$ , if

$$(3.3) \quad m \geq \frac{12}{\epsilon(\mathbf{P})} K^2(\mathbf{A}, p) s^2 \log(2n^2/\eta),$$

every  $s$ -sparse signal  $x$  is the unique solution of the  $\ell_1$  minimization problem.

The proof of this proposition is given in Appendix 2. Before going further, some remarks may be useful to explain this theoretical result.

REMARK 5. Since the constant  $K^2(\mathbf{A}, p)$  appears in Eq. (3.3), the optimal sampling distribution using Markov chains is also distribution  $\pi$ , as proven in Proposition 2.5.

REMARK 6. In contrast to Theorem 2.4, Proposition 3.1 provides uniform results, i.e. results that hold for all  $s$ -sparse vectors.

REMARK 7. Ineq. 3.3 suffers from the so-called quadratic bottleneck (i.e. an  $O(s^2 \log(n))$  bound). It is likely that this bound can be improved to  $O(s \log(n))$  by developing new concentration inequalities on matrix-valued Markov chains.

<sup>4</sup>By making this assumption, there is no burn-in period and the chain  $X$  converges more rapidly to its stationary distribution  $p$ .

REMARK 8. *More importantly, it seems however unlikely to avoid the spectral gap  $O(1/\epsilon(P))$  using the standard mechanisms for proving compressed sensing results. Indeed, all concentration inequalities obtained so far on Markov chains (see e.g. [31, 26, 36]) depend on  $1/\epsilon(P)$ . The spectral gap satisfies  $0 < \epsilon(P) \leq 1$  and corresponds to mixing properties of the chain. The closer the spectral gap to 1, the fastest ergodicity is achieved. Roughly speaking, if  $|i-j| > 1/\epsilon(P)$  then  $X_i$  and  $X_j$  are almost independent random variables. Unfortunately, the spectral gap usually depends on the dimension  $n$  [15]. In our example, it can be shown using Cheeger's inequality that  $\epsilon(P) = O\left(n^{-\frac{1}{d}}\right)$  if the stationary distribution  $\pi$  is uniform (see Appendix 3). This basically means that the number of measurements necessary to accurately reconstruct  $x$  could be as large as  $O(sn^{1/d} \log(n))$ , which strongly limits the interest of this CS approach. The only way to lower this number consists in frequently jumping since Weyl's theorem [22] ensures that  $\epsilon(P^{(\alpha)}) > \alpha$ .*

To sum up, the main drawback of random walks lies in their inability to cover the acquisition space quickly since they are based on local considerations. Keeping this in mind, it makes sense to focus on more global displacement strategies that allow a faster exploration of the whole acquisition domain. In the next section, we thus introduce this global sampling alternative based on TSP-solver. Our main contribution is the derivation of the link between a prescribed a priori sampling density and the distribution of samples located on the TSP solution so as to eventually get a VDS.

**4. Travelling salesman-based VDS.** In order to design continuous trajectories, we may think of picking points at random and join them using a travelling salesman problem (TSP) solver. Hereafter, we show how to draw the initial points in order to reach a target distribution  $p$ . In this section, the probability distribution  $p$  is assumed to be a density.

**4.1. Introduction.** The naive idea would consist in drawing some points according to the distribution  $p$  and joining them using a TSP solver. Unfortunately, the trajectory which results from joining all samples does not fit the distribution  $p$ , as shown in Fig. 4.1(b)-(d). To bring evidence to this observation, we performed a Monte Carlo study, where we drew one thousand sampling schemes, each one designed by solving the TSP on a set of independent random samples. We notice in Fig. 4.1 (d) that the empirical distribution of the points along the TSP curve, hereafter termed the final distribution, departs from the original distribution  $p$ . A simple intuition can be given to explain this discrepancy between the initial and final distributions in a  $d$ -dimensional acquisition space. Consider a small subset of the acquisition space  $\omega$ . In  $\omega$ , the number of points is proportional to  $p$ . The typical distance between two neighbors in  $\omega$  is then proportional to  $p^{-1/d}$ . Therefore, the local length of the trajectory in  $\omega$  is proportional to  $pp^{-1/d} = p^{1-1/d} \neq p$ . In what follows, we will show that the empirical measure of the TSP solution converges to a measure proportional to  $p^{1-1/d}$ .

**4.2. Definitions.** We shall work on the hypercube  $\mathcal{H} = [0, 1]^d$  with  $d \geq 2$ . In what follows,  $\{x_i\}_{i \in \mathbb{N}^*}$  denotes a sequence of points in the hypercube  $\mathcal{H}$ , independently drawn from a density  $p : \mathcal{H} \mapsto \mathbb{R}_+$ . The set of the first  $N$  points is denoted  $X_N = \{x_i\}_{i \leq N}$ .

Using the definitions introduced in Tab. 1.1, we introduce  $\gamma_N : [0, 1] \rightarrow \mathcal{H}$  the function that parameterizes  $C(X_N)$  by moving along it at constant speed  $T(X_N, \mathcal{H})$ . Then, the *distribution of the TSP solution* reads as follows:

DEFINITION 4.1. *The distribution of the TSP solution is denoted  $\tilde{P}_N$  and defined, for any Borelian  $B$  in  $\mathcal{H}$  by:*

$$\tilde{P}_N(B) = \lambda_{[0,1]}(\gamma_N^{-1}(B)).$$

REMARK 9. *The distribution  $\tilde{P}_N$  is defined for fixed  $X_N$ . It makes no reference to the stochastic component of  $X_N$ .*

REMARK 10. *A more intuitive definition of  $\tilde{P}_N$  can be given if we introduce other tools. For a subset  $\omega \subseteq \mathcal{H}$ , we denote the length of  $C(X_N) \cap \omega$  as  $T|_\omega(X_N, \mathcal{H}) = T(X_N, \mathcal{H})\tilde{P}_N(\omega)$ . Using this definition, it follows that:*

$$(4.1) \quad \tilde{P}_N(\omega) = \frac{T|_\omega(X_N, \mathcal{H})}{T(X_N, \mathcal{H})}, \quad \forall \omega.$$

Then  $\tilde{P}_N(\omega)$  is the relative length of the curve inside  $\omega$ .

**4.3. Main results.** Our main theoretical result introduced in [11] reads as follows:

THEOREM 4.2. *Define the density  $\tilde{p} = \frac{p^{(d-1)/d}}{\int_{\mathcal{H}} p^{(d-1)/d}(x)dx}$  where  $p$  is a density defined on  $\mathcal{H}$ . Then almost surely with respect to the law  $p^{\otimes \mathbb{N}}$  of the random points sequence  $\{x_i\}_{i \in \mathbb{N}^*}$  in  $\mathcal{H}$ , the distribution  $\tilde{P}_N$  converges in distribution to  $\tilde{p}$ :*

$$(4.2) \quad \tilde{P}_N \xrightarrow{(d)} \tilde{p} \quad p^{\otimes \mathbb{N}}\text{-a.s.}$$

The proof of the theorem is given in Appendix 4.

REMARK 11. *The TSP solution does not define as such a VDS, since the underlying process is finite in time. Nevertheless, since  $\tilde{P}_N$  is the occupation measure of  $\gamma_N$ , the following result holds:*

COROLLARY 4.3.  *$(\gamma_N)_{N \in \mathbb{N}}$  is a generalized  $\tilde{p}$  VDS.*

REMARK 12. *The theorem indicates that if we want to reach distribution  $p$  in  $2D$ , we have to draw the initial points with respect to a distribution proportional to  $p^2$ , and to  $p^{3/2}$  in  $3D$ . Akin to the previous Monte Carlo study illustrating the behavior of the naive approach in Fig. 4.1 (top row), we repeated the same procedure after having taken this result into account. The results are presented in Fig. 4.1(e)-(g), in which it is shown that the final distribution now closely matches the original one (compare Fig. 4.1(g) with Fig. 4.1(a)).*

REMARK 13. *Contrarily to the Markov chain approach for which we derived compressed sensing results in Proposition 3.1, the TSP approach proposed here is mostly heuristic and based on the idea that the TSP solution curve covers the space rapidly. An argument supporting this idea is the fact that in  $2D$ , the TSP curve  $C(X_N)$  does not self-intersect. This property is clearly lacking for random walks.*

REMARK 14. *One of the drawback of this approach is the TSP's NP-hardness. We believe that this is not a real problem. Indeed, there now exist very efficient approximate solvers such as the Concorde solver [2]. It finds an approximate solution with  $10^5$  cities from a few seconds to a few hours depending on the required accuracy of the solution. The computational time of the approximate solution is not a real limitation since the computation is done off-line from the acquisition procedure. Moreover,*



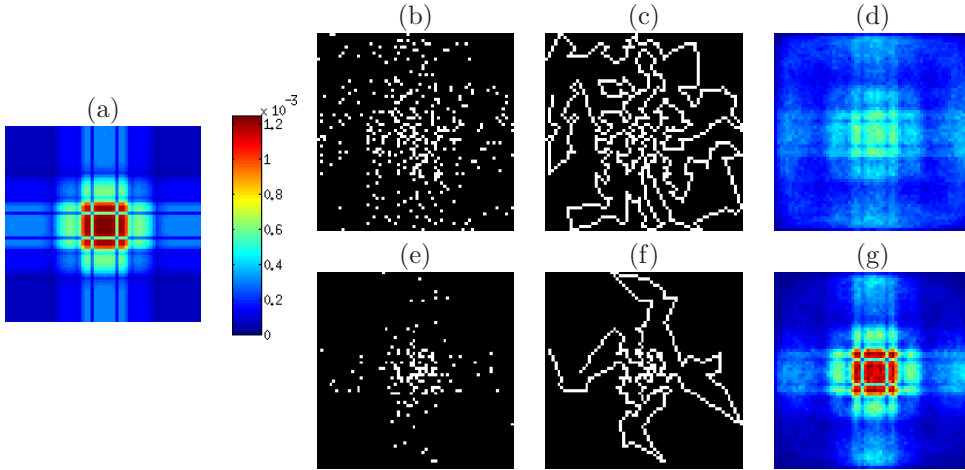


FIG. 4.1. Illustration of the TSP-based sampling scheme to reach distribution  $\pi$ . (a): distribution  $\pi$ . (b) (resp. (e)): independent drawing of points from distribution  $\pi$  (resp.  $\propto \pi^2$ ). (c) (resp (f)): solution of the TSP amongst points of (b) (resp. (e)). (d) and (g): Monte Carlo study: average scheme over one thousand drawings of sampling schemes, with the same color scale as in (a).

many solvers are actually designed in such a way that their solution also fulfil Theorem. 4.2. For example, in 2D, to reach a sampling factor of  $R = 5$  on a  $256 \times 256$  image, one need  $N \simeq 10^4$  cities, and an approximate solution is obtained in 142s. In 3D, for a  $256 \times 256 \times 256$  image,  $N \simeq 9 \cdot 10^5$  and an approximate solution is obtained in about 4 hours. In each case the solutions seem to be correctly approximated. In particular they do not self-intersect in 2D.

**5. Experimental results in MRI.** In this section, we focus on the reconstruction results by minimizing the  $\ell_1$  problem (1.1) with a simple MRI model:  $\mathbf{A} = \mathbf{F}^* \Psi$ , where  $\Psi$  denote the inverse Symmlet-10 transform<sup>5</sup>. The solution is computed using Douglas-Rachford’s algorithm [14]. We consider an MR image of size  $256 \times 256 \times 256$  as a reference, and perform reconstruction for different discrete sampling strategies. Every sampling scheme was regridded using a nearest neighbour approach to avoid data interpolation.<sup>6</sup>

**5.1. 2D-MRI.** In 2D, we focused on a single slice of the MR image and considered its discrete Fourier transform as the set of possible measurements. First, we found the best made a comparison of independent drawings with respect to various distributions in order to find heuristically the best sampling density. Then we explored the performance of the two proposed methods to design continuous schemes: random walks and Travelling Salesman Problem. We also compared our solution to classical MRI sampling schemes. In every sampling schemes, the number of measurements is

<sup>5</sup>We focused on  $\ell_1$  reconstruction since it is central in the CS theory. The reconstruction quality can be improved by considering more *a priori* knowledge on the image. Moreover we considered a simple MRI model, but our method can be extended to parallel MRI [39], or spread-spectrum techniques [20, 40].

<sup>6</sup>We provide Matlab codes to reproduce the proposed experiments here: <http://chauffertn.free.fr/codes.html>



the same and equals 20% of the number of pixels in the image, so that the *sampling factor*  $R$  is equal to 5. In cases where the sampling strategy is based on randomness (VDS, random walks, TSP...), we performed a Monte Carlo study by generating 100 sampling patterns for each variable density sampler.

**5.1.1. Variable density sampling using independent drawings.** Here, we assessed the impact of changing the sampling distribution using independent drawings. In all experiments, we sampled the Fourier space center deterministically as shown on Figure 5.1.

TABLE 5.1

*Quality of reconstruction results in terms of PSNR for 2D sampling with variable density independent drawings.*

	$\pi$	polynomial decay: $(k_x^2 + k_y^2)^{-d/2}$					
		$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
mean PSNR (dB)	35.6	<b>36.4</b>	<b>36.4</b>	36.3	36.0	35.5	35.2
std dev.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1

Table 5.1 shows that the theoretically-driven optimal distribution  $\pi$  is outperformed by the best heuristics. Amongst the latter, the distribution leading to the best reconstruction quality decays as  $1/|k|^2$ , which is the distribution used by Krahmer and Ward [29] as an approximation of  $\pi$  for Haar wavelets. The standard deviation of the PSNR is negligible compared to the mean values and for a given distribution, each reconstruction PSNR equals its average value at the precision used in Tab. 5.1.

**5.1.2. Continuous VDS.** In this part we compared various variable density samplers:

- Random walks with a stationary distribution proportional to  $1/|k|^2$  and different average chain lengths of  $1/\alpha$ ,
- TSP-based sampling with distributions proportional to  $1/|k|^2$  and  $\pi$ ,
- Classical MRI sampling strategies such as spiral, radial and radial with random angles. The choice of the spiral follows Example 5: the spiral is parameterized by  $s : [0, T] \rightarrow \mathbb{R}^2$ ,  $\theta \mapsto r(\theta/T) \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$  where  $r(t) := \frac{r(0)r(1)}{r(1)-t(r(1)-r(0))}$ , so as the spiral density decays as  $1/|k|^2$ .

The sampling schemes are presented in Fig. 5.1 and the reconstruction results in Tab. 5.2.

TABLE 5.2

*Quality of reconstruction results in terms of PSNR for continuous sampling trajectories.*

	Markovian drawing ( $\alpha$ )			TSP sampling		spiral	radial	radial random
	0.1	0.01	0.001	$\propto \pi$	$\propto 1/ k ^2$			
mean PSNR	35.7	34.6	33.5	35.6	<b>36.1</b>	35.6	34.1	33.1
std dev.	0.1	0.3	0.6	0.1	0.1			0.4
max value	36.0	35.1	34.8	35.9	36.2			34.0
in Fig. 5.1:	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)

As predicted by the theory, the shorter the chains the better the reconstructions. The optimal case corresponds to chains of length 1 ( $\alpha = 1$ ) i.e. corresponding to

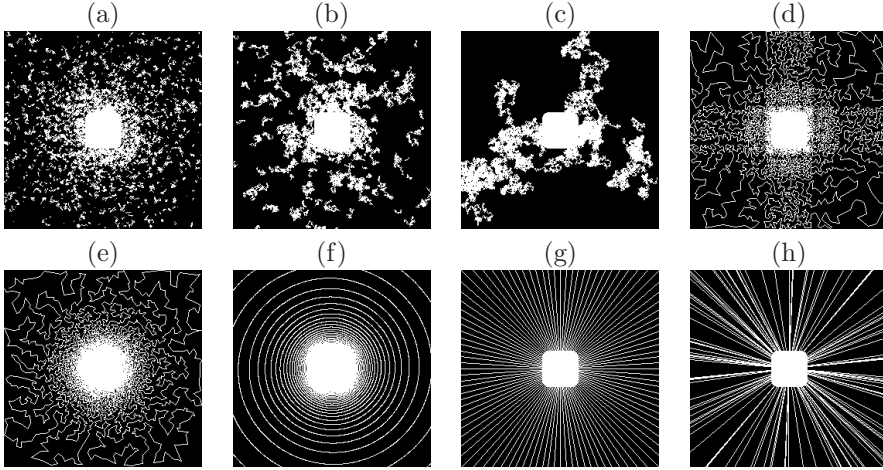


FIG. 5.1. 2D continuous sampling schemes based on random walks with  $\alpha = .1$  (a),  $\alpha = .01$  (b),  $\alpha = .001$  (c), and based on TSP solutions with distributions proportional to  $\pi$  (d) and to  $1/|k|^2$  (e). Classical sampling schemes: spiral (f), radial (g) and radial with random angles (h).

independent VDS. When the chain is too long, large  $k$ -space areas are left unexplored, and the reconstruction quality decreases.

Besides, the use of a target distribution proportional to  $1/|k|^2$  instead of  $\pi$  for TSP-based schemes provides slightly better reconstruction results.

We also considered more classical sampling scheme. We observe that the spiral scheme and the proposed ones provide more accurate reconstruction results than radial schemes. We believe that the main reason underlying these different behaviors is closely related to the sampling rate decay from low to high frequencies, which is proportional to  $1/|k|$  for radial schemes.

**5.2. 3D-MRI.** Since VDS based on Markov chains have shown rather poor reconstruction results compared to the TSP-based sampling schemes in 2D simulations, we only focus on comparing TSP-based sampling schemes to classical CS sampling schemes. Moreover, the computational load to treat 3D images being much higher than in 2D, we only perform one drawing per sampling scheme in the following experiments. Experiments in 2D suggest that the reconstruction quality is not really impacted by the realization of a particular sampling scheme, except for drawing with Markov chains or with radial with random angles, which are not considered in our 3D experiments.

**5.2.1. Variable density sampling using independent drawings.** The first step of the TSP-based approach is to identify a relevant target distribution. For doing so, we consider independent drawings as already done in 2D. The results are summarized in Tab. 5.3. In this experiment, we still use a number of measurements equal to 20% of the total amount ( $R = 5$ ).

The best reconstruction result is achieved with  $d = 2$  and not the theoretically optimal distribution  $\pi$ . This illustrates the importance of departing from the sole sparsity hypothesis under which we constructed  $\pi$ . Natural signals have a much richer structure. For instance wavelet coefficients tend to become sparser as the resolution levels increase, and this feature should be accounted for to derive optimal sampling densities for natural images (see Section 6.)

TABLE 5.3

Quality of reconstruction results in terms of PSNR for sampling schemes based on 3D variable density independent drawings, with densities  $\propto 1/k^d$  and  $\pi$ , and with 20% of measured samples.

$d$	1	2	3	4	$\pi$
PSNR (dB)	44.78	<b>45.01</b>	44.56	44.03	42.94

**5.2.2. Efficiency of the TSP sampling based strategy.** Let us now compare the reconstruction results using the TSP based method and the method proposed in the original CS-MRI paper [32]. These two sampling strategies are depicted in Fig. 5.2. For 2D independent drawings, we used the distribution providing the best reconstruction results in 2D, *i.e.* proportional to  $1/|k|^2$ . The TSP-based schemes were designed by drawing city locations independently with respect to a distribution proportional to  $p^{\frac{3}{2}}$ . According to Theorem 4.2 this is the correct way to reach distribution  $p$  after joining the cities with constant speed along the TSP solution path. The experiments were performed with  $p = \pi$  (see Fig. 2.1 (b)), and  $p \propto 1/|k|^2$ , since the latter yielded the best reconstruction results in the 3D independent VDS framework. We also compared these two continuous schemes to 3D independent drawings with respect to a distribution proportional to  $1/|k|^2$ .

Reconstruction results with an sampling rate  $R = 8.8$  are presented in Fig. 5.4, with a zoom on the cerebellum. The reconstruction quality using the proposed sampling scheme is better than the one obtained from classical CS acquisition and contains less artifacts. In particular, the branches of the cerebellum are observable with our proposed sampling scheme only. At higher sampling rate, we still observe less artifacts with the proposed schemes, as depicted in Fig. 5.5 with a sampling rate  $R = 14.9$ . Moreover, Fig. 5.3 shows that our proposed method outperforms the method proposed in [32] by up to 2dB. If one aims at reaching a fixed PSNR, we can increase  $r$  by more than 50% using the TSP based strategy. In other words, we could expect a substantial decrease of scanning time by using more advanced sampling strategies than those proposed until to now.

The two different choices of the target density  $\pi$  and  $\propto 1/|k|^2$  provide similar results. This is a bit surprising since 3D independent VDS with these two probability distributions provide very different reconstruction results (see Tab. 5.3). A potential explanation for that behavior is that the TSP tends to “smooth out” the target distribution. An independent drawing would collect very few Fourier coefficients in the blue zones of Fig. 2.1, notably the vertical and horizontal lines crossing the Fourier plane center. Sampling these zones seems to be of utmost importance since they contain high energy coefficients. The TSP approach tends to sample these zones by crossing the lines.

Perhaps the most interesting fact is that Fig. 5.3 shows that the TSP based sampling schemes provide results that are similar to independent drawings up to important sampling rates such as 20. We thus believe that the TSP solution proposed in this paper is near optimal since it provides results similar to unconstrained acquisition schemes. The price to be paid by integrating continuity constraints is thus almost null.

**6. Discussion and perspectives.** In this paper, we investigated the use of variable density sampling along continuous trajectories. Our first contribution was to provide a well-grounded mathematical definition of  $p$ -variable density samplers (VDS) as stochastic processes with a prescribed limit empirical measure  $p$ . We identified

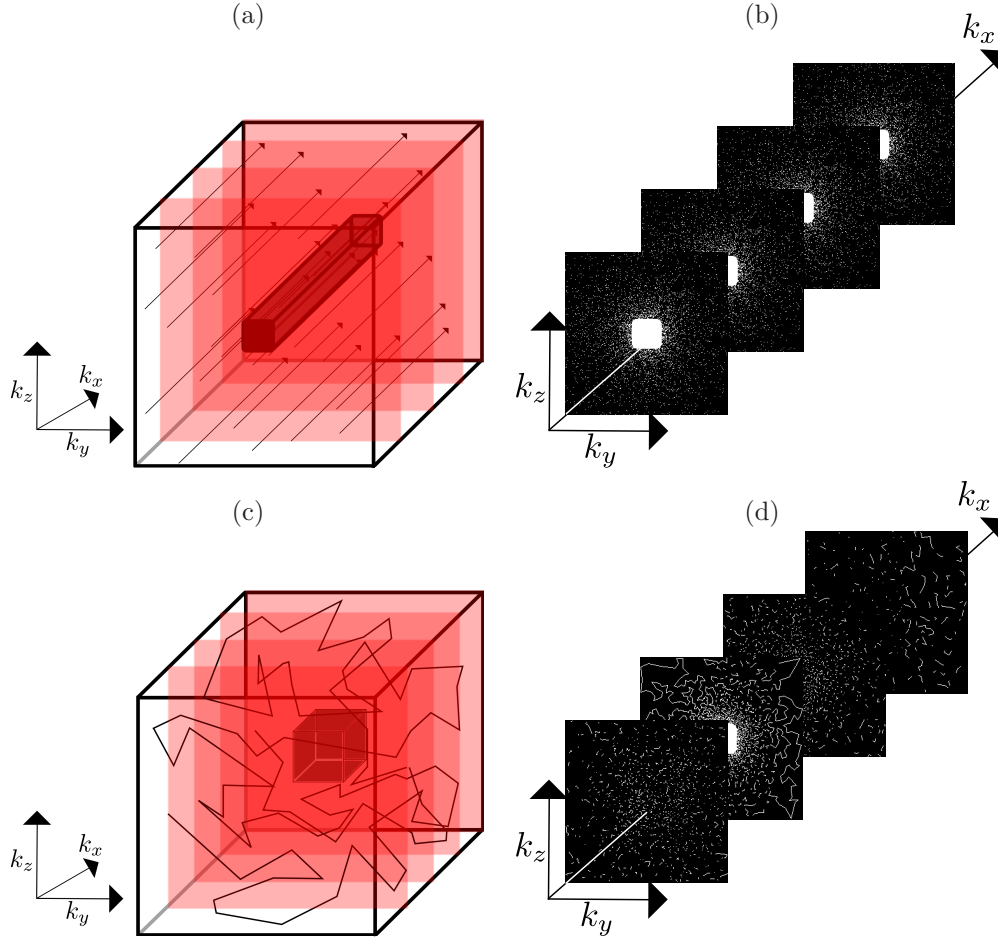


FIG. 5.2. Compared sampling strategies in 3D-MRI. **Top:** 2D independent drawing sampling schemes designed by a planar independent drawing and measurements in the orthogonal readout direction. **Bottom:** 3D TSP-based sampling scheme. **Left:** Schematic representation of the 3D sampling scheme. **Right:** Representations of 4 parallel slices.

through both theoretical and experimental results two key features characterizing their efficiency: their **empirical measure** as well as their **mixing properties**. We showed that VDS based on random walks were doomed to fail since they were unable to quickly cover the whole acquisition space. This led us to propose a two-step alternative that consists first of drawing random points independently and then joining them using a Travelling Salesman Problem solver. In contrast to what has been proposed in the literature so far, we paid attention to the manner the points have to be drawn so as to reach a prescribed empirical measure. Strikingly, our numerical results suggest that the proposed approach yields reconstruction results that are nearly equivalent to independent drawings. This suggests that adding continuity constraints to the sampling schemes might not be so harmful to derive CS results.

We believe that the proposed work opens many perspectives as outlined in what follows.

**How to select the target density?** We recalled existing theoretical results

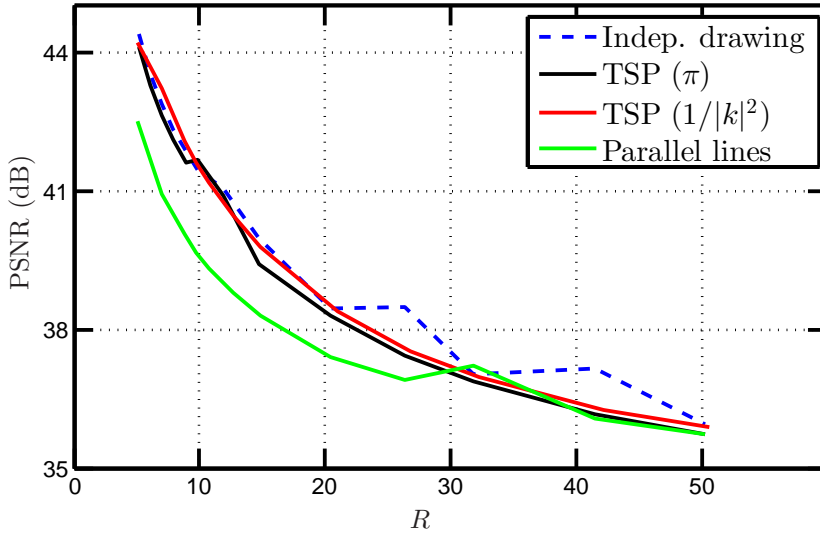


FIG. 5.3. Quality of 3D reconstructed images in terms of PSNR as a function of sampling rates  $R$  for various sampling strategies: independent drawings with respect to distribution  $\propto 1/|k|^2$  (dashed blue line), TSP-based sampling with target densities  $\pi$  (black line) and  $\propto 1/|k|^2$  (red line), and parallel lines with 2D independent drawing with respect to  $\propto 1/|k|^2$  distribution (green line) as depicted in Fig. 5.2[Top row].

to address this point in Section 2 and showed that *deterministic* sampling could reduce the total number of required measurements. The analysis we performed closely followed the proofs proposed in [42, 9] and was based solely on sparsity hypotheses on the signal/image to be reconstructed. The numerical experiments we performed indicate that heuristic densities still outperform the theoretical optimal ones. This suggests that the optimality criteria used so far to derive target sampling densities does not account for the whole structure of the sought signal/image. Although sparsity is a key feature that characterizes natural signals/images, we believe that introducing stronger knowledge like *structured sparsity* might contribute to derive a new class of optimal densities that would compete with heuristic densities.

To the best of our knowledge, the recent paper [1] is the first contribution that addresses the design of sampling schemes by accounting for a simple structured sparsity hypothesis. The latter assumes that wavelet coefficients become sparser as the resolution increases. The main conclusion of the authors is the same as that of Theorem 2.6 even though it is based on different arguments: the low frequencies of a signal should be sampled deterministically.

Finally, let us notice that the best empirical convex reconstruction techniques do not rely on the resolution of a simple  $\ell^1$  problem such as (1.1). They are based on regularization with redundant frames and total variation for instance [6]. The signal model, the target density and the reconstruction algorithm should clearly be considered simultaneously to make a substantial leap on reconstruction guarantees.

**What VDS properties govern their practical efficiency?** In Section 3, it was shown that the key feature characterizing random walks efficiency was the mixing properties of the associated stochastic transition matrix. In order to derive CS results using generic random sets rather than point processes or random walks, it seems important to us to find an equivalent notion of mixing properties.

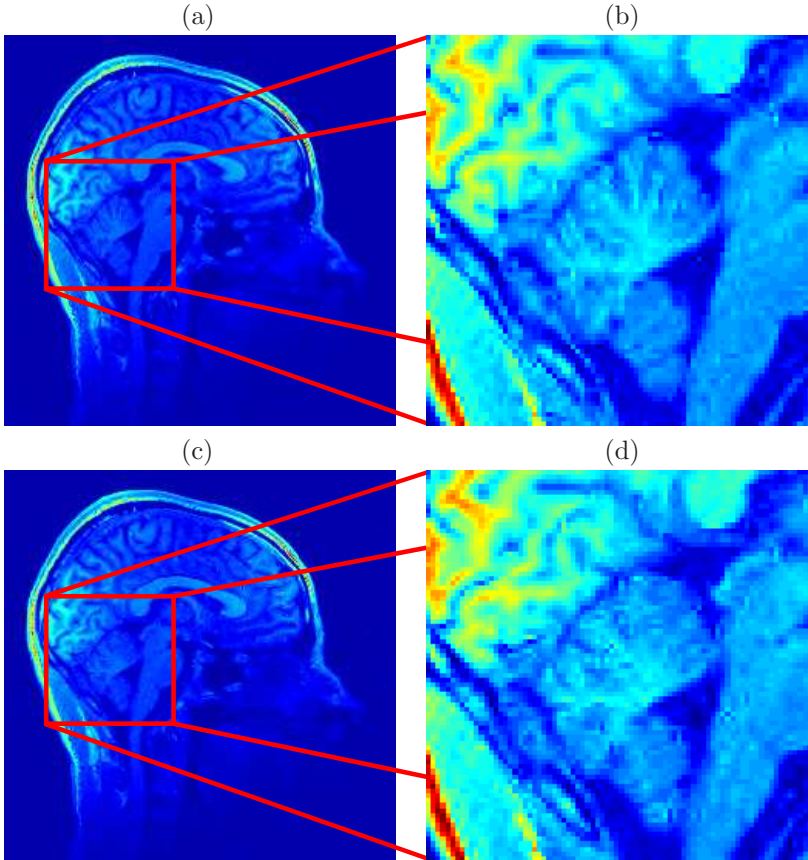


FIG. 5.4. Reconstruction results for  $R = 8.8$  for various sampling strategies. **Top row:** TSP-based sampling schemes (PSNR=42.1 dB). **Bottom row:** 2D random drawing and acquisitions along parallel lines (PSNR=40.1 dB). Sagittal view (left) and zoom on the cerebellum (right).

**How to generate VDS with higher degrees of regularity?** This is probably the most important question from a practical point of view. We showed that the TSP based VDS outperformed more conventional sampling strategies by substantial acceleration factors for a given PSNR value or recovers 3D images at an improved PSNR for a given acceleration factor. However, this approach may not really be appealing for many applications: continuity is actually not a sufficient condition for making acquisition sequences implementable on devices like MRI scanners or robot motion where additional kinematic constraints such as bounded first (gradients) and second (slew rate) derivatives should be taken into account. Papers such as [33] derive time-optimal waveforms to cross a given curve using optimal control. By using this approach, it can be shown that the angular points on the TSP trajectory have to be visited with a zero speed. This strongly impacts the scanning time and the distribution of the parametrized curve. The simplest strategy to reduce scanning time would thus consist in smoothing the TSP trajectory, however this approach dramatically changes the target distribution which was shown to be a key feature of the method. The key element to prove our TSP Theorem 4.2 was the famous Beardwood, Halton and Hammersley theorem [3]. To the best of our knowledge, extending this



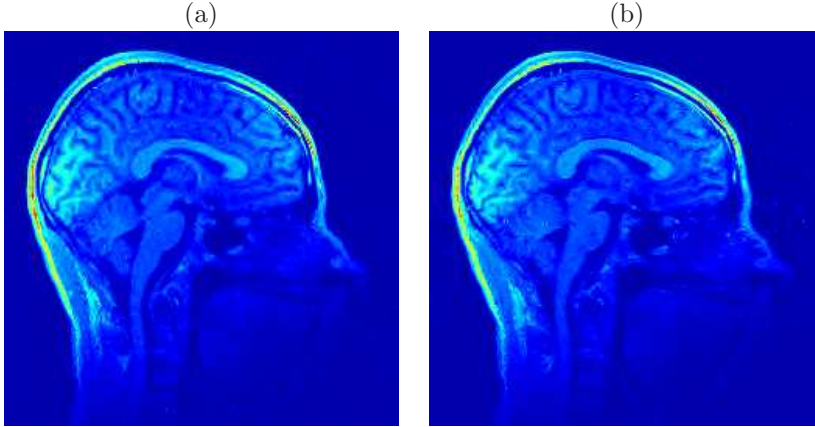


FIG. 5.5. Reconstruction results for  $R = 14.9$  for various sampling strategies. **Left:** TSP-based sampling schemes (PSNR=39.8 dB). **Right:** 2D random drawing and acquisitions along parallel lines (PSNR=38.3 dB).

result to smooth trajectories remains an open question<sup>7</sup>. Recent progresses in that direction were obtained in recent papers such as [30], but they do not provide sufficient guarantees to extend Theorem 4.2. Answering this question is beyond the scope of this paper. We believe that the work [47] based on attraction and repulsion potentials opens an appealing research avenue for solving this issue.

**Appendix 1 - proof of Theorem 2.6.** For a symmetric matrix  $M$ , we denote by  $\lambda_{\max}(M)$  its largest eigenvalue and by  $\|M\|$  the largest eigenvalue modulus. The crucial step to obtain Theorem 2.6 is Proposition 6.1 below. The rest of the proof is the same as the one proposed in [42] and we refer the interested reader to [42, Section 7.3] for further details.

**PROPOSITION 6.1.** *Let  $\Omega = \Omega_1 \cup \Omega_2 \subseteq \{1, \dots, n\}$  be a set constructed as in Theorem 2.6. Define*

$$\tilde{a}_i = \begin{cases} a_i & \text{if } i \in \Omega_1 \\ a_i/\sqrt{p_i} & \text{if } i \in \{1, \dots, n\} \setminus \Omega_1. \end{cases}$$

and

$$(6.1) \quad \tilde{\mathbf{A}} = \begin{pmatrix} \tilde{a}_{\Omega_1(1)} \\ \vdots \\ \tilde{a}_{\Omega_1(m_1)} \\ \frac{1}{\sqrt{m_2}} \tilde{a}_{\Omega_2(1)} \\ \vdots \\ \frac{1}{\sqrt{m_2}} \tilde{a}_{\Omega_2(m_2)} \end{pmatrix} \in \mathbb{C}^{m \times n}.$$

Then for all  $\delta \in [0, \frac{1}{2}]$ :

$$\mathbb{P}\left(\left\|\tilde{\mathbf{A}}^{S*} \tilde{\mathbf{A}}^S - I_s\right\| \geq \delta\right) \leq 2s \exp\left(-\frac{m_2 \delta^2}{CK_2^2 s}\right).$$

<sup>7</sup>To be precise, many crucial properties of the length of the shortest path used to derive asymptotic results are lost. The most important one is subadditivity [46].



where  $\tilde{\mathbf{A}}^S \in \mathbb{C}^{m \times s}$  is the matrix composed of the  $s$  columns of  $\tilde{\mathbf{A}}$  belonging to  $S$ .  $C = 7/3$  is a constant.

The proof of this proposition relies heavily on the matrix Bernstein inequality below [48].

**PROPOSITION 6.2** (Matrix Bernstein inequality). *Let  $\mathbf{Z}_k$  be a finite sequence of independent, random, self-adjoint matrices in  $\mathbb{C}^{d \times d}$ . Assume that each random matrix satisfies*

$$\mathbb{E}(\mathbf{Z}_k) = 0 \quad \text{and} \quad \lambda_{\max}(\mathbf{Z}_k) \leq R \quad \text{a.s.}$$

Denote  $\sigma^2 = \left\| \sum_k \mathbb{E}(\mathbf{Z}_k^2) \right\|$ . Then, for all  $t \geq 0$ :

$$\mathbb{P}\left(\left\| \sum_k \mathbf{Z}_k \right\| \geq t\right) \leq 2d \exp\left(-\frac{t^2/2}{\sigma^2 + Rt/3}\right).$$

We are now ready to prove Proposition 6.1.

*Proof.* For any vector  $v \in \mathbb{C}^n$ , denote by  $v^S \in \mathbb{C}^s$  the vector composed of the entries of  $v$  belonging to  $S \subseteq \{1, \dots, n\}$ . Consider the random sequence  $X_1, \dots, X_{m_2}$  where  $X_i = j \in \{1 \dots n\} \setminus \Omega_1$  with probability  $p_j$ , and denote by  $\Omega_2$  the set  $\{X_1, \dots, X_{m_2}\}$ . Denote by  $\mathbf{M}_1 := \sum_{i \in \Omega_1} a_i^S a_i^{S*}$ . Consider the matrices  $\mathbf{Z}_j := \mathbf{M}_1 + \tilde{a}_j^S \tilde{a}_j^{S*} - I_s$ . According to Eq. (6.1), we get by construction:

$$\tilde{\mathbf{A}}^{S*} \tilde{\mathbf{A}}^S - I_s = \frac{1}{m_2} \sum_{j \in \Omega_2} \mathbf{Z}_j.$$

Since  $I_s = \sum_{i=1}^n a_i^S a_i^{S*}$ , we notice that  $\forall i \in \{1, \dots, m_2\}$  (i)  $\mathbb{E}(\mathbf{Z}_{X_i}) = 0$ , (ii)  $\mathbb{E}(\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}) = I_s - \mathbf{M}_1$ . Moreover, we have (iii)  $0 \preceq I_s - \mathbf{M}_1 \preceq I_s$  and (iv)  $0 \preceq \mathbf{M}_1 \preceq I_s$ .

Using the identity  $(\tilde{a}_j^S \tilde{a}_j^{S*})^2 = \|\tilde{a}_j^S\|^2 \tilde{a}_j^S \tilde{a}_j^{S*}$  and the fact that  $\|\tilde{a}_i^S\| \leq \sqrt{s} \|\tilde{a}_i^S\|_\infty$ , we get  $\mathbb{E}((\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*})^2) \preceq K_2^2 s (I_s - \mathbf{M}_1)$  using (ii). We can then proceed as follows using points (iii) to (iv):

$$\begin{aligned} \mathbb{E}(\mathbf{Z}_{X_i}^2) &= \mathbf{M}_1^2 - 2\mathbf{M}_1 + I_s + \mathbb{E}((\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*})^2) + 2\mathbf{M}_1 \mathbb{E}(\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}) - 2\mathbb{E}(\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}) \\ &\leq \mathbf{M}_1^2 - 2\mathbf{M}_1 + I_s + K_2^2 s (I_s - \mathbf{M}_1) + 2\mathbf{M}_1 (I_s - \mathbf{M}_1) - 2(I_s - \mathbf{M}_1) \\ &= -(I_s - \mathbf{M}_1)^2 + K_2^2 s (I_s - \mathbf{M}_1) \\ &\preceq K_2^2 s I_s. \end{aligned}$$

Then  $\left\| \sum_{i=1}^{m_2} \mathbb{E}(\mathbf{Z}_{X_i}^2) \right\| \leq m_2 K_2^2 s$ .

By noticing that  $\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*} - I_s \preceq \mathbf{Z}_{X_i} \preceq \tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}$ , we obtain  $\|\mathbf{Z}_{X_i}\| \leq K_2^2 s$ . Finally, by applying Bernstein inequality to the sequence of matrices  $\mathbf{Z}_{X_1}, \dots, \mathbf{Z}_{X_{m_2}}$ , we derive for all  $t \geq 0$ :

$$\mathbb{P}\left(\left\| \sum_{j \in \Omega_2} \mathbf{Z}_j \right\| \geq t\right) \leq 2s \exp\left(-\frac{t^2/2}{m_2 K_2^2 s + K_2^2 s t/3}\right).$$

Plugging  $\delta := t/m_2$ , and noticing that  $\delta \leq 1/2 \Rightarrow 2(1 + \delta/3) \leq 2(1 + \delta/3) \leq 7/3$ , the announced result is shown.  $\square$

**Appendix 2 - proof of Proposition 3.1.** Our approach relies on the following perfect recovery condition introduced in [25]:

PROPOSITION 6.3 ([25]). *If  $\mathbf{A}_\Omega \in \mathbb{R}^{m \times n}$  satisfies*

$$\gamma(\mathbf{A}_\Omega) = \min_{\mathbf{Y} \in \mathbb{R}^{m \times n}} \|I_n - \mathbf{Y}^T \mathbf{A}_\Omega\|_\infty < \frac{1}{2s},$$

*all  $s$ -sparse signals  $x \in \mathbb{R}^n$  are recovered exactly by solving the  $\ell_1$  minimization problem (1.1).*

We noted  $\|A\|_\infty$  the maximal modulus of all the entries of  $A$ . This can be seen as an alternative to the *mutual coherence* [16]. We limit our proof to the real case but it could be extended to the complex case using a slightly different proof.

We aim at finding  $\mathbf{Y} \in \mathbb{R}^{m \times n}$ , such that  $\|I_n - \mathbf{Y}^T \mathbf{A}_\Omega\|_\infty < \frac{1}{2s}$ , for a given positive integer  $s$ , where  $\mathbf{A}_\Omega$  is the sensing matrix defined in Proposition 3.1. Following [24], we set  $\Theta_i = \frac{a_i a_i^T}{p_i}$  and use the decomposition  $I_n = \mathbf{A}^T \mathbf{A} = \sum_{i=1}^n p_i \Theta_i$ . We consider a realization of the Markov chain  $X_1, \dots, X_m$ , with  $X_1 \sim p$  and  $X_i \sim \mathbf{P}_{X_{i-1}}$ , for  $i > 1$ . Let us denote  $\mathbf{W}_m = \frac{1}{m} \sum_{l=1}^m \Theta_{X_l}$ . Then  $\mathbf{W}_m$  may be written as  $\mathbf{Y}^T \mathbf{A}_\Omega$ .

LEMMA 6.4.  $\forall 0 < t \leq 1$ ,

$$(6.2) \quad \mathbb{P}(\|I_n - \mathbf{W}_m\|_\infty \geq t) \leq n(n+1)e^{\frac{\epsilon(\mathbf{P})}{5}} \exp\left(-\frac{mt^2\epsilon(\mathbf{P})}{12K^2(\mathbf{A}, p)}\right).$$

Before proving the lemma, let first recall a concentration inequality for finite-state Markov chains [31].

PROPOSITION 6.5. *Let  $(\mathbf{P}, p)$  be an irreducible and reversible Markov chain on a finite set  $G$  of size  $n$  with transition matrix  $\mathbf{P}$  and stationary distribution  $p$ . Let  $f : G \rightarrow \mathbb{R}$  be such that  $\sum_{i=1}^n p_i f_i = 0$ ,  $\|f\|_\infty \leq 1$  and  $0 < \sum_{i=1}^n f_i^2 p_i \leq b^2$ . Then, for any initial distribution  $q$ , any positive integer  $m$  and all  $0 < t \leq 1$ ,*

$$\mathbb{P}\left(\frac{1}{m} \sum_{i=1}^m f(X_i) \geq t\right) \leq e^{\frac{\epsilon(\mathbf{P})}{5}} N_q \exp\left(-\frac{mt^2\epsilon(\mathbf{P})}{4b^2(1+g(5t/b^2))}\right)$$

where  $N_q = (\sum_{i=1}^n (\frac{q_i}{p_i})^2 p_i)^{1/2}$  and  $g$  is given by  $g(x) = \frac{1}{2}(\sqrt{1+x} - (1-x/2))$ .

Now, we can prove Lemma 6.4

*Proof.* By applying Proposition 6.5 to a function  $f$  and then to its opposite  $-f$ , we get:

$$\mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^m f(X_i)\right| \geq t\right) \leq 2e^{\frac{\epsilon(\mathbf{P})}{5}} N_q \exp\left(-\frac{mt^2\epsilon(\mathbf{P})}{4b^2(1+g(5t/b^2))}\right).$$

Then we set  $f(X_i) = (I_n - \Theta_{X_i})^{(a,b)}/K(\mathbf{A}, p)$  as real-valued function. Recall that  $p$  satisfies  $\sum_{i=1}^n p_i f(X_i) = 0$ . Since  $\|f\|_\infty \leq 1$ ,  $b = 1$  and  $t \leq 1$ , we deduce  $1+g(5t) < 3$ . Moreover, since the initial distribution is  $p$ ,  $q_i = p_i, \forall i$  and thus  $N_q = 1$ . Finally, resorting to a union bound enables us to extend our result for the  $(a, b)$ th entry to the whole infinite norm of the  $n \times n$  matrix  $I_n - \mathbf{W}_m$  (6.2).  $\square$

Finally, set  $s \in \mathbb{N}^*$  and  $\eta \in (0, 1)$ . If  $m$  satisfies Ineq. (3.3), then

$$\mathbb{P}\left(\|I_n - \mathbf{W}_m\|_\infty \geq \frac{1}{2s}\right) < \eta.$$

In other words, with probability at least  $1 - \eta$ , every  $s$ -sparse signal can be recovered by  $\ell_1$  minimization (1.1).

REMARK 15. *It is straightforward to derive a similar result to Theorem 2.6 and thus to justify that a partial deterministic sampling reduces the total number of measurements required for perfect recovery.*

**Appendix 3 - proof of Remark 8.** In this part, we prove that for a random walk with uniform stationary distribution  $p$ ,  $\epsilon(\mathbf{P}) = O(n^{-\frac{1}{d}})$ . We use geometric bounds known as *Cheeger's inequality* in [15] and *Conductance Bounds* in [23, 8]. Let us recall a useful result concerning finite state space irreducible reversible transition matrices  $\mathbf{P}$ .

The *capacity* of a set  $B \subset \{1, \dots, n\}$  is defined as  $p(B) := \sum_{i \in B} p(i)$  and the *ergodic flow* out of  $B$  is defined by  $F(B) := \sum_{i \in B, j \in B^c} p(i)\mathbf{P}_{i,j}$ . The *conductance* of the pair  $(\mathbf{P}, p)$  is:

$$\varphi(\mathbf{P}) := \inf_B \left( \frac{F(B)}{p(B)}; 0 < |B| < n, p(B) \leq \frac{1}{2} \right).$$

Then the following result holds (see [23] and [8, Theorem 4.3]):

PROPOSITION 6.6.

$$\frac{\varphi(\mathbf{P})^2}{2} \leq \epsilon(\mathbf{P}) \leq 2\varphi(\mathbf{P}).$$

Now, assume that  $n^{1/d} \in \mathbb{N}$  is even and construct a finite graph with  $n$  nodes representing a Euclidean grid of the unit hypercube of dimension  $d$ . Assume that the vertices of the graph at one grid point are the  $2d$  nearest nodes, with periodic boundary conditions (the graph can be seen as a  $d$ -dimensional torus). Assume that the transition probability is uniform over the neighbors, thus the stationary distribution is also the uniform one. This graph is depicted in Fig. 6.1[Left], with  $d = 2$ .

Let  $B$  be the halved graph defined by the hyperplane parallel to an axis of the grid and including its center, so that  $p(B) = 1/2$ . An illustration is given in 2D in Fig. 6.1[Right]. Since we assumed periodic boundary conditions, the number of nodes belonging to  $B$  and having a neighbor in  $B^c$  is  $2n^{(d-1)/d}$ . Each of these nodes have  $2d$  neighbors, but only one belonging to  $B^c$ . Since the stationary distribution is equal to  $1/n$  on each node, the ergodic flow is  $2n^{(d-1)/d}(\frac{1}{n} \frac{1}{2d})$ . It follows that  $\epsilon(\mathbf{P}) \leq \frac{4}{d}n^{-\frac{1}{d}}$ .

**Appendix 4 - proof of Theorem 4.2.** Let  $h \in \mathbb{N}$ . The set  $\mathcal{H} = [0, 1]^d$  will be partitioned in  $h^d$  congruent hypercubes  $(\omega_i)_{i \in I}$  of edge length  $1/h$ . The following proposition is central to obtain the proof:

PROPOSITION 6.7. *Almost surely, for all  $\omega_i$  in  $\{\omega_i\}_{1 \leq i \leq h^d}$ :*

$$(6.3) \quad \lim_{N \rightarrow \infty} \tilde{P}_N(\omega_i) = \tilde{p}(\omega_i)$$

$$(6.4) \quad = \frac{\int_{\omega_i} p^{(d-1)/d}(x) dx}{\int_{\mathcal{H}} p^{(d-1)/d}(x) dx} \quad p^{\otimes \mathbb{N}}\text{-a.s.}$$

The strategy consists in proving that  $T_{|\omega_i}(X_N, \mathcal{H})$  tends asymptotically to  $T(X_N, \omega_i)$ . The estimation of each term can then be obtained by applying the asymptotic result of Beardwood, Halton and Hammersley [3, 46]:

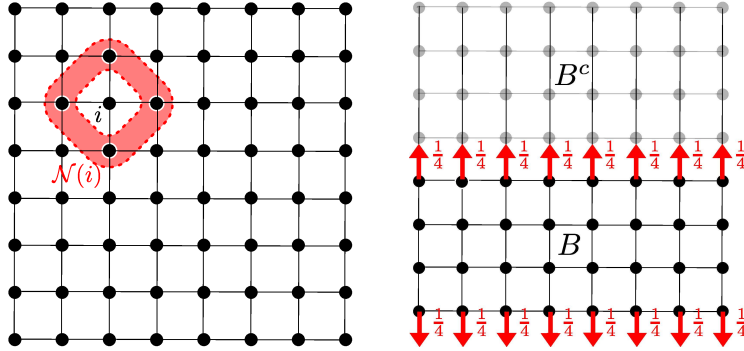


FIG. 6.1. Illustration of the proof of Remark 8 in dimension 2. **Left:** regular grid with  $n = \sqrt{n} \times \sqrt{n}$  nodes. **Right:** Graph partitioning in  $B$  and  $B^c$  with  $p(B) = 1/2$ .

**THEOREM 6.8.** *If  $R$  is a Lebesgue-measurable set in  $\mathbb{R}^d$  such that the boundary  $\partial R$  has zero measure, and  $\{y_i\}_{i \in \mathbb{N}^*}$ , with  $Y_N = \{y_i\}_{i \leq N}$  is a sequence of i.i.d. points from a density  $p$  supported on  $R$ , then, almost surely,*

$$(6.5) \quad \lim_{N \rightarrow \infty} \frac{T(Y_N, R)}{N^{(d-1)/d}} = \beta(d) \int_R p^{(d-1)/d}(x) dx,$$

where  $\beta(d)$  depends on the dimension  $d$  only.

To show Prop.6.7, we need to introduce the boundary TSP. For a set of points  $F$  and an area  $R$ , we denote by  $T_B(F, R)$  its length on the set  $F \cap R$ . The boundary TSP is defined as the shortest Hamiltonian tour on  $F \cap R$  for the metric obtained from the Euclidean metric by the quotient of the boundary of  $R$ , that is  $d(a, b) = 0$  if  $a, b \in \partial R$ . Informally, it matches the original TSP while being allowed to travel along the boundary for free. We refer to [18] for a complete description of this concept.

We shall use a set of classical results on TSP and boundary TSP, that may be found in the survey books [18] and [53].

**LEMMA 6.9.** *Let  $F$  denote a set of  $n$  points in  $\mathcal{H}$ .*

1. *The boundary TSP is superadditive, that is, if  $R_1$  and  $R_2$  have disjoint interiors.*

$$(6.6) \quad T_B(F, R_1 \cup R_2) \geq T_B(F, R_1) + T_B(F, R_2).$$

2. *The boundary TSP is a lower bound on the TSP, both globally and on subsets. If  $R_2 \subset R_1$ :*

$$(6.7) \quad T(F, R) \geq T_B(F, R)$$

$$(6.8) \quad T_{|R_2}(F, R_1) \geq T_B(F, R_2)$$

3. *The boundary TSP approximates well the TSP [53, Lemma 3.7]:*

$$(6.9) \quad |T(F, \mathcal{H}) - T_B(F, \mathcal{H})| = O(n^{(d-2)/(d-1)}).$$

4. The TSP in  $\mathcal{H}$  is well-approximated by the sum of TSPs in a grid of  $h^d$  congruent hypercubes [18, Eq. (33)].

$$(6.10) \quad |T(F, \mathcal{H}) - \sum_{i=1}^{h^d} T(F, \omega_i)| = O(n^{(d-2)/(d-1)}).$$

We now have all the ingredients to prove the main results.

*Proof.* [Proof of Proposition 6.7]

$$\begin{aligned} \sum_{i \in I} T_B(X_N, \omega_i) &\stackrel{(6.6)}{\leq} T_B(X_N, \mathcal{H}) \\ &\stackrel{(6.7)}{\leq} T(X_N, \mathcal{H}) = \sum_{i=1}^{h^d} T|_{\omega_i}(X_N, \mathcal{H}) \\ &\stackrel{(6.10)}{\leq} \sum_{i=1}^{h^d} T(X_N, \omega_i) + O(N^{(d-1)/(d-2)}) \end{aligned}$$

Let  $N_i$  be the number of points of  $X_N$  in  $\omega_i$ .

Since  $N_i \leq N$ , we may use the bound (6.9) to get:

$$(6.11) \quad \lim_{N \rightarrow \infty} \frac{T(X_N, \omega_i)}{N^{(d-1)/d}} = \lim_{N \rightarrow \infty} \frac{T_B(X_N, \omega_i)}{N^{(d-1)/d}}.$$

Using the fact that there are only finitely many  $\omega_i$ , the following equalities hold almost surely:

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^{h^d} T_B(X_N, \omega_i)}{N^{(d-1)/d}} &= \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^{h^d} T(X_N, \omega_i)}{N^{(d-1)/d}} \\ &\stackrel{(6.10)}{=} \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^{h^d} T|_{\omega_i}(X_N, \mathcal{H})}{N^{(d-1)/d}}. \end{aligned}$$

Since the boundary TSP is a lower bound (cf. Eqs. (6.8)-(6.7)) to both local and global TSPs, the above equality ensures that:

$$(6.12) \quad \begin{aligned} \lim_{N \rightarrow \infty} \frac{T_B(X_N, \omega_i)}{N^{(d-1)/d}} &= \lim_{N \rightarrow \infty} \frac{T(X_N, \omega_i)}{N^{(d-1)/d}} \\ &= \lim_{N \rightarrow \infty} \frac{T|_{\omega_i}(X_N, \mathcal{H})}{N^{(d-1)/d}} \quad p^{\otimes \mathbb{N}}\text{-a.s.}, \forall i. \end{aligned}$$

Finally, by the law of large numbers, almost surely  $N_i/N \rightarrow p(\omega_i) = \int_{\omega_i} p(x) dx$ . The law of any point  $x_j$  conditioned on being in  $\omega_i$  has density  $p/p(\omega_i)$ . By applying Theorem 6.8 to the hypercubes  $\omega_i$  and  $\mathcal{H}$  we thus get:

$$\lim_{N \rightarrow +\infty} \frac{T(X_N, \omega_i)}{N^{(d-1)/d}} = \beta(d) \int_{\omega_i} p(x)^{(d-1)/d} dx \quad p^{\otimes \mathbb{N}}\text{-a.s.}, \forall i.$$

and

$$\lim_{N \rightarrow +\infty} \frac{T(X_N, \mathcal{H})}{N^{(d-1)/d}} = \beta(d) \int_{\mathcal{H}} p(x)^{(d-1)/d} dx \quad p^{\otimes \mathbb{N}}\text{-a.s.}, \forall i.$$

Combining this result with Eqs. (6.12) and (4.1) yields Proposition 6.7.  $\square$

*Proof.* [Proof of Theorem 4.2] Let  $\varepsilon > 0$  and  $h$  be an integer such that  $\sqrt{dh}^{-d} < \varepsilon$ . Then any two points in  $\omega_i$  are at distance less than  $\varepsilon$ .

Using Theorem 6.7 and the fact that there is a finite number of  $\omega_i$ , almost surely, we get:

$\lim_{N \rightarrow +\infty} \sum_{i=1}^{h^d} \left| \tilde{P}_N(\omega_i) - \tilde{p}(\omega_i) \right| = 0$ . Hence, for any  $N$  large enough, there is a coupling  $K$  of  $\tilde{P}_N$  and  $\tilde{p}$  such that both corresponding random variables are in the same  $\omega_i$  with probability  $1 - \varepsilon$ . Let  $A \subseteq \mathcal{H}$  be a Borelian. The coupling satisfies  $\tilde{P}_N(A) = K(A \otimes \mathcal{H})$  and  $\tilde{p}(A) = K(\mathcal{H} \otimes A)$ . Define the  $\varepsilon$ -neighborhood by  $A^\varepsilon = \{X \in \mathcal{H} \mid \exists Y \in A, \|X - Y\| < \varepsilon\}$ . Then, we have:  $\tilde{P}_N(A) = K(A \otimes \mathcal{H}) = K(\{A \otimes \mathcal{H}\} \cap \{|X - Y| < \varepsilon\}) + K(\{A \otimes \mathcal{H}\} \cap \{|X - Y| \geq \varepsilon\})$ . It follows that:

$$\begin{aligned} \tilde{P}_N(A) &\leq K(A \otimes A^\varepsilon) + K(|X - Y| \geq \varepsilon) \\ &\leq K(\mathcal{H} \otimes A^\varepsilon) + \varepsilon = \tilde{p}(A^\varepsilon) + \varepsilon. \end{aligned}$$

This exactly matches the definition of convergence in the Prokhorov metric, which implies convergence in distribution.  $\square$

**Acknowledgments.** The authors wish to thank Yves Wiaux, Fabrice Gamboa, Jérémie Bigot, Laurent Miclo, Alexandre Vignaud and Claire Boyer for fruitful discussions and feedback. This research was supported by the Labex CIMI through a 3 months invitation of Philippe Ciuciu. This work was partially supported by ANR SPH-IM-3D (ANR-12-BSV5-0008), by the FMJH Program Gaspard Monge in optimization and operation research (MAORI project), and by the support to this program from EDF.

#### REFERENCES

- [1] B. Adcock, A. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing. *arXiv preprint arXiv:1302.0561*, 2013.
- [2] D. Applegate, R. Bixby, V. Chvatal, and W. Cook. Concorde TSP solver. *URL: <http://www.tsp.gatech.edu/concorde>*, 2006.
- [3] J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 55, pages 299–327. Cambridge Univ Press, 1959.
- [4] J. Bigot, C. Boyer, and P. Weiss. An analysis of block sampling strategies in compressed sensing. *arXiv preprint arXiv:1305.4446*, 2013.
- [5] P. Billingsley. *Convergence of probability measures*, volume 493. Wiley, 2009.
- [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 Science*, (in press) 2014.
- [8] P. Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer, 1999.
- [9] E. Candès and Y. Plan. A probabilistic and ripples theory of compressed sensing. *IEEE Trans. Inf. Theory*, 57(11):7235–7254, 2011.
- [10] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.
- [11] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss. Travelling salesman-based variable density sampling. In *Proc. of 10th SampTA conference*, pages 509–512, Bremen, Germany, July 2013.

- [12] 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.
- [13] N. Chauffert, P. Ciuciu, P. Weiss, and F. Gamboa. From variable density sampling to continuous sampling using Markov chains. In *Proc. of 10th SampTA conference*, pages 200–203, Bremen, Germany, July 2013.
- [14] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.
- [15] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.
- [16] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, Apr. 2006.
- [17] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. *Appl. Numer. Harmon. Anal. Birkhäuser, Boston*, 2013.
- [18] A. M. Frieze and J. E. Yukich. Probabilistic analysis of the TSP. In G. Gutin and A. P. Punnen, editors, *The traveling salesman problem and its variations*, volume 12 of *Combinatorial optimization*, pages 257–308. Springer, 2002.
- [19] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory*, 57(3):1548–1566, 2011.
- [20] J. P. Haldar, D. Hernando, and Z.-P. Liang. Compressed-sensing MRI with random encoding. *IEEE Trans. Med. Imag.*, 30(4):893–903, 2011.
- [21] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, Apr. 1970.
- [22] R. Horn and C. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991.
- [23] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM journal on computing*, 18(6):1149–1178, 1989.
- [24] A. Juditsky, F.K. Karzan, and A. Nemirovski. On low rank matrix approximations with applications to synthesis problem in compressed sensing. *SIAM J. on Matrix Analysis and Applications*, 32(3):1019–1029, 2011.
- [25] A. Juditsky and A. Nemirovski. On verifiable sufficient conditions for sparse signal recovery via  $\ell_1$  minimization. *Mathematical Programming Ser. B*, 127:89–122, 2011.
- [26] V. Kargin. A large deviation inequality for vector functions on finite reversible Markov chains. *The Annals of Applied Probability*, 17(4):1202–1221, Aug. 2007.
- [27] D. H. Kim, E. Adalsteinsson, and D. M. Spielman. Simple analytic variable density spiral design. *Magnetic resonance in medicine*, 50(1):214–219, 2003.
- [28] F. Knoll, C. Clason, C. Diwoky, and R. Stollberger. Adapted random sampling patterns for accelerated MRI. *Magma*, 24(1):43–50, 2011.
- [29] F. Krahermer and R. Ward. Beyond incoherence: stable and robust sampling strategies for compressive imaging. preprint, 2012.
- [30] J. Le Ny, E. Feron, and E. Frazzoli. On the dubins traveling salesman problem. *Automatic Control, IEEE Transactions on*, 57(1):265–270, 2012.
- [31] P. Lezaud. Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, 8(3):849–867, 1998.
- [32] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195, Dec. 2007.
- [33] M. Lustig, S. J. Kim, and J. M. Pauly. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. *IEEE Trans. Med. Imag.*, 27(6):866–873, 2008.
- [34] M. M. Marim, M. Atlan, E. Angelini, and J.-C. Olivo-Marin. Compressed sensing with off-axis frequency-shifting holography. *Optics letters*, 35(6):871–873, 2010.
- [35] J. Park, Q. Zhang, V. Jellus, O. Simonetti, and D. Li. Artifact and noise suppression in GRAPPA imaging using improved k-space coil calibration and variable density sampling. *Magnetic resonance in medicine*, 53(1):186–193, 2005.
- [36] D. Paulin. Concentration inequalities for markov chains by marton couplings. *arXiv preprint arXiv:1212.2015*, 2012.
- [37] A. C. Polak, M. F. Duarte, and D. L. Goeckel. Grouped incoherent measurements for compressive sensing. In *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, pages 732–735. IEEE, 2012.
- [38] A. C. Polak, M. F. Duarte, and D. L. Goeckel. Performance bounds for grouped incoherent measurements in compressive sensing. *arXiv preprint arXiv:1205.2118*, 2012.
- [39] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: sensitivity encoding



- for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, Jul. 1999.
- [40] G. Puy, J. P. Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux. Spread spectrum magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 31(3):586–598, 2012.
- [41] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *IEEE Signal Processing Letters*, 18(10):595–598, 2011.
- [42] H. Rauhut. Compressive Sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.
- [43] Y. Rivenson, A. Stern, and B. Javidi. Compressive fresnel holography. *Journal of Display Technology*, 6(10):506–509, 2010.
- [44] E. Y. Sidky, C.-M. Kao, and X. 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.
- [45] 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.
- [46] J. M. Steele. Subadditive euclidean functionals and nonlinear growth in geometric probability. *The Annals of Probability*, 9(3):365–376, 1981.
- [47] T. Teuber, G. Steidl, P. Gwosdek, C. Schmaltz, and J. Weickert. Dithering by differences of convex functions. *SIAM Journal on Imaging Science*, 4(1):79–108, 2011.
- [48] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, pages 1–32, Dec. 2012.
- [49] C. M. Tsai and D. G. Nishimura. Reduced aliasing artifacts using variable-density k-space sampling trajectories. *Magnetic resonance in medicine*, 43(3):452–458, 2000.
- [50] H. Wang, X. Wang, Y. Zhou, Y. Chang, and Y. Wang. Smoothed random-like trajectory for compressed sensing MRI. In *Proc. of the 34th annual IEEE EMBC*, pages 404–407, 2012.
- [51] Y. Wiaux, G. Puy, Y. Boursier, and P. Vandergheynst. Spread spectrum for imaging techniques in radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 400(2):1029–1038, 2009.
- [52] R. M. Willett. Errata: Sampling trajectories for sparse image recovery. Note, Duke University, 2011.
- [53] J. E. Yukich. *Probability theory of classical Euclidean optimization problems*. Springer, 1998.