Generalized sampling, infinite-dimensional compressed sensing, and semi-random sampling for asymptotically incoherent dictionaries

Ben Adcock*, Anders C. Hansen, Evelyn Herrholz and Gerd Teschke

***Abstract*—Recent developments in sampling in abstract Hilbert spaces have led to a new theory of compressed sensing for infinite-dimensional signals. In this paper, we continue with this theme by introducing a new type of subsampling for infinite-dimensional sparse recovery problems, known as semi-random sampling. As we demonstrate, this allows for subsampling in problems which previously had not been amenable to more conventional compressed sensing tools. More specifically, semi-random sampling allows one to overcome the so-called incoherence barrier, which limits the potential for subsampling via standard random sampling techniques. The key to this improvement is a property known as asymptotic incoherence. In the final part of this paper we provide specific estimates for this property in several important cases, and illustrate via numerical example the benefit of semi-random sampling.**

***Index Terms*—Compressed sensing, generalized sampling, incoherence, asymptotic incoherence, semi-random sampling**

I. Introduction

Compressed sensing (CS) has been one of the major developments in applied mathematics in the last decade [10], [17], [16], [25], [27]. The key to its success is that it allows one to exploit a particular signal structure, namely sparsity, to circumvent the traditional barriers of sampling theory (such as the Nyquist rate) [13], [14], [15]. In addition, efficient algorithms, based on l^1-minimization, allow the practical realization of such a theory.

However, CS deals only with the reconstruction of vectors in finite-dimensional vector spaces. In other words, it is a finite-dimensional theory. On the other hand, many problems one encounters in applications are inherently infinite-dimensional: the signal f lives in some function space (typically a Hilbert or Banach space), and its samples form some countable collection (typically in $l^2(\mathbb{N})$). Take, for example, the Magnetic Resonance Imaging (MRI) problem. Here f (the image) is modelled as an element of the Hilbert space $\mathcal{H} = L^2(-b,b)^d$, with the samples $\{\zeta_j(f)\}_{j\in\mathbb{N}}$ being pointwise evaluations of the Fourier transform of f. Suppose now that f is known to be sparse in some basis $\{\varphi_j\}_{j\in\mathbb{N}}$ for $\mathcal{H}$. It is by no means obvious how to apply current CS techniques to this problem (see §I-D for details). In particular, to use current tools, one must first somehow 'discretize' the problem, so as to reduce it to one involving finite vectors and matrices. How best to carry out such a discretization is, in general, neither clear nor straightforward.

The MRI problem also illustrates a point that is key to what follows. In many important applications the *sampling scheme* $\{\zeta_j\}_{j\in\mathbb{N}}$ is fixed. Although much work in finite-dimensional CS is devoted to designing efficient measurement systems, there are numerous applications for which one does not have the luxury to choose how to sample. This is the case in MRI, where the physical device takes pointwise measurements of the Fourier transform, and this cannot easily be altered. Thus, throughout this paper, we deal exclusively with the problem of how to reconstruct given *fixed* samples.

A. Infinite-dimensional CS

The fact that the standard CS framework can only be applied to finite-dimensional problems has been acknowledged, but an extension to infinite-dimensional problems (of the above type) was only recently introduced in [3], based on ideas from [30]. Therein it was proposed to tackle the infinite-dimensional problem directly. Rather than discretizing first, and applying existing tools (an approach that may well not be successful [3]), the idea is to formulate the sparse recovery problem in infinite-dimensions, and *then* discretize. Provided such discretization is carried out appropriately, one obtains new techniques, and indeed a whole new theory, for CS in infinite dimensions (see §II for details).

With this in mind, the purpose of this paper is to continue the development of infinite-dimensional CS. Specifically, we introduce a new type of sampling, based on so-called *asymptotic incoherence*, which improves on the techniques of [3].

Remark 1 Recently there have been a number of other attempts at generalizations of compressed sensing to infinite-dimensional (or analog) signal models. See [19], [23], [39], [40]. Other infinite-dimensional generalizations (for specific problems) have also been considered in [31], [41], [42]. We remark that these are quite different in character (i.e. they treat different types of signal models) to the generalization considered in this paper and [3]. In a sense, the generalization we propose is a natural extension of the current theory: rather than orthonormal vectors in finite-dimensional vector spaces, we consider orthonormal bases in separable Hilbert spaces.

It is also worth noting that there are number of alternatives to CS for infinite-dimensional signal models. Amongst the

B. Adcock* (corresponding author) is with the Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. A. C. Hansen is with the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom. E. Herrholz and G. Teschke are with the Hochschule Neubrandenburg, University of AppliedSciences, Brodaer Straße 2, 17033 Neubrandenburg, Germany

most well-known is the technique of finite rates of innovation, due to Vetterli et al. See [9], [18], [47] for further details.

B. Generalized sampling (GS): sampling and reconstruction in infinite-dimensions

The framework developed in [3] is part of a new approach to problems in sampling, known as *generalized sampling* (GS). First introduced in [2], [4], [6], [7], GS works directly with an infinite-dimensional signal model. That is, we assume that $f \in \mathcal{H}$ is an element of a separable Hilbert space $\mathcal{H}$, its samples

$$\zeta_j(f), \quad j \in \mathbb{N}, \tag{1}$$

are actions of linear functionals $\zeta_j : \mathcal{H} \to \mathbb{C}$ on $\mathcal{H}$, and the reconstruction system is a basis $\{\varphi_j\}_{j\in\mathbb{N}}$ for $\mathcal{H}$. As evidenced by the MRI problem, such a model is often far more representative of reality (in this case, the reconstruction basis could, for example, consist of wavelets).

Of course, although f has infinitely many samples (1), in practice we will only have access to a finite number N. Thus, we may ask the following question: how good a reconstruction can we obtain from this finite collection? What generalized sampling provides is a framework that allows one to model signals as infinite-dimensional objects, whilst giving a practical numerical algorithm for reconstruction from the given collection of finitely many measurements.

GS, as introduced in [2], [7], does not require sparsity to obtain a reconstruction. A key property is that it give estimates for the reconstruction *error* committed through reconstructing from only N measurements (naturally, an arbitrary, nonsparse f cannot typically be recovered *perfectly* from only finitely many samples). In [3], the GS framework was combined with random subsampling and convex optimization techniques to exploit sparsity, leading to so-called *generalized sampling with compressed sensing* (GS–CS).

C. Incoherence, asymptotic incoherence and semi-random sampling

A key theorem proved in [3] states that if f is k-sparse in the orthonormal basis $\{\varphi_j\}_{j\in\mathbb{N}}$, i.e.

$$f = \sum_{j=1}^{\infty} \alpha_j \varphi_j, \quad |\{j : \alpha_j \neq 0\}| = k,$$

then, under a number of assumptions (see §II) it is possible to recover f perfectly by subsampling. The amount of subsampling depends critically on the *coherence* μ of the sampling system $\{\zeta_j\}_{j\in\mathbb{N}}$ and the reconstruction basis $\{\varphi_j\}_{j\in\mathbb{N}}$:

$$\mu = \sup_{i\in\mathbb{N}} \sup_{j\in\mathbb{N}} |\zeta_i(\varphi_j)|.$$

In particular, when μ is small (see §II for specific estimates), then one requires, up to a log-factor, only $\mathcal{O}(k)$ samples to recover f exactly (with high probability).

The coherence μ is a fundamental quantity related to the sampling and reconstruction specifications. Recall that the system $\{\zeta_j\}_{j\in\mathbb{N}}$ is fixed. The reconstruction basis $\{\varphi_j\}_{j\in\mathbb{N}}$ is also typically fixed, since this is the basis we know f to be sparse in. Hence, one cannot necessarily alter either the sampling or reconstruction specification, and therefore the coherence μ is, for all intents and purposes, also *fixed*. As a result, there may well be situations where μ is large, meaning that subsampling is not always possible with GS–CS. In fact, this arises in very straightorward cases, as we shall see later.

Remark 2 This is, of course, not a problem exclusive to infinite-dimensional CS. It is well known in finite dimensions that incoherence is essentially necessary to achieve subsampling [11] using CS techniques.

This *incoherence barrier* places limitations on the amount of subsampling that can be achieved with random sampling techniques. However, many pairs $\{\{\zeta_j\}_{j\in\mathbb{N}}, \{\varphi_j\}_{j\in\mathbb{N}}\}$, whilst not incoherent, actually possess the property of so-called *asymptotic incoherence*. More specifically, if

$$\mu_n = \sup_{i\in\mathbb{N}} \sup_{j>n} |\zeta_i(\varphi_j)|,$$

then asymptotic incoherence is the property that $\mu_n \to 0$ as $n \to \infty$. In other words, for large n, the pair $\{\{\zeta_j\}_{j\in\mathbb{N}}, \{\varphi_j\}_{j>n}\}$ is incoherent.

With this in mind, the purpose of this paper is to introduce a new approach, known as *semi-random subsampling*, that exploits asymptotic incoherence to achieve subsampling, and therefore circumvent the incoherence barrier. This is a fundamentally new approach to sampling in infinite dimensions, which builds directly on the GS–CS framework. As we demonstrate, it proves to be remarkably effective in numerical examples. Problems with $\mu = \mathcal{O}(1)$ that would previously not be amenable to subsampling can now be efficiently tackled by this new approach. In addition, semi-random sampling provides an answer to a question posed by K. Gröchenig on how to subsample from Fourier measurements and achieve reconstructions in (piecewise) polynomial bases [33].

D. Example

The main example considered in this paper is that of the MRI problem. In other words, we wish to recover $f \in L^2(-b, b)$ from pointwise samples of its Fourier transform

$$\mathcal{F}f(t) = \int_{(-b,b)} f(x)\mathrm{e}^{-2\pi \mathrm{i} t x}\,\mathrm{d}x, \ t \in \mathbb{R}.$$

Specifically, we shall be interested in reconstructions from the equispaced samples

$$\zeta_j(f) = \mathcal{F}(\epsilon\rho(j)), \quad j \in \mathbb{N},$$

where $\epsilon \leq \frac{1}{2b}$ and $\rho : \mathbb{N} \to \mathbb{Z}$ is the standard re-indexing function (i.e. $\rho(1) = 0$, $\rho(2) = 1$, $\rho(3) = -1$,...). Although the famous Shannon Sampling Theorem [35], [44] guarantees that f can be recovered exactly from $\{\zeta_j(f)\}_{j\in\mathbb{N}}$, the reconstruction given by Shannon (i.e. the Fourier series of f) is often useless in applications, due to the slow decay of the coefficients $\zeta_j(f)$ as $j \to \infty$ [25], [44]. Thus, the purpose of this paper is to introduce different ways to reconstruct f, in particular, those which take advantage of sparsity.

Remark 3 The MRI problem was one of the original motivators for CS [13], and work by Lustig et al has led to

the successful use of finite-dimensional CS in this area [37], [38]. However, there are some potentially significant issues in tackling the fundamentally infinite-dimensional MRI problem with finite-dimensional tools. We shall not discuss this further, and refer to [3] for details.

E. Key results and outline

The key result of this paper is that one can use asymptotic incoherence to circumvent the incoherence barrier and achieve subsampling in situations where it is otherwise not possible. This aside we shall also give estimates for the behaviour of μ_n as $n \to \infty$. Specifically, for the problem of §I-D we show that $\mu_n = \mathcal{O}(n^{-\frac{1}{2}})$ whenever $\{\varphi_j\}_{j\in\mathbb{N}}$ consists of wavelets and $\mu_n = \mathcal{O}(n^{-\frac{1}{3}})$ for polynomials.

The outline of the remainder of this paper is as follows. In §II we recap the work of [2], [3], [4], [7] on generalized sampling and infinite-dimensional compressed sensing. The new approach of semi-random sampling is introduced in §III, and in §IV we provide estimates for asymptotic incoherence and give numerical examples illustrating the effectiveness of this approach.

II. GS and GS–CS

Let $\{\psi_j\}_{j\in\mathbb{N}}$ and $\{\varphi_j\}_{j\in\mathbb{N}}$ be orthonormal bases (the sampling and reconstruction bases respectively) for a separable Hilbert space $\mathcal{H}$ with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. Given $f \in \mathcal{H}$, we assume that we have access to the samples

$$\zeta_j(f) := \langle f, \psi_j\rangle, \quad j \in \mathbb{N}. \tag{2}$$

The task is to recover f in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$ (at the moment, we do not assume any sparsity in this basis). Since

$$f = \sum_{j=1}^{\infty} \alpha_j \varphi_j, \quad \alpha_j = \langle f, \varphi_j\rangle,$$

we wish to compute approximations to the coefficients α_j using the the given samples (2). Naturally, we only have finitely many samples in practice, hence we typically can only compute approximations

$$\tilde{\alpha}_j \approx \alpha_j, \quad j = 1, \ldots, M,$$

to the first M coefficients, where $M \in \mathbb{N}$ is finite.

There are two key issues in the computation of the values $\{\tilde{\alpha}_j\}_{j=1}^M$ that are vital to what follows:

(i) The value $\tilde{\alpha}_j$ should be a good *approximation* to α_j, $j = 1, \ldots, M$. The whole premise for reconstructing f in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$ is that we know f to have a 'nice' representation in this basis; sparsity or rapidly decaying coefficients, for example. Thus, it is vital that we approximate the first M coefficients to good accuracy.

(ii) The computation of $\{\tilde{\alpha}_j\}_{j=1}^M$ should be *numerically stable*. In practice, there is always error in computations, be it from round-off, noise or otherwise. Thus it is vital to have a stable reconstruction to avoid catastrophic propagation of errors.

A. Generalized sampling

Suppose now that we have access to the first N samples $\{\zeta_j(f)\}_{j=1}^N$ of f. A common way to reconstruct f would be to use a consistent reconstruction [20], [21], [22], [26], [45], [46]. In this technique, one recovers $M = N$ approximate coefficients $\tilde{\alpha}_j$ by enforcing the reconstruction be *consistent* with the first N measurements $\{\zeta_j(f)\}_{j=1}^N$. However, whilst this technique is useful in some circumstances, there is no guarantee in general that either of the key properties listed above will hold. In fact, it is easy to construct examples where a consistent reconstruction is exponentially unstable and divergent [2], [4], [6] (see also [20], [24], [32]).

In [2], [7] a new approach, generalized sampling (GS), was introduced as an alternative to consistent reconstructions. The key idea is very simple: one should allow the parameters N (the number of samples) and M (the number of coefficients to be reconstructed) to vary. When done correctly, via GS, one obtains both the critical properties listed above.

Let $\{e_j\}_{j\in\mathbb{N}}$ be the canonical basis for $l^2(\mathbb{N})$ and write $P_N : l^2(\mathbb{N}) \to \text{span}\{e_1, \ldots, e_n\}$ for the orthogonal projection. Define the *change-of-basis* operator $U : l^2(\mathbb{N}) \to l^2(\mathbb{N})$ by

$$U = \begin{pmatrix} u_{11} & u_{12} & u_{13} & \ldots \\ u_{21} & u_{22} & u_{23} & \ldots \\ u_{31} & u_{32} & u_{33} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{3}$$

where $u_{i,j} = \langle\varphi_j, \psi_i\rangle$. Note that infinite vector of coefficients $\alpha = \{\alpha_1, \alpha_2, \ldots\} \in l^2(\mathbb{N})$ satisfies

$$U\alpha = \zeta(f),$$

where $\zeta(f) = \{\zeta_1(f), \zeta_2(f), \ldots\} \in l^2(\mathbb{N})$. Naturally, had we been able to invert U, and had we access to all of $\zeta(f)$, we could have obtained α exactly. However, in general we do not, and hence we must discretize these equations. This is achieved by replacing U with its *uneven section*

$$P_N U P_M \in \mathbb{C}^{N\times M},$$

and then solving

$$P_M U^* P_N U P_M \tilde{\alpha} = P_M U^* P_N \zeta(f). \tag{4}$$

where $\tilde{\alpha} = \{\tilde{\alpha}_1, \ldots, \tilde{\alpha}_M\} \in \mathbb{C}^M$ is the vector of approximate coefficients. The GS reconstruction of f is now given by

$$f_{N,M} = \sum_{j=1}^{M} \tilde{\alpha}_j \varphi_j.$$

The main theorem proved in [2], [6] is as follows:

Theorem 4. *Suppose that $M \in \mathbb{N}$ is given. Then there exists an $N_0 \in \mathbb{N}$ such that, for every $N \geq N_0$, there is a unique solution $\tilde{\alpha}$ to (4). Furthermore, we have the sharp bound*

$$\|\alpha - \tilde{\alpha}\|_{l^2} \leq \frac{1}{\sqrt{1 - C_{N,M}}} \|P_M^{\perp}\alpha\|_{l^2}, \tag{5}$$

where $P_M^{\perp} = I - P_M$ and

$$C_{N,M} = \|P_M - P_M U^* P_N U P_M\|_{l^2\to l^2}. \tag{6}$$

Specifically, N_0 is the least N such that $C_{N,M} < 1$.

This theorem states the following: given M, we require $N \geq N_0$ samples to reconstruct the first M coefficients in a manner that is guaranteed to be both stable and accurate. Note that (5) implies accuracy of $\tilde{\alpha} \approx P_M \alpha$ (with $P_M \alpha$ being, of course, the best possible reconstruction of the first M coefficients $\alpha_1, \ldots, \alpha_M$). Regarding stability, one can show that the condition number of the uneven section $P_N U P_M$ behaves like $\frac{1}{\sqrt{1-C_{N,M}}}$. Thus, the same quantity $C_{N,M}$ which guarantees accuracy allow ensures numerical stability. We note that $C_{N,M} \to 0$ as $N \to \infty$ for fixed M. Thus, both the accuracy of $\tilde{\alpha}$ and the condition number of its computation can be made arbitrarily close to unity.

Remark 5 The consistent reconstruction technique [21], [22], [26], [45] actually corresponds to GS with $M = N$. Unfortunately, one can quite easily find examples for which the relevant constant $\frac{1}{\sqrt{1-C_{N,N}}}$ is exponentially large in N, making this approach completely unstable and nonconvergent. However, by allowing $N \geq M$ and using GS instead, one avoids this issue. GS has been shown to lead to a vast improvement over consistent reconstructions in numerical examples [2], [6].

Remark 6 Throughout this paper we assume that both the sampling and reconstruction systems form orthonormal bases of $\mathcal{H}$. In fact, this is not necessary: one can develop GS for both Riesz bases [6] and frames [8] with only minor modifications.

When using GS in practice, it is useful to have a means to quantify how N must scale with M to ensure stability and accuracy. In [6] the so-called *stable sampling rate* was introduced:

$$\Theta(M;\theta) = \min\left\{N \in \mathbb{N} : C_{N,M} < \theta\right\}, \quad \theta \in (0,1).$$

This quantity stipulates at what rate one must sample (i.e. how many samples to take) to obtain a stable and accurate reconstruction. Note that setting $N \geq \Theta(M;\theta)$ implies that $\frac{1}{\sqrt{1-C_{N,M}}} \leq \frac{1}{\sqrt{1-\theta}}$, and therefore stability and accuracy are guaranteed. Specific examples of the behaviour of $\Theta(M;\theta)$ are described in [2], [7].

B. Generalized sampling with compressed sensing

GS does not take advantage of any sparsity in the coefficients $\{\alpha_j\}_{j\in\mathbb{N}}$. Suppose now that f is *k-sparse* in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$, i.e. $|\Delta| = k$, where

$$\Delta = \{j : \alpha_j \neq 0\}.$$

In [3], GS was combined with convex optimization techniques to yield a framework, *generalized sampling with compressed sensing* (GS–CS), for subsampling in this setting. We now review this work.

Suppose that M is such that $\Delta \subseteq \{1, \ldots, M\}$. Let $N \in \mathbb{N}$ and pick $\Omega \subseteq \{1, \ldots, N\}$, $|\Omega| = m$ uniformly at random (note that m corresponds to the total number of samples to be used and N the range of indices from which they are drawn). The GS–CS reconstruction is now given by

$$\inf_{\eta \in l^1(\mathbb{N})} \text{ subject to } P_\Omega U P_M \eta = P_\Omega \zeta(f), \tag{7}$$

where U is as in (3). There are now three important questions:

(i) Given M and k how do we choose N? Recall that there are infinitely many samples $\{\zeta_j(f)\rangle\}_{j\in\mathbb{N}}$ to draw from. Moreover, taking $M = N$ may well not work in practice [3] (this is analogous to case $M = N$ in GS – see Remark 5).

(ii) We choose $\Omega \subseteq \{1, \ldots, N\}$ uniformly at random, but how big should $m = |\Omega|$ be to guarantee recovery with high probability? Moreover, will we recover f with probability one if $m = N$?

(iii) If f is not sparse but compressible, i.e. $f = g + h$ where g is k-sparse but $h = \sum_{j=1}^{\infty} \beta_j \varphi_j$ is not sparse but small in norm. Can we recover f from random subsamples of $\{\zeta_j(f)\}_{j\in\mathbb{N}}$? If not, can we at least do so with some error depending only on h? Moreover, how large should N and m be to guarantee this recovery?

The answers to these questions are given in the following theorems. Note that it is the notion of the *balancing property* that allows us to determine the correct relationship between N and M, k:

Definition 7. *We say that N and m satisfy the weak balancing property with respect to U, M and k if*

$$\|P_M U^* P_N U P_M - P_M\| \leq \left(4\sqrt{\log_2\left(4N\sqrt{k}/m\right)}\right)^{-1}, \tag{8}$$

and

$$\max_{\substack{|\Gamma|=k \\ \Gamma\subseteq\{1,\ldots,M\}}} \|P_M P_\Gamma^\perp U^* P_N U P_\Gamma\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{k}}, \tag{9}$$

where $\|\cdot\|_{\mathrm{mr}}$ is given by (11). If (8) holds and (9) is replaced by

$$\max_{\substack{|\Gamma|=k \\ \Gamma\subseteq\{1,\ldots,M\}}} \|P_\Gamma^\perp U^* P_N U P_\Gamma\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{k}}, \tag{10}$$

then N and m are said to satisfy the strong balancing property with respect to U, M and k.

In this definition, note that $\|\cdot\|_{\mathrm{mr}}$ is defined by

$$\|B\|_{\mathrm{mr}} = \sup_{i\in\mathbb{N}} \sqrt{\sum_{j\in\mathbb{N}} |b_{ij}|^2}, \tag{11}$$

for a bounded operator $B : l^2(\mathbb{N}) \to l^2(\mathbb{N})$, where $b_{ij} = \langle Be_j, e_i\rangle$.

Remark 8 One can show that both the weak and strong balancing properties are well defined [3, Prop. 6.2]. Also, note that (9) and (10) can be replaced by much easier to verify (although less sharp) conditions [3, Sec. 6].

The balancing property is the direct analogue of the stable sampling rate for GS–CS. Much like the former, it determines precisely how to 'discretize' the infinite change-of-basis operator U (3), the additional stipulations in the definition arising from the desire to exploit the sparsity of the problem.

Let us also define the *coherence* of U by

$$\mu(U) = \sup_{i,j\in\mathbb{N}} |u_{i,j}|, \quad u_{i,j} = \langle \varphi_j, \psi_i\rangle.$$

We are now able to state the key theorems proved in [3]:

Theorem 9 ([3] Thm 7.1)**.** *Let $M \in \mathbb{N}$, $\epsilon > 0$ and suppose that $\alpha \in l^1(\mathbb{N})$ with $\Delta = \{j : \alpha_j \neq 0\}$, where $\Delta \subseteq \{1, \dots, M\}$ and $|\Delta| = k$. Suppose that N and m satisfy the weak balancing property with respect to U, M and k, and let $\Omega \subseteq \{1, \dots, N\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = U\alpha$ then, with probability exceeding $1 - \epsilon$, the problem*

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U P_M \eta = P_\Omega \zeta, \tag{12}$$

has a unique solution ξ and this solution coincides with α, provided m exceeds

$$C \cdot N \cdot \mu^2(U) \cdot k \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{MN\sqrt{k}}{m}\right), \tag{13}$$

for some universal constant C. Furthermore, if $m = N$ then ξ is unique and $\xi = \alpha$ with probability 1.

Theorem 9 answers questions (i) and (ii) posed previously. The next result concerns (iii). Before stating this theorem, let us define the function

$$\tilde{\omega}_{M,U} : \{1, \dots, M\} \times \mathbb{R}_+ \times \mathbb{N} \to \mathbb{N}$$

by

$$\tilde{\omega}_{M,U}(r, s, N) = \left| \left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subseteq \{1,\dots,N\} \\ \Gamma_2 \subseteq \{1,\dots,M\} \\ |\Gamma_2| = r}} \|P_{\Gamma_2} U^* P_{\Gamma_1} U e_i\| > s \right\} \right|. \tag{14}$$

Observe that $s \mapsto \tilde{\omega}_{M,U}(r, s, N)$ is a decreasing function.

Theorem 10 ([3] Thm 7.3)**.** *Let $M \in \mathbb{N}$, $\epsilon > 0$ and $\alpha, h \in l^1(\mathbb{N})$ with $\{j : \alpha_j \neq 0\} = \Delta$, where $\Delta \subseteq \{1, \dots, M\}$ and $|\Delta| = k$. Define $\beta = \alpha + h$. Suppose that N and m satisfy the strong balancing property with respect to U, M and k, and let $\Omega \subseteq \{1, \dots, N\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = U\beta$ and $\xi \in \mathcal{H}$ is a minimizer of*

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \zeta, \tag{15}$$

then, with probability exceeding $1 - \epsilon$, we have that

$$\|\xi - \beta\| \leq \left(\frac{20N}{m} + 11 + \frac{m}{2N}\right) \|h\|_{l^1}, \tag{16}$$

provided m exceeds

$$C \cdot N \cdot \mu^2(U) \cdot k \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{\omega N\sqrt{k}}{m}\right), \tag{17}$$

for some universal constant C, where

$$\omega = \tilde{\omega}_{M,U}(k, s, N), \qquad s = \frac{m}{32N\sqrt{k}\log(e^4\epsilon^{-1})}$$

and $\tilde{\omega}_{M,U}$ is given by (14). If $m = N$ then (16) holds with probability 1.

The reader will no doubt have noticed that the optimization problem (15) is infinite-dimensional. In practice, we solve the finite-dimensional problem

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U P_t \eta = P_\Omega \zeta, \tag{18}$$

where $t \in \mathbb{N}$. One can show that, provided t is taken sufficiently large, any solution of (18) will approximately solve (15). Indeed,

Proposition 11 ([3] Prop. 7.4)**.** *Let $\alpha \in l^1(\mathbb{N})$ and P_Ω be a finite rank projection. Then, for all sufficiently large $t \in \mathbb{N}$, there exists an $\xi_t \in \mathcal{H}$ satisfying*

$$\|\xi_t\|_{l^1} = \inf_{\eta \in l^1(\mathbb{N})} \left\{\|\eta\|_{l^1} : P_\Omega U P_t \eta = P_\Omega U \alpha\right\}.$$

Moreover, for every $\epsilon > 0$ there is a $T \in \mathbb{N}$ such that, for all $t \geq T$, we have $\|\xi_t - \tilde{\xi}_t\|_{l^1} \leq \epsilon$, where $\tilde{\xi}_t$ satisfies

$$\|\tilde{\xi}_t\|_{l^1} = \inf_{\eta \in l^1(\mathbb{N})} \left\{\|\eta\|_{l^1} : P_\Omega U \eta = P_\Omega U \alpha\right\}. \tag{19}$$

In particular, if there exists a unique minimiser α of (19), then $\xi_t \to \alpha$ in the l^1 norm as $t \to \infty$.

C. Summary

Theorems 9 and 10 demonstrate that it is possible with GS–CS to recover sparse or compressible signals using only a small number m of their measurements. However, one notes that the number of samples m depends critically on the coherence $\mu(U)$, or, more precisely, the product $N\mu^2(U)$. Unfortunately, $\mu(U)$ is fixed. Therefore if M (and consequently N) is sufficiently large we cannot expect to be able to subsample, i.e. we must take a number of samples m equal to N. This is the so-called *incoherence barrier*.

Now, in many cases, this may not be a problem: numerical examples given in [3] illustrate sparse recovery with a good degree of subsampling. In other words, $\mu(U)$ is often sufficiently small so that the incoherence barrier is never encountered for all reasonable values of M. However, if $\mu(U)$ is 1 (or even if $\mu(U)$ is $\mathcal{O}(1)$) then it may well not be possible to subsample with this approach (specific examples are given in §IV). Fortunately, the new technique of sem-random subsampling, which we introduce in the next section, circumvents this barrier. To do so, it exploits so-called asymptotic incoherence of many typical sampling and reconstruction pairs $\{\{\psi_j\}_{j\in\mathbb{N}}, \{\phi_j\}_{j\in\mathbb{N}}\}$.

III. Semi-random sampling and GS–CS

Definition 12. *Let $\{\psi_j\}_{j\in\mathbb{N}}$ and $\{\varphi_j\}_{j\in\mathbb{N}}$ be orthonormal bases of a Hilbert space $\mathcal{H}$ and suppose that U is the corresponding change-of-basis operator (3). If*

$$\mu_n := \mu(UP_n^\perp) \to 0, \quad n \to \infty, \tag{20}$$

then the pair $\{\{\psi_j\}_{j\in\mathbb{N}}, \{\phi_j\}_{j\in\mathbb{N}}\}$ is said to be asymptotically incoherent.

We give specific examples of pairs which possess this property in §IV. For now, however, we shall explain how to exploit this property to achieve subsampling regardless of how large M is, even in the case that $\mu(U) = 1$.

A. Semi-random sampling

The key to overcoming the incoherence barrier is a very simple idea. We shall seek to sample fully those indices corresponding to the columns of U that are coherent, and subsample elsewhere.

The sparse case. Suppose first that $\alpha \in l^1(\mathbb{N})$ and

$$\Delta = \{j : \alpha_j \neq 0\} = \Delta_1 \cup \Delta_2, \tag{21}$$

where Δ_1 and Δ_2 are disjoint with

$$\Delta_1 \subseteq \{1, \dots, M_1\}, \quad \Delta_2 \subseteq \{M_1 + 1, \dots, M_2\},$$

for some $M_1, M_2 \in \mathbb{N}$. Write $k_1 = |\Delta_1|$ and $k_2 = |\Delta_2|$. To recover α we now proceed as follows:

- Let $\epsilon > 0$ and suppose that $N_2, m_2 \in \mathbb{N}$ satisfy the weak balancing property with respect to U, M_2 and k_2. Also, let m_2 exceed

$$C \cdot N_2 \cdot \mu^2(UP_{M_1}^{\perp}) \cdot k_2 \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{M_2 N_2 \sqrt{k_2}}{m_2}\right), \tag{22}$$

  for some universal constant C.
- Let N_1, $m_1 := N_1$ satisfy the weak balancing property with respect to U, M_1 and k_1 and let $\Omega_1 = \{1, \dots, N_1\}$.
- Choose $\Omega_2 \subseteq \{N_1 + 1, \dots, N_2\}$ uniformly at random with $|\Omega_2| = m_2$, and set

$$\Omega = \Omega_1 \cup \Omega_2.$$

- If $\zeta = U\alpha$, solve

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } \; P_\Omega U P_{M_2} \eta = P_\Omega \zeta. \tag{23}$$

The compressible case. Suppose now that $\beta = \alpha + h$, where $h \in l^1(\mathbb{N})$ and α satisfies (21). In this case, we make the following construction:

- Let $\epsilon > 0$ and suppose that $N_2, m_2 \in \mathbb{N}$ satisfy the strong balancing property with respect to U, M_2 and k_2. Also, let m_2 exceed

$$C \cdot N_2 \cdot \mu^2(UP_{M_1}^{\perp}) \cdot k_2 \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{\omega N_2 \sqrt{k_2}}{m_2}\right), \tag{24}$$

  where

$$\omega = \tilde{\omega}_{M_2, U}(k_2, s, N_2), \quad s = \frac{m_2}{32 N_2 \sqrt{k_2} \log(e^4 \epsilon^{-1})}.$$

- Let N_1, $m_1 := N_1$ satisfy the weak balancing property with respect to U, M_1 and k_1 and let $\Omega_1 = \{1, \dots, N_1\}$.
- Choose $\Omega_2 \subseteq \{N_1 + 1, \dots, N_2\}$ uniformly at random with $|\Omega_2| = m_2$, and set

$$\Omega = \Omega_1 \cup \Omega_2.$$

- If $\zeta = U\alpha$, solve

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } \; P_\Omega U P_t \eta = P_\Omega \zeta, \tag{25}$$

  where $t \in \mathbb{N}$ is sufficiently large.

It is beyond the scope of this paper to analyse the reconstructions (23) and (25). However, both Theorems 9 and 10 can be generalized to this case (proofs will be presented in a separate paper [5] – these are technical, but similar in nature to the arguments given in [3]). In other words, in the sparse case, one can show that with probability greater than $1 - \epsilon$ there is a unique solution ξ to (23) and this solution coincides with α. On the other hand, in the compressible case, one can prove that if ξ is a minimizer of

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } \; P_\Omega U P \eta = P_\Omega \zeta,$$

then with probability exceeding $1 - \epsilon$ we have

$$\|\xi - \beta\| \leq \left(\frac{20 N_2}{m_2} + 11 + \frac{m_2}{2N_2}\right) \|h\|_{l^1}. \tag{26}$$

Hence, using Proposition 11, one finds that the solutions of (25) will approximately satisfy (26) for sufficiently large M_2'.

B. Explanation of semi-random sampling

With semi-random subsampling we sample deterministically over those indices in $\Omega_1 = \{1, \dots, N_1\}$, where N_1 is specified through M_1 and the balancing property. We then subsample in the incoherent region of U, i.e. that corresponding to the tail $UP_{M_1}^{\perp}$. The approach avoids the potential pitfalls of the standard GS–CS technique of §II (based on random sampling) in the following manner. Unlike with the latter, the estimates (22) and (24) both decrease as M_1 increases. Thus, one can take a smaller value of m_2, and therefore subsample, even when the total coherence $\mu(U)$ is large.

To obtain a complete picture of how to perform semi-random sampling, we need estimates for the rate of convergence in (20). These estimates depend completely on the pairs of sampling and reconstruction bases, and thus must be derived on a case-by-case basis. Fortunately, as we discuss in the next section, in many important cases (e.g. reconstructions in wavelets or polynomials) one can obtain good estimates.

Remark 13 Note that the issue of fixed coherence is not just a problem in infinite dimensions. Consider the following example. Let $U_{\mathrm{df},n} \in \mathbb{C}^{n \times n}$ denote the discrete Fourier transform. Then $\mu(U_{\mathrm{df},n}) \to 0$ as $n \to \infty$. However, if we let $V_{\mathrm{dw},n} \in \mathbb{C}^{n \times n}$ denote the discrete wavelet transform corresponding to the Haar wavelet, then $\mu(U_{\mathrm{df},n} V_{\mathrm{dw},n}^{-1}) = 1$ for all n. In particular, we have the worst possible incoherence. In this case one may also solve the problem by semi-random sampling, as suggested above. We omit the details as this paper is about the infinite-dimensional case. Note, however, that in the finite-dimensional case, such a semi-random sampling technique has already been used in applications [43] (in this paper it is referred to as the "half-half strategy").

IV. Estimates for asymptotic incoherence

Recall the main example of this paper (§I-D) which concerns reconstructions from equispaced samples of the Fourier transform. The purpose of this section is to obtain estimates for (20) for this type of sampling in the cases where the reconstruction basis is consists of (i) wavelets and (ii) polynomials. Proof are found in Appendix B.

A. Wavelet bases

Let $\Omega_a = \{\varphi_j\}_{j\in\mathbb{N}}$ be an orthonormal system of wavelets such that $L^2(0,a) \subset \overline{\text{span}}\{\Omega_a\} \subset L^2(-b,b)$, where $0 < a,b < \infty$, with mother wavelet ψ and a scaling function ϕ (see Appendix A for a general construction). Note that $L^2(0,a)$ is chosen for simplicity, and this extends in an obvious way to $L^2(-a/2,a/2)$. If

$$\psi_j(\cdot) = \sqrt{\epsilon}\mathrm{e}^{2\pi\mathrm{i}\epsilon\rho(j)\cdot}, \quad j \in \mathbb{N}, \tag{27}$$

(recall ρ from §I-D), then let U denote the corresponding infinite matrix (3). We are interested in the behaviour of $\mu_n = \mu(UP_n^\perp)$. We have

Theorem 14. *Let Ω_a and U be as above with $a \geq 1$. Then*

$$\mu(UP_n^\perp) \leq \begin{cases} \sqrt{\epsilon}\max\{A,B\} & n < 2\lceil a\rceil - 1, \\ \sqrt{\epsilon}\sqrt{2}^{-\left\lfloor \log_2\left(\frac{n+3}{\lceil a\rceil}-3\right)\right\rfloor}B & n \geq 2\lceil a\rceil - 1, \end{cases} \tag{28}$$

where

$$A = \sup_{t\in\mathbb{R}} |\mathcal{F}\phi(t)|, \quad B = \sup_{t\in\mathbb{R}} |\mathcal{F}\psi(t)|.$$

In particular,

$$\mu(UP_n^\perp) = \mathcal{O}\left(\sqrt{\frac{\epsilon}{n}}\right), \qquad n \to \infty. \tag{29}$$

Corollary 15. *Suppose that $\{\varphi_j\}_{j\in\mathbb{N}}$ are the Haar wavelets spanning $\mathrm{L}^2([0,1])$, and let $\{\psi_j\}_{j\in\mathbb{N}}$ be given by (27). If U is as in (3), we have*

$$\mu(U) = \sqrt{\epsilon}, \qquad \mu(UP_n^\perp) \leq \frac{0.73\cdot\sqrt{\epsilon}}{\sqrt{2}^{\lfloor \log_2(n)\rfloor}}, \ n \in \mathbb{N}.$$

In particular, when n is a power of two, then

$$\mu(UP_n^\perp) \leq \frac{0.73\cdot\sqrt{\epsilon}}{\sqrt{n}}.$$

Theorem 14 comes as a nice surprise. First, it is true for all wavelets! Second, the result is rather striking when compared with the DFT matrix $U_{\mathrm{df},n} \in \mathbb{C}^{n\times n}$. In this case, one has $\mu(U_{\mathrm{df},n}) = \mathcal{O}(1/\sqrt{n})$; a result which is often referred to as *perfect* incoherence [12].

B. An example with wavelets

We now give an example of semi-random sampling with wavelets, using Fourier sampling described in (27) with $\epsilon = 1$. Let $f \in L^2(-\frac{1}{2},\frac{1}{2})$ be given by

$$f(x) = \sum_{j=1}^{K} \alpha_j\varphi_j(x) + (x+\tfrac{1}{2})^2 e^{-x-\frac{1}{2}},$$

where $\{\varphi_j\}_{j\in\mathbb{N}}$ is the Haar wavelet basis on $[-\frac{1}{2},\frac{1}{2}]$, $K = 800$ and $\beta_j \in (0,20]$ are some arbitrarily chosen numbers with

$$|\{j : \beta_j \neq 0\}| = 70.$$

The classical approach to this problem would be to use the Shannon Sampling Theorem and construct the partial Fourier series f_N of f. However, as commented, this gives a very poor reconstruction of f (see also Figure 1). Nonetheless, given that $f = g + h$, where g is sparse in the Haar basis, it is natural to try and recover f using the infinite-dimensional compressed sensing techniques developed in this paper. In particular, let $f_{N,m,t}$ and $f_{N_1,N_2,m_2,t}$ be the reconstructions of f obtained by applying GS–CS with firstly random (§II-B), and secondly semi-random (§III-A), subsampling. In other words,

$$f_{N,m,t} = \sum_{j=1}^{t} \alpha_j\varphi_j, \quad f_{N_1,N_2,m_2,t} = \sum_{j=1}^{t} \beta_j\varphi_j,$$

where $\alpha = \{\alpha_1,\ldots,\alpha_t\}$ and $\beta = \{\beta_1,\ldots,\beta_t\}$ are minimizers of (18) and (25) respectively.

In Figure 1 we plot the errors committed by the three reconstruction techniques. As expected, the Shannon reconstruction f_N is extremely poor. Moreover, GS–CS with random subsampling also gives a poor reconstruction. On the other hand, when semi-random subsampling is used, one obtains a significantly better reconstruction, despite using precisely the same total number of samples. The advantage of semi-random subsampling is exactly as expected, given the analysis of the previous section. In particular, by Corollary 15, $\mu(U) = 1$, whereas $\mu(UP_n^\perp) \leq \frac{0.73}{\sqrt{n}}$. Note also that both the random and semi-random reconstructions require only around 38% of the samples used in the extremely poor reconstruction f_N.

C. Polynomial bases

A convenient way to represent signals $f \in L^2(-1,1)$ (especially when f has some smoothness) is with a basis of orthogonal polynomials. Let us assume once more that f is sampled via (27). We now consider reconstructions in the orthonormal basis $\{\varphi_j\}_{j\in\mathbb{N}}$, where

$$\varphi_j(x) = \sqrt{j-\tfrac{1}{2}}p_{j-1}(x), \quad j \in \mathbb{N}, \tag{30}$$

and p_j is the j^{th} Legendre polynomial [1].

There are a number of methods to reconstruct functions in terms of polynomials from their Fourier samples (see [7], [28], [29] for details). GS, which in this specific instance turns out to be closely related to work of Hrycak & Gröchenig [34], was applied successfully to this problem in [7], and found to outperform more conventional approaches. Having said this, none of the existing techniques exploit sparsity to achieve subsampling. Indeed, Gröchenig has conjectured [33] the potential use of subsampling for this type of problem, but there does not currently exist a technique which utilizes sparsity in this way.

For this reason, we are interested in applying GS–CS to solve this problem. The question we must therefore ask is the following: is this scenario well suited to subsampling? An answer in the affirmative is provided in the next theorem:

Theorem 16. *Suppose that $\{\psi_j\}_{j\in\mathbb{N}}$ and $\{\varphi_j\}_{j\in\mathbb{N}}$ are as in (27) and (30) respectively, and let U be given by (3). Then $\mu(U) = 1$ and*

$$\mu(UP_n^\perp) = \mathcal{O}\left(n^{-\frac{1}{3}}\right), \quad n \to \infty.$$

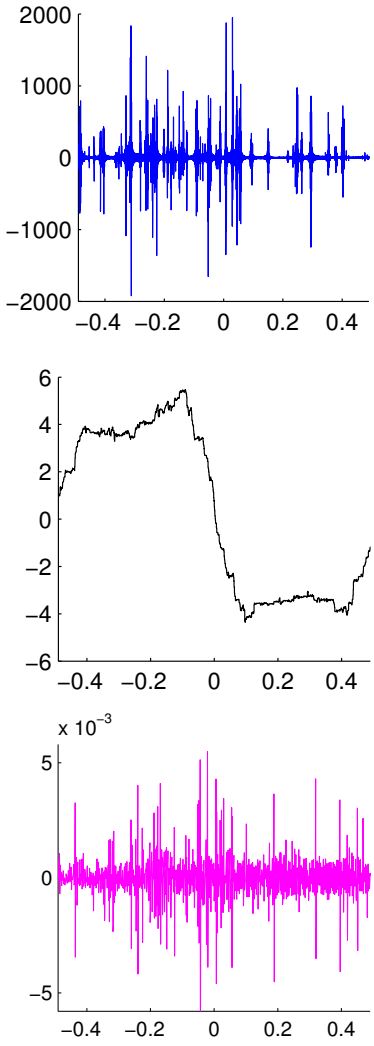

Fig. 1. The errors $f - f_N$ (top), $f - f_{N,m,t}$ (avg. 20 trials) (middle) and $f - f_{N_1,N_2,m_2,t}$ (avg. 20 trials) (bottom), where $N = N_2 = 1451$, $m = 560$, $N_1 = 260$, $m_2 = 300$. Note that $f_{N_1,N_2,m_2,t}$ uses only 38% of the samples used for f_N.

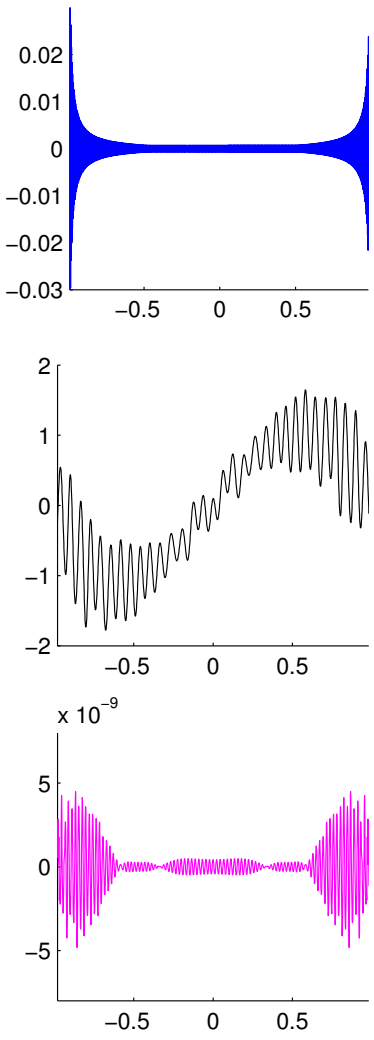

Fig. 3. The errors $f - f_N$ (top), $f - f_{N,m,t}$ (avg. 20 trials) (middle) and $f - f_{N_1,N_2,m_2,t}$ (avg. 20 trials) (bottom), where $N = N_2 = 1001$, $m = 370$, $N_1 = 170$, $m_2 = 200$. Note that $f_{N_1,N_2,m_2,t}$ uses only 37% of the samples used for f_N.

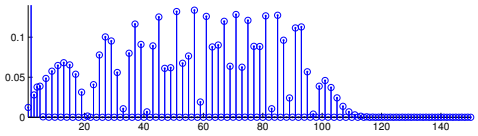

Fig. 2. The absolute value of the Legendre coefficients of f.

D. An example with polynomials

Consider the function

$$f(x) = \sinh(x) + x\sin(100x).$$

Since f is nonperiodic, its reconstruction f_N via Shannon is very poor. On the other hand, since f is smooth it has rapidly decaying coefficients in the basis of Legendre polynomials. This is illustrated in Figure 2. As can also be seen from this figure, we may actually decompose f into $g + h$, where g is sparse in the Legendre basis and h has small coefficients. This suggests that subsampling may be possible.

In Figure 3 we plot the results for the reconstructions f_N, $f_{N,m,t}$ and $f_{N_1,N_2,m_2,t}$, where the latter two correspond to GS–CS with random and semi-random subsampling. As in the wavelet example, random subsampling gives a poor reconstruction. However, semi-random approach yields a dramatic improvement. This is exactly as predicted by Theorem 16. Note that the semi-random reconstruction also significantly outperforms the Shannon reconstruction: despite using only 37% of the samples, it improves the reconstruction of f by roughly seven digits.

Remark 17 Sparse expansions in Legendre polynomials have also recently been considered by Rauhut & Ward [41]. However, [41] considers a fundamentally different problem. Therein, the user is free to choose the sampling scheme, and, as shown, if one takes pointwise samples drawn from a Chebyshev distribution, then one can successfully subsample. Herein lies the key difference with this paper. In GS–CS the sampling scheme (in this case, sampling the Fourier transform

at equispaced nodes) is considered fixed.

ACKNOWLEDGMENTS

The authors would like to thank Akram Aldroubi, Emmanuel Candès, Karlheinz Gröchenig, Michael Unser and Martin Vetterli for useful discussions and comments.

APPENDIX A
WAVELET BASES

Let us review the basic wavelet approach [2] on how to create orthonormal subsets $\{\varphi_k\}_{k\in\mathbb{N}} \subseteq \mathrm{L}^2(\mathbb{R})$ with the property that $\mathrm{L}^2(0,a) \subseteq \overline{\mathrm{span}}\{\varphi_k\}_{k\in\mathbb{N}}$ for some $a > 0$. Suppose that we are given a mother wavelet ψ and a scaling function ϕ such that $\mathrm{supp}(\psi) = \mathrm{supp}(\phi) = [0,a]$ for some $a \geq 1$. The most standard approach is to consider the following collection of functions

$$\begin{aligned}\Omega_a = \{&\phi_k, \psi_{j,k} : \mathrm{supp}(\phi_k)^o \cap [0,a] \neq \emptyset,\\ &\mathrm{supp}(\psi_{j,k})^o \cap [0,a] \neq \emptyset, j \in \mathbb{Z}_+, k \in \mathbb{Z},\},\end{aligned}$$

where

$$\phi_k = \phi(\cdot - k), \qquad \psi_{j,k} = 2^{\frac{j}{2}}\psi(2^j \cdot - k).$$

(the notation K^o denotes the interior of a set $K \subseteq \mathbb{R}$). This now gives

$$\mathrm{L}^2(0,a) \subseteq \mathrm{cl}(\mathrm{span}\{\varphi : \varphi \in \Omega_a\}) \subseteq \mathrm{L}^2(-b,b),$$

where $b > 0$ is such that $[-b,b]$ contains the support of all functions in Ω_a. Note that the inclusions may be proper (but not always, as is the case with the Haar wavelet.) It is easy to see that

$$\begin{aligned}\psi_{j,k} \notin \Omega_a &\iff \frac{a+k}{2^j} \leq 0, \quad a \leq \frac{k}{2^j},\\ \phi_k \notin \Omega_a &\iff a + k \leq 0, \quad a \leq k,\end{aligned}$$

and therefore

$$\begin{aligned}\Omega_a = &\{\phi_k : |k| = 0, \ldots, \lceil a\rceil - 1\}\\ &\cup \{\psi_{j,k} : j \in \mathbb{Z}_+, k \in \mathbb{Z}, -\lceil a\rceil < k < 2^j\lceil a\rceil\}.\end{aligned}$$

We will order Ω_a as follows:

$$\begin{aligned}\{&\phi, \phi_1, \ldots, \phi_{\lceil a\rceil-1}, \phi_{-1}, \ldots, \phi_{-\lceil a\rceil+1}, \psi_{0,0}, \psi_{0,1},\\ &\ldots, \psi_{0,\lceil a\rceil-1}, \psi_{0,-1}, \ldots, \psi_{0,-\lceil a\rceil+1}, \psi_{1,0}, \ldots\}.\end{aligned} \tag{31}$$

APPENDIX B
PROOFS

Proof of Theorem 14: By definition Note that by the definition of U it follows that

$$\begin{aligned}u_{l,m} = \zeta_l(\varphi_m) &= \int_{-\infty}^{\infty} \sqrt{\epsilon}\,\mathrm{e}^{-2\pi\mathrm{i}\rho(l)\epsilon t}\varphi_m(t)\,\mathrm{d}t\\ &= \sqrt{\epsilon}\,\mathcal{F}\varphi_m(\epsilon\rho(l)), \quad l,m \in \mathbb{N}.\end{aligned}$$

Thus,

$$\mu(UP_n^{\perp}) = \sup\{|\sqrt{\epsilon}\,\mathcal{F}\varphi_m(\epsilon\rho(l))| : l \in \mathbb{N},\, m > n\}. \tag{32}$$

Note also that

$$\mathcal{F}\phi_k(t) = \mathrm{e}^{-2\pi\mathrm{i}kt}\mathcal{F}\phi(t), \tag{33}$$

and

$$\mathcal{F}\psi_{j,k}(t) = \mathrm{e}^{-2\pi\mathrm{i}2^{-j}kt}2^{\frac{-j}{2}}\mathcal{F}\psi(2^{-j}t). \tag{34}$$

So, by the ordering in (31) as well as (32), (33) and (34) it follows that

$$\mu(UP_n^{\perp}) \leq \sqrt{\epsilon}\max\left\{\sup_{t\in\mathbb{R}}|\mathcal{F}\phi(t)|, \sup_{t\in\mathbb{R}}|\mathcal{F}\psi(t)|\right\},$$

whenever $n < 2\lceil a\rceil - 1$, which gives the first part of (28). As for the second part of (28), define the bijection

$$\theta : \mathbb{N} \setminus \{1, \ldots, 2\lceil a\rceil - 2\} \ni n \mapsto (j,k)\,\text{s.t.}\,\varphi_n = \psi_{j,k}.$$

Then, by (34), it follows that

$$\mu(UP_n^{\perp}) \leq \epsilon\, 2^{\frac{-j}{2}} \sup_{t\in\mathbb{R}}|\mathcal{F}\psi(t)|, \quad (j,k) = \theta(n+1), \tag{35}$$

for $n \geq 2\lceil a\rceil - 1$. Thus, to get the bound we only need to figure out what is j in $(j,k) = \theta(n+1)$. Note that since k ranges from $-\lceil a\rceil + 1$ to $2^j\lceil a\rceil - 1$ it follows that j must satisfy $j = \lfloor q\rfloor$ where

$$(2^q + 1)\lceil a\rceil - 1 = r,$$

for some integer r. In particular,

$$j = \left\lfloor \log_2\left(\frac{r+1}{\lceil a\rceil} - 1\right)\right\rfloor,$$

however, it is clear that we must have

$$r = n + 1 - (2\lceil a\rceil - 1).$$

Thus, by (35), we get that

$$\mu(UP_n^{\perp}) \leq \epsilon\,\sqrt{2}^{-\left\lfloor \log_2\left(\frac{n+3}{\lceil a\rceil} - 3\right)\right\rfloor} \sup_{t\in\mathbb{R}}|\mathcal{F}\psi(t)|,$$

yielding the second part of (28). ∎

Proof of Corollary 15: First, let us observe that $\phi(x) = \chi_{[0,1]}(x)$ and $\psi(x) = \chi_{[0,1/2)}(x) - \chi_{[1/2,1]}(x)$, which automatically gives

$$\mathcal{F}\phi(t) = \frac{\mathrm{i}(\mathrm{e}^{-2\pi\mathrm{i}t} - 1)}{2\pi t}, \ \mathcal{F}\psi(t) = \frac{\mathrm{i}\mathrm{e}^{-2\pi\mathrm{i}t}(1 - \mathrm{e}^{\pi\mathrm{i}t})^2}{2\pi t}.$$

Let A and B be as in Theorem 14. Then

$$A = \sup_{t\in\mathbb{R}}|\mathcal{F}\phi(t)| = 1.$$

Moreover, a simple calculus exercise yields that $|\mathcal{F}\psi(t)|$ has two global maxima at $t = \pm t_{\max}$ with $|\mathcal{F}\psi(t_{\max})| = |\mathcal{F}\psi(-t_{\max})| \leq 0.73$. Hence $B \leq 0.73$. We may now apply Theorem 14 with $a = 1$ to yield the upper bound for $\mu(UP_n^{\perp})$. Since this also holds for $n = 0$, we find that $\mu(U) \leq \sqrt{\epsilon}$. To show that $\mu(U) = \sqrt{\epsilon}$ we merely note that $|u_{1,1}| = \sqrt{\epsilon}|\mathcal{F}\phi(0)| = \sqrt{\epsilon}$. ∎

Proof of Theorem 16:
By definition,

$$u_{1,1} = \int_{-1}^{1}\frac{1}{2}\,\mathrm{d}x = 1.$$

Therefore $\mu(U) \geq 1$. However, since $\{\psi_j\}_{j\in\mathbb{N}}$ and $\{\phi_j\}_{j\in\mathbb{N}}$ are orthonormal bases, we must have that $\sum_{j=1}^{\infty}|\langle\phi_k, \psi_j\rangle|^2 =$

1 for any $k \in \mathbb{N}$. It follows that $\mu(U) = 1$. Next we consider the asymptotic bound. Note that

$$u_{1,1} = 1, \quad u_{1,k} = 0, \ k \neq 1,$$

and that

$$u_{j,k} = (-1)^k \sqrt{\frac{k-\frac{1}{2}}{\rho(j)}} J_{k-\frac{1}{2}}(\rho(j)\pi), \quad k \in \mathbb{N},$$

where, for $\nu \in \mathbb{R}_+$, J_ν denotes the Bessel function of the first kind [7]. Note that this follows directly from the integral representation

$$j_k(z) = \frac{1}{2}(-\mathrm{i})^k \int_{-1}^{1} \mathrm{e}^{\mathrm{i}zt} p_k(t)\,\mathrm{d}t, \quad \forall z \in \mathbb{C}, \ k \in \mathbb{Z}_+, \quad (36)$$

(see [1, 10.1.14]), where j_k is the spherical Bessel function of the first kind, given by

$$j_k(z) = \sqrt{\frac{\pi}{2z}} J_{k+\frac{1}{2}}(z), \quad k \in \mathbb{Z}_+.$$

In particular, for any $k \in \mathbb{N}$,

$$\sup_{j\in\mathbb{N}} |u_{j,k}| \leq \sqrt{\pi\left(k-\tfrac{1}{2}\right)} \sup_{t\in[0,\infty)} \frac{J_{k-\frac{1}{2}}(t)}{\sqrt{t}}. \quad (37)$$

Thus, to bound the right hand side of (37) we are interested in estimating $\sup_{t\in[0,\infty)} |j_\nu(t)|$ for $\nu \in \mathbb{Z}_+$. To get such estimates, let $a'_{\nu,r}$ (using notation from [1]) denote the r^{th} non-negative zero of j'_ν. In particular, for some r,

$$\sup_{t\in[0,\infty)} |j_\nu(t)| = j_\nu(a'_{\nu,r}).$$

However, it is known [48] that

$$j_\nu(a'_{\nu,1}) > j_\nu(a'_{\nu,2}) > j_\nu(a'_{\nu,3}) > \ldots, \qquad \nu \in \mathbb{Z}_+.$$

Thus, we only need to have estimates for $j_\nu(a'_{\nu,1})$. It is known (see [1, 10.1.59]) that

$$a'_{\nu,1} \sim \left(\nu+\frac{1}{2}\right) + \sum_{s=1}^{\infty} c_s \left(\nu+\frac{1}{2}\right)^{1-\frac{2s}{3}}, \quad \nu \to \infty, .$$

where $|c_s| \leq 1$. Thus, there exists a function $\theta : \mathbb{R} \to \mathbb{R}$ such that

$$a'_{\nu,1} = \theta(\nu)\left(\nu+\frac{1}{2}\right), \qquad \lim_{\nu\to\infty} \theta(\nu) = 1.$$

In particular, we have established that

$$\sup_{t\in[0,\infty)} \frac{J_{\nu+\frac{1}{2}}(t)}{\sqrt{t}} = \frac{J_{\nu+\frac{1}{2}}\left(\theta(\nu)\left(\nu+\frac{1}{2}\right)\right)}{\sqrt{\theta(\nu)\left(\nu+\frac{1}{2}\right)}}, \ \nu \in \mathbb{Z}_+. \quad (38)$$

From (37) and (38) it follows that

$$\mu(UP_n^\perp) \leq \sup_{m>n} \sqrt{\frac{\pi}{\theta(m-1)}} J_{m-\frac{1}{2}}\left(\theta(m-1)\left(m-\tfrac{1}{2}\right)\right). \quad (39)$$

In [36] it was shown that

$$|J_\nu(t)| < b\nu^{-1/3}, \qquad t \in \mathbb{R}, \quad \nu \in \mathbb{R}_+, \quad (40)$$

where $b = 0.674885\ldots$. Hence, from (39) and (40) we obtain

$$\mu(UP_n^\perp) = \mathcal{O}\left(\frac{1}{n^{1/3}}\right), \qquad n \to \infty,$$

as required. ∎

REFERENCES

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, 1974.

[2] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *J. Fourier Anal. Appl. (to appear)*, 2010.

[3] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Submitted*, 2011.

[4] B. Adcock and A. C. Hansen. Reduced consistency sampling in Hilbert spaces. In *Proceedings of the 9th International Conference on Sampling Theory and Applications*, 2011.

[5] B. Adcock and A. C. Hansen. Semi-random sampling and asymptotic incoherence in compressed sensing. *In preparation*, 2011.

[6] B. Adcock and A. C. Hansen. Sharp bounds, optimality and a geometric interpretation for generalised sampling in Hilbert spaces. *Submitted*, 2011.

[7] B. Adcock and A. C. Hansen. Stable reconstructions in Hilbert spaces and the resolution of the Gibbs phenomenon. *Appl. Comput. Harmon. Anal. (to appear)*, 2011.

[8] B. Adcock, A. C. Hansen, E. Herrholz, and G. Teschke. Generalized sampling: extension to frames and ill-posed problems. *Submitted*, 2011.

[9] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulout. Sparse sampling of signal innovations. *IEEE Signal Process. Mag.*, 25(2):31–40, 2008.

[10] E. J. Candès. An introduction to compressive sensing. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.

[11] E. J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *Submitted*, 2010.

[12] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.

[13] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[14] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[15] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[16] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to compressed sensing. In *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011.

[17] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[18] P. L. Dragotti, M. Vetterli, and T. Blu. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix. *IEEE Trans. Signal Process.*, 55(5):1741–1757, 2007.

[19] M. F. Duarte and Y. C. Eldar. Structured compressed sensing: from theory to applications. *IEEE Trans. Signal Process.*, 59(9):4053–4085, 2011.

[20] T. Dvorkind and Y. C. Eldar. Robust and consistent sampling. *IEEE Signal Process. Letters*, 16(9):739–742, 2009.

[21] Y. C. Eldar. Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors. *Journal of Fourier Analysis and Applications*, 9(1):77–96, 2003.

[22] Y. C. Eldar. Sampling without input constraints: Consistent reconstruction in arbitrary spaces. *Sampling, Wavelets and Tomography*, 2003.

[23] Y. C. Eldar. Compressed sensing of analog signals in shift-invariant spaces. *IEEE Trans. Signal Process.*, 57(8):2986–2997, 2009.

[24] Y. C. Eldar and T. Dvorkind. A minimum squared-error framework for generalized sampling. *IEEE Trans. Signal Process.*, 54(6):2155–2167, 2006.

[25] Y. C. Eldar and T. Michaeli. Beyond Bandlimited Sampling. *IEEE Signal Process. Mag.*, 26(3):48–68, 2009.

[26] Y. C. Eldar and T. Werther. General framework for consistent sampling in Hilbert spaces. *Int. J. Wavelets Multiresolut. Inf. Process.*, 3(3):347, 2005.

[27] M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.

[28] A. Gelb and J. Tanner. Robust reprojection methods for the resolution of the Gibbs phenomenon. *Appl. Comput. Harmon. Anal.*, 20:3–25, 2006.

[29] D. Gottlieb and C.-W. Shu. On the Gibbs' phenomenon and its resolution. *SIAM Rev.*, 39(4):644–668, 1997.

[30] A. C. Hansen. On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators. *J. Amer. Math. Soc.*, 24(1):81–124, 2011.

[31] E. Herrholz and G. Teschke. Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems. *Inverse Problems*, 26(12):125012, 2010.

[32] A. Hirabayashi and M. Unser. Consistent sampling and signal recovery. *IEEE Trans. Signal Process.*, 55(8):4104–4115, 2007.

[33] T. Hrycak and K. Gröchenig. Pseudospectral Fourier reconstruction with IPRM. In *8th International Conference on Sampling Theory and Applications*, 2009.

[34] T. Hrycak and K. Gröchenig. Pseudospectral Fourier reconstruction with the modified inverse polynomial reconstruction method. *J. Comput. Phys.*, 229(3):933–946, 2010.

[35] A. J. Jerri. The Shannon Sampling Theorem – its various extensions and applications: A tutorial review. *Proc. IEEE*, 65:1565–1596, 1977.

[36] L. J. Landau. Bessel functions: monotonicity and bounds. *J. London Math. Soc. (2)*, 61(1):197–215, 2000.

[37] P. E. Z. Larson, S. Hu, M. Lustig, A. B. Kerr, S. J. Nelson, J. Kurhanewicz, J. M. Pauly, and D. B. Vigneron. Fast dynamic 3D MR spectroscopic imaging with compressed sensing and multiband excitation pulses for hyperpolarized 13c studies. *Magnetic Resonance in Medicine*, 2010.

[38] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed Sensing MRI. *IEEE Signal Process. Mag.*, 25(2):72–82, March 2008.

[39] M. Mishali and Y. C. Eldar. Xampling: Analog to digital at sub-Nyquist rates. *IET Circuits, Devices, & Systems*, 5(1):8–20, 2011.

[40] M. Mishali, Y. C. Eldar, and A. J. Elron. Xampling: Signal acquisition and processing in union of subspaces. *IEEE Trans. Signal Process.*, 59(10):4719–4734, 2011.

[41] H. Rauhut and R. Ward. Sparse Legendre expansions via l1-minimization. *Submitted*, 2010.

[42] H. Rauhut and R. Ward. Sparse recovery for spherical harmonic expansions. In *Proceedings of the 9th International Conference on Sampling Theory and Applications*, 2011.

[43] V. Studer, J. Bobin, M. Chahid, H. Moussavi, E. Candès, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Submitted*, 2011.

[44] M. Unser. Sampling–50 years after Shannon. *Proc. IEEE*, 88(4):569–587, 2000.

[45] M. Unser and A. Aldroubi. A general sampling theory for nonideal acquisition devices. *IEEE Trans. Signal Process.*, 42(11):2915–2925, 1994.

[46] M. Unser and J. Zerubia. A generalized sampling theory without band-limiting constraints. *IEEE Trans. Circuits Syst. II.*, 45(8):959–969, 1998.

[47] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Process.*, 50(6):1417–1428, 2002.

[48] G. N. Watson. *A Treatise on the Theory of Bessel Functions*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1995. Reprint of the second (1944) edition.