Found Comput Math
DOI 10.1007/s10208-015-9276-6

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics



# Generalized Sampling and Infinite-Dimensional Compressed Sensing

**Ben Adcock**[1] · **Anders C. Hansen**[2]



**Abstract** We introduce and analyze a framework and corresponding method for compressed sensing in infinite dimensions. This extends the existing theory from finite-dimensional vector spaces to the case of separable Hilbert spaces. We explain why such a new theory is necessary by demonstrating that existing finite-dimensional techniques are ill suited for solving a number of key problems. This work stems from recent developments in generalized sampling theorems for classical (Nyquist rate) sampling that allows for reconstructions in arbitrary bases. A conclusion of this paper is that one can extend these ideas to allow for significant subsampling of sparse or compressible signals. Central to this work is the introduction of two novel concepts in sampling theory, the *stable sampling rate* and the *balancing property*, which specify how to appropriately discretize an infinite-dimensional problem.

**Keywords** Compressed sensing · Hilbert spaces · Generalized sampling · Uneven sections

**Mathematics Subject Classification** 94A20 · 47A99 · 42C40 · 15B52

Communicated by Emmanuel Candès.

✉ Ben Adcock
ben_adcock@sfu.ca

[1] Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

[2] DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Rd, Cambridge CB3 0WA, UK

Published online: 20 August 2015

## 1 Introduction

Compressed sensing (CS) has, with little doubt, been one of the great successes of applied mathematics in the last decade [16,20,21,23,25,26,33]. It allows one to sample at rates dramatically lower than conventional wisdom suggests—namely the Nyquist rate—provided the signal to be recovered is sparse in a particular basis, and the sampling vectors are incoherent.

However, the standard theory of CS is finite dimensional. It concerns the recovery of vectors in some finite-dimensional space (usually $\mathbb{R}^N$ or $\mathbb{C}^N$) whose nonzero entries with respect to a particular basis are small in number in comparison with $N$. Herein lies a problem. Real-world signals are often analog, or continuous-time, and thus are modeled more faithfully in infinite-dimensional function spaces [14]. Any finite-dimensional model may therefore not be well suited to such problems.

Although this issue has been quite widely recognized [24,56,60], there have been few attempts made thus far to extend the existing theory to infinite-dimensional models (see Sect. 1.4). The purpose of this paper is to provide such a generalization.

To do this, we move away from the usual matrix-vector model and consider the following scenario. A signal $f$ is viewed as an element of a separable Hilbert space $\mathcal{H}$, and its measurements are modeled as a sequence of linear functionals $\zeta_j : \mathcal{H} \to \mathbb{C}$, $j \in \mathbb{N}$, giving rise to the countable collection

$$\zeta_1(f), \zeta_2(f), \zeta_3(f), \dots \tag{1.1}$$

of samples of $f$. Suppose now that the signal $f$ is sparse or compressible in some orthonormal basis $\{\varphi_j\}_{j\in\mathbb{N}}$ of $\mathcal{H}$. The main questions we address in this paper are the following: can we recover $f$ by subsampling from the collection (1.1), and if so, how can this realized by a numerical algorithm? In doing so, we obtain a framework for so-called *infinite-dimensional* compressed sensing in Hilbert spaces, valid for a large class of sampling schemes $\{\zeta_j\}_{j\in\mathbb{N}}$ and reconstruction bases $\{\varphi_j\}_{j\in\mathbb{N}}$.

This work stems from recent developments in classical sampling of signals. In [2,3,5], a sampling theory, known as *generalized sampling*, was introduced for stable reconstructions of signals in arbitrary bases $\{\varphi_j\}_{j\in\mathbb{N}}$ from their samples (1.1) (see Sects. 1.3 and 3 for further details). The contribution of this paper is a continuation of this work in which sparsity is exploited to allow for substantial subsampling.

### 1.1 An Example

Magnetic resonance imaging (MRI) was one of the original motivations for CS [20]. Developed extensively by the work of Lustig et al. [46], the application of CS in MRI is now a subject of intensive research.

However, the MRI problem is inherently infinite dimensional, as are a number of related applications, such as X-ray CT and electron microscopy [8]. In these problems, an image, modeled as a function $f \in \mathcal{H} = L^2(-1, 1)^2$, is measured by taking pointwise samples of its *continuous* Fourier transform. If the samples are assumed to be taken on the usual Cartesian lattice, then the collection of measurements $\{\zeta_j(f)\}_{j\in\mathbb{N}}$

are precisely the continuous Fourier coefficients of $f$. To put this example into the above formulation, one usually assumes that $f$ is approximately sparse in an appropriate orthonormal wavelet basis $\{\varphi_j\}_{j\in\mathbb{N}}$.

On the other hand, the standard means of applying CS techniques in this setting is based on discretization. Namely, one replaces $f$ by a finite vector (or array) of pixel values and replaces the continuous Fourier transform by its discrete analog [34]. This gives a finite-dimensional recovery problem amenable to standard CS tools. Yet, as we explain further in Sect. 2.1, modeling the inherently infinite-dimensional problem in this way can quite easily lead to problems due to samples mismatch, even in simple examples (see Sect. 2.1). Moreover, such approaches are also susceptible to the well-known "inverse crime" [39].

Note that the above continuous/infinite-dimensional formulation of the MRI reconstruction problem has recently been employed by Guerquin-Kern et al. [38,39] (for earlier work, see Fessler et al. [58]). However, there are currently no recovery guarantees for this problem. The general results we prove in this paper seek to address this gap. For a demonstration of the advantages of the continuous formulation in electron microscopy, we refer to [53].

This aside, the MRI problem also illustrates another aspect of this paper. In many problems of interest, the samples (1.1) are *fixed* and cannot be altered. In MRI, this is due to the particular design of the physical scanner. Although much of research in CS has been devoted to the topic of designing good sampling systems [16,33], for many problems one does not necessarily have this luxury. Thus we require theorems and techniques for infinite-dimensional CS that allows one to work with fixed measurements.

### 1.2 Discretizing Infinite-Dimensional Problems

At this stage, it is worth asking whether or not a new framework is needed. In order to solve the above problem, one must at some stage discretize. It may therefore seem plausible that finite-dimensional CS techniques and theory could be applied once one had restricted the problem from the Hilbert space $\mathcal{H}$ to a suitable finite-dimensional space. If $f$ is sparse in an, albeit countably infinite, basis $\{\varphi_j\}_{j\in\mathbb{N}}$, it might seem reasonable to treat the corresponding recovery problem using solely existing finite-dimensional CS tools.

In some limited cases, this may be possible. However, as we discuss in Sect. 2, in general this problem cannot be tackled in such a way. Indeed, 'discretizing' the problem so as to make it amenable to computations is a subtle process (see Sect. 2). The most obvious discretizations may well destroy the original structure of the problem. This means that exact (or, more generally, stable) recovery may not be possible since the key structure that allows for subsampling is not carried over to the discretization

In this paper, we provide new techniques for discretizing the infinite-dimensional reconstruction problem. Loosely speaking, these are based on the following guiding principle: *Seek to retain as much of the infinite-dimensional structure and key properties of the original problem as possible when discretizing.* By following this principle, we obtain a framework for infinite-dimensional CS that overcomes these potential

issues. Specifically, we first devise an appropriate infinite-dimensional formulation of the problem and then truncate carefully in the second step. This leads to a finite-dimensional problem which retains the key features of the original problem, but which can be solved numerically.

It is worth noting that this principle is quite general and is by no means unique to this particular problem. Whilst found in many areas of numerical analysis, most relevantly for this paper it was recently employed in [40,41] to solve the long-standing computational spectral problem. A number of ideas in this article stem from [41] and the contributions of this paper may be viewed as a continuation of this work. Note that similar versions of this principle have also been advanced by Stuart et al. for solving inverse problems [57].

### 1.3 Generalized Sampling (GS)

The framework we propose in this paper is based on recent developments in classical sampling known as *generalized sampling* (GS) [2,3,5] (see also [10,29,43,62,63]). GS is a novel sampling theory which incorporates the key issues of approximation and stability via the so-called *stable sampling rate* [5]. This allows for guaranteed recovery of any signal in an arbitrary basis from a collection of its samples in a manner which is both numerically stable and, in a certain sense, optimal. In this paper, we extend this work by showing that when the signal to be reconstructed is sparse or compressible, reconstruction can also be performed with significant subsampling. We refer to the corresponding technique as *generalized sampling with compressed sensing* (GS–CS).

One important instance of both GS and GS–CS is recovery from Fourier samples (the MRI problem, in particular). Although classical Fourier analysis allows for reconstruction in terms of an infinite series of complex exponentials, the slow convergence of this series and the appearance of the Gibbs phenomenon [44] renders such approximations impractical. Nonetheless, in many circumstances it is well known that the given signal can be well represented (i.e., it is sparse or has rapidly decaying coefficients) in a new basis of functions, be they splines, wavelets, curvelets, etc. [31]. GS and GS–CS allow one to reconstruct in such a basis in manner that is both accurate, numerically stable and, in the case of the latter, amenable to subsampling when the signal is sparse or compressible.

### 1.4 Relation to Other Work and Contributions of the Paper

There have been a number of recent attempts to generalize CS to infinite dimensions. In [30,47,48], an infinite-dimensional CS approach is described for analog-to-digital conversion based on a union of subspaces signal model, which is related to previous research of finite rates of innovation [14,27,64]. In [42], the approach of [30,47,48] was applied to inverse and ill-posed problems. The application of CS techniques to the recovery of functions was considered by Rauhut and Ward. By devising an appropriate sampling distribution to ensure a restricted isometry property, they prove near-optimal recovery guarantees for the reconstruction of sparse sums of Legendre polynomials [52] or spherical harmonics [51]. Note that the sampling mechanism in this work is

limited to pointwise samples of the function itself, as opposed to its Fourier transform. Hence, it is not applicable to the MRI problem, for example.

Besides medical imaging, infinite-dimensional problems are found in other applications including radar, sonar and remote sensing [56]. Use of standard, finite-dimensional CS in such problems is susceptible to gridding error (also known as basis mismatch) [22]. Although the setting here is somewhat different to that which we consider in this paper, the same issue arises: Poor discretization of an infinite-dimensional problem leads to inferior reconstructions. Recent works [32,60] have sought to address this by applying essentially the same guiding principle followed in this paper. Closely related to this is the work of Candès and Fernandez-Granda on super-resolution [17], wherein an analog model is employed for reconstruction from low-bandwidth Fourier samples using convex optimization.

Note that most of the above works describe CS in infinite dimensions for some particular class of problem and do not address the general scenario of arbitrary sampling schemes $\{\zeta_j\}_{j\in\mathbb{N}}$ and reconstruction bases $\{\varphi_j\}_{j\in\mathbb{N}}$ in a Hilbert spaces $\mathcal{H}$. Our GS–CS framework aims to do this. It is therefore a natural way to extend standard finite-dimensional CS. In this light, it should not come as a surprise that certain finite-dimensional CS results (specifically, those related to the incoherence-based theory of CS [18,34]) become corollaries of our main theorems.

Having said this, we remark that the intent of this paper is to present only a first step toward extending CS fully to the Hilbert space setting. Finite-dimensional CS has been the subject of many papers over the last decade (see [34] for a comprehensive treatment). Unsurprisingly, there are some key aspects of the finite-dimensional theory whose extensions to infinite dimensions are either beyond the scope of this paper or currently unknown. See Sects. 6 and 10 for further details.

**Notation** In the remainder of the paper, we use the following notation. If $l^2(\mathbb{N})$ is the standard space of square summable sequences in $\mathbb{C}$, we write $\|\cdot\|$ for its standard norm (all other norms will be specified). Let $\{e_j\}_{j\in\mathbb{N}}$ be the natural basis on $l^2(\mathbb{N})$, and, for $\Gamma \subset \mathbb{N}$ define $P_\Gamma$ to be the orthogonal projection onto $\mathrm{cl}(\mathrm{span}\{e_j : j \in \Gamma\})$. If $\Gamma = \{1, \dots, N\}$, then we simply write $P_N$. We shall also use similar notation when working in finite-dimensional vector spaces. Finally, if $U$ is a bounded operator on either $l^2(\mathbb{N})$ or $\mathbb{C}^N$, we write $\|U\|$ for its induced operator norm.

## 2 The Need for an Infinite-Dimensional Framework

Consider the following simple model problem, which will form the primary example throughout this paper:

**Problem 2.1** Suppose that $f \in L^2(\mathbb{R})$ has support contained in $[-1, 1]$, and let $\{\varphi_j\}_{j\in\mathbb{N}}$ be the orthonormal basis of Haar wavelets on $L^2(-1, 1)$. Define

$$\zeta_j(f) = \mathcal{F}f(j/(2T)), \quad j \in \mathbb{Z}, \tag{2.1}$$

to be the Fourier coefficients of $f$, where $\mathcal{F}f$ denotes the Fourier transform of $f$, and $T \geq 1$ is arbitrary. Throughout, we shall take $T = 1$. Assume that $f$ is sparse, or

compressible, in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$. Then, the problem is to recover $f$ from a small number of the measurements (2.1).

Recall that $f$ can be recovered from the infinite collection $\{\zeta_j(f)\}_{j\in\mathbb{Z}}$ via its Fourier series. However, since $f$ is known to be sparse in the Haar wavelet basis $\{\varphi_j\}_{j\in\mathbb{N}}$, this raises the possibility of reconstructing $f$ from a small and finite number of its samples, which is the focus of this paper.

## 2.1 The Discrete Model

Let us consider the simplest possible case, where $f = \chi_{[0,1/2)} - \chi_{[1/2,1)}$ is the Haar mother wavelet and is by definition one sparse in the Haar basis. The usual approach to recover $f$ used in applications such as sparse MRI [46] involves two steps. First, one discretizes $f$ to an equispaced grid of $2N$ points and replaces the infinite collection of samples (2.1) by the finite vector

$$y = \zeta(f) = \{\zeta_j(f) : j = -N+1, \dots, N\}, \quad N \in \mathbb{N}. \tag{2.2}$$

Second, one uses a combination of the discrete Fourier and discrete wavelet transforms (DFT and DWT, respectively) to formulate the corresponding discrete recovery problem. Specifically, let $U_{\mathrm{df}}, V_{\mathrm{dw}} \in \mathbb{C}^{2N\times 2N}$ be the matrices of these transforms. The classical discrete approximation to the problem of inverting the Fourier transform is then given by $y = U_{\mathrm{df}}x$, where $x$ is a vector approximating the grid values of $f$. Since $f$ is sparse in the Haar basis, one may think that $x_0 = V_{\mathrm{dw}}x$ is also sparse, and therefore, we could recover $f$ perfectly from only relatively few of its samples $y = \zeta(f)$ by using standard CS techniques. In particular, if $\Omega \subset \{1, \dots, 2N\}$, $|\Omega| = m < 2N$ is chosen uniformly at random, one solves

$$\min_{\eta\in\mathbb{C}^{2N}} \|\eta\|_{l^1} \text{ subject to } P_\Omega U_{\mathrm{df}} V_{\mathrm{dw}}^{-1}\eta = P_\Omega y. \tag{2.3}$$

If $\xi$ is a minimizer of this problem, then one might expect $\xi$ to agree with the vector $x_0$ with high probability, and hence, we could recover $f$ via $x = V_{\mathrm{dw}}^{-1}x_0$.

To test whether this is the case or not, we consider an example with $2N = 256$ and $m = 130$, i.e., we use nearly 50 % of the measurements in the range $-N+1, \dots, N$. Figure 1 displays the reconstruction of $f$ by $f_{N,m}$, where $f_{N,m} = \sum_{j=1}^{2N} \xi_j\varphi_j$ and $\xi$ is a minimizer of (2.3). As is evident, $f$ is not recovered anywhere near exactly, and the reconstruction $f_{N,m}$ computed from (2.3) commits a rather large error. This occurs despite the fact that $f$ is one-sparse Haar wavelet basis, and we used $m = 130$ Fourier samples.

### *2.1.1 The DFT Destroys Sparsity*

The source of the poor reconstruction in (2.3) is the discretization used to approximate the continuous Fourier transform of a function by the discrete Fourier transform a vec-

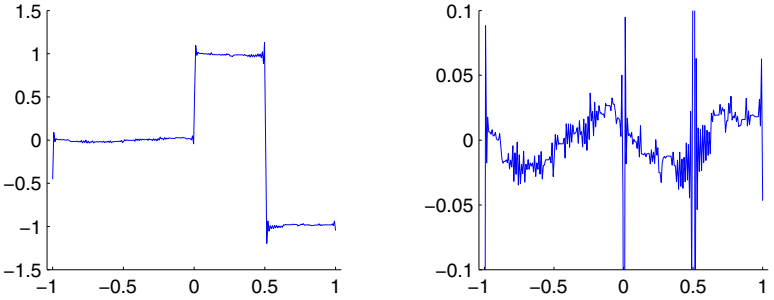


**Fig. 1** Graphs of $f_{N,m}(t)$ (*left*) and $f(t) - f_{N,m}(t)$ (*right*) against $t$ for $2N = 256$ and $m = 130$, where $f_{N,m}(t) = \sum_{j=1}^{2N} \xi_j \varphi_j(t)$ and $\xi = \{\xi_j\}_{j=1}^{2N}$ is a minimizer of (2.3)

tor. The result is a mismatch between the data, which are continuous Fourier samples, and their modeling as discrete Fourier samples.

To explain in more detail, consider the matrix $U_{\mathrm{df}}^{-1}$, which maps the vector of Fourier coefficients $\zeta(f)$ of a function $f$ to a vector consisting of pointwise values on an equispaced $2N$-grid in $[-1, 1]$. This mapping commits an error: For an arbitrary function $f$, the result is only an *approximation* to the grid values of $f$. Indeed, consider the vector $x \in \mathbb{C}^{2N}$ defined by $U_{\mathrm{df}} x = \zeta(f)$. It is straightforward to see that $x$ consists precisely of the values of the function

$$f_N(t) = \frac{1}{2} \sum_{j=-N+1}^{N} \mathcal{F} f(j/2) \mathrm{e}^{\pi \mathrm{i} j t}, \tag{2.4}$$

on the equispaced $2N$-grid. This function is nothing more than the truncated Fourier series of $f$, and therefore, the approximation resulting from modeling the continuous Fourier transform with $U_{\mathrm{df}}$ is equivalent to replacing a function $f$ by its partial Fourier series $f_N$.

Now consider the discrete wavelet transform $x_0 = V_{\mathrm{dw}} x \in \mathbb{C}^{2N}$ of $x$. The right-hand side of the equality constraint in (2.3) then reads $P_\Omega U_{\mathrm{df}} V_{\mathrm{dw}}^{-1} x_0,$, and therefore, for the method (2.3) to be successful we require $x_0 = V_{\mathrm{dw}} x$ to be a sparse vector. However, this can never happen. Sparsity of $x_0$ is equivalent to stipulating that the partial Fourier series $f_N$ be sparse in the Haar wavelet basis. Yet the function $f_N$ consists of smooth complex exponentials and thus cannot have a sparse representation in a basis of piecewise smooth functions. In other words, by using the DFT to discretize the problem, we have destroyed some of its key structure—namely the sparsity—thus going against the guiding principle of Sect. 1.2.

### *2.1.2 The DFT Leads to the Gibbs Phenomenon*

Given that $\eta$ is not sparse in Haar wavelets, suppose now, as an exercise, we forgo any subsampling and let $m = 2N$. The problem (2.3) then has a unique solution $\eta$. However, the entries of $\eta$ are not the Haar wavelet coefficients of $f$, but rather those

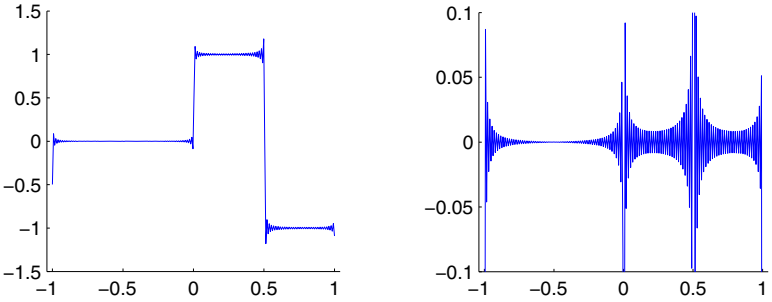


**Fig. 2** Graphs of $f_N(t)$ (*left*) and $f(t) - f_N(t)$ (*right*) against $t$ for $2N = 256$

of $f_N$. Thus, by solving (2.3) (both with and without subsampling) we are not actually computing Haar wavelet coefficients of $f$, but those of the partial Fourier series $f_N$ instead. Thus, we cannot expect to obtain a better (i.e., more accurate) reconstruction of $f$ than $f_N$.

Unfortunately, since $f$ is piecewise smooth, $f_N$ is a very poor approximation to $f$. As $N \to \infty$, $f_N$ does not converge uniformly to $f$ and only converges very slowly in the weaker $L^2$ norm. One also witnesses the Gibbs phenomenon, with its characteristic $\mathcal{O}(1)$ oscillations, near each jump in $f$. These effects are visualized in Fig. 2. The fact that (2.3) leads to a Haar wavelet approximation to $f_N$, as opposed to $f$, can be observed by noting the similarities between the left panels in Figs. 1 and 2, respectively.

### *2.1.3 Relation to the Inverse and Wavelet Crimes*

The poor reconstruction seen above is due to a mismatch between the data and the model. Had the data been simulated using the DFT, then no such problems would have occurred, and one would have seen perfect recovery. However, this improvement is artificial and an example of the well-known inverse crime [39]. That is to say, unrealistic simulation of the data leads to spuriously good reconstructions, but when realistic data are used (i.e., arising from the continuous Fourier transform) such as in the above examples, the reconstruction quality substantially declines.

Besides this, the above approach also incurs the wavelet crime. Recall that the discrete wavelet transform is an infinite-dimensional operator that takes as input the coefficients of the expansion of the function corresponding to the scaling function. The output of the discrete wavelet transform are the wavelet coefficients as well as the scaling coefficients corresponding to the next level. In the discretization model above, the vector $x$ contains approximate pointwise samples of the function $f$. Thus, at best we can interpret $x$ as the coefficient vector corresponding to an expansion using the scaling function of the Haar wavelet (which is the step function). However, in all other cases of Daubechies wavelets (where the scaling is different from the step function), the vector $x_0 = V_{\mathrm{dw}}x$ is just an approximation to the actual wavelet coefficients of $f$. This effect is referred to as the "wavelet crime" by Strang and Nguyen [55, p. 232].

### 2.2 Discretization Via Finite Sections

Since discretization via the DFT caused the above problems, it may at first seem reasonable to overcome these issues by replacing the DFT-DWT matrix $U_{\mathrm{df}} V_{\mathrm{dw}}^{-1}$ by the new measurement matrix

$$U_N = \begin{pmatrix} \zeta_1(\varphi_1) & \cdots & \zeta_1(\varphi_{2N}) \\ \vdots & \ddots & \vdots \\ \zeta_{2N}(\varphi_1) & \cdots & \zeta_{2N}(\varphi_{2N}) \end{pmatrix}. \tag{2.5}$$

If $\Omega \subset \{1, \ldots, 2N\}$, $|\Omega| = m$, is chosen uniformly at random, one now finds a minimizer $\xi$ to the problem

$$\min_{\eta \in \mathbb{C}^{2N}} \|\eta\|_{l^1} \text{ subject to } P_\Omega U_N \eta = P_\Omega \zeta(f), \tag{2.6}$$

and forms the reconstruction $f_{N,m} = \sum_{j=1}^{2N} \xi_j \varphi_j$ (note that in this case we have, for convenience, reindexed the Fourier samples $\{\zeta_j\}_{j \in \mathbb{N}}$ over $\mathbb{N}$ rather than $\mathbb{Z}$). Clearly this approach preserves the sparsity of the original problem, unlike (2.3).

We now consider an example of (2.6) where $N = 384$ and $f = \sum_{j=1}^{2N} \alpha_j \varphi_j$ is such that $|\mathrm{supp}(f)| = |\{\alpha_j : \alpha_j \neq 0\}| = 5$. In Fig. 3, we display the reconstruction given by (2.6) using $m = 760$ samples. Unfortunately, this reconstruction is still poor. Despite using 98 % of its Fourier samples in the range $1, \ldots, 2N$, the function $f$ is recovered nowhere near exactly by (2.6). Repeating the experiment with different random draws of $\Omega$ yields very similar results.

Why does this happen? On the face of it, (2.6) looks like a standard CS problem: A measurement matrix is formed by taking inner products of the first $2N$ elements of two orthonormal bases (the Fourier and Haar wavelet bases in this case) and then one subsamples randomly from its rows. The issue lies with the fact that $U_N$ is not an isometry: In fact, its condition number in this instance is at least $10^{16}$ in magnitude.

This lack of isometric structure can be traced to the underlying infinite dimensionality of the problem. In general, a matrix $U_N$ of the form (2.5) will only be an isometry if and only if the $N$ basis elements span the same space. This is clearly not the case in (2.5), where the sampling and reconstruction bases consist of the first $N$ (smooth) complex exponentials and (piecewise constant) Haar wavelets, respectively. In other words, simply thinking (since $f$ has only finitely many nonzero Haar wavelet coefficients) that the problem can be viewed as a finite-dimensional one in $\mathbb{C}^N$ is problematic. As noted above, $U_N$ is not an isometry, whereas the "infinite change of basis matrix"

$$U = \begin{pmatrix} \zeta_1(\varphi_1) & \zeta_1(\varphi_2) & \cdots \\ \zeta_2(\varphi_1) & \zeta_2(\varphi_2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \tag{2.7}$$

formed by combining the full countably infinite bases is. It is precisely the loss of this structure when "discretizing" $U$ via $U_N$ that is the source of the poor reconstruction observed above.

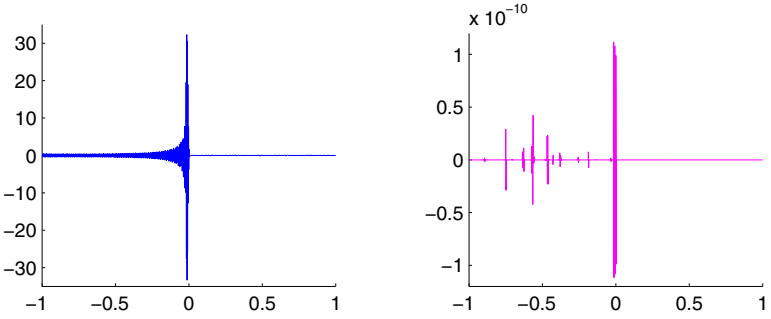


**Fig. 3** Errors $f(t) - f_{N,m}(t)$ against $t$ (*left*) and $f(t) - g_{\tilde{N},\tilde{m}}(t)$ (*right*). Note that $f_{N,m}$ requires $m = 760$ samples, whereas $g_{\tilde{N},\tilde{m}}$ requires only $\tilde{m} = 50$ samples

### 2.3 A New Approach

With these two examples in mind, in the remainder of this paper we introduce a new approach that allows for successful recovery by discretizing in a way that preserves the two key properties highlighted (sparsity and the isometric structure). This approach will be introduced formally in the next sections. However, we first give a brief numerical demonstration. Let $f(t) = \sum_{j=1}^{M} \alpha_j \varphi_j(t)$ be as in Sect. 2.2, and let $U$ be given by (2.7). For $\tilde{N} \in \mathbb{N}$, we now choose $\Omega \subset \{1, \dots, \tilde{N}\}$, $|\Omega| = \tilde{m}$, uniformly at random and compute a minimizer $\xi$ to

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega P_{\tilde{N}} U P_M \eta = P_\Omega y, \quad y = \{\zeta_1(f), \zeta_2(f), \dots\}, \tag{2.8}$$

where $M \in \mathbb{N}$, and let $g_{\tilde{N},\tilde{m}} = \sum_{j=1}^{M} \xi_j \varphi_j$ be the reconstructed approximation to $f$. In Fig. 3, we apply this algorithm with $\tilde{N} = 1351$, $\tilde{m} = 50$ and $M = 768$. Note the significant improvement over the approach of Sect. 2.2. In particular, when averaged over 50 trials, the error $\|f - g_{\tilde{N},\tilde{m}}\|_{L^2}$ is found to be roughly $1.15 \times 10^{-11}$ in comparison with $\|f - f_{N,m}\|_{L^2} \approx 2.43$ for the previous approach. Moreover, this new reconstruction uses fewer than 4 % of the Fourier coefficients, whereas $f_{N,m}$ used roughly 98 % and was still a poor approximation.

The purpose of remainder of this paper is to explain precisely why (2.8) leads to such a marked improvement. As we explain, the key to this is the judicious choice of the parameter $\tilde{N}$, which is selected according to what we refer to as the *balancing property* (see Sect. 5.1). This property guarantees a faithful discretization of the operator $U$, which, unlike in (2.5), ensures that the isometry structure is approximately preserved when discretizing with the finite matrix $P_{\tilde{N}} U P_M$.

We note in passing that the notion of keeping the structure when discretizing infinite-dimensional operators is of course not new, and the delicate issues are indeed widely recognized. See, for example [11,12], although this framework is not specific to the CS setting.

## 3 Generalized Sampling: Signal Recovery in Infinite Dimensions

Suppose that $\mathcal{H}$ is a separable Hilbert space over $\mathbb{C}$. Let $\{\varphi_j\}_{j\in\mathbb{N}}$ be an orthonormal basis, $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$ and suppose that we have access to the countable collection of samples

$$\zeta_1(f), \zeta_2(f), \zeta_3(f), \ldots, \tag{3.1}$$

where $\zeta_j : \mathcal{H} \to \mathbb{C}$ are continuous linear functionals on $\mathcal{H}$. Our aim is to recover the coefficients $\{\alpha_j\}_{j\in\mathbb{N}}$, and therefore $f$, from the samples (3.1). Before discussing how to recover infinite-dimensional sparse or compressible signals, it is first necessary to address the classical case where no sparsity is assumed. Only once this problem has been solved can one consider the issue of subsampling.

To do this, we shall use the technique of *generalized sampling* (GS) [2–5], which we now recap. Under some assumptions on $\{\zeta_j\}_{j\in\mathbb{N}}$ (e.g., each $\zeta_j$ is continuous and $\zeta_j(\cdot) = \langle \cdot, \psi_j \rangle$ where $\{\psi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{H}$), we can view the full recovery problem as the infinite system of equations

$$U\alpha = \zeta(f), \tag{3.2}$$

where $\alpha = \{\alpha_1, \alpha_2, \ldots\}, \zeta(f) = \{\zeta_1(f), \zeta_2(f), \ldots\}$ and $U$ is the infinite measurement matrix (2.7). Clearly, if we were able to invert $U$, and provided we had access to all samples of $f$, then we could recover $\alpha$ (and hence $f$) exactly. Of course, this is never the case in practice and so we instead consider truncations of (3.2) and look to compute approximations $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_N$ to the first $N$ coefficients of $\alpha$.

### 3.1 Uneven Sections and Generalized Sampling (GS)

The most obvious approach for discretizing (3.2) follows by taking a *finite section* $U_N$, i.e., the leading $N \times N$ submatrix of $U$ and solving the resulting $N \times N$ linear system. However, finite sections can be extremely problematic in practice, leading in general to both poor approximations and numerical instability [2,5]. As noted in Sect. 2.2, this is due to the loss of isometric structure when discretizing $U$ via $U_N$.

Fortunately, there is a simple way to overcome the problems associated with the finite section method, based on taking rectangular, as opposed to square, sections of $U$ (see [2,35] and references therein). In particular, we replace the finite section method with

$$A\tilde{\alpha}^{[M]} = P_M U^* P_N \zeta(f), \quad A = P_M U^* P_N U P_M, \tag{3.3}$$

where $M \in \mathbb{N}$ (the number of coefficients $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_M$ computed) is appropriately chosen (typically $M \leq N$). The result is known as *generalized sampling* (GS). Note that $A = (P_N U P_M)^* P_N U P_M$, where $P_N U P_M$ is the $N \times M$ *uneven* section of $U$.

The main idea now is to allow $M$ and $N$ to vary independently of each other—in particular, selecting $M \leq N$ sufficiently small—to ensure a numerically stable and accurate reconstruction of the first $M$ coefficients $\alpha_1, \ldots, \alpha_M$. To this end, the main theorem proved in [2,5] is as follows:

**Theorem 3.1** *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and suppose that $\alpha \in l^2(\mathbb{N})$ satisfies $U\alpha = \zeta(f)$ for some $\zeta(f) \in l^2(\mathbb{N})$. Let $N_0$ be the least $N$ such that $C_{N,M} < 1$, where*

$$C_{N,M} = \|P_M - P_M U^* P_N U P_M\|. \tag{3.4}$$

*Then, there is a unique solution $\tilde{\alpha}^{[M]}$ to (3.3), and we have the sharp bound*

$$\|\alpha - \tilde{\alpha}^{[M]}\| \leq \frac{1}{\sqrt{1 - C_{N,M}}} \|P_M^{\perp} \alpha\|. \tag{3.5}$$

It can be shown that the quantity $C_{N,M}$—a measure of how close the uneven section $P_N U P_M$ is to an isometry—tends to zero as $N \to \infty$, for any fixed $M$. Thus, $\tilde{\alpha}^{[M]}$ can be made arbitrarily close to $P_M \alpha$ (the best approximation to $\alpha$ from $P_M(l^2(\mathbb{N}))$) by varying $N$ suitably. Furthermore, the method is also stable, since the condition number of the matrix $A$ scales like $1/\sqrt{1 - C_{N,M}}$ [5]. That is to say, precisely the same quantity that ensures that the isometric structure of $U$ is approximately retained by the discretization $P_N U P_M$ also guarantees accuracy and numerical stability of the reconstruction.

In practice, we need a way in which to quantify the required scaling of $N$ and $M$. To do this, in [5] the *stable sampling rate*

$$\Theta(M; \theta) = \min \left\{ N \in \mathbb{N} : C_{N,M} < 1 - \theta^2 \right\}, \quad \theta \in (0, 1), \tag{3.6}$$

was introduced. In particular, sampling at a rate $N \geq \Theta(M; \theta)$ ensures that $\sqrt{1 - C_{N,M}} \geq \theta$, and therefore stability and accuracy of $\tilde{\alpha}^{[M]}$ up to the magnitude of $\theta$.

*Remark 3.2* Note that $\Theta(M; \theta)$ can be computed numerically [5], since $C_{N,M}$ is just the 2-norm of an $M \times M$ matrix (see 3.4). Hence, the conditions of Theorem 3.1 can be verified *a priori* via a straightforward calculation. Analytical bounds are also possible [2,3,5,6]. In [6], it was proved that $\Theta(M; \theta) \sim c(\theta)M$ for Fourier sampling with wavelets as the reconstruction system, the principal example of this paper. Typically, $c(\theta)$, whilst greater than one, is not too large. However, any attempt to sample much below this rate necessarily fails. In [6], it was also shown that setting $N = M$ (this corresponds to the finite section), or in fact $N = \tilde{c}M$ for any $\tilde{c}$ less than some critical threshold $c_0 > 1$, leads to exponential instability and divergence. For some earlier related results, see [35].

### 3.2 Generalized Sampling with Fourier Samples

Suppose $\{\zeta_j(f)\}_{j \in \mathbb{Z}}$ correspond to the Fourier samples (2.1). In practice, the Fourier series (2.4) based on the first $N$ samples leads to a very poor reconstruction of $f$. However, suppose now we know another basis $\{\varphi_j\}_{j \in \mathbb{N}}$ in which $f$ is well represented. We can then apply GS to obtain an improved reconstruction

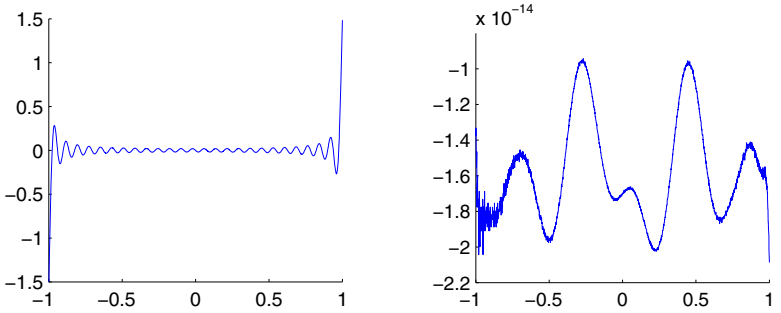


**Fig. 4** Errors $f(t) - f_N(t)$ (*left*) $f(t) - f_{N,M}(t)$ (*right*) against $t$ for $N = 51$ and $M = 12$

$$f_{N,M} = \sum_{j=1}^{M} \tilde{\alpha}_j \varphi_j, \tag{3.7}$$

in this basis, where the coefficients $\tilde{\alpha}_j$ are the solution of (3.3). The key point is that, if we know that $f$ is well represented in $\{\varphi_j\}_{j\in\mathbb{N}}$, then we can recover $f$ optimally (up to a multiplicative constant) in terms of the first $M$ basis function $\varphi_1, \ldots, \varphi_M$ using only its first $N$ Fourier coefficients.

To demonstrate, consider the function $f(t) = t^5 \mathrm{e}^{-t}$, $t \in [-1, 1]$. Suppose we can sample the Fourier coefficients of $f$, namely $\{\mathcal{F}f(j/2)\}_{j\in\mathbb{Z}}$. To reconstruct $f$ from these samples, we will consider two different techniques. First, we test the truncated Fourier series $f_N$ defined in (2.4). Due to the fact that $f$ is not periodic we cannot expect rapid convergence of $f_N$ to $f$. However, GS allows us to reconstruct in any basis. Thus, (due to analyticity of $f$) we will choose the reconstruction basis $\{\varphi_j\}_{j\in\mathbb{N}}$ consisting of orthonormal Legendre polynomials on $[-1, 1]$ and define $f_{N,M}$ as in (3.7). In Fig. 4, we have displayed the errors $f - f_N$ and $f - f_{N,M}$. Note that both reconstructions, $f_N$ and $f_{N,M}$, use the same samples, yet the improvement of $f_{N,M}$ over $f_N$ is significant: An $\mathcal{O}(1)$ error is reduced to roughly machine precision. We remark that for this choice of reconstruction basis the stable sampling rate $\Theta(M; \theta)$ is quadratic in $M$ [3]. Moreover, a lower scaling (in particular, $N = M$) necessarily results in extreme ill-conditioning [9].

## 4 Infinite-Dimensional Compressed Sensing

We are now in a position to introduce sparsity and subsampling into the reconstruction problem. The infinite-dimensional compressed sensing approach we introduce next is based on the ideas of generalized sampling, and we refer to it as *generalized sampling with compressed sensing* (GS–CS).

### 4.1 Sparsity and Compressibility in Infinite Dimensions

First, we need to notions of sparsity and compressibility in infinite dimensions. We shall say that $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$ is sparse in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$ if there exists an $M \in \mathbb{N}$ such that

$$\Delta = \mathrm{supp}(f) \subset \{1, \dots, M\} \quad \mathrm{supp}(f) = \{j \in \mathbb{N} : \alpha_j \neq 0\}, \tag{4.1}$$

Note that we do not know $\Delta$, but we usually have estimates for $|\Delta|$ and $M$. If $|\Delta| = r$, we say that $f$ is $(r, M)$-sparse in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$.

It is important in this definition that the nonzero entries of $f$ are restricted to some finite bandwidth $M$. We cannot expect stable recovery from a finite number of samples if the $|\Delta|$ nonzero entries are allowed to have arbitrary locations in $\mathbb{N}$, regardless of the reconstruction algorithm used [15]. This is one of several ways in which infinite-dimensional CS differs from its finite-dimensional counterpart, and means that in practice so-called *uniform recovery* is not achievable in infinite dimensions without bandwidth restrictions. We discuss this point further in Sect. 10.

Typically, $f$ will not be perfectly sparse, and moreover, exact knowledge of the effective sparsity $|\Delta|$ and bandwidth $M$ may be lacking. In finite-dimensional CS, it is standard to address this by considering *compressible* signals, i.e., those whose $r$-term approximation error decays rapidly. In the infinite-dimensional setting, we require a slightly different notion that takes into account the bandwidth $M$. To this end, we let

$$\sigma_{r,M}(\alpha) = \min\{\|\alpha - \eta\|_{l^1} : \eta \text{ is } (r, M)\text{-sparse}\},$$

be the error of the best approximation of $f$ by a $(r, M)$-sparse vector. Loosely speaking, we shall say that $f$ is compressible if this term is small.

### 4.2 Models

Having defined sparsity, we now introduce the signal models we consider in the remainder of this paper:

(i) *Semi-infinite-dimensional model* Here we assume $f$ is either sparse with bandwidth $M$, or that

$$f = g + h, \quad \Delta = \mathrm{supp}(g) \subset \{1, \dots, M\}, \quad \mathrm{supp}(h) = \{1, \dots, M\}. \tag{4.2}$$

In other words, $f$ is $(r, M)$-compressible for some $r$ and $\sigma_{M,M}(x) = 0$. This model is semi-infinite dimensional: Although $f$ has only finite support in $\{\varphi_j\}_{j\in\mathbb{N}}$, we may draw samples from the countable collection (3.1).

(ii) *Fully infinite-dimensional model* Here we consider the significantly more general setting:

$$f = g + h, \quad \Delta = \mathrm{supp}(g) \subset \{1, \dots, M\}, \quad \mathrm{supp}(h) \subseteq \mathbb{N}. \tag{4.3}$$

This model is termed fully infinite-dimensional since the support of $f$ can be infinite.

### 4.3 Generalized Sampling with Compressed Sensing (GS–CS)

Suppose now that $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$ is sampled via $\{\zeta_j\}_{j\in\mathbb{N}}$, and let $\Omega \subseteq \mathbb{N}$ of size $|\Omega| = m \in \mathbb{N}$ be the index set of the measurements taken. We first propose the

infinite-dimensional optimization problem

$$\min_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \zeta(f), \tag{4.4}$$

where $U$ is the infinite matrix (2.7) and $\zeta(f) = \{\zeta_1(f), \zeta_2(f), \ldots\}$ is the infinite vector of samples.

Recall that GS relies on a well-posed infinite-dimensional recovery problem (3.2) before discretization can proceed. Seeking similar notions for (4.4), we are led to the following questions:

(i) How do we choose $\Omega$? Obviously there is no unique choice, but it makes sense to choose $\Omega$ uniformly at random from $\{1, \ldots, N\}$, where $N \in \mathbb{N}$ is some fixed number. This raises the question following question: How large must $N$ be?

(ii) Suppose that $\eta$ is a minimizer of (4.4) (note that $\eta$ need not be unique). How large is $\|\eta - \alpha\|$, where $\alpha$ is the infinite vector of coefficients of $f$ in the basis $\{\varphi_j\}_{j \in \mathbb{N}}$. In particular, how does $\|\eta - \alpha\|$ depend on both $m$ (the total number of samples) and $N$ (the range from which the samples are drawn)?

(iii) If $f$ is exactly sparse in $\{\varphi_j\}_{j \in \mathbb{N}}$, do we recover its coefficient vector $\alpha$ exactly (with high probability) from (4.4), and what are the conditions on $m$ and $N$ that ensure this recovery?

Suppose for the moment that we have answers to these questions. Besides special circumstances, we cannot solve (4.4) numerically; hence, we must discretize. For this, we follow the same ideas that lead to generalized sampling. Thus, we introduce a parameter $k \in \mathbb{N}$ and consider the finite-dimensional optimization problem

$$\min_{\eta \in \mathbb{C}^M} \|\eta\|_{l^1} \text{ subject to } P_\Omega U P_k \eta = P_\Omega \zeta(f). \tag{4.5}$$

We refer to this as *generalized sampling with compressed sensing* (*GS–CS*).

This formulation leads to a further set of questions:

(iv) When will (4.5) have a solution? Note that (4.5) need not have a solution for all $k$, since $P_\Omega \zeta(f)$ need not be in the range of $P_\Omega U P_k$ (although, as we shall show, this is always the case for sufficiently large $k$). However, this raises the following question: Will solutions of (4.5) converge to solutions of (4.4) as $k \to \infty$?

(v) If $f$ is not sparse but compressible, how large is the error $\|\eta - \alpha\|$ when $\eta$ is a solution to (4.5) and $\alpha$ is the vector of coefficients of $f$? In particular, if $f$ belongs to either of the models (4.2) or (4.3), can $\|\eta - \alpha\|$ be bounded above in terms of $\sigma_{r,M}(f)$ for appropriate $r$ and $M$?

Answers to these questions will be provided in Sect. 6, where we state the main results of this paper.

*Remark 4.1* The reader may at first be concerned that replacing (4.5) by (4.4) is problematic since there is no possibility with (4.5) of recovering signals for which $\text{supp}(f)$ is not fully contained in the range $\{1, \ldots, k\}$. However, recall that we cannot, with any algorithm, expect stable recovery of coefficients with arbitrary locations in $\mathbb{N}$. So there is no issue with replacing (4.4) by (4.5) provided $k$ is chosen sufficiently large (see question (iv)).

## 5 Additional Notation and Definitions

We now introduce some additional notation that will be used in the remainder of this paper. From now on, let $\mathcal{H} = l^2(\mathbb{N})$, and if $\xi \in \mathcal{H}$ and $j \in \mathbb{N}$, then write $\xi(j) = \xi_j = \langle \xi, e_j \rangle$. For $\Gamma \subset \mathbb{N}$, we denote the natural embedding operator from $l^2(\Gamma)$ to $\mathcal{H}$ by $\iota_\Gamma$. Note that $\iota_\Gamma^* \eta = \eta|_\Gamma$ for $\eta \in \mathcal{H}$. For any vector $\xi \in \mathcal{H}$ we write $\mathrm{supp}(\xi) = \{j \in \mathbb{N} : \xi(j) \neq 0\}$, and we define the sign $\mathrm{sgn}(\xi) \in l^\infty(\mathbb{N})$ of $\xi \in l^\infty(\mathbb{N})$ as follows:

$$\mathrm{sgn}(\xi)(j) = \begin{cases} \xi(j)/|\xi(j)| & \text{if } \xi(j) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

For an operator $U \in \mathcal{B}(\mathcal{H})$, we let

$$\upsilon(U) = \sup_{i,j \in \mathbb{N}} |u_{ij}|, \quad u_{ij} = \langle U e_j, e_i \rangle, \tag{5.1}$$

be the coherence parameter, i.e., the max norm of the operator $U$ with respect to $\{e_j\}_{j \in \mathbb{N}}$. Also, we define the maximum row norm of $U$ by $\|U\|_{\mathrm{mr}} = \sup_{i \in \mathbb{N}} \left( \sum_{j \in \mathbb{N}} |u_{ij}|^2 \right)^{1/2}$. This quantity forms a vector space norm on the vector space of all infinite matrices (although not an algebra norm). Finally, for convenience, we define the following function that will be used frequently in the exposition. For $M \in \mathbb{N}$ and $U \in \mathcal{B}(\mathcal{H})$, let $\tilde{\omega}_{M,U} : \{1, \ldots, M\} \times \mathbb{R}_+ \times \mathbb{N} \to \mathbb{N} \cup \{0\}$ be given by

$$\tilde{\omega}_{M,U}(r, s, N) = \left| \left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\ldots,M\}, |\Gamma_1| = r \\ \Gamma_2 \subset \{1,\ldots,N\}}} \left\| P_{\Gamma_1} U^* P_{\Gamma_2} U e_i \right\| > s \right\} \right|. \tag{5.2}$$

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

$$\|P_{\Gamma_1} U^* P_{\Gamma_2} U e_i\| \leq \|P_{\Gamma_1} U^* P_{\Gamma_2}\| \|P_{\Gamma_2} U e_i\| \leq \|P_M U^* P_N\| \|P_N U P_{i-1}^\perp\|,$$

where the term $\|P_N U P_{i-1}^\perp\|$ is decreasing in $i \to \infty$, we see that can be bounded above by the simpler, but admittedly less sharp quantity

$$\tilde{\omega}_{M,U}(r, s, N) \leq \min \left\{ i \in \mathbb{N} : \|P_M U^* P_N\| \|P_N U P_i^\perp\| \leq s \right\}. \tag{5.3}$$

### 5.1 The Balancing Property

For GS, the stable sampling rate (3.6) determines how to discretize the recovery problem in line with the principle of Sect. 1.2. For GS–CS, we require an analogous quantity, known as the balancing property.

**Definition 5.1** Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry and $M, |\Delta| \in \mathbb{N}$. Then $N$ and $m$ satisfy the weak balancing property with respect to $U$, $M$ and $|\Delta|$ if

$$\|P_M U^* P_N U P_M - P_M\| \leq \left(4\sqrt{\log_2\left(4N\sqrt{|\Delta|}/m\right)}\right)^{-1}, \tag{5.4}$$

$$\max_{|\Gamma|=|\Delta|,\Gamma\subset\{1,\dots,M\}} \|P_M P_\Gamma^\perp U^* P_N U P_\Gamma\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}. \tag{5.5}$$

We say that $N$ and $m$ satisfy the strong balancing property with respect to $U$, $M$ and $|\Delta|$ if (5.4) holds, and (5.5) is replaced by

$$\max_{|\Gamma|=|\Delta|,\Gamma\subset\{1,\dots,M\}} \|P_\Gamma^\perp U^* P_N U P_\Gamma\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}. \tag{5.6}$$

The balancing property dictates how large a bandwidth $M$ the $|\Delta|$ nonzero coefficients can lie, given $m$ measurements in the range $\{1,\dots,N\}$. Note that there is a strong analogy between this and the stable sampling rate (3.6), the main differences being due to the additional complications that enter when subsampling. Indeed, condition (5.4) ensures that $P_N U P_M$ is close to an isometry, and is very similar to (3.6).

The following proposition establishes that the balancing property is well defined:

**Proposition 5.2** *If $U$, $M$ and $|\Delta|$ are as in Definition 5.1, then there always exists integers $N$ and $m$ that satisfy the weak and strong balancing properties with respect to $U$, $M$ and $|\Delta|$.*

*Proof* First let $m = cN$ for some $0 < c < 1$. Now note that since $P_N \to I$ strongly as $N \to \infty$ we have that $P_N U \to U$ strongly. However, for any $\Gamma \subset \mathbb{N}$ with $|\Gamma| < \infty$ we have that $P_N U P_\Gamma \to U P_\Gamma$ in norm as $N \to \infty$ by compactness. Notice also that $\|V\|_{\mathrm{mr}} \leq \|V^*\|$ for any $V \in \mathcal{B}(\mathcal{H})$. The fact that $U$ is an isometry now shows that the left-hand sides of (5.4)–(5.6) can be made arbitrarily small by taking $N$ sufficiently large. □

*Remark 5.3* The inequality in (5.5) is somewhat inconvenient. However, it can be replaced by the far simpler, although stronger, condition

$$\|P_M U^* P_N U P_M - \mathrm{diag}(P_M U^* P_N U P_M)\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}. \tag{5.7}$$

Here $\mathrm{diag}(B)$ denotes the diagonal part of the matrix $B$. In particular, condition (5.7) is the requirement on the magnitude of the off-diagonal entries of the matrix $P_M U^* P_N U P_M$. In much the same manner, (5.6) can also be replaced by the much more convenient (however stronger) condition $\|U^* P_N U P_M - \mathrm{diag}(U^* P_N U P_M)\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}$.

## 6 Main Results

We now present our main results. Proofs of these results form the content of the remainder of this paper. To avoid pathological examples, we from now on assume that the sparsity $r = |\Delta| \geq 3$.

### 6.1 The Semi-Infinite-Dimensional Model

The first results concern the semi-infinite-dimensional model (see Sect. 4.2):

**Theorem 6.1** *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $M \in \mathbb{N}$, $\epsilon > 0$ and suppose that $x_0 \in l^\infty(\mathbb{N})$ with $\mathrm{supp}(x_0) = \Delta$, where $\Delta \subset \{1, \dots, M\}$. Suppose that $N$ and $m$ satisfy the weak balancing property with respect to $U$, $M$ and $|\Delta|$, and let $\Omega \subset \{1, \dots, N\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = Ux_0$ then, with probability exceeding $1 - \epsilon$, the problem*

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega U P_M \eta = P_\Omega \zeta, \tag{6.1}$$

*has a unique solution $\xi$ and this solution coincides with $x_0$, provided that $m$ satisfies*

$$m \geq C \cdot N \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{MN\sqrt{|\Delta|}}{m}\right), \tag{6.2}$$

*for some universal constant $C$. Furthermore, if $m = N$, then $\xi$ is unique and $\xi = x_0$ with probability* 1.

This theorem states that a sparse signal $x_0$ can be recovered perfectly (with high probability) by subsampling from the coefficients $\zeta$, provided (5.4), (5.5) and (6.2) hold. The main estimate (6.2) is similar to a standard finite-dimensional CS bound, in that it relates the number of measurements $m$ to the coherence $\upsilon(U)$, the sparsity $|\Delta|$ and several log factors. The primary difference is that the parameter $N$ in the estimate (the range from which the samples $\Omega$ are drawn) differs from the term $M$ (the range of $\Delta$). The precise relation between the two is given by (5.4) and (5.5), i.e., the weak balancing property. Note that this result addresses the questions (i) and (iii) posed in Sect. 4.

*Remark 6.2* Much like the stable sampling rate (see Remark 3.2), the weak and strong balancing properties depend completely on the matrix $U$. In particular, there is no universal choice (for all isometries $U$) of $N$ given $M$. For analysis of the balancing properties in the case of the Fourier/wavelets matrix (in which case $N$ can be taken to be linear in $M$, up to log factors), we refer to [7,50].

Recall that the second scenario in the semi-infinite-dimensional model corresponds to signals $y_0 = x_0 + h$, where $x_0$ is sparse and $\mathrm{supp}(h) \subset \{1, \dots, M\}$. The following theorem concerns this case:

**Theorem 6.3** *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $M \in \mathbb{N}$, $\epsilon > 0$ and suppose that $y_0 \in l^\infty(\mathbb{N})$ with $\mathrm{supp}(y_0) \subset \{1, \dots, M\}$. Suppose that $N$ and $m$ satisfy the weak balancing property with respect to $U$, $M$ and $|\Delta|$, and let $\Omega \subset \{1, \dots, N\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = Uy_0$ and $\xi \in \mathcal{H}$ is a minimizer of (6.1) then, with probability exceeding $1 - \epsilon$, we have that*

$$\|\xi - y_0\| \leq 8\left(1 + \frac{2N}{m}\right)\sigma_{r,M}(y_0), \quad r = |\Delta|, \tag{6.3}$$

*provided $m$ satisfies (6.2). If $m = N$ then (6.3) holds with probability* 1.

This theorem demonstrates recovery for compressible signals of the form (4.2). Specifically, we witness perfect recovery, up to an error determined by the best $(r, M)$-term approximation and a constant proportional to $N/m$ (see Remark 6.7 below). In particular, this result answers question (v) for the semi-infinite-dimensional model.

### 6.2 The Fully Infinite-Dimensional Model

The semi-infinite-dimensional model (4.2) places the restriction that the support of the nonsparse term $h$ is contained in $\{1, \dots, M\}$. As discussed in Sect. 3, this assumption is quite rare in practice, and a more realistic setting is provided by the fully infinite-dimensional model in which we assume that $y_0 = x_0 + h$, where $x_0$ is sparse and $|\mathrm{supp}(h)|$ is infinite.

To address this setting, it is first necessary to consider the infinite-dimensional optimization (4.4):

**Theorem 6.4** *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $\epsilon > 0$ and suppose that $y_0 \in l^1(\mathbb{N})$. Suppose that $N$ and $m$ satisfy the strong balancing property with respect to $U$, $M$ and $|\Delta|$ and let $\Omega \subset \{1, \dots, N\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = U y_0$ and $\xi \in \mathcal{H}$ is a minimizer of*

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega U \eta = P_\Omega \zeta, \tag{6.4}$$

*then*

$$\|\xi - y_0\| \leq 8 \left(1 + \frac{2N}{m}\right) \sigma_{r,M}(y_0), \quad r = |\Delta|, \tag{6.5}$$

*with probability exceeding $1 - \epsilon$, provided*

$$m \geq C \cdot N \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{\omega N \sqrt{|\Delta|}}{m}\right), \tag{6.6}$$

*where $\omega = \tilde{\omega}_{M,U}(|\Delta|, s, N)$ (recall 5.2), $s = \frac{m}{128 N \sqrt{|\Delta|} \log(\mathrm{e}^4 \epsilon^{-1})}$ and $C$ is a universal constant. If $m = N$ then (6.5) holds with probability 1.*

*Remark 6.5* Using (5.3), the quantity $\omega$ in (6.6) can also be replaced by a much more convenient (and of correspondingly less sharp) estimate. In particular, we have that $\omega \leq \tilde{M}$, where

$$\tilde{M} = \min\left\{ r \in \mathbb{N} : \|P_N U P_r^\perp\| \leq \frac{m}{128 N \sqrt{|\Delta|} \log(\mathrm{e}^4 \epsilon^{-1})} \right\}.$$

Observe that $\|P_N U P_r^\perp\|^2 = \|P_N U U^* P_N - P_N U P_r U^* P_N\|$, which is similar to the quantity $C_{r,N}$ introduced in (3.4). Hence, $\tilde{M}$ can be estimated in much the same way the stable sampling rate is estimated in GS (see Remark 3.2). For the case of the Fourier/wavelets matrix, we refer to [7,50].

This theorem, much like Theorem 6.3, confirms recovery of $y_0$ up to an error determined solely by $\sigma_{r,M}(y_0)$. Note that it provides answers to questions (i)–(iii) posed in Sect. 4. However, observe that the optimization problem (6.4) is infinite-dimensional. In practice, one always replaces (6.4) with the finite-dimensional problem

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega U P_k \eta = P_\Omega \zeta, \tag{6.7}$$

where $k \in \mathbb{N}$ is suitably chosen. The obvious question now arises: How do solutions of (6.7) compare to those of (6.4) as $k \to \infty$? For this we have the following:

**Proposition 6.6** *Let $U \in \mathcal{B}(\mathcal{H})$, $x_0 \in l^1(\mathbb{N})$ and $P_\Omega$ be a finite rank projection. Then, for each sufficiently large $k \in \mathbb{N}$ there exists a $\xi_k \in \mathcal{H}$ such that*

$$\xi_k \in \operatorname*{argmin}_{\eta \in l^1(\mathbb{N})} \left\{ \|\eta\|_{l^1} : P_\Omega U P_k \eta = P_\Omega U x_0 \right\}.$$

*Moreover, for every $\epsilon > 0$, there is a $K \in \mathbb{N}$ such that, for all $k \geq K$, we have $\|\xi_k - \tilde{\xi}_k\|_{l^1} \leq \epsilon$, where*

$$\tilde{\xi}_k \in \operatorname*{argmin}_{\eta \in l^1(\mathbb{N})} \left\{ \|\eta\|_{l^1} : P_\Omega U \eta = P_\Omega U x_0 \right\}. \tag{6.8}$$

In other words, solutions of (6.7) will be approximate minimizers of (6.4) for all sufficiently large $k$ and, in particular, will approximately satisfy (6.5). Note that when (6.8) has a unique minimizer $z$ it is straightforward to show that $\xi_k = z$ for all large $k$. We do not include this result since it is a rarity in practice for minimizers to be unique. Conversely, this proposition demonstrates that having multiple minimizers is not problematic. We remark also that this proposition does not show how to determine $k$ in a signal-independent way. Yet this can be done and is discussed further in [1,7].

We now make several further comments on the above theorems:

*Remark 6.7* The error bounds (6.3) and (6.5) are close to optimal in the sense that they involve the best approximation error $\sigma_{r,M}(y_0)$, yet multiplied by a factor proportional to $N/m$. Such a factor is not found in analogous finite-dimensional results. However, this term is the reciprocal subsampling percentage and in practice will usually not be much larger than 100 in magnitude (this would correspond to 1 % subsampling). Also unlike finite-dimensional results, we do not address the issue of noisy data in this work.[1]

*Remark 6.8* Neither the bandwidth $M$ nor the sparsity $r = |\Delta|$ need be known in Theorems 6.3 or 6.4. Specifically, these results state the following: Given $m$ and $N$ (the parameters of the sampling), any vector $y_0$ is recovered up to an error proportional to $\sigma_{r,M}(y_0)$, where $r$ and $M$ are determined implicitly through the balancing property and (6.6). This is typical in applications such as MRI, where the sampling resolution

[1] Since writing this paper, it has subsequently been shown that the term $N/m$ can be removed and that noise can be incorporated in the data and recovery guarantees. See [7] for details.

$N$ is fixed (due to the physical limitations of the scanner), as is the total number of samples $m$.

*Remark 6.9* The amount of subsampling depends on the coherence parameter $\upsilon(U)$. For a given operator $U$, this is a fixed (although arbitrarily small) number, and this suggests that subsampling will not be possible when $M$ is large. That is, we must take $m = N$. However, if $U$ has the property that $\upsilon(U P_k^{\perp}), \upsilon(P_k^{\perp} U) \to 0$ as $k \to \infty$, then one can actually circumvent this problem using so-called *multilevel random subsampling* techniques. This is not within the scope of this paper but is treated in detail in [7].

### 6.3 Theorems on Finite-Dimensional CS

As mentioned, GS–CS extends finite-dimensional CS from finite-dimensional vector spaces to separable Hilbert spaces. It is therefore unsurprising, but important to note nonetheless, that similar results to existing theorems for finite-dimensional CS can be obtained as straightforward corollaries of Theorems 6.1–6.4. In particular, we have

**Theorem 6.10** *Let $U \in \mathbb{C}^{n\times n}$ be an isometry, and suppose that $x_0 \in \mathbb{C}^n$ with $\operatorname{supp}(x_0) = \Delta$. For $\epsilon > 0$ suppose that $m \in \mathbb{N}$ is such that*

$$m \geq C \cdot n \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log(\epsilon^{-1}) + 1\right) \cdot \log n, \tag{6.9}$$

*for some universal constant $C$, and let $\Omega \subset \{1, \ldots, n\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = U x_0$, then, with probability exceeding $1 - \epsilon$, the problem*

$$\min_{\eta\in\mathbb{C}^n} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \zeta,$$

*has a unique solution $\xi$ and this solution coincides with $x_0$.*

**Theorem 6.11** *Let $U \in \mathbb{C}^{n\times n}$ be an isometry, and suppose that $y_0 = x_0 + h \in \mathbb{C}^n$ with $\operatorname{supp}(x_0) = \Delta$. For $\epsilon > 0$ suppose that $m \in \mathbb{N}$ satisfies (6.9), and let $\Omega \subset \{1, \ldots, n\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = U y_0$, then, with probability exceeding $1 - \epsilon$, any minimizer $\xi$ of the problem*

$$\min_{\eta\in\mathbb{C}^n} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \zeta,$$

*satisfies*

$$\|\xi - y_0\| \leq 8 \left(1 + \frac{2n}{m}\right) \|h\|_{l^1}.$$

*Proof of Theorems 6.10 and 6.11* The matrix $U$ extends in the obvious way to a partial isometry $\widetilde{U}$ on $\mathcal{H}$. Note that $(\widetilde{U})^* P_N \widetilde{U} = P_N$, for $N = n$. We may now, in an obvious

way, extend $\widetilde{U}$ to an isometry $\widehat{U}$ on $\mathcal{H}$ such that $\upsilon(\widehat{U}) = \upsilon(U)$. Therefore, the weak balancing property is automatically satisfied for $M = N$ and any $m \in \mathbb{N}$. We now apply Theorems 6.1 or 6.3. □

These results are similar to existing nonuniform recovery results for finite-dimensional CS proved recently by Candès and Plan [18]. The main difference is the larger factor $n/m$ in the error bound and the absence of noise in the data (see Remark 6.9).

## 7 Numerical Examples

Before giving proofs of these theorems, it is useful to present some further examples of GS–CS. In doing so, we will demonstrate the main premises of this paper. Firstly, if it is known that a function $f$ has a good representation in terms of a different basis, then one can obtain a far better reconstruction of $f$ than that stemming from its Fourier series. Secondly, by using GS–CS it is possible to get a substantial improvement over finite-dimensional CS techniques.

Consider the problem of reconstructing $g = \mathcal{F}f$ and $f$ from the samples $\{\zeta_j(f)\}_{j\in\mathbb{N}}$ where $\zeta_j(f) = \mathcal{F}f(\rho(j)/2)$ and $\rho$ is defined by

$$\rho(1) = 0, \rho(2) = 1, \rho(3) = -1, \rho(4) = 2, \rho(5) = -2 \dots.$$

We now compare three methods for approximating $f$ and $g$:

(i) The partial Fourier series $f_N$ (see 2.4) and its Fourier transform $g_N = \mathcal{F}f_N$.
(ii) The GS reconstructions $f_{N,M}$ (see 3.7) and $g_{N,M} = \mathcal{F}f_{N,M}$.
(iii) The GS–CS reconstructions

$$f_{N,m,k}(t) = \sum_{j=1}^{k} \alpha_j \varphi_j(t), \quad g_{N,m,k}(t) = \sum_{j=1}^{k} \alpha_j \mathcal{F}\varphi_j(t),$$

where $\alpha = \{\alpha_1, \dots, \alpha_k\}$ is computed via the convex optimization problem (4.5). Note that $f_{N,M}$ and $g_{N,M}$ use exactly the same samples as $f_N$ and $g_N$, yet $f_{N,m,k}$ and $g_{N,m,k}$ use only a subset of those samples.

If $f$ is sparse or has rapidly decaying coefficients in Haar wavelets, then we expect (i) to give a very poor reconstruction. However, both the GS and GS–CS methods should give very good reconstructions, with the latter taking advantage of the sparsity to reduce the number of Fourier coefficients sampled (recall that GS does not exploit any sparsity—it offers guaranteed recovery for all functions $f$ by using the full range of Fourier coefficients). Note that in the first example below the samples in the case of GS–CS are chosen such that half of them are fixed (from the first indices) and the other half is chosen uniformly at random. This is to improve results because of coherence issues (see Remark 6.9).

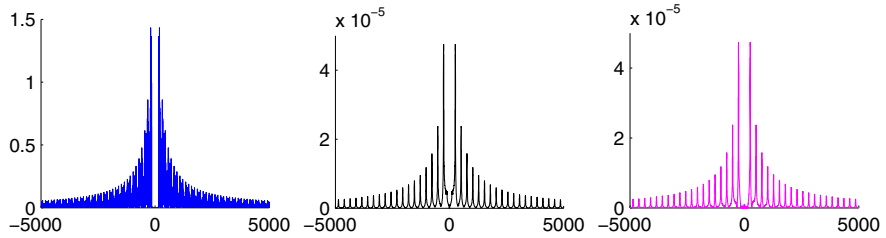


**Fig. 5** Errors $|g(t) - g_N(t)|$ (*left*), $|g(t) - g_{N,M}(t)|$ (*middle*) and $|g(t) - g_{N,m,k}(t)|$ (*right*) against $t$, for $N = 601$, $M = 200$, $m = 230$ and $k = 650$

**Table 1** Errors for the reconstructions $g_N$, $g_{N,M}$ and $g_{N,m,k}$

| $N$ | $\|g - g_N\|_{L^\infty}$ | $\|g - g_{N,M}\|_{L^\infty}$ | $\|g - g_{N,m,k}\|_{L^\infty}$ (avg. 20 trls) |
|---|---|---|---|
| 601 | 1.43 | $4.74 \times 10^{-5}$, $(M = 200)$ | $4.73 \times 10^{-5}$, $(m = 230, k = 550)$ |
| 1201 | 0.85 | $2.36 \times 10^{-5}$, $(M = 400)$ | $2.38 \times 10^{-5}$, $(m = 460, k = 1400)$ |

### 7.1 First Example

As a first example, let us consider the function $g = \mathcal{F}f$, where

$$f(t) = \sum_{j=1}^{200} \alpha_j \varphi_j(t) + \cos(2\pi t)\chi_{[\frac{1}{2},\frac{9}{16}]}(t), \quad t \in [0,1], \tag{7.1}$$

$\{\varphi_j\}_{j\in\mathbb{N}}$ are Haar wavelets on $[0, 1]$ and $\chi_{[\frac{1}{2},\frac{9}{16}]}$ is the indicator function of the interval $[\frac{1}{2}, \frac{9}{16}]$. Suppose that $|\{j : \alpha_j \neq 0\}| = 25$, so that $f$ can be decomposed into a sparse component and a remainder. Note that the remainder has infinite support in the Haar wavelet basis, so this function belongs to the fully infinite-dimensional model (see Sect. 4.2).

In Fig. 5, we display the errors committed by the approximations (i)–(iii) for this function. As expected, the expansion $g_N$ gives an extremely poor reconstruction, whereas both the GS and GS–CS give far better approximations, both reducing the error by a factor of roughly 10,000. Moreover, and also as expected, the GS–CS approximation attains the same numerical error as the GS approximation using only around 38 % of the Fourier samples. These observations are given in Table 1.

Whilst GS and GS–CS give very similar numerical errors, it is important to notice that the reconstructions are typically very different. In particular, in GS one reconstructs approximately the first $M$ Haar wavelet coefficients $\alpha_1, \ldots, \alpha_M$, where $M < N$. On the other hand, in GS–CS one computes $k$ such coefficients, where typically (although not always) $k > N$. This difference can be explained by examining Eqs. (3.3) and (4.5). In GS, which corresponds to (3.3), one requires $M < N$ to ensure invertibility of the operator $A$. On the other hand, unless $k$ is taken sufficiently large, (4.5) need not have a solution, since the right-hand side $P_\Omega \zeta(f)$ may not lie in the range of the (finite-dimensional) section $P_\Omega U P_k : \mathbb{C}^k \to \mathbb{C}^{|\Omega|}$. In particular, this

**Table 2** The error $\|g - g_{N,m,k}\|_{L^\infty}$ for different values of $N$, $m$ and $k$ [the notation $E_{N,m,k} = \infty$ means that (4.5) does not have a solution]

| $N$ | $E_{N,m,k} = \|g - g_{N,m,k}\|_{L^\infty}$ (avg. 20 trials) | | | |
|---|---|---|---|---|
| 601 | $E_{N,230,200} = \infty$ | $E_{N,230,350} = \infty$ | $E_{N,230,550} = 4.759 \times 10^{-5}$ | $E_{N,230,850} = 4.727 \times 10^{-5}$ |
| 1201 | $E_{N,460,400} = \infty$ | $E_{N,460,500} = \infty$ | $E_{N,460,1000} = 2.384 \times 10^{-5}$ | $E_{N,460,1300} = 2.392 \times 10^{-5}$ |

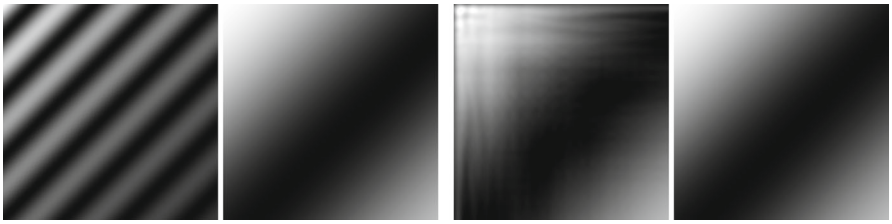

**Fig. 6** *Left to right* (i) Original, (ii) zoomed original, (iii) finite-dimensional CS reconstruction using periodic Daubechies-6 wavelets, (iv) infinite-dimensional CS reconstruction using boundary Daubechies-6 wavelets

may well be the case whenever $k < N$. Fortunately, as shown in Proposition 6.6, this cannot happen if $k$ is sufficiently large. The effect of increasing $k$ for the example (7.1) is illustrated in Table 2. As is evident, once $k$ is sufficiently large, the problem (4.5) has a solution, and the error drops accordingly.

### 7.2 Second Example

This example was first introduced in [53] and demonstrates the difference between finite-dimensional CS techniques and the proposed infinite-dimensional GS-CS techniques in a practical electron microscopy setup. In this setting, the sampling procedure is completely dictated by the physics behind the microscope and corresponds to radial line sampling in Fourier space. Note that we use exactly the same sampled data in both cases, and seek to recover the function $f(x, y) = \exp(-x - y)\cos^2(x + y)$ from $m = 16120$ (6.15 %) continuous Fourier samples taken radially from a $512 \times 512$ grid. We commence with the finite-dimensional approach (see Sect. 2.1) and solve

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \text{ subject to } P_\Omega U_{\text{df}} V_{\text{dw}}^{-1} z = P_\Omega y,$$

where $U_{\text{df}}$ and $V_{\text{dw}}^{-1}$ denote the discrete Fourier and discrete wavelet transforms, respectively. Next we solve

$$\min_{z \in P_k(\ell^2(\mathbb{N}))} \|z\|_1 \text{ subject to } P_\Omega U P_k z = P_\Omega y,$$

where $U$ is as in (2.7). As Fig. 6 displays, the infinite-dimensional approach shows a substantial improvement over the finite-dimensional technique. The reason behind the

superior reconstruction is, as discussed in Sect. 2.1, the use of $U_{\mathrm{df}} V_{\mathrm{dw}}^{-1}$ implies that the error is dominated by the truncated Fourier series, whereas in the GS-CS approach the discretization is faithful to the guiding principle of Sect. 1.2, and hence the error is dominated by the actual wavelet decay, which is much faster than the Fourier decay.

## 8 Infinite-Dimensional Optimization

The remainder of this paper is largely devoted to proving the results of Sect. 6, which will be the specific focus of Sect. 9. Before doing so, however, we first discuss the topic of infinite-dimensional optimization in a little more detail. As the informed reader will have noticed, all our results are really questions of infinite-dimensional optimization: for example, in the case of Proposition 6.6, showing the existence of minimizers to the finite rank discretizations of an infinite-dimensional optimization problem, and their convergence to minimizers of that problem. For this reason, we now recap some of the basics of this field.

### 8.1 Background

The field of infinite-dimensional convex optimization is certainly not new [28,49]. However, it is much less standard than the more thoroughly investigated topic of finite-dimensional convex optimization. We now cover some of the basic tools that will subsequently prove useful.

We consider complex vector spaces. Standard optimization theory is usually considered over the reals, and this is also the case in [28] (the main reference we consider herein for the field of infinite-dimensional optimization). To be able to quote [28] freely, we use the standard trick and consider any complex Banach space $X$ as a real vector space. In particular, if $\widetilde{X}$ is the real Banach space induced by $X$, then

$$\widetilde{X}^* = \{\mathrm{Re}(x^*) : x^* \in X^*\}.$$

This follows by the observation that if $x^* \in X^*$ and $u = \mathrm{Re}(x^*)$ then $u$ is a real linear functional. Also, if $u \in \widetilde{X}^*$ and $x^* : X \to \mathbb{C}$ is defined by $x^*(x) = u(x) - \mathrm{i}u(\mathrm{i}x)$, then $x^* \in X^*$. To avoid unnecessary clutter, we will (with slight abuse of notation) use $X$ as the notation for $\tilde{X}$.

**Definition 8.1** Let $X$ be a Banach space and let $F : X \to \overline{\mathbb{R}}$. The polar function $F^* : X^* \to \overline{\mathbb{R}}$ is defined by

$$F^*(x^*) = \sup_{x \in X} \{\mathrm{Re}(x^*(x)) - F(x)\},$$

where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

**Definition 8.2** Let $X$ be a Banach space, $F : X \to \overline{\mathbb{R}}$ be convex and consider the following problem

$$(P): \quad \inf\{F(x) : x \in X\}.$$

If $Y$ is a Banach space and $\Phi : X \times Y \to \mathbb{R} \cup \{\infty\}$ is a convex lower semi-continuous function such that $\Phi(x, 0) = F(x)$ for all $x \in X$, then the dual problem $P^*$ with respect to $\Phi$ is defined by

$$(P^*): \quad \sup\{-\Phi^*(0, y^*) : y^* \in Y^*\}.$$

If $\Phi$ is not specified, we will say that $(P^*)$ is a dual problem for $(P)$.

Let $X$ and $Y$ be Banach spaces and suppose that $T \in \mathcal{B}(X, Y)$ and $y_0 \in Y$. Consider the problem

$$(P_1): \quad \inf\{\|x\| : x \in X,\ Tx = y_0\}.$$

Note that $(P_1)$ can be written as the equivalent convex optimization problem:

$$(\widetilde{P_1}): \quad \inf\{F(x) + G(Tx), x \in X\}, \tag{8.1}$$

where $F(x) = \|x\|$ and $G : Y \to \mathbb{R} \cup \{\infty\}$ is defined by $G(z) = \delta_{\{0\}}(z - y_0)$. Here the function $\delta_C : Y \to \mathbb{R} \cup \{\infty\}$, where $C \subset Y$ is convex, is defined by $\delta_C(z) = 0$ if $z \in C$ and $\delta_C(z) = \infty$ if $z \notin C$. Moreover, by letting $\Phi : X \times Y \to \mathbb{R} \cup \{\infty\}$ be defined by

$$\Phi(x, y) = F(x) + G(Tx + y), \tag{8.2}$$

and observing that

$$\Phi^*(x^*, y^*) = F^*(x^* - T'y^*) + G^*(y^*),$$

where $T' : Y^* \to X^*$ denotes the dual mapping, we also obtain the following dual problem with respect to $\Phi$:

$$(P_1^*): \quad \sup\{-F^*(-T'y^*) - G^*(y^*) : y^* \in Y^*\}.$$

Much like $(P_1)$ and $(\tilde{P}_1)$, the dual problem $(P_1^*)$ also has an equivalent form. In fact, since $F^*(x^*) = 0$ if $\|x^*\|_{X^*} \leq 1$ and $F^*(x^*) = \infty$ if $\|x^*\|_{X^*} > 1$, together with the observation that

$$G^*(y^*) = \sup\{\mathrm{Re}(y^*(y)) - \delta_{\{0\}}(y - y_0), y \in Y\} = \mathrm{Re}(y^*(y_0)),$$

we find that

$$(P_1^*): \quad \sup\{\mathrm{Re}(y^*(y_0)) : \|T'y^*\|_{X^*} \leq 1,\ y^* \in Y^*\}.$$

Using these ideas, we obtain the following well-known result [28]:

**Proposition 8.3** *Let $X$ and $Y$ be Banach spaces and suppose that $T \in \mathcal{B}(X, Y)$ and $y_0 \in Y$. If $T$ is onto, then*

$$\inf\{\|x\| : x \in X,\ Tx = y_0\} = \sup\{\mathrm{Re}(y^*(y_0)) : \|T'y^*\|_{X^*} \leq 1,\ y^* \in Y^*\}.$$

### 8.2 Stability Analysis for Infinite-Dimensional Convex Optimization

We now consider the issue of stability of the optimization problem

$$\min_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega U (x_0 + h), \tag{8.3}$$

and its finite-dimensional approximations

$$\min_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U P_k \eta = P_\Omega U (x_0 + h), \tag{8.4}$$

where $x_0, h \in l^1(\mathbb{N})$ and $h$ is small in norm. Note that this is the first step toward a proof of Theorems 6.3 an 6.4 concerning the recovery of compressible signals which are described by the semi/fully infinite-dimensional models Sect. 3. However, at this moment we do not consider either sparsity or randomness. This comes in Sect. 9, in which the results proved in this and the previous section are applied to the sparse recovery problems (6.1) and (6.4) to yield proofs of Theorems 6.1–6.4.

Stability turns out to be a rather subtle issue. We now illustrate why.

**Definition 8.4** Let $\Omega, \Delta$ be finite subsets of $\mathbb{N}$, $U \in \mathcal{B}(l^2(\mathbb{N}))$ and let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function such that $\lim_{t \to 0} f(t) = 0$. Suppose that $\xi \in \mathcal{H}$, $\text{supp}(\xi) = \Delta$, is the unique minimizer of

$$\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \xi\}. \tag{8.5}$$

If, for any $\epsilon > 0$ and $\zeta \in \mathcal{H}$ such that $\|\xi - \zeta\|_{l_1} \leq \epsilon$, we have that $\|x - \xi\|_{l_1} \leq f(\epsilon)$, where $x$ is a minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \zeta\}$, then we say that $\{U, \Omega, \Delta\}$ is locally $f$-stable at $\xi$. If $f(t) = Ct$ for some constant $C > 0$, then $\{U, \Omega, \Delta\}$ is said to be locally linearly stable at $\xi$. We say that $\{U, \Omega, \Delta\}$ is globally $f$-stable (linearly stable) if the above statements hold for all $\xi \in l^2(\mathbb{N})$, $\text{supp}(\xi) = \Delta$, such that $\xi$ is the unique minimizer of (8.5).

**Proposition 8.5** *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be unitary and real, and let $\Omega, \Delta$ be finite subsets of $\mathbb{N}$. Suppose that $\{U, \Omega, \Delta\}$ is globally $f$-stable. Suppose also that there exists a real $x \in l^2(\mathbb{N})$, $\text{supp}(x) = \Delta$, such that $x$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U x\}$. Then, if $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta|_{P_\Delta l^2(\mathbb{N})}$ is invertible, and $y \in \mathcal{H}$, $\text{supp}(y) = \Delta$, is arbitrary, then $y$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U y\}$.*

**Proposition 8.6** *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be unitary and real, and let $\Omega, \Delta$ be finite subsets of $\mathbb{N}$. Suppose that for any real $\xi \in l^2(\mathbb{N})$, $\text{supp}(\xi) = \Delta$, it holds that $\xi$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \xi\}$, and also that $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta|_{P_\Delta l^2(\mathbb{N})}$ is invertible. Then, $\{U, \Omega, \Delta\}$ is globally linearly stable when the vector space $l^2(\mathbb{N})$ is considered over the reals.*

These results establish the relationship between global stability and the existence of unique minimizers (proofs are given in the Appendix). In particular, existence of unique minimizers for all $y$ with $\text{supp}(y) = \Delta$ is (almost) equivalent to global stability. Thus,

global stability is a rather strict condition and may be difficult to achieve. However, we will be concerned with a fixed signal to recover and hence global stability may not be necessary. Conditions in order to establish local stability are the topic in the next section.

### 8.3 The Key Result

The key result of this section, which will later lead to the proofs of Theorems 6.1–6.4, is the following (see, e.g., [34, Thm. 4.33] for a related finite-dimensional version):

**Proposition 8.7** *Let $U \in \mathcal{B}(\mathcal{H})$ with $\|U\| \leq 1$, and suppose that $\Delta$ and $\Omega$ are finite subsets of $\mathbb{N}$. Let $x_0, h \in \mathcal{H}$ be such that $\mathrm{supp}(x_0) = \Delta$, $\mathrm{supp}(h) \cap \Delta = \emptyset$ and $\|h\|_{l^1} < \infty$, and suppose that $\Delta \subset \{1, \dots, M\}$ for some $M \in \mathbb{N}$. Let $\xi, \xi_M \in \mathcal{H}$ such that*

$$\begin{aligned} \xi &\in \operatorname*{argmin}_{\eta \in l^1(\mathbb{N})} \left\{ \|\eta\|_{l^1} : P_\Omega U \eta = P_\Omega U (x_0 + h) \right\}, \\ \xi_M &\in \operatorname*{argmin}_{\eta \in l^1(\mathbb{N})} \left\{ \|\eta\|_{l^1} : P_\Omega U P_M \eta = P_\Omega U (x_0 + P_M h) \right\}. \end{aligned} \tag{8.6}$$

*If there exists $\rho \in \mathrm{ran}(U^* P_\Omega)$ and $q > 0$ with the following properties:*

(i) $\|(q^{-1} P_\Delta U^* P_\Omega U P_\Delta)^{-1}\| \leq 2$,
(ii) $\|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq q/8$,
(iii) $\|P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2$,

*then*

$$\|\xi - x_0\| \leq \left( \frac{16}{q} + 7 \right) \|h\|_{l^1}. \tag{8.7}$$

*Also, if (i) and (ii) hold and (iii) is replaced with $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2$ then*

$$\|\xi_M - x_0\| \leq \left( \frac{16}{q} + 7 \right) \|P_M h\|_{l^1}. \tag{8.8}$$

*Proof* Note that (8.6) and (i) yield

$$\begin{aligned} & P_\Omega U (x_0 - P_\Delta \xi) = P_\Omega U (P_\Delta^\perp \xi - h) \\ \Rightarrow \quad & P_\Delta U^* P_\Omega U (x_0 - P_\Delta \xi) = P_\Delta U^* P_\Omega U (P_\Delta^\perp \xi - h) \\ \Rightarrow \quad & x_0 - P_\Delta \xi = (P_\Delta U^* P_\Omega U P_\Delta)^{-1} P_\Delta U^* P_\Omega U (P_\Delta^\perp \xi - h). \end{aligned} \tag{8.9}$$

(note that (i) implies that $P_\Delta U^* P_\Omega U P_\Delta$ is invertible). Hence, from (i) and (8.9), and by using the fact that $\|U\| \leq 1$, we obtain

$$\|x_0 - P_\Delta \xi\| \leq 2/q \|P_\Delta^\perp \xi - h\|. \tag{8.10}$$

Thus,

$$\|x_0 - \xi\| \leq \frac{2}{q}\|P_\Delta^\perp \xi - h\| + \|P_\Delta^\perp \xi\| \leq \left(\frac{2}{q} + 1\right)\|P_\Delta^\perp \xi\|_{l^1} + \frac{2}{q}\|h\|_{l^1}. \quad (8.11)$$

The rest of the proof is therefore devoted to showing that $\|P_\Delta^\perp \xi\|_{l^1}$ is bounded by a constant times $\|h\|_{l^1}$.

The fact that $\rho \in \mathrm{ran}(U^* P_\Omega)$ and $P_\Omega U(\xi - (x_0 + h)) = 0$ implies that $\langle \xi, \rho\rangle = \langle x_0 + h, \rho\rangle$. Thus, it follows, by appealing to (iii), that

$$\mathrm{Re}\big(\langle x_0, \rho\rangle\big) + \mathrm{Re}\big(\langle h, \rho\rangle\big) = \mathrm{Re}\big(\langle \xi, \rho\rangle\big) \leq \mathrm{Re}\big(\langle \xi, P_\Delta \rho\rangle\big) + \frac{1}{2}\sum_{j \in \Delta^c} |\xi(j)|. \quad (8.12)$$

Thus, since $\mathrm{supp}(h) \cap \Delta = \emptyset$, we have

$$\begin{aligned}\mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho\rangle &= \mathrm{Re}\,\langle x_0, \rho\rangle - \mathrm{Re}\,\langle \xi, P_\Delta \rho\rangle \leq -\mathrm{Re}\,\langle h, \rho\rangle + \frac{1}{2}\|P_\Delta^\perp \xi\|_{l^1} \\ &= -\mathrm{Re}\,\langle h, P_\Delta^\perp \rho\rangle + \frac{1}{2}\|P_\Delta^\perp \xi\|_{l^1} \leq \frac{1}{2}\left(\|h\|_{l^1} + \|P_\Delta^\perp \xi\|_{l^1}\right).\end{aligned}$$

We will return to this equation, but for the meantime we will continue to investigate the quantity $\mathrm{Re}(\langle x_0 - \xi, P_\Delta \rho\rangle)$. Observe that

$$\begin{aligned}\mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho\rangle &= \mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho - \mathrm{sgn}(x_0)\rangle + \|x_0\|_{l^1} - \mathrm{Re}\,\langle \xi, \mathrm{sgn}(x_0)\rangle \\ &\geq \mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho - \mathrm{sgn}(x_0)\rangle + \|x_0\|_{l^1} - \|P_\Delta \xi\|_{l^1} \\ &= \mathrm{Re}\,\langle x_0 - P_\Delta \xi, P_\Delta \rho - \mathrm{sgn}(x_0)\rangle + \|x_0\|_{l^1} - \|\xi\|_{l^1} + \|P_\Delta^\perp \xi\|_{l^1}.\end{aligned}$$

Since $\|x_0 + h\|_{l^1} \geq \|\xi\|_{l^1}$, we obtain

$$\mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho\rangle \geq \mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho - \mathrm{sgn}(x_0)\rangle - \|h\|_{l^1} + \|P_\Delta^\perp \xi\|_{l^1}. \quad (8.13)$$

Moreover, using (ii) and (8.10), we get $|\langle x_0 - P_\Delta \xi, P_\Delta \rho - \mathrm{sgn}(x_0)\rangle| \leq \frac{1}{4}\|P_\Delta^\perp \xi - h\|$. Hence, substituting this into (8.13) now gives

$$\begin{aligned}\mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho\rangle &\geq -\frac{1}{4}\|P_\Delta^\perp \xi - h\| - \|h\|_{l^1} + \|P_\Delta^\perp \xi\|_{l^1} \\ &\geq -\frac{5}{4}\|h\|_{l^1} + \frac{3}{4}\|P_\Delta^\perp \xi\|_{l^1}. \quad (8.14)\end{aligned}$$

Combining (8.13) and (8.14) and rearranging now gives $\|P_\Delta^\perp \xi\|_{l^1} \leq 7\|h\|_{l^1}$. Substituting this into (8.11) now yields (8.7). The proof of (8.8) is almost identical, and we omit the details. □

### 8.4 Proof of Proposition 6.6

To end this section, we now present a proof of Proposition 6.6. We first require the following:

**Lemma 8.8** *Let $U \in \mathcal{B}(\mathcal{H})$ and $P$ be a finite rank projection. Then, for every $\chi \in \mathrm{Ran}(PU)$, there exists $\xi \in \mathcal{H}$ such that*

$$\xi \in \operatorname*{argmin}_{\eta \in l^1(\mathbb{N})} \left\{ \|\eta\|_{l^1} : PU\eta = \chi \right\}.$$

*Proof* Recall that $(c_0)^* = l^1$. By weak* compactness there is a sequence $\{\xi_k\} \subset l^1$ and a $\xi \in l^1$ such that $PU\xi_k = \chi$, $\|\xi_k\|_{l^1} \searrow \inf\{\|\eta\|_{l^1} : PU\eta = \chi\}$ and $\langle \xi_k, e_j \rangle \to \langle \xi, e_j \rangle$ as $k \to \infty$ for all $j \in \mathbb{N}$. Hence $\|\xi\|_{l^1} \leq \lim_{k\to\infty} \|\xi_k\|_{l^1}$. Since $\xi_k \to \xi$ weakly as elements in $\mathcal{H}$, it follows by the fact that $PU$ is compact (since $P$ is of finite rank) that $PU\xi_k \to PU\xi$. Thus $PU\xi = \chi$ as required. □

*Proof of Proposition 6.6* To see the existence of $\xi_k$ for large $k$, it suffices to observe that $\mathrm{Ran}(P_\Omega U)$ and $\mathrm{Ran}(P_\Omega U P_k)$ coincide for all sufficiently large $k$, since $P_\Omega$ has finite rank.

For the second part of the proposition, it is easy to see that it suffices to show that every subsequence of $\{\xi_k\}_{k\in\mathbb{N}}$ has a convergent subsequence in the $l^1$ norm with limit $\xi$ satisfying

$$\|\xi\|_{l^1} = \inf_{\eta \in \mathcal{H}} \left\{ \|\eta\|_{l^1} : P_\Omega U\eta = P_\Omega U x_0 \right\}. \tag{8.15}$$

Therefore. let $\{\xi_k\}_{k\in\mathbb{N}}$ be a subsequence of the original sequence (we use the same notation for simplicity). Since $\|\xi_k\|_{l^1} \geq \|\xi_{k+1}\|_{l^1}$ for all large $k$ it follows that $\{\xi_k\}$ is bounded. So by weak* compactness of the $l^1$ ball we have that, by possibly passing to a subsequence, there is a $\xi \in \mathcal{H}$ such that $\xi_k \to \xi$ weakly (as elements in $\mathcal{H}$) as $k \to \infty$. By compactness of $P_\Omega U$, we find that $P_\Omega U\xi_k \to P_\Omega U\xi$ as $k \to \infty$, and, since $P_\Omega U\xi_k = P_\Omega U x_0$, it follows that $P_\Omega U\xi = P_\Omega U x_0$.

To see that $\xi$ satisfies (8.15), we argue as follows. We claim that for any $\lambda > 0$ we have

$$\|\xi_k\|_{l^1} \leq \inf_{\eta \in \mathcal{H}} \{\|\eta\|_{l^1} : P_\Omega U\eta = P_\Omega U x_0\} + \lambda, \tag{8.16}$$

for all sufficiently large $k$. Let $r = \dim(\mathrm{Ran}(P_\Omega U)) < \infty$, and let $\hat{e}_1, \ldots, \hat{e}_r$ be coordinate vectors such that $\mathrm{span}\{P_\Omega U\hat{e}_j\}_{j=1}^r = \mathrm{Ran}(P_\Omega U)$. Then every $\eta \in \mathrm{Ran}(P_\Omega U)$ with $\|\eta\| = 1$ can be written as $\eta = c_1 P_\Omega U\hat{e}_1 + \ldots + c_r P_\Omega U\hat{e}_r$, where the $c_j$s are bounded by, say, $1 \leq c < \infty$. Now let $\tilde{\xi}$ be a minimizer of (8.15) (the existence of such a minimizer is guaranteed by Lemma 8.8), and choose $k$ so large that $\{\hat{e}_j\}_{j=1}^r \subset \mathrm{Ran}(P_k)$, $\|P_\Omega U P_k^\perp \tilde{\xi}\| \leq \lambda/(2cr)$ and $\|P_k^\perp \tilde{\xi}\| \leq \lambda/2$. Let $c_1, \ldots, c_r$ be chosen such that $P_\Omega U P_k^\perp \tilde{\xi} / \|P_\Omega U P_k^\perp \tilde{\xi}\| = c_1 P_\Omega U\hat{e}_1 + \ldots + c_r P_\Omega U\hat{e}_r$, and set $\tilde{\eta} = P_k\tilde{\xi} + (c_1\hat{e}_1 + \ldots c_r\hat{e}_r)\|P_\Omega U P_k^\perp \tilde{\xi}\|$. It follows that $P_\Omega U\tilde{\eta} = P_\Omega U\tilde{\xi} = P_\Omega U x_0$, $\|\tilde{\eta}\|_{l^1} \leq \|\tilde{\xi}\|_{l^1} + \lambda$ and $\tilde{\eta} \in \mathrm{Ran}(P_k)$. Hence $\|\xi_k\|_{l^1} \leq \|\tilde{\xi}\|_{l^1} + \lambda$ and we have shown (8.16). Now choose $m \in \mathbb{N}$ such that $\|P_m^\perp \xi\|_{l^1} \leq \lambda$. Then $\|\xi\|_{l^1} \leq \|P_m\xi\|_{l^1} + \|P_m^\perp \xi\|_{l^1}$.

But $P_m\xi_k \to P_m\xi$ and $\xi_k$ satisfies (8.16), thus $\|\xi\|_{l^1} \leq \inf_{\eta\in\mathcal{H}}\{\|\eta\|_{l^1} : P_\Omega U\eta = P_\Omega U x_0\} + 2\lambda$ for any $\lambda > 0$. Therefore $\xi$ satisfies (8.15), as required.

For the final part of the proof, we are required to show that $\|\xi_k - \xi\|_{l^1} \to 0$ as $k \to \infty$. By possibly passing to another subsequence, it follows by (8.16) that

$$\|\xi_k\|_{l^1} \leq \inf_{\eta\in\mathcal{H}}\{\|\eta\|_{l^1} : P_\Omega U\eta = P_\Omega U x_0\} + 1/k. \tag{8.17}$$

Note also that, for fixed $m \in \mathbb{N}$, we have $P_m(\xi_k - \xi) \to 0$ as $k \to \infty$. But by (8.17) we also have $\|P_m\xi_k\|_{l^1} + \|P_m^\perp\xi_k\|_{l^1} \leq \|P_m\xi\|_{l^1} + \|P_m^\perp\xi\|_{l^1} + 1/k$. So

$$\lim_{m\to\infty}\limsup_{k\to\infty}\|P_m^\perp\xi_k\|_{l^1} = 0.$$

It thus follows that $\xi_k \to \xi$ (in $l^1$) as $k \to \infty$, and the proof is complete. □

## 9 Proofs of the Main Results

### 9.1 Key Ideas

Before we present proofs of Theorems 6.1–6.4, we would like to sketch the key ideas. Our approach is to use Proposition 8.7 to show the existence of a $\rho \in \mathrm{ran}(U^*P_\Omega)$ with the following properties

$$\begin{aligned}&(i)\ \|\theta^{-1}P_\Delta U^*P_\Omega U P_\Delta - P_\Delta\| \leq 1/2, \quad (ii)\ \|P_\Delta\rho - \mathrm{sgn}(x_0)\| \leq \theta/8\\ &(iii)\ \|P_M P_\Delta^\perp\rho\|_{l^\infty} \leq 1/2,\end{aligned}$$

for some $\theta > 0$ (recall the setup in Theorems 6.1 and 6.3).

Throughout the paper, we will be concerned with randomly choosing a set $\Omega \subset \{1,\dots,N\}$. In our models, we will choose $\Omega$ uniformly at random; however, in some of the proofs we will also use another approach that renders the analysis possible, whilst not affecting the model unduly. Therein, we take a sequence $\{\delta_1,\dots\delta_N\}$ of independent identically distributed Bernoulli random variables taking values 0 and 1 with $\mathbb{P}(\delta_j = 1) = q$ for all $j$. We then set $\Omega = \{j : \delta_j = 1\}$. We will refer to this type of random selection of $\Omega$ as the Bernoulli model, and we will denote such a procedure by $\{N,\dots,1\} \supset \Omega \sim \mathrm{Ber}(q)$.

Note that transitioning from the uniform sampling model to the Bernoulli sampling model in this way has become standard approach in the literature. In particular, one can show that the Bernoulli model implies (up to a constant) the uniform sampling model in each of the conditions in Proposition 8.7. We refer to [19,20,34] for details.

From now on, we thus consider (without loss of generality) the Bernoulli sampling scheme. We assume that $\{N,\dots,1\} \supset \Omega \sim \mathrm{Ber}(\theta)$, for some finite $N \in \mathbb{N}$. However, we will construct $\Omega$ in an equivalent, but slightly different way. Namely, we let

$$\Omega = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_\mu, \quad \Omega_j \sim \mathrm{Ber}(q_j),$$

where the specific value of $\mu$ will be determined later. Note that as long as the $q_j$s are chosen according to $\theta$ this is equivalent to letting $\Omega \sim \text{Ber}(\theta)$. Indeed, we have that $\Omega \sim \text{Ber}(\theta)$ is equivalent to $\Omega^c \sim \text{Ber}(1-\theta)$. So, for $k \in \{1, \ldots, N\}$, we have

$$\mathbb{P}(k \in \Omega^c) = (1-\theta),$$

where $\Omega^c = \{1, \ldots, N\} \backslash \Omega$. But

$$\mathbb{P}(k \in (\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_\mu)^c) = (1-q_1)(1-q_2)\cdots(1-q_\mu).$$

Thus, if we let

$$(1-q_1)(1-q_2)\cdots(1-q_\mu) = (1-\theta) \tag{9.1}$$

it is easy to see (by independence) that the two models are equivalent. Note that there might be overlaps between the $\Omega_j$'s. This automatically gives us the following:

$$q_1 + q_2 + \ldots + q_\mu \geq \theta.$$

This observation will be used several times in the arguments that follow.

### 9.2 The Golfing Scheme

We can now present the golfing scheme. Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry and let $\{N, \ldots, 1\} \supset \Omega_j \sim \text{Ber}(q_j)$ for $j = 1, \ldots, \mu$ and some $\mu \in \mathbb{N}$ where the $q_j$s satisfy (9.1) for some $0 < \theta \leq 1$. Suppose also that $x_0 \in \mathcal{H}$. Define the operator

$$E_{\Omega_j} = U^* P_{\Omega_j} U, \quad j = 1, \ldots, \mu.$$

The construction of $\rho$ is based on the following idea. Let

$$\begin{aligned} \rho = Y_\mu, \quad Y_i &= \sum_{j=1}^{i} q_j^{-1} E_{\Omega_j} Z_{j-1} \\ Z_i = \text{sgn}(x_0) - P_\Delta Y_i, \quad Z_0 &= \text{sgn}(x_0), \end{aligned} \tag{9.2}$$

where the specific value of $\mu$ will be determined later. The construction suggested in (9.2) will be referred to as the golfing scheme and is a variant of the original scheme introduced in [36] by D. Gross. The actual construction will differ slightly from the one suggested here; however, this should give the reader an idea about the approach.

Before we can prove the main results we need to establish some ancillary results that will be crucial in the construction of $\rho$.

### 9.3 The Proofs

We first require the following three results. Proofs are found in the Appendix. For related finite-dimensional versions, see [18] or [34, Chpt. 12].

**Proposition 9.1** *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $\{N, \dots, 1\} \supset \Omega \sim \mathrm{Ber}(q)$ for some $0 < q \leq 1$, and $\Delta \subset \mathbb{N}$ with $|\Delta| < \infty$. Also, let $M \in \mathbb{N}$ be sufficiently large so that $\Delta \subset \{1, \dots, M\}$ and define $E_\Omega = U^* P_\Omega U$. Then, for $\eta \in \mathcal{H}$ and $t, \gamma > 0$*

$$\mathbb{P}\left( \|q^{-1} P_M P_\Delta^\perp E_\Omega P_\Delta \eta\|_{l^\infty} > (t + \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}}) \|\eta\| \right) \leq \gamma \tag{9.3}$$

*provided*

$$q \geq \left( \frac{4}{t^2} + \frac{2\sqrt{2}}{3t} \sqrt{|\Delta|} \right) \cdot \log\left( \frac{4}{\gamma} |\Delta^c \cap \{1, \dots, M\}| \right) \cdot \upsilon^2(U).$$

*Also,*

$$\mathbb{P}\left( \|q^{-1} P_\Delta^\perp E_\Omega P_\Delta \eta\|_{l^\infty} > (t + \|P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}}) \|\eta\| \right) \leq \gamma \tag{9.4}$$

*whenever*

$$q \geq \left( \frac{4}{t^2} + \frac{2\sqrt{2}}{3t} \sqrt{|\Delta|} \right) \cdot \log(4\omega/\gamma) \cdot \upsilon^2(U),$$

*where $\omega = \tilde{\omega}_{M,U}(|\Delta|, tq, N)$ and $\tilde{\omega}_{M,U}$ is as in (5.2). In addition, if $q = 1$, the left-hand sides of (9.3) and (9.4) are equal to zero.*

**Proposition 9.2** *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $\Delta \subset \mathbb{N}$ with $|\Delta| < \infty$ and $\{N, \dots, 1\} \supset \Omega \sim \mathrm{Ber}(q)$ for some $0 < q \leq 1$. Then, for fixed $\eta \in \mathcal{H}$ and $0 < t, \gamma \leq 1$, we have*

$$\mathbb{P}\left( \left\| \left( q^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta \right) \eta \right\| > \left( t + \|P_\Delta U^* P_N U P_\Delta - P_\Delta\| \right) \|\eta\| \right) \leq \gamma,$$

*provided*

$$q(1-q)^{-1} \geq 4t^{-2} \cdot \upsilon^2(U) \cdot |\Delta|,$$

*and*

$$\log\left(1 + \frac{t}{4}\right) \geq \frac{2K}{t} \max\{q^{-1} - 1, 1\} \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log\left(\frac{3}{\gamma}\right),$$

*where $K$ is the constant in Talagrand's Theorem (Theorem 11.2).*

**Theorem 9.3** *There exists a constant $C > 0$ with the following property. Suppose that $U \in \mathcal{B}(\mathcal{H})$ is an isometry, $\Delta$ is a finite subset of $\mathbb{N}$ and $\{N, \dots, 1\} \supset \Omega \sim \mathrm{Ber}(\theta)$ for some $0 < \theta \leq 1$. Then, for $\epsilon > 0$ and $\gamma > 1$ we have*

$$\mathbb{P}\left( \left\| \theta^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta \right\| \geq \frac{1}{\gamma} + \|P_\Delta U^* P_N U P_\Delta - P_\Delta\| \right) \leq \epsilon, \tag{9.5}$$

*provided*

$$\theta \geq C \cdot \gamma \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log(|\Delta|),$$
$$\theta \geq C \cdot \gamma \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log(C\epsilon^{-1}) \cdot \left(\log\left(1+\frac{1}{4\gamma}\right)\right)^{-1}. \tag{9.6}$$

*If* $\theta = 1$ *then the left-hand side of* (9.5) *is equal to zero.*

With these results in hand, we can now give the proofs of the main results:

*Proof of Theorems 6.1 and 6.3* The set $\Omega \subset \{1, \dots, N\}$ is chosen uniformly at random with $|\Omega| = m$. By Proposition 8.7, it suffices to show that there exists a $\rho \in \mathrm{ran}(U^* P_\Omega)$ such that

$$(i)\ \|\theta^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\| \leq 1/2, \quad (ii)\ \|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq \theta/8,$$
$$(iii)\ \|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2, \tag{9.7}$$

with high probability. As discussed, we may (without loss of generality) replace this way of choosing $\Omega$ with the model that $\{N, \dots, 1\} \supset \Omega \sim \mathrm{Ber}(\theta)$ for $\theta = m/N$ ($\theta$ will have this value throughout the proof). Doing so may only change the constant $C$ in (6.2). Note that, as discussed in Sect. 9.1, the model $\{N, \dots, 1\} \supset \Omega \sim \mathrm{Ber}(\theta)$ is equivalent to choosing $\Omega$ as

$$\Omega = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_\mu, \quad \Omega_j \sim \mathrm{Ber}(q_j),$$

for some $\mu \in \mathbb{N}$ with

$$(1-q_1)(1-q_2)\cdots(1-q_\mu) = (1-\theta). \tag{9.8}$$

The latter model is the one we will use throughout the proof and the specific value of $\mu$ will be chosen later. The theorems will follow if we can show that the conditions in (9.7) occur with probability exceeding $1-\epsilon$, and what follows is a setup to ensure this eventuality. We will focus on (ii) and (iii) in (9.7) first and deal with (i) at the end of the proof. The proof proceeds in a number of steps.

Step I (The construction of $\rho$)**:** Let $\nu$ be a positive number such that $\nu \leq \mu$ and let $\{\alpha_1, \dots, \alpha_\mu\}$ and $\{\beta_1, \dots, \beta_\mu\}$ be sequences of positive numbers. The values of $\mu$, $\nu$, $\{\alpha_i\}_{i=1}^\mu$ and $\{\beta_i\}_{i=1}^\mu$ will be carefully chosen later in the proof. Consider now the following construction of $\rho$ : let

$$Z_0 = \mathrm{sgn}(x_0),$$

and define recursively the sequences $\{Z_i\}_{i=0}^\mu \subset \mathcal{H}$, $\{Y_i\}_{i=1}^\mu \subset \mathcal{H}$ and $\{\Theta_i\}_{i=1}^\mu \subset \mathbb{N}$ as follows. First, let

$$Z_i = \mathrm{sgn}(x_0) - P_\Delta Y_i, \quad Y_i = \sum_{j=1}^{i} q_j^{-1} E_{\Omega_j} Z_{j-1}, \quad i = 1, 2,$$

where $E_{\Omega_j} = U^* P_{\Omega_j} U$, and $\{q_1, \ldots, q_\mu\}$ stem from (9.8). The precise values of the $q_j$'s will be chosen later. Let also $\Theta_1 = \{1\}$ and $\Theta_2 = \{1, 2\}$. Then, for $i \geq 3$, define recursively the following:

$$\Theta_i = \begin{cases} \Theta_{i-1} \cup \{i\} & \text{if } \left\| \left( P_\Delta - q_i^{-1} P_\Delta E_{\Omega_i} P_\Delta \right) Z_{i-1} \right\| \leq \alpha_i \, \| Z_{i-1} \|, \\ & \text{and } \left\| q_i^{-1} P_M P_\Delta^\perp E_{\Omega_i} P_\Delta Z_{i-1} \right\|_{l^\infty} \leq \beta_i \| Z_{i-1} \|, \\ \Theta_{i-1} & \text{otherwise,} \end{cases}$$

$$Y_i = \begin{cases} \sum_{j \in \Theta_i} q_j^{-1} E_{\Omega_j} Z_{j-1} & \text{if } i \in \Theta_i, \\ Y_{i-1} & \text{otherwise,} \end{cases}$$

$$Z_i = \begin{cases} \operatorname{sgn}(x_0) - P_\Delta Y_i & \text{if } i \in \Theta_i, \\ Z_{i-1} & \text{otherwise.} \end{cases} \tag{9.9}$$

Now, let $\{A_i\}_{i=1}^2$ and $\{B_i\}_{i=1}^4$ denote the following events

$$\begin{aligned} A_i : &\quad \left\| \left( P_\Delta - q_i^{-1} P_\Delta E_{\Omega_i} P_\Delta \right) Z_{i-1} \right\| \leq \alpha_i \, \| Z_{i-1} \|, \quad i = 1, 2, \\ B_i : &\quad \left\| q_i^{-1} P_M P_\Delta^\perp E_{\Omega_i} P_\Delta Z_{i-1} \right\|_{l^\infty} \leq \beta_i \| Z_{i-1} \|, \quad i = 1, 2, \\ B_3 : &\quad |\Theta_\mu| \geq \nu, \\ B_4 : &\quad (\cap_{i=1}^2 A_i) \cap (\cap_{i=1}^3 B_i), \end{aligned} \tag{9.10}$$

where $|\Theta_\mu|$ denotes the length of $\Theta_\mu$. Also, let $\tau(j)$ denote the $j$th element in $\Theta_\mu$ (e.g., $\tau(1) = 1$, $\tau(2) = 2$ etc, we also let $\tau(0) = 0$). Finally, define $\rho$ by

$$\rho = \begin{cases} Y_{\tau(\nu)} & \text{if } B_4 \text{occurs,} \\ \operatorname{sgn}(x_0) & \text{otherwise.} \end{cases}$$

Note that $\rho \in \operatorname{ran}(U^* P_\Omega)$ if $B_4$ occurs. Now make the following observations. Since $Z_0 = \operatorname{sgn}(x_0)$ yields, for $i \leq |\Theta_\mu|$, we have

$$\begin{aligned} & Z_{\tau(i)} \\ &= \operatorname{sgn}(x_0) - P_\Delta \left( q_{\tau(1)}^{-1} E_{\Omega_{\tau(1)}} \operatorname{sgn}(x_0) + q_{\tau(2)}^{-1} E_{\Omega_{\tau(2)}} Z_1 + \ldots + q_{\tau(i)}^{-1} E_{\Omega_{\tau(i)}} Z_{\tau(i-1)} \right) \\ &= Z_{\tau(i-1)} - q_{\tau(i)}^{-1} P_\Delta E_{\Omega_{\tau(i)}} P_\Delta Z_{\tau(i-1)} = (P_\Delta - q_{\tau(i)}^{-1} P_\Delta E_{\Omega_{\tau(i)}} P_\Delta) Z_{\tau(i-1)}. \end{aligned} \tag{9.11}$$

Hence, if the event $B_4$ occurs, then

$$\| P_\Delta \rho - \operatorname{sgn}(x_0) \| = \| Z_{\tau(\nu)} \| \leq \sqrt{|\Delta|} \prod_{i=1}^{\nu} \alpha_{\tau(i)}, \tag{9.12}$$

$$\| P_M P_\Delta^\perp \rho \|_{l^\infty} \leq \sum_{i=1}^{\nu} \| q_{\tau(i)}^{-1} P_M P_\Delta^\perp E_{\Omega_{\tau(i)}} Z_{\tau(i-1)} \|_{l^\infty}$$

$$\leq \sum_{i=1}^{v} \beta_{\tau(i)} \|Z_{\tau(i-1)}\| \leq \sqrt{|\Delta|} \sum_{i=1}^{v} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)}, \quad (9.13)$$

and $\rho \in \mathrm{ran}(U^* P_\Omega)$ (note that in the above equation we interpret $\alpha_0 = 1$). We will now show that with a certain choice of parameters $v$, $\{\beta_j\}_{j=1}^{\mu}$ and $\{\alpha_j\}_{j=1}^{\mu}$ then (ii) and (iii) in (9.7) are satisfied when the event $B_4$ occurs. We delay specifying a the value for $\mu$ until Step IV. Let $L \geq 2$ (we will give a value for $L$ in a moment) and

$$\alpha_1 = \alpha_2 = \frac{1}{2 \log_2^{1/2}(L)}, \quad \alpha_i = 1/2, \quad 3 \leq i \leq \mu,$$
$$\beta_1 = \beta_2 = \frac{1}{4\sqrt{|\Delta|}}, \quad \beta_i = \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{4\sqrt{|\Delta|}}, \quad 3 \leq i \leq \mu.$$

It follows that

$$\sqrt{|\Delta|} \prod_{i=1}^{v} \alpha_{\tau(i)} = \frac{\sqrt{|\Delta|}}{2^v \log_2(L)}.$$

Hence, if

$$v = \left\lceil \log_2 \left( 8\theta^{-1} \sqrt{|\Delta|} \right) \right\rceil, \quad (9.14)$$

then it follows by (9.12) that

$$\|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq \theta/8$$

(recall that $L \geq 2$) yielding (ii) in (9.7). Also, after inserting the values of $v$, $\{\beta_j\}_{j=1}^{\mu}$ and $\{\alpha_j\}_{j=1}^{\mu}$ into (9.13) we get:

$$\begin{aligned} \sqrt{|\Delta|} \sum_{i=1}^{v} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)} &= \frac{1}{4} \Bigg( 1 + \frac{1}{2} \frac{1}{\log_2^{1/2}(L)} + \frac{1}{4} \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{\log_2(L)} \\ &\quad + \frac{1}{8} \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{\log_2(L)} + \ldots + \frac{1}{2^{v-1}} \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{\log_2(L)} \Bigg) \\ &\leq \frac{1}{2}, \end{aligned}$$

if we let $L = 4\theta^{-1}\sqrt{|\Delta|}$. Thus, by (9.13) we have

$$\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2,$$

yielding (iii) in (9.7). In particular, we have showed that if $v$, $\{\beta_j\}_{j=1}^{\mu}$ and $\{\alpha_j\}_{j=1}^{\mu}$ are chosen as above, then (ii) and (iii) are satisfied when $B_4$ occurs.

Thus, we have now obtained a means to show that (ii) and (iii) in (9.7) hold with a certain probability. To do this, we will make a careful choice of $\mu$ and then provide

bounds on $\mathbb{P}(B_4^c)$. The way this latter step is carried out is by giving estimates for $\mathbb{P}(A_1^c \cup A_2^c)$, $\mathbb{P}(B_1^c \cup B_2^c)$ and $\mathbb{P}(B_3^c)$. This is the content of Steps II–IV.

Step II: We claim that if $\gamma > 0$ then $\mathbb{P}(A_1^c \cup A_2^c) \leq 2\gamma$, provided $N$, $q_1$, $q_2$ are chosen such that

$$\|P_\Delta U^* P_N U P_\Delta - P_\Delta\| \leq \frac{1}{4\log_2^{1/2}(4\theta^{-1}\sqrt{|\Delta|})}, \tag{9.15}$$

and

$$q_1 = q_2 \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1}\right) + 1\right) \cdot \log\left(\theta^{-1}\sqrt{|\Delta|}\right), \tag{9.16}$$

for some universal constant $C > 0$. Also, if $q_1 = q_2 = 1$, then $\mathbb{P}(A_1^c \cup A_2^c) = 0$.

To deduce the claim, we first observe that by Proposition 5.2 these requirements are well defined. Now note that Proposition 9.2 gives, for $i = 1, 2$ and $0 < t, \gamma \leq 1$ that

$$\mathbb{P}\left(\left\|\left(q_i^{-1} P_\Delta U^* P_{\Omega_i} U P_\Delta - P_\Delta\right) Z_{i-1}\right\| > \left(t + \|P_\Delta U^* P_N U P_\Delta - P_\Delta\|\right) \|Z_{i-1}\|\right) \leq \gamma, \tag{9.17}$$

if

$$q_i(1-q_i)^{-1} \geq 4t^{-2} \cdot \upsilon^2(U) \cdot |\Delta|, \tag{9.18}$$

and

$$\log\left(1 + \frac{t}{4}\right) \geq \frac{2K}{t} \max\{q^{-1} - 1, 1\} \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log\left(\frac{3}{\gamma}\right), \tag{9.19}$$

where $K$ is the constant in Talagrand's Theorem (Theorem 11.2). Thus, by (9.17), (9.18) and (9.19) (and a small computation using Taylor's Theorem), we can choose $t = \alpha_i/2$ and deduce the first assertion in Step II. As for the second assertion, clearly, if $q_1 = q_2 = 1$ then the right-hand side of (9.17) is zero as required.

Step III: We claim that $\mathbb{P}(B_1^c \cup B_2^c) \leq 2\gamma$ for $\gamma > 0$ if $N$, $q_1$ and $q_2$ are chosen such that

$$\|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}, \tag{9.20}$$

and

$$q_1 = q_2 \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1} M\right) + 1\right), \tag{9.21}$$

for some universal constant $C > 0$. Also, if $q_1 = q_2 = 1$, then $\mathbb{P}(B_1^c \cup B_2^c) = 0$.

To prove the claim, recall that Proposition 9.1 gives, for $i = 1, 2$ and $t, \gamma > 0$, that

$$\mathbb{P}\left(\left\|q_i^{-1} P_M P_\Delta^\perp E_{\Omega_i} P_\Delta Z_{i-1}\right\|_{l^\infty} > \left(t + \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}}\right) \|Z_{i-1}\|\right) \leq \gamma,$$

if

$$q_i \geq \left(\frac{4}{t^2} + \frac{2\sqrt{2}}{3t}\sqrt{|\Delta|}\right) \cdot \log\left(\frac{4}{\gamma}|\Delta^c \cap \{1, \ldots, M\}|\right) \cdot \upsilon^2(U).$$

Choosing $t = \beta_i/2$ automatically yields the first assertion in Step III. Also, the fact that $\mathbb{P}(B_1^c \cup B_2^c) = 0$, when $q_1 = q_2 = 1$, follows automatically from Proposition 9.1.

Step IV: We claim that $\mathbb{P}(B_3^c) \leq \gamma$, for $\gamma > 0$ if $\mu$ (recall $\mu$ and $\nu$ from Step I), $N$ and $\{q_3, \ldots, q_\mu\}$ are chosen such that

$$\mu = 8\lceil 3\nu + \log(\gamma^{-1})\rceil, \tag{9.22}$$

$$\|P_\Delta U^* P_N U P_\Delta - P_\Delta\| \leq 1/4, \tag{9.23}$$

and

$$\|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{8\sqrt{|\Delta|}}, \tag{9.24}$$

and also $q_3 = q_4 = \ldots = q_\mu = q$, where

$$q \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\frac{\log(M)}{\log_2(4\theta^{-1}\sqrt{|\Delta|})} + 1\right), \tag{9.25}$$

for some universal constant $C > 0$. Also, if $q_3 = q_4 = \ldots = q_\mu = 1$, then $\mathbb{P}(B_3^c) = 0$.

To prove the claim, we start by determining the condition (9.22) on $\mu$. Define the random variables $X_1, \ldots X_{\mu-2}$ by

$$X_j = \begin{cases} 0 & Z_{j+2} \neq Z_{j+1}, \\ 1 & Z_{j+2} = Z_{j+1}. \end{cases}$$

We immediately observe that

$$\mathbb{P}(B_3^c) = \mathbb{P}(|\Theta_\mu| < \nu) = \mathbb{P}(X_1 + \ldots + X_{\mu-2} > \mu - \nu).$$

Unfortunately, the random variables $X_1, \ldots X_{\mu-2}$ are not independent, which prevents the use of standard tools such as Chernoff bounds or Hoeffding's inequality. Note that the use of such tools has become popular when dealing with the golfing scheme; however, the dependency issue prevents this in the general case. See also [37] for a discussion of this phenomenon. To deal with the dependency issue, we proceed as follows:

$$\begin{aligned} &\mathbb{P}(X_1 + \ldots + X_{\mu-2} > \mu - \nu) \\ &\leq \sum_{l=1}^{\binom{\mu-2}{\mu-\nu}} \mathbb{P}(X_{\pi(l)_1} = 1, X_{\pi(l)_2} = 1, \ldots, X_{\pi(l)_{\mu-\nu}} = 1) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{l=1}^{\binom{\mu-2}{\mu-\nu}} \mathbb{P}(X_{\pi(l)_{\mu-\nu}} = 1 \mid X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-1}} = 1) \\
&\quad \times \mathbb{P}(X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-1}} = 1) \\
&= \sum_{l=1}^{\binom{\mu-2}{\mu-\nu}} \mathbb{P}(X_{\pi(l)_{\mu-\nu}} = 1 \mid X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-1}} = 1) \\
&\quad \times \mathbb{P}(X_{\pi(l)_{\mu-\nu-1}} = 1 \mid X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-2}} = 1) \cdots \mathbb{P}(X_{\pi(l)_1} = 1) \quad (9.26)
\end{aligned}
$$

where $\pi : \{1, \ldots, \binom{\mu-2}{\mu-\nu}\} \to \mathbb{N}^{\mu-\nu}$ ranges over all $\binom{\mu-2}{\mu-\nu}$ ordered subsets of $\{1, \ldots, \mu-2\}$ of size $\mu-\nu$. Let $P > 0$ (a specific value for $P$ will be assigned later) be such that

$$
\begin{aligned}
&P \geq \mathbb{P}(X_{\pi(l)_1} = 1), \quad P \geq \mathbb{P}(X_{\pi(l)_{\mu-\nu-j}} = 1 \mid X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-(j+1)}} = 1), \\
&\quad l = 1, \ldots, \binom{\mu-2}{\mu-\nu}, \quad j = 0, \ldots, \mu-\nu-2, \qquad (9.27)
\end{aligned}
$$

then, by (9.26),

$$
\mathbb{P}(X_1 + \ldots + X_{\mu-2} > \mu-\nu) \leq \binom{\mu-2}{\mu-\nu} P^{\mu-\nu}. \qquad (9.28)
$$

We now choose $P = 1/4$ and claim that

$$
\mathbb{P}(X_1 + \ldots + X_{\mu-2} > \mu-\nu) \leq \exp\left(-2(\mu-2)t^2 + (\mu-\nu)\log\left(\frac{(\mu-2)e}{\mu-\nu}\right)\right), \qquad (9.29)
$$

where $t = (\mu-\nu)/(\mu-2) - P \geq 0$. To see this, note that it is a straightforward calculus exercise to show that

$$
\frac{\mu-\nu}{\mu-2}\log(P) \leq -2\left(\frac{\mu-\nu}{\mu-2} - P\right)^2, \quad \frac{\mu-\nu}{\mu-2} \in [P, 1].
$$

By using the fact that $\binom{\mu-2}{\mu-\nu} \leq \left(\frac{(\mu-2)e}{\mu-\nu}\right)^{(\mu-\nu)}$, we immediately get

$$
\binom{\mu-2}{\mu-\nu} P^{\mu-\nu} \leq \exp\left(-2(\mu-2)t^2 + (\mu-\nu)\log\left(\frac{(\mu-2)e}{\mu-\nu}\right)\right),
$$

which, after recalling (9.28), yields (9.29). So, by (9.29) we have that

$$
\mathbb{P}\left(\sum_{i=1}^{\mu-2} X_i \geq \mu-\nu\right) \leq \gamma \qquad (9.30)
$$

whenever

$$\mathrm{e}^{-2(\mu-2)t^2+(\mu-\nu)(\log(\frac{\mu-2}{\mu-\nu})+1)} \leq \gamma.$$

Hence, after observing that $\log((\mu-2)/(\mu-\nu))+1 \leq (\mu-2)/(\mu-\nu)$, we deduce that (9.30) is satisfied whenever

$$\mu \geq x, \quad (x-2)\left(\frac{x-\nu}{x-2}-P\right)^2 - \log\left(\gamma^{-1/2}\right) - \frac{x-2}{2} = 0 \tag{9.31}$$

where $x$ is the largest root satisfying (9.31). In particular, we have shown that $\mathbb{P}(B_3^c) \leq \gamma$ when (9.31) is satisfied. The choice of $P$ yields $x \leq 8\lceil 3\nu + \log(\gamma^{-1/2})\rceil$. Hence (9.22) yields (9.31).

For the rest of the proof of Step IV, we need to determine the conditions on $N$ and $\{q_3, \ldots, q_\mu\}$ such that (9.27) is satisfied with $P = 1/4$. Note that $X_k = 1$ if and only if one of the following events occur:

$$\begin{aligned} D_1 &: \quad \left\| \left(P_\Delta - q_j^{-1} P_\Delta E_{\Omega_j} P_\Delta\right) Z_{j-1} \right\| > \alpha_j \left\| Z_{j-1} \right\|, \quad j = k+2, \\ D_2 &: \quad \left\| q_j^{-1} P_M P_\Delta^\perp E_{\Omega_j} P_\Delta Z_{j-1} \right\|_{l^\infty} > \beta_j \| Z_{j-1} \|, \quad j = k+2. \end{aligned} \tag{9.32}$$

Observe that we may argue exactly as in the proof of Step II (via Proposition 9.2) and, regardless of the vector $Z_{j-1}$, we deduce that $\mathbb{P}(D_1) \leq 1/8$ when $N$ and $q_j$ are chosen such that

$$\begin{aligned} &\| P_\Delta U^* P_N U P_\Delta - P_\Delta \| \leq \alpha_j/2, \\ &\quad q_j \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \alpha_j^{-2} \cdot (\log(24)+1), \quad j = k+2, \end{aligned} \tag{9.33}$$

for some universal constant $C > 0$. Observe also that we may argue exactly as in the proof of Step III (via Proposition 9.1) and, regardless of the vector $Z_{j-1}$, we may deduce that $\mathbb{P}(D_2) \leq 1/8$ when $N$ and $q_j$ are chosen such that

$$\begin{aligned} &\| P_M P_\Delta^\perp U^* P_N U P_\Delta \|_{\mathrm{mr}} \leq \beta_j/2, \\ &\quad q_j \geq C \cdot \upsilon^2(U) \cdot \left(\frac{1}{\beta_j^2} + \frac{1}{\beta_j}\sqrt{|\Delta|}\right) \cdot (\log(32M)+1), \quad j = k+2, \end{aligned} \tag{9.34}$$

for some universal constant $C > 0$. Thus, for $l = 1, \ldots, \binom{\mu-2}{\mu-\nu}$ and $i = 0, \ldots, \mu-\nu-2$, by letting $k = \pi(l)_{\mu-\nu-i}$, we find that

$$\begin{aligned} &\mathbb{P}(X_{\pi(l)_{\mu-\nu-i}} = 1 \,|\, X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-(i+1)}} = 1) \\ &\quad \leq \mathbb{P}(D_1 \cup D_2 \,|\, X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-(i+1)}} = 1) \leq P, \end{aligned}$$

and similarly, by letting $k = \pi(l)_1$ we get that

$$\mathbb{P}(X_{\pi(l)_1} = 1) \leq \mathbb{P}(D_1 \cup D_2) \leq P,$$

whenever (9.33) and (9.34) are satisfied. In particular, (9.33) and (9.34) imply (9.27). But (9.33) and (9.34) follow from (9.23), (9.24) and (9.25) (with a possibly different, however universal, constant $C$) and hence the first part of the claim is proved. The fact that if $q_3 = q_4 = \ldots = q_\mu = 1$ then $\mathbb{P}(B_3^c) = 0$ follows from Propositions 9.2 and 9.1.

Step V: We claim that

$$\mathbb{P}\left(\|P_\Delta \rho - \operatorname{sgn}(x_0)\| > \theta/8 \cup \|P_M P_\Delta^\perp \rho\|_{l^\infty} > 1/2\right) \leq 5\gamma, \tag{9.35}$$

for $\gamma > 0$ when $N \in \mathbb{N}$ and $\theta > 0$ are chosen according to (5.4), (5.5) and

$$\theta \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1}\right) + 1\right) \cdot \log\left(M\theta^{-1}\sqrt{|\Delta|}\right), \tag{9.36}$$

for some universal constant $C > 0$. Also, if $\theta = 1$ then the left-hand side of (9.35) is equal to zero.

To prove this, recall the events $A_1, A_2, B_1, B_2, B_3, B_4$ from Step I. We have already established in Step I that if the event $B_4$ occurs then $\|P_\Delta \rho - \operatorname{sgn}(x_0)\| \leq \theta/8$ and $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2$. It therefore suffices to show that

$$\mathbb{P}\left(B_4^c\right) \leq 5\gamma. \tag{9.37}$$

given the conditions (5.4), (5.5) and (9.36). To do this, we begin by making some observations. First

$$\mathbb{P}\left(B_4^c\right) \leq \mathbb{P}(A_1^c \cup A_2^c) + \mathbb{P}(B_1^c \cup B_2^c) + \mathbb{P}(B_3^c), \tag{9.38}$$

and second

$$q_1 + q_2 + \ldots + q_\mu \geq \theta. \tag{9.39}$$

Recall from Step II we have that $\mathbb{P}(A_1^c \cup A_2^c) \leq 2\gamma$ whenever (9.15) and (9.16) are satisfied. Also, by Step III, $\mathbb{P}(B_1^c \cup B_2^c) \leq 2\gamma$ whenever (9.20) and (9.21) are fulfilled. Finally, from Step IV we have that $\mathbb{P}(B_3^c) \leq \gamma$ provided

$$\mu = 8\left\lceil \log(\gamma^{-1}) + 3\left\lceil \log_2\left(8\theta^{-1}\sqrt{|\Delta|}\right)\right\rceil\right\rceil, \tag{9.40}$$

and (9.23), (9.24) and (9.25) are satisfied. In particular, using (9.38) we find that (9.37) follows from (9.15), (9.16), (9.20), (9.21), (9.23), (9.24) and (9.25). Thus, we must then show that these equations follow from (5.4), (5.5) and (9.36). Now let $q_1 = q_2 = \theta/4$. Then, by (9.36), we have that (9.16) follows (with a possibly different constant), and

similarly (9.21) follows. Let $q = q_3 = \ldots = q_\mu$. By (9.39) and (9.40), we have

$$16q \left\lceil \log(\gamma^{-1}) + 3 \left\lceil \log_2 \left( 8\theta^{-1} \sqrt{|\Delta|} \right) \right\rceil \right\rceil \geq \theta,$$

and hence (9.25) follows. The only thing left to do is to deal with the requirements on $N$. In particular, we need to show that (9.15), (9.20), (9.23) and (9.24) follow when (5.4) and (5.5) are satisfied. Note that (9.23) and (9.24) are weaker than (9.15) and (9.20). Thus, we only need to concentrate on (9.15) and (9.20). To see that (5.4) and (5.5) imply (9.15) and (9.20), observe that

$$P_\Delta U^* P_N U P_\Delta - P_\Delta = P_\Delta (P_M U^* P_N U P_M - P_M) P_\Delta,$$

(since $\Delta \subseteq \{1, \ldots, M\}$) and therefore

$$\| P_\Delta U^* P_N U P_\Delta - P_\Delta \| \leq \| P_M U^* P_N U P_M - P_M \|.$$

Hence (9.15) follows from (5.4). The fact that (9.20) follows from (5.5) is clear. Also, the fact that the left-hand side of (9.35) is equal to zero when $\theta = 1$ follows from Steps II–IV and the observation that when $\theta = 1$ we have $q_1 = \ldots = q_\mu = 1$.

Step VI: We claim that, for $\gamma > 0$,

$$\mathbb{P} \left( \| \theta^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta \| > 1/2 \right) \leq \gamma, \tag{9.41}$$

when $N \in \mathbb{N}$ and $\theta > 0$ are chosen such that

$$\| P_\Delta U^* P_N U P_\Delta - P_\Delta \| \leq 1/4, \quad \theta \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left( \log \left( \gamma^{-1} |\Delta| \right) + 1 \right),$$

for some universal constant $C$. Also, if $\theta = 1$ then the left hand side of (9.41) is equal to zero.

To prove this claim note that, by Theorem 9.3, there is a $K > 0$ such that

$$\mathbb{P} \left( \left\| \theta^{-1} (P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta \right\| \geq \frac{1}{4} + \| P_\Delta U^* P_N U P_\Delta - P_\Delta \| \right) \leq \gamma,$$

provided

$$\theta \geq 4K \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log(|\Delta|),$$

and

$$\theta \geq 4K \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log(C\gamma^{-1}) \cdot \left( \log \left( 1 + \frac{1}{16} \right) \right)^{-1}.$$

This yields the asserted claim. The fact that the left-hand side of (9.41) is equal to zero when $\theta = 1$ is clear.

Step VII: In this final step, we will patch the different parts of the proof together. Recall that our initial goal was to show that (9.7) follows with probability exceeding $1-\epsilon$. Note that in Step V we have shown that if $\gamma > 0$, then (ii) and (iii) in (9.7) are satisfied with probability exceeding $1-5\gamma$, provided (5.4), (5.4) and (9.36) are satisfied. We are thus only left to show that (i) follows with a certain probability. However, we immediately recognize that the conditions in Step VI follow from (5.5) and (9.36), and hence (i) in (9.7) follows with probability exceeding $1-\gamma$. This implies that (i), (ii) and (iii) in (9.7) hold with probability exceeding $1-6\gamma$. By choosing $\gamma$ such that $6\gamma = \epsilon$ we observe that (9.36) follows (with possibly a different $C$) from the conditions in Theorems 6.1 and 6.3. Hence we have finally proved the first assertions in these theorems. The second assertions follow from the fact that $\theta = 1$ when $m = N$ (and hence also $q_1 = \ldots = q_\mu = 1$), and Steps V and VI. □

*Proof of Theorem 6.4* We will follow the recipe from the of proof of Theorem 6.3 almost word for word, pointing out only where the main differences lie. The first such difference is the set of conditions provided by Proposition 8.7. In particular, we must show that there exists a $\rho \in \mathrm{ran}(U^* P_\Omega)$ such that

$$\begin{aligned} &(i)\ \|\theta^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\| \leq 1/2, \quad (ii)\ \|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq \theta/8 \\ &(iii)\ \|P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2, \end{aligned} \tag{9.42}$$

is true with probability exceeding $1-\epsilon$ (note that only (iii) is different from the proof of Theorem 6.3).

Step I: This is almost as in the proof of Theorem 6.3, except that (9.9) is replaced by

$$\Theta_i = \begin{cases} \Theta_{i-1} \cup \{i\} & \text{if } \left\| \left( P_\Delta - q_i^{-1} P_\Delta E_{\Omega_i} P_\Delta \right) Z_{i-1} \right\| \leq \alpha_i \|Z_{i-1}\|, \\ & \text{and } \left\| q_i^{-1} P_\Delta^\perp E_{\Omega_i} P_\Delta Z_{i-1} \right\|_{l^\infty} \leq \beta_i \|Z_{i-1}\|, \\ \Theta_{i-1} & \text{otherwise,} \end{cases}$$

and the events $B_1$ and $B_2$ in (9.10) are now

$$B_j: \quad \left\| q_j^{-1} P_\Delta^\perp E_{\Omega_j} P_\Delta Z_{j-1} \right\|_{l^\infty} \leq \beta_j \|Z_{j-1}\|, \qquad j = 1, 2.$$

Also, (9.13) must be changed to

$$\begin{aligned} \|P_\Delta^\perp \rho\|_{l^\infty} &\leq \sum_{i=1}^{\nu} \|q_{\tau(i)}^{-1} P_\Delta^\perp E_{\Omega_{\tau(i)}} Z_{\tau(i-1)}\|_{l^\infty} \\ &\leq \sum_{i=1}^{\nu} \beta_{\tau(i)} \|Z_{\tau(i-1)}\| \leq \sqrt{|\Delta|} \sum_{i=1}^{\nu} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)}. \end{aligned}$$

Step II: Exactly as in the proof of Theorem 6.1.

Step III: We claim that, for $\gamma > 0$, then $\mathbb{P}(B_1^c \cup B_2^c) \leq 2\gamma$, if $N$, $q_1$ and $q_2$ are chosen such that

$$\|P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}, \tag{9.43}$$

and

$$q_1 = q_2 \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1}\omega_1\right) + 1\right), \tag{9.44}$$

where

$$\omega_1 = \tilde{\omega}_{M,U}(|\Delta|, q_1(8\sqrt{|\Delta|})^{-1}, N),$$

(recall $\tilde{\omega}_{M,U}$ from (5.2)) for some universal constant $C > 0$. Also, if $q_1 = q_2 = 1$, then $\mathbb{P}(B_1^c \cup B_2^c) = 0$. The claim follows exactly as in the proof of Step III in the proof of Theorem 6.1 by using the last part of Proposition 9.1.

Step IV: We claim that, for $\gamma > 0$, then $\mathbb{P}(B_3^c) \leq \gamma$, if $\mu$, (recall $\mu$ and $\nu$ from Step I) $N$ and $\{q_3, \ldots, q_\mu\}$ are chosen according to (9.22), (9.23) and

$$\|P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{8\sqrt{|\Delta|}}, \tag{9.45}$$

and also that $q_3 = q_4 = \ldots = q_\mu = q$, where

$$q \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\frac{\log(\omega_2)}{\log_2(4\theta^{-1}\sqrt{|\Delta|})} + 1\right), \tag{9.46}$$

and

$$\omega_2 = \tilde{\omega}_{M,U}\left(|\Delta|, q\frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{8\sqrt{|\Delta|}}, N\right),$$

(recall $\tilde{\omega}_{M,U}$ from 5.2) for some universal constant $C > 0$. Also, if $q_3 = q_4 = \ldots = q_\mu = 1$, then $\mathbb{P}(B_3^c) = 0$. The argument is almost the same as in the proof of Theorem 6.3, except that the last part of (9.32) should read

$$D_2: \quad \left\|q_j^{-1} P_\Delta^\perp E_{\Omega_j} P_\Delta Z_{j-1}\right\|_{l^\infty} > \beta_j \|Z_{j-1}\|, \qquad j = k+2,$$

and (9.34) should be

$$\begin{aligned} &\|P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \beta_j/2, \\ &q_j \geq C \cdot \upsilon^2(U) \cdot \left(\frac{1}{\beta_j^2} + \frac{1}{\beta_j}\sqrt{|\Delta|}\right) \cdot (\log(32\omega_2) + 1), \quad j = k+2. \end{aligned}$$

Step V: We claim that, for $\gamma > 0$,

$$\mathbb{P}\left(\|P_\Delta \rho - \mathrm{sgn}(x_0)\| > \theta/8 \cup \|P_\Delta^\perp \rho\|_{l^\infty} > 1/2\right) \leq 5\gamma, \tag{9.47}$$

when $N \in \mathbb{N}$ and $\theta > 0$ are chosen according to (5.4), (5.6) and

$$\theta \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1}\right) + 1\right) \cdot \log\left(\omega\theta^{-1}\sqrt{|\Delta|}\right), \tag{9.48}$$

where

$$\omega = \tilde{\omega}_{M,U}(|\Delta|, s, N), \quad s = \frac{\theta}{128\sqrt{|\Delta|}\log(\mathrm{e}^4\gamma^{-1})},$$

and $\tilde{\omega}_{M,U}$ is defined in (5.2), for some universal constant $C > 0$. Also, if $\theta = 1$ then the left-hand side of (9.47) is equal to zero.

The strategy is almost as in the proof of Step V in Theorem 6.1. In particular, we argue by using Steps II–IV that $\mathbb{P}\left(B_4^c\right) \leq 5\gamma$ when (9.15), (9.16), (9.43), (9.44), (9.23), (9.45) and (9.46) are satisfied, and thus (9.47) follows. We then need to show that these equations follow from (5.4), (5.6) and (9.48). To do this, let $q_1 = q_2 = \theta/4$. Then, by (9.48), we have that (9.16) follows with a possibly different constant. To show that (9.44) is implied by (9.48), it suffices to show that $\omega \geq \omega_1$. This will follow by the definition (5.2) of $\tilde{\omega}_{M,U}$ (recall that the mapping $s \mapsto \tilde{\omega}_{M,U}(|\Delta|, s, N)$ is a decreasing function), and by observing that

$$q_1(8\sqrt{|\Delta|})^{-1} > s = \theta\left(128\sqrt{|\Delta|}\log\left(\mathrm{e}^4\gamma^{-1}\right)\right)^{-1}.$$

To show that (9.46) follows from (9.48), it suffices to show that $\omega \geq \omega_2$. To do this (as argued above), it is sufficient to prove that

$$q\frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{8\sqrt{|\Delta|}} \geq s. \tag{9.49}$$

To see why the latter inequality is true, note that

$$q_1 + q_2 + \ldots + q_\mu \geq \theta.$$

So, by recalling the value of $\mu$ (from 9.22) from Step IV and observing that $q = q_3 = \ldots = q_\mu$, we get

$$16q\left\lceil \log(\gamma^{-1}) + 3\left\lceil \log_2\left(8\theta^{-1}\sqrt{|\Delta|}\right)\right\rceil\right\rceil \geq \theta.$$

In particular, it follows that

$$q\log_2\left(4\theta^{-1}\sqrt{|\Delta|}\right) \geq \frac{\theta\log_2\left(4\theta^{-1}\sqrt{|\Delta|}\right)}{16\left(\log\left(\gamma^{-1}\right) + 3\log_2\left(8\theta^{-1}\sqrt{|\Delta|}\right) + 1\right)} > \frac{\theta}{8\log\left(\mathrm{e}^4\gamma^{-1}\right)}. \tag{9.50}$$

Thus, we have shown (9.49).

We are now left with the task of showing that (9.15), (9.43), (9.44), (9.23) and (9.45) follow from (5.4) and (5.6), and this follows by arguing exactly as in Step V in the proof of Theorem 6.1

Step VI and Step VII: Exactly as in the proof of Theorem 6.1. □

## 10 Conclusions and Challenges

This paper provides a new framework for infinite-dimensional CS in Hilbert spaces based on the ideas of generalized sampling. Although we have presented a mathematical analysis, there are several issues not addressed in this paper as well as a number of avenues for further investigations. First, as mentioned, we have also not treated the issue of noise in this paper. Fortunately, this can be done and follows by extending the techniques introduced in this paper. We refer to [7] for details.

Second we note that the results in this paper are nonuniform, i.e., they hold for a fixed signal of a given sparsity rather than for all signals of that sparsity. As mentioned in Sect. 4 and discussed in [15], uniform recovery of sparse signals with nonzero coefficients taking arbitrary locations is not possible in infinite dimensions. Moreover, as explained in [7], this can become an issue even when using finite-dimensional CS techniques. In problems which are inherently infinite-dimensional, arising in applications such as MRI, X-ray CT, electron microscopy and radio interferometry, uniform recovery of all sparse vectors is unrealistic.

Nevertheless, it is of interest to see if the results in this paper can be strengthened to uniform recovery results over a smaller class of signals, such as sparse signals with some fixed sparsity bandwidth $M$, or the sparsity-in-levels class introduced in [7]. Such results could obviously not be based on the standard definition of the restricted isometry property (RIP). However, alternatives such as the RIP in levels [13] may be more suitable for extension to the infinite-dimensional setting.

This aside, another strong motivation for infinite-dimensional CS is the desire to reconstruct functions from a small number of pointwise samples. Although a number of works have studied this from a finite-dimensional perspective [51,52], the underlying infinite dimensionality of this problem can lead to a number of issues in practice [1]. Fortunately, these can be overcome by tackling the infinite-dimensional problem directly using similar ideas to those introduced in this paper. See [1] for details.

**Acknowledgments** The authors would like to thank Akram Aldroubi, Emmanuel Candès, Ron DeVore, David Donoho, Karlheinz Gröchenig, Gerd Teschke, Joel Tropp, Martin Vetterli, Christopher White, Pengchong Yan and Özgür Yilmaz for useful discussions and comments. They would also like to thank Clarice Poon for helping to improve several of the arguments in the proofs, Bogdan Roman for producing the example in Sect. 7.2 and the anonymous referees for their many useful comments and suggestions.

## 11 Appendix

This appendix contains all the proofs not given so far. Before we do this, there are two results that will be crucial. The first is a due to Rudelson [54].

**Lemma 11.1** *(Rudelson) Let $\eta_1, \dots, \eta_M \in \mathbb{C}^n$ and let $\varepsilon_1, \dots \varepsilon_M$ be independent Bernoulli variables taking values $1, -1$ with probability $1/2$. Then*

$$\mathbb{E}\left(\left\|\sum_{i=1}^{M} \varepsilon_i \bar{\eta}_i \otimes \eta_i\right\|\right) \leq \frac{3}{2}\sqrt{p} \max_{i \leq M} \|\eta_i\| \sqrt{\left\|\sum_{i=1}^{M} \bar{\eta}_i \otimes \eta_i\right\|}$$

*where $p = \max\{2, 2\log(n)\}$.*

Note that the original lemma in [54] does not apply in this case. Actually, we need the complex version proved in [61]. We will, however, still refer to it as Rudelson's Lemma. The following theorem is also indispensable:

**Theorem 11.2** *(Talagrand [45, 59]) There exists a number $K$ with the following property. Consider $n$ independent random variables $X_i$ valued in a measurable space $\Omega$. Let $\mathcal{F}$ be a (countable) class of measurable functions on $\Omega$ and consider the random variable $Z = \sup_{f \in \mathcal{F}} \sum_{i \leq n} f(X_i)$. Let*

$$S = \sup_{f \in \mathcal{F}} \|f\|_\infty, \quad V = \sup_{f \in \mathcal{F}} \mathbb{E}\left(\sum_{i \leq n} f(X_i)^2\right).$$

*If $\mathbb{E}(f(X_i)) = 0$ for all $f \in \mathcal{F}$ and $i \leq n$, then, for each $t > 0$, we have*

$$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq t) \leq 3\exp\left(-\frac{1}{K}\frac{t}{S}\log\left(1 + \frac{tS}{V + S\mathbb{E}(\overline{Z})}\right)\right),$$

*where $\overline{Z} = \sup_{f \in \mathcal{F}} |\sum_{i \leq n} f(X_i)|$.*

Note that we deliberately forgo the use of any vector/matrix Bernstein inequalities in the proofs that follow, and instead use Talagrand's result. This allows for more flexibility in the infinite-dimensional setting.

We next present the proofs of Propositions 8.5 and 8.6. For this, it is useful to have a result about the existence of unique minimizers. The finite-dimensional version of the following proposition has become standard for showing existence of unique minimizers for finite-dimensional problems found in CS (see, e.g., [20, Lem. 2.1] or [34, Thm. 4.26]). Fortunately, the extension to infinite dimensions is rather straightforward:

**Proposition 11.3** *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be unitary and let $\Omega, \Delta \subset \mathbb{N}$ be such that $|\Omega|, |\Delta| < \infty$. Suppose that $x_0 \in \mathcal{H}$ with $\operatorname{supp}(x_0) = \Delta$ and consider the optimization problem*

$$\inf_{\eta \in \mathcal{H}} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega U x_0. \tag{11.1}$$

*Suppose that there exists a vector $\rho \in \mathcal{H}$ such that*

(i) $\rho = U^* P_\Omega \eta$ *for some* $\eta \in \mathcal{H}$
(ii) $\langle \rho, e_j \rangle = \langle \operatorname{sgn}(x_0), e_j \rangle$, $j \in \Delta$

(iii) $|\langle \rho, e_j \rangle| < 1$, $j \notin \Delta$,

*and in addition that $P_\Omega U P_\Delta : P_\Delta \mathcal{H} \to P_\Omega \mathcal{H}$ has full rank, then $x_0$ is the unique minimizer of (11.1). If $U$ and $x_0$ are real the converse is also true.*

*Proof* By assumption, there is a $\rho \in l^\infty(\mathbb{N})$ such that $\rho = U^* P_\Omega y$ for some $y \in P_\Omega \mathcal{H}$ and $\|\rho\|_{l^\infty} \leq 1$. Also, by (ii)

$$\mathrm{Re}\big(\langle P_\Omega U P_\Delta x_0, y\rangle\big) = \mathrm{Re}\big(\langle x_0, P_\Delta \rho\rangle\big) = \sum_{j\in\Delta} \mathrm{sign}\big(\langle x_0, e_j\rangle\big)\langle x_0, e_j\rangle = \|x_0\|_{l^1}.$$

Thus, by using duality (recall Proposition 8.3), in particular the fact that $P_\Omega U : \mathcal{H} \to P_\Omega \mathcal{H}$ is onto (this follows since $U$ is unitary) and that

$$\inf\{\|x\|_{l^1} : P_\Omega U x = P_\Omega U x_0\} = \sup\{\mathrm{Re}\big(\langle P_\Omega U x_0, y\rangle\big) : \|U^* P_\Omega y\|_{l^\infty} \leq 1\},$$

it follows that $x_0$ is a minimizer. But $\langle \rho, e_j\rangle < 1$ for $j \notin \Delta$ so if $\xi$ is another minimizer then $\mathrm{supp}(\xi) = \Delta$. However, $P_\Omega U P_\Delta$ has full rank, so $\xi = x_0$.

As for the converse in the real case, suppose that $x_0$ is the unique minimizer. Then, for all sufficiently large $n$, $x_0$ is the unique minimizer to the finite-dimensional optimization problem $\inf\{\|x\|_{l^1} : x \in P_n\mathcal{H},\ P_\Omega U P_n x = P_\Omega U x_0\}$. Proposition 11.3 is well known to be true in finite dimensions [34]. It follows that there is a $y_n$ such that, for $\rho_n = P_n U^* P_\Omega y_n$, we have $\langle \rho_n, e_j\rangle < 1$ when $j \notin \Delta$ and $j \leq n$, and $\langle \rho_n, e_j\rangle = \mathrm{sgn}(\langle x_0, e_j\rangle)$ for $j \in \Delta$. It is easy to see that there is a constant $M < \infty$ such that $\|y_n\|_{l^\infty} \leq M$ for all large $n$. Now we can define $\rho = U^* P_\Omega y_n$. Then $\rho = \rho_n + P_n^\perp U^* P_\Omega y_n$, and thus $\rho$ satisfies the requirements (i), (ii) and (iii) for large $n$. □

*Proof of Proposition 8.5* Let $\alpha = |\Delta|$ and also $\omega = \{\omega_j\}_{j=1}^\alpha$, where $\omega_j \in \mathbb{C}$. Now define

$$V_\omega = I_{\Delta^c} \oplus S_\omega : P_\Delta^\perp \mathcal{H} \oplus P_\Delta \mathcal{H} \to P_\Delta^\perp \mathcal{H} \oplus P_\Delta \mathcal{H}, \tag{11.2}$$

where $S_\omega = \mathrm{diag}\left(\{\omega_j\}_{j=1}^\alpha\right)$ on $P_\Delta\mathcal{H}$ and $I_{\Delta^c}$ is the identity on $P_\Delta^\perp\mathcal{H}$. Define $U(\omega) = U V_\omega$. Note that to prove the proposition it suffices to show that $V_\omega x$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U(\omega) x\}$ for all $\omega$, where

$$\omega \in \Lambda = \left\{\left(\mathrm{e}^{i\theta_1}, \ldots, \mathrm{e}^{i\theta_\alpha}\right) \in \mathbb{C}^\alpha : \theta_j \in [0, 2\pi), 1 \leq j \leq \alpha\right\}. \tag{11.3}$$

Indeed, if the assertion is true, Proposition 11.3 yields that every real $\tilde{x} \in l^2(\mathbb{N})$ with $\mathrm{supp}(\tilde{x}) = \Delta$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \tilde{x}\}$. Thus, for any $y \in l^2(\mathbb{N})$ such that $\mathrm{supp}(y) = \Delta$ choose $\omega \in \Lambda$ and a real $\tilde{x} \in l^2(\mathbb{N})$ such that $y = V_\omega \tilde{x}$. Then, by using the assertion above for $\tilde{x}$ we have proved the proposition.

To prove the assertion, note that if $\omega \in \Lambda$, then $V_\omega$ is clearly unitary and also an isometry on $l^1(\mathbb{N})$. Thus, it is easy to see that $V_\omega \zeta$ is a minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U(\omega) x\}$ if and only if $\zeta$ is a minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U(\omega)\eta = P_\Omega U(\omega) x\}$. We will therefore consider the latter minimization problem and show that

$x$ is the unique minimizer of that problem for all $\omega \in \Lambda$. To do that, it suffices, by Proposition 11.3 and the fact that $U(\omega)$ is unitary to show that there exists a vector $\rho \in l^2(\mathbb{N})$ such that

$$P_{\Omega^c} U(\omega)\rho = 0, \quad P_\Delta \rho = \operatorname{sgn}(x), \quad \|P_{\Delta^c}\rho\|_{l^\infty} < 1. \tag{11.4}$$

Now, for $\epsilon > 0$ (we will specify the value of $\epsilon$ later), define the function $\varphi : \cup_{a\in\Lambda}\mathcal{B}(a,\epsilon) \to \mathbb{R}_+$, where $\mathcal{B}(a,\epsilon)$ denotes the $\epsilon$-ball around $a$, in the following way. Let

$$W = I_\Delta \oplus P_{\Omega^c} U P_{\Delta^c} : P_\Delta \mathcal{H} \oplus P_{\Delta^c}\mathcal{H} \to P_\Delta \mathcal{H} \oplus P_{\Omega^c}\mathcal{H},$$

and define

$$\varphi(\omega) = \inf\{\|P_{\Delta^c}\rho\|_{l^\infty} : W\rho = \iota_\Delta^* \operatorname{sgn}(x) \oplus -P_{\Omega^c} U(\omega) P_\Delta \iota_\Delta^* \operatorname{sgn}(x)\},$$

where $\iota_\Delta : P_\Delta l^2(\mathbb{N}) \to l^2(\mathbb{N})$ is the inclusion operator. Then (11.4) is satisfied if and only if $\varphi(\omega) < 1$. Thus, to show (11.4) we must show that $\varphi(\omega) < 1$ for all $\omega \in \Lambda$.

Suppose for the moment that $\epsilon$ is chosen such that $\varphi$ is defined on its domain. We will show that $\varphi$ is continuous. For this, it suffices to show that $\varphi$ is continuous on $\mathcal{B}(a,\epsilon)$ for $a \in \Lambda$. Note that, by the fact that $\mathcal{B}(a,\epsilon)$ is open it is enough to show that $\varphi$ is convex. To see that $\varphi$ is convex, let $\omega_1, \omega_2 \in \mathcal{B}(a,\epsilon)$ and $t \in (0,1)$. Also let $\xi, \eta \in l^2(\mathbb{N})$ be such that

$$\begin{aligned} W\xi &= \iota_\Delta^* \operatorname{sgn}(x) \oplus -P_{\Omega^c} U(\omega_1) P_\Delta \iota_\Delta^* \operatorname{sgn}(x), \\ W\eta &= \iota_\Delta^* \operatorname{sgn}(x) \oplus -P_{\Omega^c} U(\omega_2) P_\Delta \iota_\Delta^* \operatorname{sgn}(x). \end{aligned}$$

Note that the existence of such vectors is guaranteed by the assumption that $\varphi$ is defined on its domain. Now

$$\varphi(t\omega_1 + (1-t)\omega_2) \le \|P_{\Delta^c}(t\xi + (1-t)\eta)\|_{l^\infty} \le t\|P_{\Delta^c}\xi\|_{l^\infty} + (1-t)\|P_{\Delta^c}\eta\|_{l^\infty}.$$

Thus, taking infimum on the right-hand side yields $\varphi(t\omega_1 + (1-t)\omega_2) \le t\varphi(\omega_1) + (1-t)\varphi(\omega_2)$, as required. Returning to the question of the domain of $\varphi$, note that if $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta|_{P_\Delta l^2(\mathbb{N})}$ is invertible, then $(P_\Omega U(\omega) P_\Delta)^* P_\Omega U(\omega) P_\Delta|_{P_\Delta l^2(\mathbb{N})}$ is invertible if $\|U(\tilde{\omega}) - U(\omega)\|$ is small and $\tilde{\omega} \in \Lambda$. Letting

$$\rho = U(\omega)^* P_\Omega U(\omega) P_\Delta ((P_\Omega U(\omega) P_\Delta)^* P_\Omega U(\omega) P_\Delta|_{P_\Delta l^2(\mathbb{N})})^{-1} \operatorname{sgn}(x)$$

we get

$$P_{\Omega^c} U P_{\Delta^c} \rho = -P_{\Omega^c} U(\omega) P_\Delta \operatorname{sgn}(x).$$

Thus, $\varphi$ is defined on its domain for small $\epsilon$.

Let $\Gamma$ denote the subset of all $\omega \in \Lambda$ such that $x$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U(\omega)\eta = P_\Omega U(\omega)x\}$. Note that $\Gamma$ is closed. Indeed, if $\omega \in \overline{\Gamma}$ and

$\{\omega_n\} \subset \Gamma$ is a sequence such that $\omega_n \to \omega$ then $\omega \in \Gamma$. To see that, observe that since $\{U, \Omega, \Delta\}$ is weakly $f$ stable, it follows that for $\xi \in l^2(\mathbb{N})$ satisfying

$$\|\xi\|_{l^1} = \inf\{\|\eta\|_{l_1} : P_\Omega U(\omega)\eta = P_\Omega U(\omega)x\}$$

we have

$$\|\xi - x\|_{l^1} \leq f(\|\omega - \omega_n\|_{l^\infty}), \quad \forall n \in \mathbb{N}.$$

Thus, $\xi = x$ and hence $\omega \in \Gamma$.

Note also that $\Gamma$ is open. Indeed, for if $\tilde{\omega} \in \Gamma$ then there exist $\rho \in \mathcal{H}$ such that $\rho$ satisfies (11.4) (with $\omega$ replaced by $\tilde{\omega}$) e.g., $\varphi(\tilde{\omega}) < 1$. But, by continuity of $\varphi$ it follows that $\varphi$ is strictly less than one on a neighborhood of $\tilde{\omega}$. Since $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta|_{P_\Delta l^2(\mathbb{N})}$ is invertible, it is easy to see that $P_\Omega U(\omega) P_\Delta)^* P_\Omega U(\omega) P_\Delta|_{P_\Delta l^2(\mathbb{N})}$ is invertible, for all $\omega \in \Lambda$. Thus it follows by Proposition 11.3 that (11.4) is satisfied for all $\omega \in \Lambda$ in a neighborhood of $\tilde{\omega}$ and hence $\Gamma$ is open.

The fact that $\Gamma$ is open and closed yields that either $\Gamma = \emptyset$ or $\Gamma = \Lambda$. The fact that $\{1, \ldots, 1\} \in \Gamma$ by assumption yields the theorem. □

*Proof of Proposition 8.6* Let $V_\omega$ and $\Lambda$ be defined as in (11.2) and (11.3), respectively. Suppose that $y \in l^2(\mathbb{N})$ is real with $\text{supp}(y) = \Delta$. Then, by assumption, $V_\omega y$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U V_\omega y\}$, when $V_\omega$ is real. Thus, by Proposition 11.3 it follows that there exists a $\rho_\omega \in l^2(\mathbb{N})$ such that

$$P_{\Omega^c} U \rho_\omega = 0, \quad P_\Delta \rho_\omega = \text{sgn}(V_\omega y), \quad \|P_{\Delta^c} \rho_\omega\|_{l^\infty} < 1. \tag{11.5}$$

Let $\beta = \max_{\omega \in \Lambda}\{\|P_{\Delta^c}\rho_\omega\|_{l^\infty}, \omega \text{ is real}\}$. It is clear that $\beta < 1$. Thus, for every $y \in \mathcal{H}$ with $\text{supp}(y) = \Delta$ there exists $\rho_\omega \in l^2(\mathbb{N})$ satisfying (11.5) where $\|P_{\Delta^c}\rho_\omega\|_{l^\infty} \leq \beta$. It is now easy to show that (see the proof of Lemma 2.1 in [21]) there exists a constant $C > 0$ (depending on $\beta$) such that, if $\xi \in l^2(\mathbb{N})$, $\text{supp}(\xi) = \Delta$, is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \xi\}$, $\zeta \in l^2(\mathbb{N})$ and $x$ is a minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \zeta\}$ then $\|P_{\Delta^c} x\|_{l_1} \leq C\|\xi - \zeta\|_{l_1}$. Thus, since

$$P_\Omega U P_\Delta(x - \xi) = P_\Omega U(\zeta - \xi) - P_\Omega U P_{\Delta^c} x,$$

and $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta|_{P_\Delta \mathcal{H}}$ is invertible, the proposition follows. □

*Proof of Proposition 9.1* Without loss of generality, we may assume that $\|\eta\| = 1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = q$. We will split the proof into two steps, where we will prove the finite-dimensional part of the proposition in Step I, and then tweak these ideas to fit the infinite-dimensional part of the proposition in Step II.

Step I: We start by noting that, since $U$ is an isometry, we have

$$q^{-1} P_M P_\Delta^\perp E_\Omega P_\Delta \eta = q^{-1} \sum_{j=1}^N P_M P_\Delta^\perp U^* \delta_j (e_j \otimes e_j) U P_\Delta \eta$$

$$= q^{-1} \sum_{j=1}^{N} P_M P_\Delta^\perp U^*(\delta_j - q)(e_j \otimes e_j) U P_\Delta \eta$$
$$+ P_M P_\Delta^\perp U^* P_N^\perp U P_\Delta \eta. \tag{11.6}$$

Our goal is to eventually use Bernstein's inequality, and the following is therefore a setup to do so. For $1 \le j \le N$, define the random variables

$$Y_j = q^{-1} P_M P_\Delta^\perp U^*(\delta_j - q)(e_j \otimes e_j) U P_\Delta \eta,$$
$$X_j^i = \langle q^{-1} U^*(\delta_j - q)(e_j \otimes e_j) U P_\Delta \eta, e_i \rangle, \quad i \in \Delta^c \cap \{1, \dots, M\}.$$

Thus, for $s > 0$ it follows from (11.6) that

$$\mathbb{P}\left( \left\| q^{-1} P_M P_\Delta^\perp E_\Omega P_\Delta \eta \right\|_{l^\infty} > s \right) = \mathbb{P}\left( \left\| \sum_{j=1}^{N} Y_j + P_M P_\Delta^\perp U^* P_N^\perp U P_\Delta \eta \right\|_{l^\infty} > s \right)$$
$$\le \sum_{i \in \Delta^c \cap \{1,\dots,M\}} \mathbb{P}\left( \left| \sum_{j=1}^{N} X_j^i + \langle P_M P_\Delta^\perp U^* P_N^\perp U P_\Delta \eta, e_i \rangle \right| > s \right)$$
$$\le \sum_{i \in \Delta^c \cap \{1,\dots,M\}} \mathbb{P}\left( \left| \sum_{j=1}^{N} X_j^i \right| > s - \| P_M P_\Delta^\perp U^* P_N U P_\Delta \|_{\mathrm{mr}} \right),$$

where we have used the fact that $U$ is an isometry and hence

$$P_M P_\Delta^\perp U^* P_N U P_\Delta = -P_M P_\Delta^\perp U^* P_N^\perp U P_\Delta.$$

Thus, by choosing $s = t + \| P_M P_\Delta^\perp U^* P_N U P_\Delta \|_{\mathrm{mr}}$ we deduce that

$$\mathbb{P}\left( \left\| q^{-1} P_M P_\Delta^\perp E_\Omega P_\Delta \eta \right\|_{l^\infty} > t + \| P_M P_\Delta^\perp U^* P_N U P_\Delta \|_{\mathrm{mr}} \right)$$
$$\le \sum_{i \in \Delta^c \cap \{1,\dots,M\}} \mathbb{P}\left( \left| \sum_{j=1}^{N} X_j^i \right| > t \right). \tag{11.7}$$

To estimate the right-hand side of (11.7), we shall use Bernstein's inequality, and in order to do that, we need a couple of observations. First note that

$$\mathbb{E}\left( |X_j^i|^2 \right) = q^{-2} \mathbb{E}\left( |\langle U P_\Delta \eta, (\delta_j - q)(e_j \otimes e_j) U e_i \rangle|^2 \right)$$
$$= q^{-2} \mathbb{E}\left( (\delta_j - q)^2 \right) |\langle U P_\Delta \eta, e_j \rangle \langle U e_i, e_j \rangle|^2$$
$$= q^{-1}(1-q) |\langle U P_\Delta \eta, e_j \rangle \langle U e_i, e_j \rangle|^2, \quad i \in \Delta^c \cap \{1, \dots, M\}.$$

Thus

$$\sum_{j=1}^{N} \mathbb{E}\left(|X_j^i|^2\right) \leq q^{-1}(1-q)\|\eta\|^2 \upsilon^2(U) = q^{-1}(1-q)\upsilon^2(U), \quad i \in \Delta^c \cap \{1,\ldots,M\}. \tag{11.8}$$

Also, observe that

$$|X_j^i| = q^{-1}|(\delta_j - q)||\langle \eta, P_\Delta U^*(e_j \otimes e_j)U e_i\rangle| \leq \max\{(1-q)/q, 1\}\upsilon^2(U)\sqrt{|\Delta|}, \tag{11.9}$$

for $1 \leq j \leq N$ and $i \in \Delta^c \cap \{1,\ldots,M\}$. Now applying Bernstein's inequality to $\mathrm{Re}(X_1^i),\ldots,\mathrm{Re}(X_N^i)$ and $\mathrm{Im}(X_1^i),\ldots,\mathrm{Im}(X_N^i)$, we get that

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{j=1}^{N} X_j^i\right| > t\right)\\ &\quad\leq 4\exp\left(-\frac{t^2/4}{q^{-1}(1-q)\upsilon^2(U) + \max\{q^{-1}(1-q),1\}\upsilon^2(U)\sqrt{|\Delta|}t/3\sqrt{2}}\right),\end{aligned} \tag{11.10}$$

for all $i \in \Delta^c \cap \{1,\ldots,M\}$. Thus, by invoking (11.10) and (11.7) it follows that

$$\mathbb{P}\left(\left\|q^{-1}P_M P_\Delta^\perp E_\Omega P_\Delta \eta\right\|_{l^\infty} > t + \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}}\right) \leq \gamma$$

when

$$q \geq \left(\frac{4}{t^2} + \frac{2\sqrt{2}}{3t}\sqrt{|\Delta|}\right)\log\left(\frac{4}{\gamma}|\Delta^c \cap \{1,\ldots,M\}|\right)\upsilon^2(U).$$

The first part of the proposition now follows. The fact that the left-hand side of (9.3) is zero when $q = 1$ is clear from (11.8) and (11.9).

Step II: To prove the second part of the proposition, we will use the same ideas; however, we are now faced with the problem that $P_\Delta^\perp E_\Omega P_\Delta \eta$ (contrary to $P_M P_\Delta^\perp E_\Omega P_\Delta \eta$) has infinitely many components. This is an obstacle since the proof of the bound on $P_M P_\Delta^\perp E_\Omega P_\Delta \eta$ was based on bounding the probability of the deviation of every component of $P_M P_\Delta^\perp E_\Omega P_\Delta \eta$, and thus, if there are infinitely many components to take care of, the task would be impossible. To overcome this obstacle, we proceed as follows. Note that, just as argued in the previous case in Step I, we have that

$$\begin{aligned}q^{-1}P_\Delta^\perp E_\Omega P_\Delta \eta &= \sum_{j=1}^{N} \widetilde{Y}_j + P_\Delta^\perp U^* P_N^\perp U P_\Delta \eta,\\ \widetilde{Y}_j &= q^{-1}P_\Delta^\perp U^*(\delta_j - q)(e_j \otimes e_j)U P_\Delta \eta.\end{aligned} \tag{11.11}$$

Define (as we did above) the random variables

$$X_j^i = \langle q^{-1} U^*(\delta_j - q)(e_j \otimes e_j) U P_\Delta \eta, e_i \rangle, \quad i \in \Delta^c.$$

Note that we now have infinitely many $X_j^i$'s. However, suppose for a moment that for every $t > 0$ there exists a nonempty set $\Lambda_t \subset \mathbb{N}$ such that

$$\mathbb{P}\left( \sup_{i \in \Lambda_t} \left| \sum_{j=1}^N X_j^i \right| > t \right) = 0 \quad \left| \Delta^c \setminus \Lambda_t \right| < \infty. \tag{11.12}$$

Then, if that was the case, we would immediately get (by arguing as in Step I and using (11.11) and the assumption that $\|\eta\| = 1$) that

$$\begin{aligned}
&\mathbb{P}\left( \left\| q^{-1} P_\Delta^\perp E_\Omega P_\Delta \eta \right\|_{l^\infty} > t + \| P_\Delta^\perp U^* P_N U P_\Delta \|_{\mathrm{mr}} \right) \\
&= \mathbb{P}\left( \left\| \sum_{j=1}^N \widetilde{Y}_j + P_\Delta^\perp U^* P_N^\perp U P_\Delta \eta \right\|_{l^\infty} > t + \| P_\Delta^\perp U^* P_N U P_\Delta \|_{\mathrm{mr}} \right) \\
&\leq \sum_{i \in |\Delta^c \setminus \Lambda_t|} \mathbb{P}\left( \left| \sum_{j=1}^N X_j^i \right| > t \right),
\end{aligned}$$

Thus, we could use the analysis provided above, via (11.10), and deduce that

$$\mathbb{P}\left( \left\| q^{-1} P_\Delta^\perp E_\Omega P_\Delta \eta \right\|_{l^\infty} > t + \| P_\Delta^\perp U^* P_N U P_\Delta \|_{\mathrm{mr}} \right) \leq \gamma$$

when

$$q \geq \left( \frac{4}{t^2} + \frac{2\sqrt{2}}{3t} \sqrt{|\Delta|} \right) \log\left( \frac{4}{\gamma} \left| \Delta^c \setminus \Lambda_t \right| \right) \upsilon^2(U). \tag{11.13}$$

Hence, if we could show the existence of $\Lambda_t$ and provide a bound on $|\Delta^c \setminus \Lambda_t|$, we could appeal to (11.11) and (11.13) and complete the proof. To do that, define

$$\Lambda_t = \left\{ i \notin \Delta : \mathbb{P}\left( \left\| \sum_{j=1}^N P_\Delta U^* \delta_j (e_j \otimes e_j) U e_i \right\| \leq tq \right) = 1 \right\}.$$

Note that $(e_j \otimes e_j) U e_i \to 0$ as $i \to \infty$ for all $j \leq N$. Thus, $\Lambda_t \neq \emptyset$. Moreover, we also immediately deduce that $|\Delta^c \setminus \Lambda_t| < \infty$. Note also that (11.12) follows by the fact that $X_j^i = \langle \eta, q^{-1} P_\Delta U^* \delta_j (e_j \otimes e_j) U e_i \rangle$ and the Cauchy–Schwarz inequality. With the existence of $\Lambda_t$ established, we now continue with the task of estimating $|\Delta^c \setminus \Lambda_t|$. Note that to estimate $|\Delta^c \setminus \Lambda_t|$ we need information about the location of $\Delta$ which is not assumed. We only assume the knowledge of some $M \in \mathbb{N}$ such that

$P_M \geq P_\Delta$. Thus (although an estimate of $|\Delta^c \setminus \Lambda_t|$ would be sharper than what we will eventually obtain), we define

$$\tilde{\Lambda}_q(|\Delta|, M, t) = \left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\dots,M\}, |\Gamma_1| = |\Delta| \\ \Gamma_2 \subset \{1,\dots,N\}}} \left\| P_{\Gamma_1} U^* P_{\Gamma_2} U e_i \right\| \leq tq \right\}.$$

Note that it is straightforward to show that $\tilde{\Lambda}_q(|\Delta|, M, t) \subset \Lambda_t$. Also, $\tilde{\Lambda}_q(|\Delta|, M, t)$ depends only on known quantities. Observe that, clearly, for any $\Gamma_1 \subset \{1, \dots, M\}$ and $\Gamma_2 \subset \{1, \dots, N\}$ then $\left\| P_{\Gamma_1} U^* P_{\Gamma_2} U e_i \right\| \to \infty$ as $i \to \infty$. Thus, $|\Delta^c \setminus \Lambda_q(|\Delta|, M, t)| < \infty$ and since $\Lambda_q(|\Delta|, M, t) \subset \Lambda_t$ it follows that

$$\left|\Delta^c \setminus \Lambda_q(\Delta, t)\right| \leq \left| \left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\dots,M\}, |\Gamma_1| = |\Delta| \\ \Gamma_2 \subset \{1,\dots,N\}}} \left\| P_{\Gamma_1} U^* P_{\Gamma_2} U e_i \right\| > tq \right\} \right|.$$

This gives the second part of the proposition. The fact that the left-hand side of (9.4) is zero when $q = 1$ is clear from (11.8) and (11.9). □

*Proof of Proposition 9.2* Without loss of generality, we may assume that $\|\eta\| = 1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = q$. Also, for $k \in \mathbb{N}$, let $\xi_k = (U P_\Delta)^* e_k$. Observe that, since $U$ is an isometry,

$$q^{-1}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta = \sum_{k=1}^N q^{-1} \delta_k \xi_k \otimes \bar{\xi}_k, \quad P_\Delta = \sum_{k=1}^\infty \xi_k \otimes \bar{\xi}_k, \tag{11.14}$$

and

$$\begin{aligned} \left\| \left( \frac{1}{q} (P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta \right) \eta \right\| \leq & \left\| \left( \sum_{k=1}^N (q^{-1}\delta_k - 1) \xi_k \otimes \bar{\xi}_k \right) \eta \right\| \\ & + \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\eta\|, \end{aligned} \tag{11.15}$$

where the infinite series in (11.14) converges in operator norm. To get the desired result, we first focus on obtaining bounds on $\|(\sum_{k=1}^N (q^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k)\eta\|$. The goal is to use Talagrand's formula, and what follows is a setup for that. In particular, let $\zeta \in \mathcal{H}$ be a unit vector and denote the mapping $\mathcal{H} \ni \xi \mapsto \mathrm{Re}(\langle \xi, \zeta \rangle)$ by $\hat{\zeta}$. Let $\mathcal{F}$ be a countable collection of unit vectors such that for any $\xi \in \mathcal{H}$ we have $\|\xi\| = \sup_{\zeta \in \mathcal{F}} \hat{\zeta}(\xi)$, and now define

$$Z = \|X\|, \quad X = \sum_{k=1}^N Z_k, \quad Z_k = ((q^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k)\eta.$$

Note that the following is clear (and note how this immediately gives us the setup for Talagrand's Theorem)

$$Z = \left\| \left( \sum_{k=1}^{N} (q^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k \right) \eta \right\| = \sup_{\zeta \in \mathcal{F}} \hat{\zeta} \left( \sum_{k=1}^{N} Z_k \right) = \sup_{\zeta \in \mathcal{F}} \sum_{k=1}^{N} \hat{\zeta}(Z_k).$$

To use Talagrand's theorem, we must estimate the following quantities:

$$V = \sup_{\zeta \in \mathcal{F}} \mathbb{E}\left( \sum_{k=1}^{N} \hat{\zeta}(Z_k)^2 \right), \quad S = \sup_{\zeta \in \mathcal{F}} \|\hat{\zeta}\|_\infty, \quad R = \mathbb{E}\left( \left\| \sum_{k=1}^{N} Z_k \right\| \right).$$

For $V$ we have the following estimate:

$$\begin{aligned} \sup_{\zeta \in \mathcal{F}} \mathbb{E}\left( \sum_{k=1}^{N} \hat{\zeta}(Z_k)^2 \right) &\leq \sup_{\zeta \in \mathcal{F}} \mathbb{E}\left( \sum_{k \leq N} \left( q^{-1}\delta_k - 1 \right)^2 |\langle \xi_k, \zeta \rangle|^2 |\langle \xi_k, \eta \rangle|^2 \right) \\ &\leq q^{-1}(1-q) \sum_{k \leq N} \|\xi_k\|^2 |\langle e_k, U P_\Delta \eta \rangle|^2 \\ &\leq q^{-1}(1-q) \upsilon^2(U) |\Delta|, \end{aligned}$$

where we have used the fact that $U$ is an isometry in the step going from the second to the third inequality. The $S$ term can be estimated as follows. Note that

$$\hat{\zeta}(Z_k) = |\left( q^{-1}\delta_k - 1 \right) |\langle \xi_k, \zeta \rangle| |\langle \xi_k, \eta \rangle| \leq \max\{q^{-1} - 1, 1\} \upsilon^2(U) |\Delta|, \quad k \leq N, \tag{11.16}$$

thus

$$S \leq \max\{q^{-1} - 1, 1\} \upsilon^2(U) |\Delta|. \tag{11.17}$$

Finally, we can estimate $R$ as follows:

$$\begin{aligned} \mathbb{E}\left( \left\| \sum_{k=1}^{N} Z_k \right\|^2 \right) &= \sum_{k=1}^{N} \mathbb{E}(\|Z_k\|^2) + \sum_{k \neq j} \mathbb{E}(\langle Z_k, Z_j \rangle) \\ &\leq q^{-1}(1-q) \sum_{k \leq N} \|\xi_k\|^2 |\langle e_k, U P_\Delta \eta \rangle|^2 \\ &\leq q^{-1}(1-q) \upsilon^2(U) |\Delta|, \end{aligned}$$

again using the fact that $U$ is an isometry. Therefore,

$$\mathbb{E}\left( \left\| \sum_{k \leq N} Z_k \right\| \right) \leq \sqrt{\mathbb{E}\left( \left\| \sum_{k \leq N} Z_k \right\|^2 \right)} \leq \sqrt{q^{-1}(1-q) \upsilon^2(U) |\Delta|}. \tag{11.18}$$

With the estimates on $V$, $S$ and $R$ now established we may appeal to Theorem 11.2 and deduce that there is a constant $K > 0$ such that, for $\theta > 0$,

$$\mathbb{P}\left(\left\|\left(\sum_{k=1}^{N}(q^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\right)\eta\right\| \geq \theta + \sqrt{q^{-1}(1-q)\upsilon^2(U)|\Delta|}\right)$$
$$\leq 3\exp\left(-\frac{\theta}{K}\left(\max\{q^{-1}-1,1\}\upsilon^2(U)|\Delta|\right)^{-1}\log\left(1+\frac{\theta}{2}\right)\right). \quad (11.19)$$

provided $q$ is chosen such that the right-hand side of (11.18) is bounded by 1 (this is guaranteed by the assumptions of the proposition). But by (11.15) it follows that for any $r > 0$, we have

$$\mathbb{P}\left(\left\|\left(\frac{1}{q}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta\right)\eta\right\| \geq r\right)$$
$$\leq \mathbb{P}\left(\left\|\left(\sum_{k=1}^{N}(q^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\right)\eta\right\| \geq r - \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\|\right). \quad (11.20)$$

Therefore, by appealing to (11.20) and (11.19) we obtain that

$$\mathbb{P}\left(\left\|\left(\frac{1}{q}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta\right)\eta\right\| \geq \theta + \sqrt{q^{-1}(1-q)\upsilon^2(U)|\Delta|} + \Xi\right)$$
$$\leq 3\exp\left(-\frac{\theta}{K}\left(\max\{q^{-1}-1,1\}\upsilon^2(U)|\Delta|\right)^{-1}\log\left(1+\frac{\theta}{2}\right)\right),$$

where $\Xi = \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\|$. Choosing $\theta = t/2$ yields the proposition. □

*Proof of Theorem 9.3* The proof is quite similar to the proof of Proposition 9.2. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \theta$. Note that we may argue as in (11.14) and observe that

$$\left\|\theta^{-1}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta\right\| \leq \left\|\sum_{k=1}^{N}(\theta^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\right\|$$
$$+ \left\|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\right\|, \quad (11.21)$$

where $\xi_k = (U P_\Delta)^* e_k$. To get the desired result, we first focus on getting bounds on $\|\sum_{k=1}^N(\theta^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\|$. As in the proof of Proposition 9.2, the goal is to use Talagrand's thereom and the first step to do so is to estimate $\mathbb{E}(\|Z\|)$, where $Z = \sum_{k=1}^N(\theta^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k$.

Claim: We claim that

$$\mathbb{E}(\|Z\|)^2 \leq 48\max\{\log(|\Delta|), 1\}\theta^{-1}\upsilon^2(U)|\Delta|, \quad (11.22)$$

when

$$\theta \geq 18 \max\{\log(|\Delta|), 1\}\upsilon^2(U)|\Delta|.$$

To prove the claim, we simply rework the techniques used in [54]. This is now standard and has also been used in [19,61]. We start by letting $\tilde{\delta} = \{\tilde{\delta}_k\}_{k=1}^N$ be independent copies of $\delta = \{\delta_k\}_{k=1}^N$. Then

$$\begin{aligned}\mathbb{E}_\delta\left(\|Z\|\right) &= \mathbb{E}_\delta\left(\left\|Z - \mathbb{E}_{\tilde{\delta}}\left(\sum_{k=1}^N \left(\theta^{-1}\tilde{\delta}_k - 1\right)\xi_k \otimes \bar{\xi}_k\right)\right\|\right) \\ &\leq \mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|Z - \sum_{k=1}^N \left(\theta^{-1}\tilde{\delta}_k - 1\right)\xi_k \otimes \bar{\xi}_k\right\|\right)\right), \end{aligned} \tag{11.23}$$

by Jensen's inequality. Let $\varepsilon = \{\varepsilon_j\}_{j=1}^N$ be a sequence of Bernoulli variables taking values $\pm 1$ with probability $1/2$. Then, by (11.23), symmetry, Fubini's Theorem and the triangle inequality, it follows that

$$\begin{aligned}\mathbb{E}_\delta\left(\|Z\|\right) &\leq \mathbb{E}_\varepsilon\left(\mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|\sum_{k=1}^N \varepsilon_k\left(\theta^{-1}\delta_k - \theta^{-1}\tilde{\delta}_k\right)\xi_k \otimes \bar{\xi}_k\right\|\right)\right)\right) \\ &\leq 2\mathbb{E}_\delta\left(\mathbb{E}_\varepsilon\left(\left\|\sum_{k=1}^N \varepsilon_k\theta^{-1}\delta_k\xi_k \otimes \bar{\xi}_k\right\|\right)\right). \end{aligned} \tag{11.24}$$

Now, by Lemma 11.1 we get that

$$\begin{aligned}&\mathbb{E}_\varepsilon\left(\left\|\sum_{k=1}^N \varepsilon_k\theta^{-1}\delta_k\xi_k \otimes \bar{\xi}_k\right\|\right) \\ &\leq 3\sqrt{\max\{2\log(|\Delta|), 2\}\theta^{-1}}\max_{1\leq k\leq N}\|\xi_k\|\sqrt{\left\|\sum_{k=1}^N \theta^{-1}\delta_k\xi_k \otimes \bar{\xi}_k\right\|}. \end{aligned} \tag{11.25}$$

And hence, by using (11.24) and (11.25), it follows that

$$\mathbb{E}_\delta\left(\|Z\|\right) \leq 3\sqrt{\max\{2\log(|\Delta|), 2\}\theta^{-1}\upsilon^2(U)|\Delta|}\sqrt{\mathbb{E}_\delta\left(\left\|Z + \sum_{k=1}^N \xi_k \otimes \bar{\xi}_k\right\|\right)}.$$

Thus, by using the easy calculus fact that if $r > 0$, $c \leq 1$ and $r \leq c\sqrt{r+1}$ then $r \leq c(1+\sqrt{5})/2$, and the fact that $U$ is an isometry (so that $\|\sum_{k=1}^N \xi_k \otimes \bar{\xi}_k\| \leq 1$), it is easy to see that the claim follows.

To be able to use Talagrand's formula, there are now some preparations that have to be done. First write

$$Z = \sum_{k=1}^{N} Z_k, \quad Z_k = \left(\theta^{-1}\delta_k - 1\right) \xi_k \otimes \bar{\xi}_k.$$

Clearly, since $Z$ is self-adjoint, we have that $\|Z\| = \sup_{\eta\in\mathcal{F}} |\langle Z\eta, \eta\rangle|$, where $\mathcal{G}$ is a countable set of unit vectors. For $\eta \in \mathcal{G}$, let the mappings $\mathcal{B}(\mathcal{H}) \ni T \mapsto \langle T\eta, \eta\rangle$ and $\mathcal{B}(\mathcal{H}) \ni T \mapsto -\langle T\eta, \eta\rangle$ be denoted by $\hat{\eta}_1$ and $\hat{\eta}_2$, respectively. Letting $\mathcal{F}$ denote the family of all these mappings we have that $\|Z\| = \sup_{\hat{\eta}_i\in\mathcal{F}} \sum_{k\le N} \hat{\eta}_i(Z_k)$, and the setup for Talagrand's theorem is complete.

For $k = 1, \ldots, N$ note that

$$|\hat{\eta}_i(Z_k)| = \left|\left(\theta^{-1}\delta_k - 1\right)\right| |\langle \left(\xi_k \otimes \bar{\xi}_k\right) \eta, \eta\rangle| \le \theta^{-1}\|\xi\|^2.$$

Thus, after restricting $\hat{\eta}_i$ to the ball of radius $\theta^{-1} \max_{k\le N} \|\xi_k\|^2$ it follows that

$$S = \sup_{\eta_i\in\mathcal{F}} \|\hat{\eta}_i\|_\infty \le \theta^{-1} \max_{k\le N} \|\xi_k\|^2 \le \theta^{-1}\upsilon^2(U)|\Delta|. \tag{11.26}$$

Also, note that

$$\begin{aligned} V &= \sup_{\hat{\eta}_i\in\mathcal{F}} \mathbb{E}\left(\sum_{k\le N} \hat{\eta}(Z_k)^2\right) \le \sup_{\hat{\eta}\in\mathcal{F}} \mathbb{E}\left(\sum_{k\le N} \left(\theta^{-1}\delta_k - 1\right)^2 |\langle \xi_k, \eta\rangle|^4\right) \\ &\le \max_{k\le N} \|\xi_k\|^2 \left(\theta^{-1} - 1\right) \sup_{\hat{\eta}\in\mathcal{F}} \sum_{k\le N} |\langle e_k, U P_\Delta \eta\rangle|^2 \\ &\le \left(\theta^{-1} - 1\right) \max_{k\le N} \|\xi_k\|^2 \le \left(\theta^{-1} - 1\right) \upsilon^2(U)|\Delta|, \end{aligned} \tag{11.27}$$

where the third inequality follows from the fact that $U$ is an isometry. It follows by Talagrand's inequality (Theorem 11.2), the earlier claim (and requiring that the right hand side of (11.22) is bounded by one, which is guarantied by the assumption of the theorem), the first part of the assumed (9.6), (11.26) and (11.27), that there is a constant $K > 0$ such that for $t > 0$

$$\begin{aligned} &\mathbb{P}\left(\left\|\sum_{k=1}^{N} (\theta^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\right\| \ge t + 48\log(|\Delta|)\theta^{-1}\upsilon^2(U)|\Delta|\right) \\ &\quad \le 3\exp\left(-\frac{t}{K}(\theta^{-1}\upsilon^2(U)|\Delta|)^{-1} \log\left(1 + \frac{t}{2}\right)\right). \end{aligned} \tag{11.28}$$

But by (11.21) we have

$$\begin{aligned} &\mathbb{P}\left(\left\|\frac{1}{\theta}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta\right\| \ge r\right) \\ &\quad \le \mathbb{P}\left(\left\|\sum_{k=1}^{N} (\theta^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\right\| \ge r - \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\|\right). \end{aligned} \tag{11.29}$$

for any $r > 0$. Therefore, by appealing to (11.29) and (11.28) we obtain

$$\mathbb{P}\left(\left\|\frac{1}{\theta}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta\right\| \geq t + 48\log(|\Delta|)\theta^{-1}\upsilon^2(U)|\Delta| + \Xi\right)$$
$$\leq 3\exp\left(-\frac{t}{K}(\theta^{-1}\upsilon^2(U)|\Delta|)^{-1}\log\left(1+\frac{t}{2}\right)\right), \quad \Xi = \|(P_\Delta U^* P_N U P_\Delta . - P_\Delta)\|.$$

for $t > 0$. Choosing $t = \frac{1}{2\gamma}$ yields the first part of the theorem. The last statement of the theorem is clear. □

## References


1. B. Adcock. Infinite-dimensional $\ell^1$ minimization and function approximation from pointwise data. arXiv:1503.02352, 2015.
2. B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *J. Fourier Anal. Appl.*, 18(4):685–716, 2012.
3. B. Adcock and A. C. Hansen. Stable reconstructions in Hilbert spaces and the resolution of the Gibbs phenomenon. *Appl. Comput. Harmon. Anal.*, 32(3):357–388, 2012.
4. B. Adcock, A. C. Hansen, E. Herrholz, and G. Teschke. Generalized sampling: extension to frames and inverse and ill-posed problems. *Inverse Problems*, 29(1):015008, 2013.
5. B. Adcock, A. C. Hansen, and C. Poon. Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem. *SIAM J. Math. Anal.*, 45(5):3114–3131, 2013.
6. B. Adcock, A. C. Hansen, and C. Poon. On optimal wavelet reconstructions from Fourier samples: linearity and universality of the stable sampling rate. *Appl. Comput. Harmon. Anal.*, 36(3):387–415, 2014.
7. B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: a new theory for compressed sensing. arXiv:1302.0561, 2014.
8. B. Adcock, A. C. Hansen, B. Roman, and G. Teschke. Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. *Advances in Imaging and Electron Physics*, 182:187–279, 2014.
9. B. Adcock, A. C. Hansen, and A. Shadrin. A stability barrier for reconstructions from Fourier samples. *SIAM J. Numer. Anal.*, 52(1):125–139, 2014.
10. A. Aldroubi. Oblique projections in atomic spaces. *Proc. Amer. Math. Soc.*, 124(7):2051–2060, 1996.
11. A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli, and Y. Shkolnisky. A framework for discrete integral transformations. I. The pseudopolar Fourier transform. *SIAM J. Sci. Comput.*, 30(2):764–784, 2008.
12. A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli, Y. Shkolnisky, and I. Sedelnikov. A framework for discrete integral transformations. II. The 2D discrete Radon transform. *SIAM J. Sci. Comput.*, 30(2):785–803, 2008.
13. A. Bastounis and A. C. Hansen. On the absence of the RIP in real-world applications of compressed sensing and the RIP in levels. arXiv:1411.4449, 2014.
14. 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.
15. A. Bourrier, M. E. Davies, T. Peleg, P. Pérez, and R. Gribonval. Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems. *IEEE Trans. Inform. Theory*, 60(12):7928–7946, 2014.
16. E. J. Candès. An introduction to compressive sensing. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.
17. E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math. (to appear)*, 67(6):906–956, 2014.
18. E. J. Candes and Y. Plan. A Probabilistic and RIPless Theory of Compressed Sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.

19. E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.
20. 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.
21. 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.
22. Y. Chi, L. L. Scharf, A. Pezeshki, and A. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.*, 59(5):2182–2195, 2011.
23. A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best $k$-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.
24. 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.
25. D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
26. D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.*, 22(1):1–53, 2009.
27. 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.
28. I. Ekeland and T. Turnbull. *Infinite-dimensional optimization and convexity*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1983.
29. Y. Eldar. Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors. *J. Fourier Anal. Appl.*, 9(1):77–96, 2003.
30. Y. C. Eldar. Compressed sensing of analog signals in shift-invariant spaces. *IEEE Trans. Signal Process.*, 57(8):2986–2997, 2009.
31. Y. C. Eldar and T. Michaeli. Beyond Bandlimited Sampling. *IEEE Signal Process. Mag.*, 26(3):48–68, 2009.
32. A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM J. Imaging Sci.*, 5:179–202, 2012.
33. M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.
34. S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhauser, 2013.
35. K. Gröchenig, Z. Rzeszotnik, and T. Strohmer. Quantitative estimates for the finite section method and Banach algebras of matrices. *Integr Equat Oper Th.*, 67(2):183–202, 2011.
36. D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory*, 57(3):1548–1566, 2011.
37. D. Gross, F. Krahmer, and R. Kueng. A partial derandomization of phaselift using spherical designs. arXiv:1310.2267, 2014.
38. M. Guerquin-Kern, M. Häberlin, K. P. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging*, 30(9):1649–1660, 2011.
39. M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic analytical phantoms for parallel Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging*, 31(3):626–636, 2012.
40. A. C. Hansen. On the approximation of spectra of linear operators on Hilbert spaces. *J. Funct. Anal.*, 254(8):2092–2126, 2008.
41. 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.
42. E. Herrholz and G. Teschke. Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems. *Inverse Problems*, 26(12):125012, 2010.
43. T. Hrycak and K. Gröchenig. Pseudospectral Fourier reconstruction with the modified inverse polynomial reconstruction method. *J. Comput. Phys.*, 229(3):933–946, 2010.
44. T. W. Körner. *Fourier Analysis*. Cambridge University Press, 1988.
45. M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.
46. 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.
47. M. Mishali and Y. C. Eldar. Xampling: Analog to digital at sub-Nyquist rates. *IET Circuits, Devices, & Systems*, 5(1):8–20, 2011.
48. 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.

49. S. K. Mitter. Convex optimization in infinite dimensional spaces. In *Recent Advances in Learning and Control, volume 371 of Lecture Notes in Control and Information Science*, pages 161–179. Springer, London, 2008.
50. C. Poon. A consistent and stable approach to generalized sampling. *J. Fourier Anal. Appl.*, 20(5):985–1019, 2014.
51. H. Rauhut and R. Ward. Sparse recovery for spherical harmonic expansions. In *Proceedings of the 9th International Conference on Sampling Theory and Applications*, 2011.
52. H. Rauhut and R. Ward. Sparse Legendre expansions via l1-minimization. *J. Approx. Theory*, 164(5):517–533, 2012.
53. B. Roman, A. Hansen, and B. Adcock. On asymptotic structure in compressed sensing. arXiv:1406.4178, 2014.
54. M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999.
55. G. Strang and T. Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, 1997.
56. T. Strohmer. Measure what should be measured: progress and challenges in compressive sensing. *IEEE Signal Proc Let.*, 19(12):887–893, 2012.
57. A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.
58. B. P. Sutton, D. C. Noll, and J. A. Fessler. Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities. *IEEE Trans. Med. Imaging*, 22(2):178–188, 2003.
59. M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.
60. G. Tang, B. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Trans. Inform. Theory.*, 59(11):7465–7490, 2013.
61. J. A. Tropp. On the conditioning of random subdictionaries. *Appl. Comput. Harmon. Anal.*, 25(1):1–24, 2008.
62. M. Unser. Sampling–50 years after Shannon. *Proc. IEEE*, 88(4):569–587, 2000.
63. M. Unser and A. Aldroubi. A general sampling theory for nonideal acquisition devices. *IEEE Trans. Signal Process.*, 42(11):2915–2925, 1994.
64. M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Process.*, 50(6):1417–1428, 2002.