Generalized sampling and infinite-dimensional compressed sensing

Ben Adcock
Department of Mathematics
Purdue University
West Lafayette, IN 47907
USA

Anders C. Hansen
DAMTP, Centre for Mathematical Sciences
University of Cambridge
Wilberforce Rd, Cambridge CB3 0WA
United Kingdom

Abstract

We introduce and analyze an abstract framework, and corresponding method, for compressed sensing in infinite dimensions. This extends the existing theory from signals in finite-dimensional vectors spaces to the case of separable Hilbert spaces. We explain why such a new theory is necessary, and demonstrate that existing finite-dimensional techniques are ill-suited for solving a number of important problems.

This work stems from recent developments in generalized sampling theorems for classical (Nyquist rate) sampling that allows for reconstructions in arbitrary bases. The main conclusion of this paper is that one can extend these ideas to allow for significant subsampling of sparse or compressible signals. The key to these developments is the introduction of two new concepts in sampling theory, the *stable sampling rate* and the *balancing property*, which specify how to appropriately discretize the fundamentally infinite-dimensional reconstruction problem.

1 Introduction

Compressed sensing (CS) has, with little doubt, been one of the great successes of applied mathematics in the last decade [13, 17, 18, 20, 23, 24, 31]. 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 relatively 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 to N. Herein lies a problem. Real-world signals are typically analog, or continuous-time, and thus are modelled more faithfully in infinite-dimensional function spaces [10]. Any finite-dimensional model may therefore not be well suited to such problems.

Although this issue has been widely recognized [22, 51, 55], there have been few attempts made thus far to extend the existing theory to a more realistic, infinite-dimensional model (see §1.5). The purpose of this paper is to provide such a generalization.

The first step in our approach is a move away from the usual matrix-vector model. In particular, we consider the following more realistic scenario. A signal f is viewed as an element of a separable Hilbert space $\mathcal{H}$, and its measurements are modelled as a sequence of linear functionals $\zeta_j : \mathcal{H} \to \mathbb{C}$. This gives rise to the countable collection

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

of samples of f. Now suppose that the signal f is sparse or compressible in some orthonormal basis $\{\varphi_j\}_{j\in\mathbb{N}}$ of $\mathcal{H}$. The main question we answer in this paper are the following: can we recover f by subsampling from the collection (1.1), and how can this realized by a numerical algorithm? In doing so, we obtain a full theory for so-called *infinite-dimensional* compressed sensing, valid for (almost) arbitrary pairs of sampling schemes $\{\zeta_j\}_{j\in\mathbb{N}}$ and reconstruction bases $\{\varphi_j\}_{j\in\mathbb{N}}$.

The framework we introduce in this paper stems from recent developments in classical sampling of signals. In [1, 2, 5] a new 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 §1.4 and §4 for further details). The contribution of this paper is a continuation of this work in which sparsity is exploited to allow for substantial subsampling.

Before explaining the necessity for this work, let us first illustrate our infinite-dimensional CS framework with an example of the type of result we prove in this paper:

Theorem 1.1. *For $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$ write $\Delta = \{j : \alpha_j \neq 0\} \subseteq \mathbb{N}$ and suppose that $\Delta \subset \{1, \ldots, M\}$ for some $M \in \mathbb{N}$. Let $\epsilon > 0$ be arbitrary. Then there exists an integer $N \in \mathbb{N}$ (specific estimates will be given in §7) depending on M and $|\Delta|$ only such that the following holds. If $\Omega \subset \{1, \ldots, N\}$, $|\Omega| = m$, is chosen uniformly at random, then, with probability greater than $1 - \epsilon$, f can be recovered exactly from the samples $\{\zeta_j(f) : j \in \Omega\}$ whenever m is proportional to $|\Delta| \cdot \log(\epsilon^{-1} + 1) \cdot \log(NM\sqrt{|\Delta|})$.*

As we explain in §6 and §7, the values m, N are specified by a system of inequalities involving $|\Delta|$, M and ϵ. The somewhat surprising result is that taking $N = M$ will typically not be sufficient. As we demonstrate in §2, there are straightforward examples where $|\Delta|$ may be very small, but choosing $N = M$ will give disastrous results. However, choosing a particular value $N > M$ (specific bounds will be given later) will allow for substantial subsampling (see §2.3).

Remark 1.1 The framework we propose in this paper for infinite-dimensional CS involves constructing an appropriate measurement matrix and solving the resulting convex optimization problem. The informed reader may well think it possible to establish a Restricted Isometry Property (RIP) for such matrices, and therefore exploit existing finite-dimensional CS results to prove Theorem 1.1, for example. However, the stumbling block is that the RIP is notoriously hard to verify. In addition, even in finite dimensions, the best RIP results give sparse recoverability estimates which are known in some notable cases not to be sharp [15, 31]. On the other hand, a so-called *RIPless* theory of finite-dimensional CS has recently been proposed by Candès & Plan, which gives improved sparse recovery results using only the so-called, and easy to verify, *coherence* of a measurement system [15]. The developments in this paper are in a similar spirit: we forgo the RIP in favour of incoherence, and therefore obtain sharp theorems for subsampling with readily verifiable conditions. Note that several of our results generalize those given in [15, 17] to the infinite-dimensional case.

1.1 An example

Magnetic Resonance Imaging (MRI) was one of the original motivations for CS [17]. Developed extensively by the work of Lustig et al. [42], the application of CS techniques in MRI is now a subject of intensive research.

However, the MRI problem is inherently infinite-dimensional. In MRI, an image, which is most faithfully modelled as a function $f \in \mathcal{H} := L^2(\mathbb{R}^2)$, 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, a common approach when applying CS techniques to MRI is based on discretization. Namely, one considers f as a finite vector (or array) of pixel values, and replaces the continuous Fourier transform by its discrete analogue [32]. By doing so, obtains a finite-dimensional recovery problem which can be addressed with existing CS tools. However, as we explain in §2.1, modelling the fundamentally infinite-dimensional problem in this way can quite easily lead to problems, even in extremely simple examples (see §2.1). This is an instance of the well-known "inverse crime" [35].

We note that the above continuous/infinite-dimensional formulation of the MRI reconstruction problem, whilst uncommon in the CS literature, has recently been used to great effect by Guerquin-Kern, Haberlin, Pruessmann & Unser [34, 35]. However, there are currently no recoverability guarantees for this problem along the lines of Theorem 1.1 above. The general results we prove in this paper seek to address this gap. Note that this type of continuous formulation was previously used successfully by Fessler et al. [53].

The MRI problem also illustrates another key point in this paper. Namely, in many problems of interest, the samples $\{\zeta_j(f)\}_{j\in\mathbb{N}}$ are *fixed*, and cannot be altered. In MRI this is due to the particular design of the physical scanning device. Although much of research in finite-dimensional CS has been devoted to the topic of designing good sampling systems [13, 31], for many important problems one does not necessarily have this luxury. Thus we require a theory, as well as techniques, for infinite-dimensional CS that allows one to work with *fixed* measurements.

1.2 The need for a new general theory

Before going further, it worth first asking whether or not such a new theory is actually necessary. Given the problem described above, one must at some stage discretize. It therefore seems plausible that finite-dimensional CS techniques could be readily applied once one had restricted the problem from the underlying Hilbert space $\mathcal{H}$ to a suitable finite-dimensional space. In particular, if f is sparse in an, albeit countably-infinite, basis $\{\varphi_j\}_{j\in\mathbb{N}}$ (i.e. it only has finitely many nonzero coefficients in this basis), it seems plausible that the corresponding sparse recovery problem is inherently finite-dimensional. In some limited cases this is indeed the case: one may treat the problem solely in finite dimensions with existing CS tools. However, as we discuss in §2, in general this problem cannot be tackled in such a way.

Indeed, 'discretizing' the problem—that is, reducing the infinite amounts of information contained in the samples $\{\zeta_j(f)\}_{j\in\mathbb{N}}$—so as to make it amenable to computations is fraught with difficulties (see §2). The most obvious, and ultimately most naïve, discretizations may well destroy the original structure of the problem. This means that exact recovery may well not be possible with finite-dimensional techniques, since the key structure that allows for subsampling is not carried over to the discretization. New techniques which tackle the infinite-dimensional problem directly are therefore necessary, and this is what we shall provide.

1.3 Discretizing infinite-dimensional problems

In general, discretizing infinite-dimensional problems is a difficult and subtle issue which cannot be carried out successfully without an understanding of the particular problem at hand. Unless done carefully, it is quite possible to end up with a discretization whose properties contrast starkly with those of the original problem, and consequently a numerical method that is neither stable nor convergent. With this in mind, our approach is based on the following fundamental philosophy:

Keep the infinite-dimensional structure and crucial properties of the original problem when discretizing. (Ph)

By correctly following this principle we obtain a framework for infinite-dimensional CS.

Notice that the approach we propose in this paper is somewhat at odds with the usual procedure in CS, in which the original problem is often modelled as finite dimensional. However, this can easily lead to an inverse crime when applied to real data. On the other hand, our initial step is to devise an appropriate infinite-dimensional formulation of the sparse recovery problem. with truncation being carried out in the second step. This leads to a finite-dimensional problem which retains the key features of the original problem, and which can be solved numerically.

It is also worth mentioning that (Ph) is by no means unique to this particular problem. Whilst it is often followed in numerical ODEs and PDEs, most relevantly for this paper it was recently employed in [37] to solve the long-standing computational spectral problem. A number of ideas in this article stem from [37], and the contributions of this paper may be viewed as a continuation of this work. Note that similar versions of (Ph) have also been used by Stuart et al. for solving inverse problems [52].

1.4 Generalized sampling (GS)

The framework we propose in this paper has its direct origins in recent developments in classical, i.e. Nyquist rate, sampling. Namely, the fundamental question of how to recover signals (not necessarily sparse or compressible) in arbitrary bases from their samples (1.1).

By employing (Ph), a new approach to this problem, known as *generalized sampling* (GS), was developed in [1, 2, 5]. GS is a new type of sampling theory which incorporates the critical issues of approximation and stability, culminating in the so-called *stable sampling rate* [5]. The resulting numerical method 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 build on this work in the following way: we show that, in the case that the signal to be reconstructed is sparse or compressible, reconstruction can also be performed with significant subsampling. We refer to the corresponding method as *generalized sampling with compressed sensing* (GS–CS).

One important instance of both GS and GS–CS is the recovery of a function from its Fourier samples (the MRI problem, in particular). Although the classical Shannon Sampling Theorem [39, 57] allows for reconstruction in terms of an infinite series of sinc functions or complex exponentials, the slow convergence of these series, as well as the appearance of the Gibbs phenomenon [40], means that such a reconstruction is often not practical. Consequently, Shannon's theorem may not be used in practice [57], even for Nyquist-rate sampling. 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 [29]. GS allows one to reconstruct in such a basis in manner that is both accurate and numerically stable. The method we develop in this paper, GS–CS, permits one to undersample whenever the signal is sparse or compressible.

1.5 Relation to other work and contributions of the paper

There have been a number of recent attempts to generalize CS to infinite dimensions. In [28, 44, 45], Eldar et al. describe an infinite-dimensional CS approach for analog-to-digital conversion based on a union of subspace signal model. This research is related to the previous work of finite rates of innovation by Vetterli et al. [10, 25, 59]. In [38], the approach of Eldar et al. was applied to inverse and ill-posed problems. The application of CS techniques to the recovery of functions was considered by Rauhut & Ward. By devising an appropriate sampling distribution to ensure a restricted isometry property, they prove nearly-optimal recoverability results for functions which are sparse in Legendre polynomial [48] or spherical harmonic [47] bases. 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, continuous-time/analog problems are found other in applications including radar, sonar, and remote sensing [51]. Use of standard, finite-dimensional CS in these problems is plagued by the phenomenon of gridding error (or basis mismatch) [19]. Although the setting here is somewhat different to that which we consider in this paper, the same issue arises: naïve discretization of the infinite-dimensional problem leads to inferior reconstructions. Recent works [30, 55] have sought to address this by applying essentially the same principle (Ph). Closely related to this is the work of Candès & Fernandez–Granda on super-resolution [14], wherein an analog model is used in the recovery of signals from low bandwidth Fourier samples.

Note that most of the above works describe infinite-dimensional CS approaches for some particular class of problems, and do not address the fundamental problem of reconstructing the coefficients $\{\alpha_j\}_{j\in\mathbb{N}}$ of a function $f = \sum_{j\in\mathbb{N}} \alpha_j \varphi_j$ from fixed, but arbitrary, linear samples $\{\zeta_j(f)\}_{j\in\mathbb{N}}$ (this is precisely the issue in the infinite-dimensional MRI model in [34, 35]). Our GS–CS framework does precisely this. It is therefore both rather general, and, in many senses, a natural and fundamental extension of finite-dimensional CS theory. Specifically, we generalize the finite-dimensional setup of orthonormal bases in vector spaces to that of separable Hilbert spaces. It should therefore not come as a surprise that results concerning finite-dimensional CS are corollaries of our main theorems.

2 Why do we need a new approach?

Consider the following very 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*

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

to be the Fourier coefficients of f (this is an example of the type of sampling one encounters in MRI). Here $\mathcal{F}f$ denotes the Fourier transform of f, and $\epsilon \leq \frac{1}{2}$ is arbitrary. Throughout, we shall take $\epsilon = \frac{1}{2}$. Assume that f is sparse, or compressible, in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$ of Haar wavelets. Thus, the problem is to recover f from the measurements (2.1).

Recall that the classical Shannon Sampling Theorem (see §4) gives that f can be recovered exactly (in the L^2 sense) from the infinite collection $\{\zeta_j(f)\}_{j\in\mathbb{Z}}$. However, since f is known to be sparse in the Haar wavelet basis $\{\varphi_j\}_{j\in\mathbb{N}}$, the question is as follows: is there a way to use this additional information to allow f to be recovered from only a finite number of its samples? Moreover, if this is true, how many such samples are required? These are questions we shall answer in this paper.

2.1 First example: the discrete model

Let us consider the simplest possible example: namely, let $f = \chi_{[0,1/2)} - \chi_{[1/2,1)}$ be the Haar mother wavelet. By definition, f is extremely sparse in the Haar basis. To recover f exactly from (2.1), at some stage one needs to discretize, so as to reduce the infinite amounts of information to something finite. The usual approach in sparse MRI [42] involves two steps. First, one replaces the infinite collection of samples with the finite vector

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

Second, one uses a combination of the discrete Fourier and discrete wavelet transforms (DFT and DWT respectively) to formulate the corresponding measurement matrix.

To this end, let $U_{\mathrm{df}}, V_{\mathrm{dw}} \in \mathbb{C}^{2N \times 2N}$ be the matrices of these transforms. We proceed as follows. Note first that the classical discrete approximation to the problem of inverting the Fourier transform is given by

$$y = U_{\mathrm{df}} x,$$

where x is a vector approximating pointwise values of f on an equispaced grid in $[-1, 1]$. Since f is sparse in the Haar basis it is very tempting to think that

$$x_0 = V_{\mathrm{dw}} x$$

is also sparse, where V_{dw} is the discrete Haar transform. One should therefore be able to recover f perfectly from only relatively few of its samples $y = \zeta(f)$ by using standard CS techniques. Specifically, if $\Omega \subset \{1, \ldots, 2N\}$, $|\Omega| = m < 2N$ is chosen uniformly at random, then it is standard practice to solve the convex optimization problem

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

or some variant thereof in the case of noisy data. Here $P_\Omega : \mathbb{C}^{2N} \to \mathbb{C}^{2N}$ denotes orthogonal the projection onto $\mathrm{span}\{e_j : j \in \Omega\}$ and $\{e_j\}_{j=1}^{2N}$ is the canonical basis for $\mathbb{C}^{2N}$. If ξ is a minimizer of this problem, then one could hope that ξ agrees with the vector x_0 with high probability, and hence we could recover $x = V_{\mathrm{dw}}^{-1} x_0$.

To test (2.3), let us consider the case where $2N = 256$ and $m = 130$, i.e. we use nearly 50% of the measurements in the range $-N+1, \ldots, N$. Write $f_{N,m} = \sum_{j=1}^{2N} \xi_j \varphi_j$, where ξ is a minimizer of (2.3). Note that $f_{N,m}$ takes the values of the vector $V_{\mathrm{dw}}^{-1} \xi$ at the grid points.

As Figure 1 demonstrates, the result of applying (2.3) to this example is extremely disappointing. The function f is not recovered anywhere near exactly, and the reconstruction $f_{N,m}$ computed from (2.3) commits rather large errors, especially near the jumps in f, i.e. $x = -1, 0, \frac{1}{2}, 1$. To be more precise, even though f only has one nonzero coefficient in the Haar wavelet basis, and despite the fact that we use $m = 130$ Fourier samples of f, we do not get anywhere near to perfect recovery.

This now begs the following question: what went wrong? We give the answer in the next section.

2.1.1 The DFT destroys sparsity

The source of the failure of (2.3) is the discretization employed: namely, the DFT. The problem is that the DFT is not exact. As a result, there is a mismatch between the data, which are continuous Fourier samples, and their modelling as discrete Fourier samples.

To explain this mismatch and its effect, consider U_{df}^{-1}. This matrix 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 $[-1, 1]$. However, this mapping commits an error: for an arbitrary function f, the result is only an *approximation* to the grid values of f. The question is, how large is this error, and how does it affect (2.3) and its solutions?

Consider the vector $x \in \mathbb{C}^{2N}$ defined by

$$U_{\mathrm{df}} x = \zeta(f).$$

It is simple to see that x consists precisely of the values of the function

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

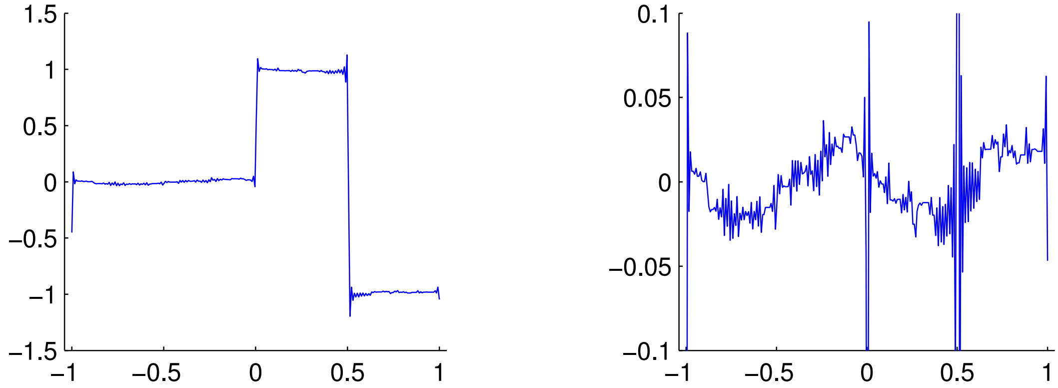

Figure 1: The figure shows the rather disappointing $f_{N,m}(t)$ (left) and $f(t) - f_{N,m}(t)$ (right) against t for $2N = 256$ and $m = 130$, where $f_{N,m}(t) = \sum_{j=1}^{2N} \xi_j \varphi_j(t)$ and $\xi = \{\xi_j\}_{j=1}^{2N}$ is a minimizer of (2.3).

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

Let us now consider the discrete wavelet transform $x_0 \in \mathbb{C}^{2N}$ of x:

$$x_0 = V_{\mathrm{dw}} x.$$

The right-hand side of the equality constraint in (2.3) now reads

$$P_\Omega U_{\mathrm{df}} V_{\mathrm{dw}}^{-1} x_0.$$

Thus, 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. The function f_N consists of smooth complex exponentials, and thus cannot have a sparse representation in a basis of piecewise smooth functions. Therefore, although it has a unitary and incoherent measurement matrix, (2.3) is not a sparse recovery problem. Consequently, there is little or no hope of recovering the sparse vector α of Haar wavelet coefficients of f from (2.3). This explains the complete failure witnessed in Figure 1.

From this argument, we now conclude the following. By forming the approximation (2.4), we have *destroyed important structure* of the original problem: namely, the sparsity. In particular, we have violated the guiding principle (Ph).

Remark 2.1 Note that this loss of sparsity is not exclusive to the Haar wavelet basis. In fact, if f is sparse in any basis of compactly supported wavelets then, by insisting on using the Shannon Sampling Theorem, we also witness the same problem: namely, f_N can never be sparse in the same basis.

2.1.2 The DFT leads to the Gibbs phenomenon

Whilst the loss of sparsity is significant, there is another important issue with this setup. Given that η is not sparse in Haar wavelets, suppose now, as an exercise, we forgo any subsampling. That is, we let $m = 2N$. The problem (2.3) now has a unique solution η. However, by the arguments given above, the entries of η are not the Haar wavelet coefficients of f, but rather coefficients of the approximation 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 hope to obtain* a better (i.e. more accurate) reconstruction of f than f_N.

The question is, how good an approximation is f_N? Since f is piecewise smooth, it turns out to be very poor. In fact, 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 unpleasant Gibbs phenomenon, with its associated $\mathcal{O}(1)$ oscillations, near each jump in f. These effects are visualized in Figure 2. The fact that (2.3) leads to a Haar

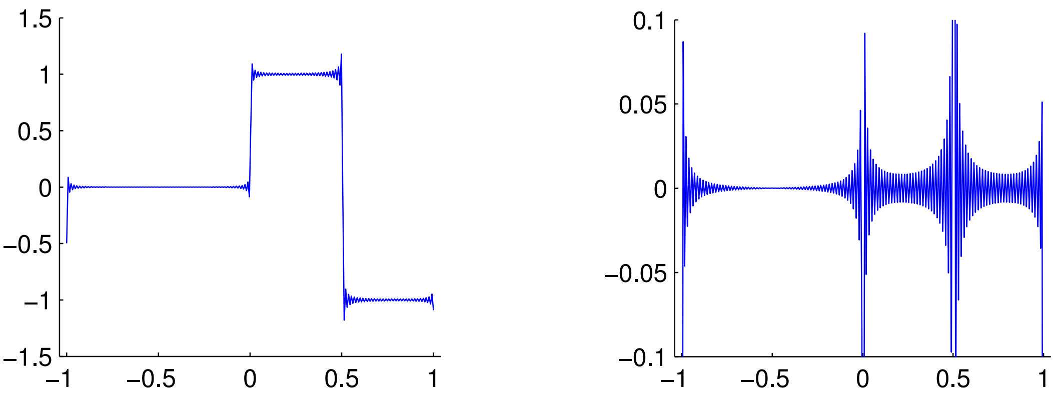

Figure 2: The figure shows $f_N(t)$ (left) and $f(t) - f_N(t)$ (right) against t for $2N = 256$.

wavelet approximation to f_N, as opposed to f, can be observed by comparing the left panels in Figures 2 and 1 respectively.

Of course, one may attempt various remedies to this problem, such as increasing N or using the total variation norm instead. However, the key point is, regardless of how clever we are, if we insist on performing the discretization via U_{df}, then we cannot hope to obtain any better than the (extremely poor) approximation f_N, the partial Fourier series.

2.1.3 Relation to the inverse crime

As mentioned, the poor reconstruction witnessed above is due to basis mismatch, i.e. continuous Fourier data being modelled as discrete Fourier data. Had this data actually been simulated using the discrete Fourier transform, then no such problems would have occurred, and one would have seen a vastly improved reconstruction (in this case, perfect recovery). However, this improvement is artificial and an example of the well-known inverse crime [35]. That is to say, inappropriate simulation of data leads to spuriously good reconstructions, but when applied to real data, such as in the above examples, the reconstruction quality substantially declines.

2.1.4 The operation $x_0 = V_{\mathrm{dw}}x$ may commit the wavelet crime

The Discrete wavelet transform is an infinite-dimensional operation 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 $x_0 = V_{\mathrm{dw}}x$ has therefore very little to do with the actual wavelet coefficients of f. This is referred to as the “wavelet crime” by Strang and Nguyen [50, p. 232]. Note that there is no inverse or wavelet crime in the new infinite-dimensional setup introduced in §2.3.

2.2 Second example: a common pitfall

The failure of (2.3) can be interpreted as a violation of the fundamental principle (Ph). In particular, the crucial property that f is sparse in the Haar wavelet basis is destroyed when the DFT is applied. With this in mind, it may seem to the reader that, since f is a finite sum of Haar wavelets, there is a simple remedy to this problem. Specifically, replace the DFT and DWT by the measurement matrix

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

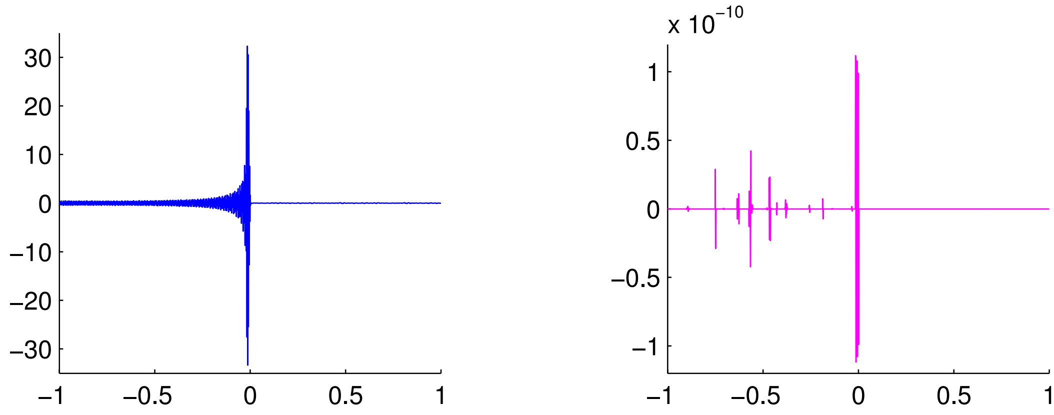

Figure 3: The figure shows the disastrous error $f(t) - f_{N,m}(t)$ against t (left) as well as the much more pleasant error $f(t) - g_{\tilde{N},\tilde{m}}(t)$ (right). Note that $f_{N,m}$ requires $m = 760$ samples whereas $g_{\tilde{N},\tilde{m}}$ requires only $\tilde{m} = 50$ samples.

randomly sample $\Omega \subset \{1, \dots, 2N\}$ with $|\Omega| = m$, obtain a minimizer ξ 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 form the reconstruction $f_{N,m} = \sum_{j=1}^{2N} \xi_j \varphi_j$ (note that in this case we have, for convenience, reindexed in the natural way the Fourier samples $\{\zeta_j\}_{j\in\mathbb{N}}$ over $\mathbb{N}$ rather than $\mathbb{Z}$). Clearly this approach, which corresponds to replacing the DFT and DWT by their continuous analogues, preserves the sparsity of the original problem, unlike (2.3).

Let us consider an example of (2.6). Suppose that $f(t) = \sum_{j=1}^{2N} \alpha_j \varphi_j(t)$, where $2N = 768$. We have chosen a function f such that $|\mathrm{supp}(f)| = |\{\alpha_j : \alpha_j \neq 0\}| = 5$. Note that f is very sparse in the Haar wavelet basis. In Figure 3 we display the reconstruction given by (2.6) using $m = 760$. As is evident, f is recovered extremely poorly by (2.6): although we have used nearly 98% of its Fourier samples in the range $1, \dots, 2N$, the reconstruction error $\|f - f_{N,m}\|$ is very large (roughly 2.43 and $\|f\| = 3.21$). We repeated the experiment fifty times with the same outcome.

This example is disastrous. Despite altering the standard CS approach (2.3) to ensure that sparsity is preserved, we still obtain an extremely poor reconstruction. Indeed, since f is sparse in the Haar basis and we sample using nearly all its Fourier samples in the range $1, \dots, 2N$, one may have reasonably hoped to recover f perfectly in this example. However, as seen in Figure 3, this is certainly not the case.

From a conventional CS viewpoint, this failure appears quite surprising. We have formed a measurement matrix in a standard way by taking inner products of one orthonormal basis (the complex exponentials $\{e^{2\pi i k\epsilon \cdot}\}_{k=-N+1}^{N}$) with another (Haar wavelets). Surely the standard finite-dimensional results should apply? However, they clearly do not, as evidenced by Figure 3. The question is why?

As it transpires, the explanation is simple, and lies with the change of basis matrix (2.5):

- The matrix U_N is not an isometry. Nor is it close to an isometry: its condition number is at least 10^{16}.
- In general, a matrix U_N of the form (2.5) is 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.

This lack of isometric structure highlights the underlying infinite-dimensionality of the problem. In particular, one needs infinitely many complex exponentials to span a space large enough so that it will contain finitely many Haar wavelets (or vice versa).

From this discussion, we now draw the following conclusion. Simply thinking (since f has only finitely many non-zero Haar wavelet coefficients) that the problem can somehow be embedded in $\mathbb{C}^N$ and solved using finite-dimensional CS is incorrect. In fact this failure can also be interpreted as a violation of (Ph). 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 does have this property (both Haar wavelets and complex exponentials form orthonormal bases for the infinite-dimensional Hilbert space $L^2(-1,1)$). It is precisely the loss of this structure when 'discretizing' U via U_N that is the source of the failure observed above.

Remark 2.2 The reader at this stage may wonder whether the absence of an isometry, and the resulting poor reconstruction, is a phenomenon isolated to the particular choice of reconstruction system (Haar wavelets). Perhaps a different choice, e.g. a basis of smoother wavelets, leads to a matrix U_N which is closer to an isometry? This is not the case. In [4] it was proved that the matrix (2.5) is has an exponentially large condition number as $N \to \infty$ for essentially *any* wavelet basis. An analogous result for polynomial bases was proved in [7].

2.3 Third example: a new approach

With these examples in mind, the purpose of the remainder of this paper is to describe a new approach for infinite-dimensional CS, known as *generalized sampling with compressed sensing* (GS–CS), which overcomes the aforementioned failings. This brings us to the purpose of this section, and really the essence of this paper: namely, why infinite dimensions? Put simply, this is because the search for the coefficients $\alpha = \{\alpha_1, \alpha_2, \ldots\}$ of f results in an infinite system of equations. By formulating reconstruction directly in an infinite-dimensional way, and *then discretizing* (as opposed to discretizing first), we are able to completely avoid the pitfalls described above.

This new approach will be explained in detail in the next sections. However, we first demonstrate that it overcomes the above issues. Let $f(t) = \sum_{j=1}^{M} \alpha_j \varphi_j(t)$ be as in §2.2, and let U be given by (2.7). Let P_K denote the projection onto $\mathrm{span}\{e_1, \ldots, e_N\}$, where $\{e_k\}_{k \in \mathbb{N}}$ is the canonical basis for $l^2(\mathbb{N})$. For $\tilde{N} \in \mathbb{N}$, we now choose $\Omega \subset \{1, \ldots, \widetilde{N}\}$ uniformly at random, with $|\Omega| = \tilde{m}$, and numerically compute a minimizer ξ to

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

where $M \in \mathbb{N}$, and let $g_{\widetilde{N},\tilde{m}} = \sum_{j=1}^{M} \xi_j \varphi_j$ be the reconstructed approximation to f. In Figure 3 we apply this algorithm with $\widetilde{N} = 1351$ and $\tilde{m} = 50$, and $M = 768$. Note the significant improvement over the approach of §2.2. In particular, when averaged over 50 trials, the error $\|f - g_{\widetilde{N},\tilde{m}}\|_{L^2}$ is found to be 1.15×10^{-11} in comparison to roughly 2.43 for the previous approach. Moreover, the GS–CS reconstruction uses less than 7% of the number of sampled Fourier coefficients that were used to form the extremely poor reconstruction in (2.6).

The aim of remainder of this paper is to explain why (2.8) leads to such a marked improvement. As we shall see, 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 §6.2). As we discuss, this property ensures a faithful discretization of the operator U. We shall present further numerical examples illustrating the effectiveness of this new approach in §8.

3 Infinite-dimensional compressed sensing

Before developing our theory of infinite-dimensional compressed sensing, let us now formally introduce the problem we shall solve in this paper, as well as the types of signal models we shall consider.

Suppose that $\mathcal{H}$ is a separable Hilbert space over $\mathbb{C}$ (throughout this paper we shall work with complex spaces). Let $\{\varphi_j\}_{j \in \mathbb{N}}$ be an orthonormal basis, and let $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$ be the signal we wish to recover. 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 $\mathcal{H}$. The problem throughout this paper will be to recover f in terms of $\{\varphi_j\}_{j \in \mathbb{N}}$ from a small number of the samples (3.1).

3.1 Sparsity and compressibility

Compressed sensing is a theory concerning sparse signals. In infinite dimensions, we say that f is sparse in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$ if there exists an $M \in \mathbb{N}$ such that

$$\Delta = \operatorname{supp}(f) \subset \{1, \ldots, M\}, \tag{3.2}$$

where

$$\operatorname{supp}(f) = \{j \in \mathbb{N} : \alpha_j \neq 0\},$$

denotes the support of the infinite vector α (throughout this paper, when is meaning is clear, we shall write Δ for $\operatorname{supp}(f)$). If $|\Delta| = r$, we say that f is (r, M)-sparse in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$. Naturally, we do not know Δ, however, we may have information about M.

In practice, the assumption that f is perfectly sparse is often unrealistic. In finite-dimensional CS, it is standard to consider *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. Let

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

correspond to 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.

3.2 Models

The models we consider in this paper are as follows:

(i) *Semi-infinite dimensional model.* Here we assume f is either sparse with bandwidth M, or that $f = g + h$, where

$$f = g + h, \qquad \Delta = \operatorname{supp}(g) \subset \{1, \ldots, M\}, \quad \operatorname{supp}(h) = \{1, \ldots, M\}. \tag{3.3}$$

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 draw samples from the countable collection (3.1).

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

$$f = g + h, \qquad \Delta = \operatorname{supp}(g) \subset \{1, \ldots, M\}, \quad |\operatorname{supp}(h)| = \infty. \tag{3.4}$$

This model is termed fully infinite-dimensional since both the set of samples and the support of f have infinite cardinality.

Note that whenever f is exactly sparse (i.e. the first case in model (i)) the goal is to recover f perfectly by subsampling the measurements (3.1). In all other cases, one cannot expect perfect recovery with subsampling. As is typical in finite-dimensional CS, the aim here is to show to show perfect recovery up to an error proportional to $\sigma_{r,M}(\alpha)$. Note that this is substantially more challenging in case (ii), since the support of the nonsparse component h is infinite.

4 Generalized sampling: guaranteed recovery in arbitrary bases

Before discussing how to recover infinite-dimensional sparse or compressible signals, it is first necessary to consider the more classical case where no sparsity is assumed. The question is, how does one actually reconstruct an arbitrary signal f from its measurements $\{\zeta_j(f)\}_{j\in\mathbb{N}}$? Or in other words, if $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$, how do we recover the infinite vector $\alpha = \{\alpha_1, \alpha_2, \ldots\}$ from the samples $\zeta_1(f), \zeta_2(f), \ldots$? Only once this problem has been solved can one tackle the issue of subsampling. Fortunately, the technique of *generalized sampling* (GS) was developed precisely for this problem [1, 2, 5, 3]. We now recap this approach.

Under some assumptions on $\{\zeta_j\}_{j\in\mathbb{N}}$ (e.g. each ζ_j is continuous and $\{\zeta_j(f)\}_{j\in\mathbb{N}} \in l^2(\mathbb{N})$, $\forall f \in \mathcal{H}$), we can view the full recovery problem as the infinite-dimensional system of linear equations

$$U\alpha = \zeta(f), \tag{4.1}$$

where $\alpha = \{\alpha_1, \alpha_2, \ldots\}$, $\zeta(f) = \{\zeta_1(f), \zeta_2(f), \ldots\}$ and U is the infinite measurement matrix

$$U = \begin{pmatrix} \zeta_1(\varphi_1) & \zeta_1(\varphi_2) & \cdots \\ \zeta_2(\varphi_1) & \zeta_2(\varphi_2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{4.2}$$

Clearly, if we were able to invert U, and provided we had access to all samples of f, then we could recover α (and hence f) exactly. However, this is never the case in practice. Instead, we must consider truncations of (4.1), and look to compute approximations $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_N$ to the first N coefficients of α.

It goes without saying that whatever strategy we use for computing such approximate coefficients, it must be stable in the presence of noise. Moreover, the resulting reconstruction $\tilde{\alpha}^{[N]} = \{\tilde{\alpha}_1, \ldots, \tilde{\alpha}_N\} \in \mathbb{C}^N$ must be a good approximation to the first N exact coefficients $\alpha_1, \ldots, \alpha_N$. Recall that the whole premise for recovering f in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$ is that we know that f is well represented in this basis. In other words, the coefficients $\{\alpha_j\}_{j\in\mathbb{N}}$ decay rapidly as $j \to \infty$, or, in the case where f is sparse, only a finite number are nonzero. Therefore, whichever method we use for solving (4.1), it is important that the error $\|P_N\alpha - \tilde{\alpha}^{[N]}\|_{l^2}$, be small (here and later, for convenience, we shall not make a distinction between the finite vector $\alpha^{[N]} = \{\tilde{\alpha}_1, \ldots, \tilde{\alpha}_N\} \in \mathbb{C}^N$ and its imbedding $\{\tilde{\alpha}_1, \ldots, \tilde{\alpha}_N, 0, 0, \ldots\}$ in $l^2(\mathbb{N})$). Clearly, the examples of §2 violate this condition.

4.1 Finite sections: a warning from spectral theory

The most obvious approach for discretizing (4.1) follows from taking finite sections of U. In other words, if $P_N : l^2(\mathbb{N}) \to \mathrm{span}\{e_j : j = 1, \ldots, N\}$ is the orthogonal projection, we consider solutions $\tilde{\alpha}^{[N]}$ to the $N \times N$ system of equations

$$P_N U P_N \tilde{\alpha}^{[N]} = P_N \zeta(f). \tag{4.3}$$

Note that

$$P_N U P_N = \begin{pmatrix} \zeta_1(\varphi_1) & \cdots & \zeta_1(\varphi_N) \\ \vdots & \ddots & \vdots \\ \zeta_N(\varphi_1) & \cdots & \zeta_N(\varphi_N) \end{pmatrix},$$

is nothing more than the leading $N \times N$ submatrix (i.e. the finite section) of U.

Finite sections are extremely widely used in practice [11, 12, 36]. However, for general operators U there is no guarantee that $\tilde{\alpha}^{[N]}$ need either exist, or that $\tilde{\alpha}^{[N]}$ (if it exists) actually converges to α as $N \to \infty$. In fact, it is easy to devise pairs of bases $\{\varphi_j\}_{j\in\mathbb{N}}$ and sampling schemes $\{\zeta_j\}_{j\in\mathbb{N}}$ for which the error $\|\alpha - \tilde{\alpha}^{[N]}\|_{l^2}$ blows up as $N \to \infty$, whenever $\tilde{\alpha}^{[N]}$ is the result of the finite section method [1, 5].

Another significant issue is that the finite section matrix $P_N U P_N \in \mathbb{C}^{N\times N}$ may be extremely poorly conditioned, even though U and its inverse U^{-1} are bounded. Examples of operators U whose finite sections exhibit exponentially poor conditioning were given in [1, 5]. In particular, the measurement matrix formed by sampling in the Fourier basis and reconstructing in Haar wavelets (the principal example of this paper) suffers from this phenomenon. As a result, the numerical method based on finite sections is not just nonconvergent, it is also extremely unstable and highly sensitive to noise.

The failure of the finite section method for solving (4.1) can be viewed as a violation of the principle (Ph). Finite sections have been studied extensively from the viewpoint of computational spectral theory. Therein one typically wishes to gain information about the spectrum of U by considering discretizations of the form $P_N U P_N$ [9, 36]. The main conclusion is that, unless U satisfies some very stringent restrictions (such as positive self-adjointness), its finite sections $P_N U P_N$ may have wildly different (spectral) properties. This violates (Ph), and thus makes finite sections typically unsuitable for solving (4.1).

The key structure that operator U of the form (4.2) possess is that they are isometries. Specifically, if measurements $\zeta_j(f) = \langle f, \psi_j \rangle$ give rise to an orthonormal basis $\{\psi_j\}_{j\in\mathbb{N}}$ (as is the case with Fourier sampling) the matrix U is a isometry on $l^2(\mathbb{N})$. Yet this structure is not usually inherited by the finite sections $P_N U P_N$, in violation of (Ph).

Note that the finite section $P_N U P_N$ is precisely the measurement matrix (2.5) encountered in the finite-dimensional CS approach (2.6). As commented in §2.2, the loss of the isometry property of U accounts for the failure seen therein.

4.2 Uneven sections and generalized sampling (GS)

Fortunately, there is a simple, albeit far less common, way to overcome the failure of finite section method, based on taking rectangular, as opposed to square, sections of U. In [1, 2] it was proposed to replace (4.3) with

$$A\tilde{\alpha}^{[M]} = P_M U^* P_N \zeta(f), \qquad A = P_M U^* P_N U P_M, \tag{4.4}$$

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 $A \equiv (P_N U P_M)^* P_N U P_M$, where $P_N U P_M$ is the $N \times M$ *uneven* section of U.

The main idea is that, by allowing M and N to vary independently—in particular, selecting $M \leq N$ sufficiently small—one can obtain both a numerically stable and accurate reconstruction of the first M coefficients $\alpha_1, \ldots, \alpha_M$. Note that this means that we typically recover fewer of the coefficients $\alpha_1, \alpha_2, \ldots$ than in the finite section method. However, unlike the latter, it is possible to guarantee both the stability and accuracy of this approach. In other words, by being less greedy in the number of coefficients we seek to recover, we actually obtain a far better result.

The main theorem proved in [1, 5] is as follows:

Theorem 4.1. *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 $\zeta(f) \in l^2(\mathbb{N})$. Then for each $M \in \mathbb{N}$ there exists an $N_0 \in \mathbb{N}$ independent of $\alpha \in l^2(\mathbb{N})$ such that, for every $N \geq N_0$, there is a unique solution $\tilde{\alpha}^{[M]}$ to (4.4). Furthermore, we have the sharp bound*

$$\|\alpha - \tilde{\alpha}^{[M]}\| \leq \frac{1}{\sqrt{1 - C_{N,M}}} \|P_M^{\perp} \alpha\|, \tag{4.5}$$

where

$$C_{N,M} = \|P_M - P_M U^* P_N U P_M\|. \tag{4.6}$$

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

It can be shown that the quantity $C_{N,M} \to 0$ as $N \to \infty$, for any fixed M. Thus, one deduces from (4.5) that $\tilde{\alpha}^{[M]}$ can be made arbitrarily close to $P_M \alpha$—the best approximation to α from $P_M(l^2(\mathbb{N}))$—by varying N suitably. Hence, a good reconstruction can always be guaranteed with this approach. Furthermore, the resulting method is also stable. The condition number of the matrix A scales like $\frac{1}{\sqrt{1-C_{N,M}}}$ [5]. That is to say, precisely the same quantity that ensures accuracy of the reconstruction also guarantees numerical stability.

To connect generalized sampling with the principle (Ph), note that the uneven section $P_N U P_M$ inherits the structure of U, whenever M and N are chosen suitably. In fact, for fixed M,

$$P_M U^* P_N U P_M \to P_M U^* U P_M = P_M, \quad N \to \infty,$$

where $I : l^2(\mathbb{N}) \to l^2(\mathbb{N})$ is the identity. Thus, $P_N U P_M$ is also an isometry on the range of P_M in the limit $N \to \infty$.

Remark 4.1 Theorem 4.1 is stated in a way such that M is fixed and N is varied, as are the results we give later in §7. Depending on the particular application, one may prefer to consider N, the total number of samples, being fixed and M being varied. Mathematically, however, this is completely equivalent.

4.3 The stable sampling rate

Theorem 4.1 proves that stable recovery is possible, provided N is chosen sufficiently large in comparison to M (or M sufficiently small in comparison to N). In practice, we need a way in which to quantify this scaling. In [5], the *stable sampling rate*

$$\Theta(M; \theta) = \min\left\{N \in \mathbb{N} : C_{N,M} < 1 - \theta^2\right\}, \quad \theta \in (0, 1), \tag{4.7}$$

was introduced. The stable sampling rate determines how tall to take the uneven section $P_N U P_M$, for a given width, to ensure that (Ph) holds. 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 θ.

Note that $\Theta(M; \theta)$ can be easily computed numerically [5], since $C_{N,M}$ is just the norm of an $M \times M$ matrix (see (4.6)). Hence the conditions of Theorem 4.1 can be verified *a priori* via a straightforward calculation. Having said this, in numerous circumstances of interest one can also obtain analytical bounds [1, 2, 4, 5].

4.4 A generalized Shannon Sampling Theorem

One of the main instances of GS, and a central reason for its development, is the case where $\{\zeta_j(f)\}_{j\in\mathbb{Z}}$ corresponds to the Fourier samples (2.1) (we replace the index set $\mathbb{N}$ with $\mathbb{Z}$ in this instance). Although the famous Shannon Sampling Theorem ensures that both f and its Fourier transform $\mathcal{F}f$ can be recovered exactly via the infinite sums

$$f(\cdot) = \epsilon \sum_{j\in\mathbb{Z}} \mathcal{F}f(j\epsilon)\mathrm{e}^{2\pi\mathrm{i}\epsilon j\cdot}, \qquad \mathcal{F}f(t) = \sum_{j\in\mathbb{Z}} \mathcal{F}f(j\epsilon)\mathrm{sinc}\left(\frac{t+j\epsilon}{\epsilon}\right),$$

(note that the first converges in L^2, whereas the second converges both uniformly and in L^2), in practice one has to truncate these series, leading to the approximations

$$f_N(t) = \epsilon \sum_{j=-\lfloor N/2\rfloor}^{\lfloor N/2\rfloor} \mathcal{F}f(j\epsilon)\mathrm{e}^{2\pi\mathrm{i}\epsilon jt}, \qquad \mathcal{F}f_N(t) = \sum_{j=-\lfloor N/2\rfloor}^{\lfloor N/2\rfloor} \mathcal{F}f(j\epsilon)\mathrm{sinc}\left(\frac{t+j\epsilon}{\epsilon}\right). \tag{4.8}$$

As discussed, these are typically very poor reconstructions of f and $\mathcal{F}f$ respectively.

However, suppose now we know another basis $\{\varphi_j\}_{j\in\mathbb{N}}$ in which f is well represented. Then we can apply GS to obtain an improved reconstruction in this basis. This leads to the following generalization of Shannon's theorem:

Theorem 4.2 (Generalized Sampling Theorem [1])**.** *Let $\mathcal{F}$ denote the Fourier transform on $\mathbb{R}$, and suppose that $\{\varphi_j\}_{j\in\mathbb{N}}$ is an orthonormal set in $L^2(\mathbb{R})$ satisfying $\mathrm{supp}(\varphi_j) \subset [-T,T]$ for all $j\in\mathbb{N}$ and some $T>0$. For $0<\epsilon\leq \frac{1}{2T}$ let*

$$\zeta_j(f) = \sqrt{\epsilon}\mathcal{F}f(\rho(j)\epsilon), \quad j\in\mathbb{Z}, \quad f\in L^2(\mathbb{R}),$$

where $\rho:\mathbb{N}\to\mathbb{Z}$ is some bijection, and suppose that U is given by (4.2). Then, for each $M\in\mathbb{N}$ there is an $N\in\mathbb{N}$ such that there exists a unique solution $\tilde{\alpha}^{[M]}\in\mathbb{C}^M$ to (4.4), for any $f\in\overline{\mathrm{span}}\{\varphi_j\}_{j\in\mathbb{N}}$. Moreover, if

$$f_{N,M} = \sum_{j=1}^{M} \tilde{\alpha}_j\varphi_j, \qquad g_{N,M} = \sum_{j=1}^{M} \tilde{\alpha}_j\mathcal{F}\varphi_j, \tag{4.9}$$

then

$$\|f-f_{N,M}\|_{L^2(\mathbb{R})} \leq \frac{1}{\sqrt{1-C_{N,M}}}\|\mathcal{P}_M^{\perp}f\|_{L^2(\mathbb{R})},$$

and

$$\|g-g_{N,M}\|_{L^2(\mathbb{R})} \leq \frac{\sqrt{2T}}{\sqrt{1-C_{N,M}}}\|\mathcal{P}_M^{\perp}f\|_{L^2(\mathbb{R})},$$

where $g=\mathcal{F}f$, $C_{N,M}$ is given by (4.6) and $\mathcal{P}_M$ denotes the projection onto $\mathrm{span}\{\varphi_1,\ldots,\varphi_M\}$.

Note that this theorem is just the special case of Theorem 4.1 corresponding to Fourier samples. It is also a straightforward exercise to extend it to the multivariate setting, where $\mathcal{F}$ corresponds to the Fourier transform on $L^2(\mathbb{R}^d)$ [1]. This theory extends the classical Shannon Sampling Theorem as well as a number of its fundamental generalizations [8, 27, 57, 58]. 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.

An important issue that was addressed in [1, 2, 4, 5] is how the stable sampling rate $\Theta(M;\theta)$ behaves in this setting. In [4] it was proved that $\Theta(M;\theta)\sim c(\theta)M$ for Fourier sampling with wavelets as the reconstruction system, i.e. 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 [4] it was also shown that setting $N=M$ (this corresponds to the finite section), or in fact $N=\tilde{c}M$ for any c less than some critical threshold $c_0>1$, leads to exponential instability and divergence.

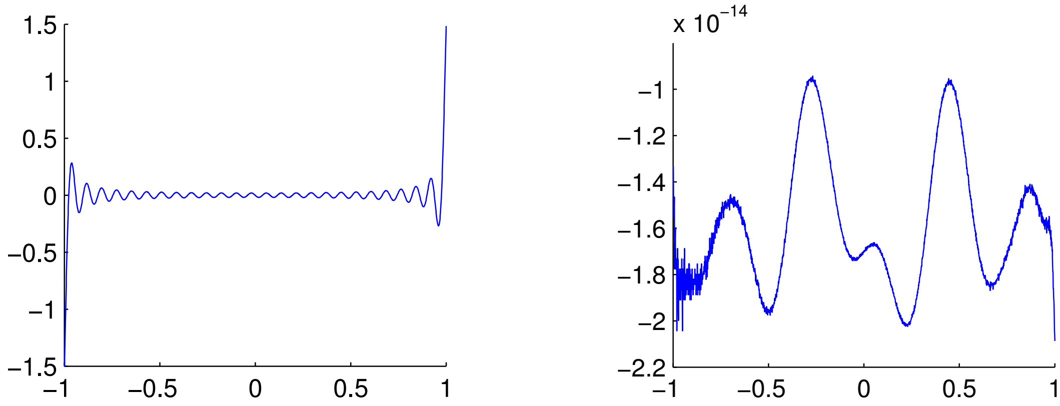

Figure 4: The figure shows the disappointing error $f(t) - f_N(t)$ against t (left) and the more pleasant error $f(t) - f_{N,M}(t)$ (right) for $N = 51$ and $M = 12$. Note that f_N and $f_{N,M}$ use exactly the same samples.

4.5 Example: the effectiveness of generalized sampling

To demonstrate the use of generalized sampling let us consider the following function

$$f(t) = t^5 e^{-t}, \qquad t \in [-1, 1].$$

Suppose we can sample the Fourier coefficients of f: in particular, we have access to $\{\mathcal{F}f(j\epsilon)\}_{j\in\mathbb{Z}}$ for $\epsilon = 1/2$. To reconstruct f from these samples we will consider two different techniques. First, we test the truncated Fourier series f_N defined in (4.8). Due to the fact that f is not periodic we cannot expect rapid convergence of f_N to f. However, the Generalized Sampling Theorem 4.2 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 $[-1, 1]$. In particular, we define $f_{N,M}$ as in (4.9), where $\rho : \mathbb{N} \to \mathbb{N}$ is given by

$$\rho(1) = 0, \rho(2) = 1, \rho(3) = -1, \rho(4) = 2, \rho(5) = -2 \ldots. \tag{4.10}$$

In Figure 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 $f_{N,M}$ compared to f_N is dramatic. In particular, we go from an $\mathcal{O}(1)$ error to roughly machine precision. We remark that for this choice of reconstruction basis the stable sampling rate $\Theta(M;\theta)$ is quadratic in M [2]. Moreover, a lower scaling (in particular, $N = M$) results in extreme ill-conditioning [7].

We now repeat the experiment above with the function

$$f(t) = \sin(5t) + \sum_{j=1}^{L} \alpha_j \psi_j(t), \qquad t \in [-1, 1],$$

where $\{\psi_j\}_{j\in\mathbb{N}}$ are the Haar wavelets on $[-1, 1]$, $L = 1700$ and the α_j's are some arbitrarily chosen coefficients. We will assume, as above, that we can sample the Fourier coefficients $\{\mathcal{F}f(j\epsilon)\}_{j\in\mathbb{Z}}$ of f (with $\epsilon = \frac{1}{2}$ once more). Due to the vast number of discontinuities of f we cannot expect the truncated Fourier series f_N to be a good approximation to f. However, by the Generalized Sampling Theorem 4.2 we can choose the reconstruction basis $\{\varphi_j\}_{j\in\mathbb{N}}$ to be the Haar wavelets, and construct $f_{N,M}$ as in (4.9). In Figure 5 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 $f_{N,M}$ compared to f_N is substantial.

5 Generalized sampling with compressed sensing (GS–CS)

An immediate consequence of Theorems 4.1 and 4.2 is that if we know that f is sparse in $\{\varphi_j\}_{j\in\mathbb{N}}$—i.e. $\mathrm{supp}(f) \subset \{1, \ldots, M\}$—then we can recover f perfectly from its first N samples, whenever $N \geq M$ is taken according to the stable sampling rate. However, by combining the same ideas with standard CS

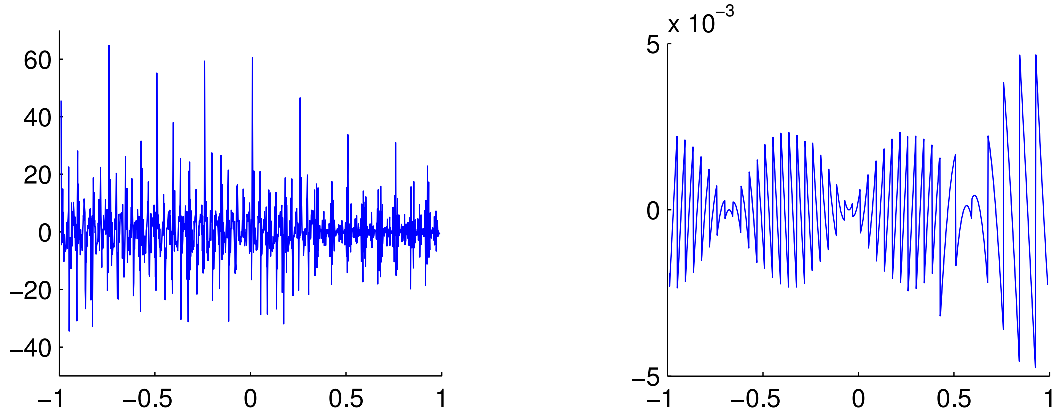

Figure 5: The figure shows the large error $f(t) - f_N(t)$ against t (left) as well as the substantially smaller error $f(t) - f_{N,M}(t)$ (right) for $N = 2401$ and $M = 1750$. Note that f_N and $f_{N,M}$ use exactly the same samples.

techniques, we can actually achieve perfect recovery using far fewer measurements. The key is to follow a similar approach, again based on uneven sections, to formulate the reconstruction appropriately. The resulting method is known as *generalized sampling with compressed sensing* (GS–CS).

Let us suppose that $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$ is sampled via $\{\zeta_j\}_{j\in\mathbb{N}}$. As opposed to the failed approaches of §2, which were loosely based on discretizing first, the technique we now propose involves first formulating the sparse recovery problem in infinite dimensions. To this end, let $\Omega \subset \mathbb{N}$ be of size $|\Omega| = m \in \mathbb{N}$ and consider the (infinite-dimensional) optimization problem

$$\min_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \zeta(f), \tag{5.1}$$

where U is the infinite measurement matrix (4.2) 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 (4.1) before discretization can proceed. Seeking similar notions for (5.1), we are led to the following questions:

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

(ii) Suppose that η is a minimizer of (5.1) (note that η need not be unique). How large is $\|\eta - \alpha\|$, where α 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 α exactly (with high probability) from (5.1), and what are the conditions on m and N that ensure this recovery?

Let us suppose for the moment that we have answers to these questions. Besides some special circumstances, we cannot solve (5.1) 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

$$\min_{\eta \in \mathbb{C}^M} \|\eta\|_{l^1} \text{ subject to } P_\Omega U P_k \eta = P_\Omega \zeta(f). \tag{5.2}$$

We shall refer to this as *generalized sampling with compressed sensing*. This formulation of course leads to another set of questions:

(iv) When will (5.2) have a solution? Note that (5.2) 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 (5.2) converge to solutions of (5.1)?

(v) If f is not sparse but compressible, how large is the error $\|\eta - \alpha\|$ when η is a solution to (5.2) and α is the vector of coefficients of f? In particular, if f belongs to either of the models (3.3) or (3.4), can $\|\eta - \alpha\|$ be bounded above in terms of $\sigma_{r,M}(f)$?

Answers to these questions will be provided in §7, where we state the main results of this paper.

6 Notation and definitions

6.1 Notation

We now introduce some additional notation that will be used in the remainder of this paper. Let $\mathcal{H} = l^2(\mathbb{N})$ be the standard space of square-summable sequences in $\mathbb{C}$, and let $\|\cdot\|$ be the standard norm on $\mathcal{H}$. All other norms will be specified. Let $\{e_j\}_{j\in\mathbb{N}}$ be the natural basis on $l^2(\mathbb{N})$, and, if $\Gamma \subset \mathbb{N}$, define P_Γ to be the orthogonal projection onto $\mathrm{cl}(\mathrm{span}\{e_j : j \in \Gamma\})$. If $\Gamma = \{1, \ldots, N\}$, then we simply write P_N.

If $\xi \in \mathcal{H}$ and $j \in \mathbb{N}$, then $\xi(j) = \langle \xi, e_j \rangle$ (we will also sometimes use the notation ξ_j). For $\Gamma \subset \mathbb{N}$, we denote the natural embedding operator by $\iota_\Gamma : l^2(\Gamma) \to \mathcal{H}$. Note that $\iota_\Gamma^* \eta = \eta|_\Gamma$ for $\eta \in \mathcal{H}$. For any vector $\xi \in \mathcal{H}$ we write $\mathrm{supp}(\xi) = \{j \in \mathbb{N} : \xi(j) \neq 0\}$. We also 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 define coherence parameter

$$\upsilon(U) = \sup_{i,j\in\mathbb{N}} |u_{ij}|, \qquad u_{ij} = \langle U e_j, e_i \rangle, \tag{6.1}$$

i.e. the max norm of the operator U with respect to $\{e_j\}_{j\in\mathbb{N}}$. Also, if $U = \{u_{ij}\}_{ij\in\mathbb{N}}$ is an infinite matrix, we define the maximum row norm of U by

$$\|U\|_{\mathrm{mr}} = \sup_{i\in\mathbb{N}} \sqrt{\sum_{j\in\mathbb{N}} |u_{ij}|^2}.$$

This quantity forms a vector space norm on the vector space of all infinite matrices (although not an algebra norm). Finally, for convenience, we will define the following crucial function that will be used frequently in the exposition. For $M \in \mathbb{N}$ and $U \in \mathcal{B}(\mathcal{H})$ let $\tilde{\omega}_{M,U} : \{1, \ldots, M\} \times \mathbb{R}_+ \times \mathbb{N} \to \mathbb{N}$ be given by

$$\tilde{\omega}_{M,U}(r, s, N) = \left| \left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\ldots,M\}, |\Gamma_1| = r \\ \Gamma_2 \subset \{1,\ldots,N\}}} \|P_{\Gamma_1} U^* P_{\Gamma_2} U e_i\| > s \right\} \right|. \tag{6.2}$$

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

6.2 Key definition: the balancing property

For GS, the stable sampling rate (4.7) determines how to discretize the recovery problem in line with (Ph), by determining how to choose N for a given M. For GS–CS we require an analogous quantity, known as the balancing property:

Definition 6.1. *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry. Then N and m satisfy the weak Balancing Property with respect to U, M and $|\Delta|$ if*

$$\|P_M U^* P_N U P_M - P_M\| \leq \left(4\sqrt{\log_2\left(4N\sqrt{|\Delta|}/m\right)} \right)^{-1}, \tag{6.3}$$

$$\max_{|\Gamma|=|\Delta|, \Gamma\subset\{1,\ldots,M\}} \|P_M P_\Gamma^\perp U^* P_N U P_\Gamma\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}. \tag{6.4}$$

We say that N and m satisfy the strong Balancing Property with respect to U, M and $|\Delta|$ if (6.3) holds, and (6.4) is replaced by

$$\max_{|\Gamma|=|\Delta|,\Gamma\subset\{1,\dots,M\}} \|P_\Gamma^\perp U^* P_N U P_\Gamma\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}. \tag{6.5}$$

Remark 6.1 The inequality in (6.4) is somewhat inconvenient. However, it can be replaced by the far simpler, although stronger, condition

$$\|P_M U^* P_N U P_M - \mathrm{diag}(P_M U^* P_N U P_M)\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}. \tag{6.6}$$

Here $\mathrm{diag}(B)$ denotes part of the matrix B. In particular, condition (6.6) 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, (6.5) can also be replaced by the much more convenient (however much stronger) condition

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

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

Proposition 6.2. *If U, M and $|\Delta|$ are as in Definition 6.1, then there always exists integers N and m that satisfy the weak and strong Balancing Properties with respect to U, M and $|\Delta|$.*

Proof. 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 by compactness that $P_N U P_\Gamma \to U P_\Gamma$ in norm as $N \to \infty$. The fact that U is an isometry yields the assertion. $\square$

Before stating our main results, let us briefly comment on the balancing property. Condition (6.3) ensures that $P_N U P_M$ is close to an isometry, and is very similar to the stable sampling rate in the non-subsampled case. Since results proved in [4] establish that the stable sampling rate cannot be avoided for essentially any recovery algorithm, (6.3) is a natural, and most likely necessary condition. Having said this it may be possible to slightly improve the right-hand side. Condition (6.4) states that the matrix U should be close to an isometry on the support of all sparse vectors. Again, we suspect that this condition cannot be removed, although it may be possible to weaken it somewhat.

7 Main results

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

7.1 The semi-infinite dimensional model

The first results concern the semi-infinite dimensional model (see §3.2):

Theorem 7.1. *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $M \in \mathbb{N}$, $\epsilon > 0$ and suppose that $x_0 \in l^1(\mathbb{N})$ 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*

$$\inf_{\eta\in l^1(\mathbb{N})} \|\eta\|_{l^1} \quad \textit{subject to} \quad P_\Omega U P_M \eta = P_\Omega \zeta, \tag{7.1}$$

has a unique solution ξ and this solution coincides with x_0, provided that m satisfies

$$m \geq C \cdot N \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{MN\sqrt{|\Delta|}}{m}\right), \tag{7.2}$$

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

The main conclusion of this theorem is as follows: a sparse signal x_0 can be recovered perfectly (with high probability) by subsampling from the coefficients ζ, provided (6.3), (6.4) and (7.2) hold. Note that this result gives answers to the questions (i) and (iii) posed in §5. Moreover, Theorem 7.1 establishes Theorem 1.1 of §1.

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 $\mathrm{supp}(h) \subset \{1, \ldots, M\}$. The following theorem concerns this case:

Theorem 7.2. *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $M \in \mathbb{N}$, $\epsilon > 0$ and suppose that $y_0 \in l^1(\mathbb{N})$ with $\mathrm{supp}(y_0) \subset \{1, \ldots, M\}$. Suppose that N and m satisfy the weak Balancing Property with respect to U, M and $|\Delta|$, and let $\Omega \subset \{1, \ldots, N\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = Uy_0$ and $\xi \in \mathcal{H}$ is a minimizer of (7.1) then, with probability exceeding $1 - \epsilon$, we have that*

$$\|\xi - y_0\| \leq 8 \left(1 + \frac{2N}{m}\right) \sigma_{r,M}(y_0), \qquad r = |\Delta|, \tag{7.3}$$

provided m satisfies (7.2). If $m = N$ then (7.3) holds with probability 1.

This theorem demonstrates recovery for compressible signals of the form (3.3). Specifically, we witness perfect recovery, up to an error determined by the best (r, M)-term approximation. In particular, this result answers part of question (v) posed previously.

Remark 7.1 It is important to notice that there need not be a unique solution to (7.1). However, this is not an issue. Theorem 7.2, and, in particular, equation (7.3), states that *all* solutions to (7.1) will be close to y_0 in norm.

7.2 The fully infinite-dimensional model

Recall that the semi-infinite dimensional model (3.3) places the restriction that the support of the nonsparse term h is contained in $\{1, \ldots, M\}$. As discussed in §3, this assumption is quite rare in practice, and a more realistic setting is provided by the fully infinite-dimensional model. Here 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 scrutinize an infinite-dimensional optimization problem of the form (5.1):

Theorem 7.3. *Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry, $M \in \mathbb{N}$, $\epsilon > 0$ 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 $|\Delta|$, for some M and Δ, and let $\Omega \subset \{1, \ldots, N\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = Uy_0$ and $\xi \in \mathcal{H}$ is a minimizer of*

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega U \eta = P_\Omega \zeta, \tag{7.4}$$

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

$$\|\xi - y_0\| \leq 8 \left(1 + \frac{2N}{m}\right) \sigma_{r,M}(y_0), \qquad r = |\Delta|, \tag{7.5}$$

provided that m satisfies

$$m \geq C \cdot N \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\epsilon^{-1}\right) + 1\right) \cdot \log\left(\frac{\omega N \sqrt{|\Delta|}}{m}\right), \tag{7.6}$$

$$\omega = \tilde{\omega}_{M,U}(|\Delta|, s, N), \qquad s = \frac{m}{128N\sqrt{|\Delta|}\log(e^4\epsilon^{-1})}$$

for some universal constant C (recall $\tilde{\omega}_{M,U}$ from (6.2)). If $m = N$ then (7.5) holds with probability 1.

Remark 7.2 The quantity ω in (7.6) can also be replaced by a much more convenient (and of course much less sharp) estimate. In particular we have that $\omega \leq \widetilde{M}$, where

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

Note that $\widetilde{M}$ is finite, since $\|P_N U P_r^\perp\| \to 0$ as $r \to \infty$. for fixed N.

This theorem, much like Theorem 7.2, 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 §5. However, observe that the optimization problem (7.4) is infinite-dimensional. In practice, one always replaces (7.4) with the finite-dimensional problem

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega U P_k \eta = P_\Omega \zeta, \tag{7.7}$$

where $k \in \mathbb{N}$ is suitably chosen. The obvious question now arises: how do solutions of (7.7) compare to those of (7.4) as $k \to \infty$? For this we have the following:

Proposition 7.4. *Let $U \in \mathcal{B}(\mathcal{H})$, $x_0 \in l^1(\mathbb{N})$ and P_Ω be a finite rank projection. Then, for all sufficiently large $k \in \mathbb{N}$, there exists an $\xi_k \in \mathcal{H}$ satisfying*

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

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 $\tilde{\xi}_k$ satisfies

$$\|\tilde{\xi}_k\|_{l^1} = \inf_{\eta \in l^1(\mathbb{N})} \left\{ \|\eta\|_{l^1} : P_\Omega U \eta = P_\Omega U x_0 \right\}. \tag{7.8}$$

In particular, if there exists a unique minimizer z of (7.8), then $\xi_k \to z$ in the l^1 norm.

This proposition states that the computed solutions of (7.7) will be approximate minimizers of (7.4) for all sufficiently large k. In particular, computed solutions will approximately satisfy (7.5). Note that it addresses question (iv) posed in §5.

Let us now make several further remarks on these theorems:

Remark 7.3 The error bounds (7.3) and (7.5) are nearly optimal in the sense that they involve the best approximation error $\sigma_{r,M}(y_0)$ multiplied by a factor proportional to N/m. The latter 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). We suspect, however, that it is possible to remove this term altogether.

Remark 7.4 Neither the bandwidth M nor the sparsity $r = |\Delta|$ need be known in either Theorem 7.2 or 7.3. 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 (7.6). This is typical in applications such as MRI, where the sampling resolution N is fixed (due to the physical limitations of the scanner), as is the total number of samples m.

Remark 7.5 In all the theorems, the amount of subsampling depends on the coherence parameter $\upsilon(U)$. For a specific operator U this is fixed, although it can be arbitrarily small. The fact that it is fixed suggests that for large enough M and N subsampling will not be possible—i.e. we must take $m = N$. However, if U has the property that $\upsilon(UP_k^\perp), \upsilon(P_k^\perp U) \to 0$ as $k \to \infty$, then one can actually circumvent this problem. This is achieved via *multilevel subsampling* techniques. This is not within the scope of this paper but will be treated elsewhere [6]. Note that [6] will also address the issue of noisy measurements. Again, this is a topic outside the scope of this paper.

7.3 Theorems on finite-dimensional CS

As mentioned, GS–CS extends standard finite-dimensional CS to an infinite-dimensional setting. It is therefore unsurprising, but important to note nonetheless, that results concerning the latter can be obtained as straightforward corollaries of Theorems 7.1–7.3. In particular, we have

Theorem 7.5. *Let $U \in \mathbb{C}^{n \times n}$ be an isometry, and suppose that $x_0 \in \mathbb{C}^n$ with $\mathrm{supp}(x_0) = \Delta$. For $\epsilon > 0$ suppose that $m \in \mathbb{N}$ is such that*

$$m \geq C \cdot n \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log(\epsilon^{-1}) + 1\right) \cdot \log n, \tag{7.9}$$

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 ξ and this solution coincides with x_0.

Theorem 7.6. *Let $U \in \mathbb{C}^{n\times n}$ be an isometry, and suppose that $y_0 = x_0 + h \in \mathbb{C}^n$ with $\mathrm{supp}(x_0) = \Delta$. For $\epsilon > 0$ suppose that $m \in \mathbb{N}$ satisfies (7.9), and let $\Omega \subset \{1, \dots, n\}$ be chosen uniformly at random with $|\Omega| = m$. If $\zeta = Uy_0$ then, with probability exceeding $1-\epsilon$, any minimizer ξ of the problem*

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

satisfies

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

Proof of Theorems 7.5 and 7.6. 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, 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 Theorem 7.1 or Theorem 7.2. $\square$

Remark 7.6 Similar results on finite-dimensional CS has recently been proved by Candès & Plan [15]. The main contribution of this paper is to extend this to infinite dimensions by the judicious use of uneven section techniques and the key concept of the balancing property.

8 Numerical examples

Before giving proofs of these theorems, it is useful to present some further examples of GS–CS. We will demonstrate the main premise of this paper in practice: namely, provided one knows that the function f has a good representation in terms of a different basis then one can obtain a far better reconstruction of f than that guaranteed by the Shannon Sampling Theorem. Consider the problem of reconstructing $g = \mathcal{F}f$ and f from the samples $\{\zeta_j(f)\}_{j\in\mathbb{N}}$ where $\zeta_j(f) = \mathcal{F}f(\rho(j)\epsilon)$, $\epsilon > 0$ (we will use $\epsilon = 1/2$) and ρ is defined in (4.10). We now compare three methods for approximating f and g:

(i) The partial Fourier series f_N and g_N (see (4.8)).
(ii) The GS reconstructions $f_{N,M}$ and $g_{N,M}$ (see Theorem 4.2).
(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 (5.2).

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 both examples 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 incoherence issues (see Remark 7.5).

8.1 First example

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

$$f(t) = \sum_{j=1}^{200} \alpha_j \varphi_j(t) + \cos(2\pi x)\chi_{[\frac{1}{2},\frac{9}{16}]}(t), \quad t \in [0,1], \tag{8.1}$$

where $\{\varphi_j\}_{j\in\mathbb{N}}$ 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

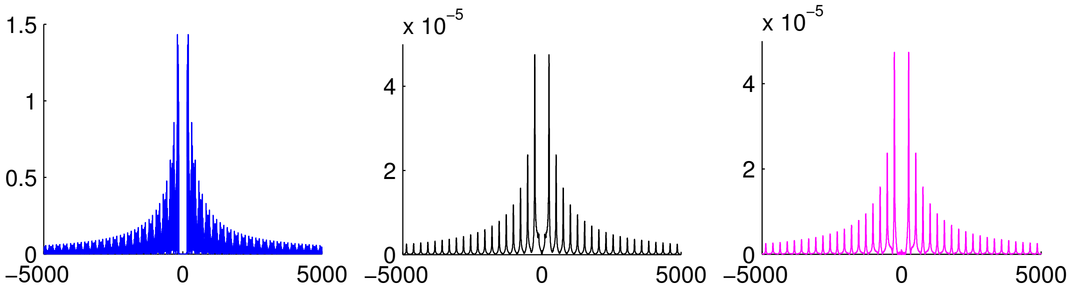

Figure 6: The figure displays the errors $|g(t)-g_N(t)|$ (left), $|g(t)-g_{N,M}(t)|$ (middle) and $|g(t)-g_{N,m,k}(t)|$ (right) against t, for $N=601$, $M=200$, $m=230$ and $k=650$.

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	4.74×10^{-5}, $(M=200)$	4.73×10^{-5}, $(m=230, k=550)$
1201	0.85	2.36×10^{-5}, $(M=400)$	2.38×10^{-5}, $(m=460, k=1400)$

Table 1: The tables displays the errors for the reconstructions g_N, $g_{N,M}$ and $g_{N,m,k}$.

that the remainder has infinite support in the Haar wavelet basis, so this function belongs to the fully-infinite dimensional model (see §3.2).

In Figure 6 we display the errors committed by the approximations (i)–(iii) for this function. As expected, the expansion in sinc functions (i) gives an extremely poor reconstruction, whereas both the GS and GS–CS give far better approximations. Specifically, by replacing the sinc series (i) with either (ii) or (iii) one reduces 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 confirmed in Table 1.

Whilst the GS and GS–CS methods give very similar numerical errors it is important to notice that the reconstructions $f_{N,M}$ and $f_{N,m,k}$ 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 discrepancy can be explained by examining the equations (4.4) and (5.2). In GS, which corresponds to (4.4), one requires $M<N$ to ensure invertibility of the operator A. On the other hand, unless k is taken sufficiently large, (5.2) 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 may well be the case whenever $k<N$. Fortunately, as shown in Proposition 7.4, this cannot happen if k is sufficiently large. The effect of increasing k for the example (8.1) is illustrated in Table 2: once k is sufficiently large, the problem (5.2) has a solution, and this error drops accordingly.

8.2 Second example

Consider the function

$$f(t)=\sum_{j=1}^{500}\alpha_j\varphi_j(t), \tag{8.2}$$

where $|\{j:\alpha_j\neq 0\}|=100$. This function is sparse in the Haar wavelet basis (and hence is an example of the semi-infinite dimensional model). The task is to reconstruct f from its Fourier samples.

Unlike the previous example, we expect exact reconstruction of this function using both the GS and GS–CS approaches, provided the parameters are chosen correctly. This is confirmed in Table 3. Observe that the Fourier series of f requires over $50,000$ Fourier samples to achieve four digits of accuracy. Conversely, the GS approximation recovers f exactly using only 1501 such samples. Furthermore, the GS–CS approximation improves over GS by a factor of three: it requires only 450 Fourier samples in total.

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	$E_{N,460,400} = \infty$	$E_{N,460,500} = \infty$	$E_{N,460,1000} = 2.384 \times 10^{-5}$	$E_{N,460,1300} = 2.392 \times 10^{-5}$

Table 2: The table shows 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 (5.2) does not have a solution).

N	$\|f - f_N\|_{L^2}$	$\|\tilde{f} - f_{N,M}\|_{L^2}$	$\|f - f_{N,m,k}\|_{L^2}$ (avg. 20 trials)
1001	4.19	$8.47 \times 10^{-2}, (M = 500)$	$5.53 \times 10^{-1}, (m = 450, k = 900)$
1501	1.43	$4.74 \times 10^{-15}, (M = 500)$	$1.06 \times 10^{-10}, (m = 450, k = 900)$
2001	1.39	$4.33 \times 10^{-15}, (M = 500)$	$1.99 \times 10^{-10}, (m = 450, k = 900)$
3001	1.37	$4.45 \times 10^{-15}, (M = 500)$	$1.98 \times 10^{-10}, (m = 450, k = 900)$
50001	2.84×10^{-4}		

Table 3: The table shows the error corresponding to the reconstructions f_N, $f_{N,M}$ and $f_{N,m,k}$ of the function (8.2).

8.3 Interlude: the remainder of the paper

In the first half of this paper we introduced the new framework GS–CS for compressed sensing in infinite dimensions, and explained why it is needed. The main recovery results concerning GS–CS were stated in §7. We devote the remainder of the paper to the proofs of these results.

9 Infinite-dimensional optimization and Proposition 7.4

We begin the second half of this paper with a proof of Proposition 7.4. As the informed reader will have noticed, this is really an question of infinite-dimensional optimization: in particular, 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. The well-informed reader may proceed directly to Proposition 9.4. Also, some of the results below, although new, are included only for completeness, and the reader only interested in the proof of the main theorems may go directly to Proposition 10.4.

9.1 Infinite-dimensional optimization

The field of infinite-dimensional convex optimization is certainly not new [26, 46]. However, it is much less standard than the more thoroughly investigated topic of finite-dimensional convex optimization. We will now cover some of the basic tools that will subsequently prove useful.

In this paper we will consider complex vector spaces. Standard optimization theory is usually considered over the reals, and this is also the case in [26] (the main reference we consider herein for the field of infinite-dimensional optimization). To be able to able to quote [26] 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 $\widetilde{X}$.

Definition 9.1. *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 9.2. *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 Φ is defined by

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

If Φ 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:

$$(\widetilde{P}_1): \quad \inf\{F(x) + G(Tx), x \in X\}, \tag{9.1}$$

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

$$\Phi(x, y) = F(x) + G(Tx + y), \tag{9.2}$$

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 Φ:

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

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 [26]:

Proposition 9.3. *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*

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

Proof. Let F, G and Φ be as in (9.1) and (9.2) respectively, and define $h : Y \to \mathbb{R} \cup \{\infty\}$ by

$$h(y) = \inf_{x \in X} \Phi(x, y).$$

Then h is convex and, since T is onto, h is finite for all $y \in Y$. Therefore, by convexity, h is also continuous, and, in particular continuous at zero. The result follows now by [26, Prop. 3.3.5]. $\square$

9.2 Proof of Proposition 7.4

We are now in a position to prove Proposition 7.4. We first require the following:

Lemma 9.4. *Let $U \in \mathcal{B}(\mathcal{H})$ and P be a finite rank projection. Then, for every $\chi \in \mathrm{Ran}(PU)$, there exists $\xi \in \mathcal{H}$ satisfying*

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad PU\eta = \chi.$$

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 $j \in \mathbb{N}$. It follows that $\|\xi\|_{l^1} \leq \lim_{k\to\infty} \|\xi_k\|_{l^1}$. 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 $PU\xi_k \to PU\xi$. Hence, $PU\xi = \chi$, as required. $\square$

We now give a proof of Proposition 7.4:

Proof of Proposition 7.4. To see the existence of ξ_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_Ω 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 ξ satisfying

$$\|\xi\|_{l^1} = \inf_{\eta\in\mathcal{H}} \{\|\eta\|_{l^1} : P_\Omega U\eta = P_\Omega U x_0\}. \tag{9.3}$$

Let therefore $\{\xi_k\}_{k\in\mathbb{N}}$ 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 ξ satisfies (9.3) we argue as follows. We claim that for any $\lambda > 0$ we have

$$\|\xi_k\|_{l^1} \leq \inf_{\eta\in\mathcal{H}} \{\|\eta\|_{l^1} : P_\Omega U\eta = P_\Omega U x_0\} + \lambda, \tag{9.4}$$

for all sufficiently large k. Let $r = \dim(\mathrm{Ran}(P_\Omega U)) < \infty$, and let $\hat{e}_1, \dots, \hat{e}_r$ 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_js are bounded by, say, $1 \leq c < \infty$. Now let $\tilde{\xi}$ be a minimizer of (9.3) (the existence of such a minimizer is guarantied by Proposition 9.4), 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 $c_1, \dots, c_r$ 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 (9.4). 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 ξ_k satisfies (9.4), 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 ξ satisfies (9.3), 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 (9.4) that

$$\|\xi_k\|_{l^1} \leq \inf_{\eta\in\mathcal{H}} \{\|\eta\|_{l^1} : P_\Omega U\eta = P_\Omega U x_0\} + 1/k. \tag{9.5}$$

Note also that, for fixed $m \in \mathbb{N}$, we have $P_m(\xi_k - \xi) \to 0$ as $k \to \infty$. But by (9.5) 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 we are done. $\square$

9.3 Existence of unique minimizers

In what follows it will be useful to have several results on the existence of unique minimizers of such problems. The finite-dimensional version of the following proposition has become standard for showing existence of unique minimizers for finite-dimensional problems found in CS [17]. Fortunately, the extension to infinite dimensions is rather straightforward:

Proposition 9.5. *Let $U \in \mathcal{B}(\mathcal{H})$ be unitary and let $\Omega, \Delta \subset \mathbb{N}$ be such that $|\Omega|, |\Delta| < \infty$. Suppose that $x_0 \in \mathcal{H}$ and that $\mathrm{supp}(x_0) = \Delta$. Consider the optimization problem*

$$\inf_{\eta\in\mathcal{H}} \|\eta\|_{l^1} \text{ subject to } P_\Omega U\eta = P_\Omega U x_0. \tag{9.6}$$

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) $\langle \rho, e_j\rangle = \langle \mathrm{sgn}(x_0), e_j\rangle, \quad j \in \Delta$

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

and in addition $P_\Omega U P_\Delta : P_\Delta \mathcal{H} \to P_\Omega \mathcal{H}$ *has full rank, then* x_0 *is the unique minimizer of (9.6).*

This proposition (for a proof, see the Appendix) may be a little hard to work with in practice. However, a more convenient result with somewhat relaxed assumptions can also be obtained.

Proposition 9.6. *Let* $U \in \mathcal{B}(\mathcal{H})$ *with* $\|U\| \leq 1$ *and suppose that* Δ *and* Ω *are finite subsets of* $\mathbb{N}$. *Let* $x_0 \in \mathcal{H}$ *such that* $\mathrm{supp}(x_0) = \Delta$. *Let* $M \in \mathbb{N}$ *and suppose that* M *is so large that* $\Delta \subset \{1, \ldots, M\}$. *Let* $\xi, \xi_M \in \mathcal{H}$ *such that*

$$\|\xi\|_{l^1} = \inf_{\eta \in \mathcal{H}} \{\|\eta\|_{l^1} : P_\Omega U \eta = P_\Omega U x_0\},$$
$$\|\xi_M\|_{l^1} = \inf_{\eta \in \mathcal{H}} \{\|\eta\|_{l^1} : P_\Omega U P_M \eta = P_\Omega U x_0\}.$$

Suppose that there is a $\rho \in \mathrm{ran}(U^* P_\Omega)$ *and a* $q > 0$ *with the following properties*

(i) $\|q^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\| \leq 1/2,$

(ii) $\|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq \sqrt{q}/4,$

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

then $\xi = x_0$. *Also, if (i) and (ii) are satisfied and (iii) is replaced with* $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2$ *then* $\xi_M = x_0$

Proof. Let $\zeta = \xi - x_0$. We will show that $\zeta = 0$. We begin by showing that $\|P_\Delta \zeta\| \leq \sqrt{\frac{2}{q}} \|P_\Delta^\perp \zeta\|$. This follows from some simple observations. First note that by a small computation and (i) we have

$$\|P_\Omega U P_\Delta \zeta\|^2 \geq q(1 - \|q^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\|)\|P_\Delta \zeta\|^2 \geq \frac{q}{2}\|P_\Delta \zeta\|^2.$$

Also, by assumption, we obviously have $\|P_\Delta^\perp \zeta\| \geq \|P_\Omega U P_\Delta^\perp \zeta\|$. Thus, if $\|P_\Delta \zeta\| > \sqrt{\frac{2}{q}} \|P_\Delta^\perp \zeta\|$ we get

$$\|P_\Omega U P_\Delta \zeta\| > \|P_\Delta^\perp \zeta\| \geq \|P_\Omega U P_\Delta^\perp \zeta\|.$$

Since $P_\Omega U \zeta = 0$ this is a contradiction.

Let us now note the following: for $j \in \Delta$ we have

$$|(x_0 + \zeta)(j)| = ||(x_0)(j)| + \zeta(j)\overline{\mathrm{sgn}(x_0)(j)}| \geq |(x_0)(j)| + \mathrm{Re}(\zeta(j)\overline{\mathrm{sgn}(x_0)(j)}).$$

Since $\mathrm{supp}(x_0) = \Delta$ we obtain

$$\|x_0 + \zeta\|_{l^1} \geq \|x_0\|_{l^1} + \mathrm{Re}\langle \zeta, \mathrm{sgn}(x_0)\rangle + \sum_{j \in \Delta^c} |\zeta(j)|, \tag{9.7}$$

where $\Delta^c = \mathbb{N}\backslash\Delta$. Also, by the assumption that $\rho \in \mathrm{ran}(U^* P_\Omega)$ and the fact that $P_\Omega U \zeta = 0$, it follows that $\zeta \perp \rho$. Thus, using (9.7) we obtain (by applying (ii), (iii), Hölder's inequality and finally the observation $\|P_\Delta \zeta\| \leq \sqrt{\frac{2}{q}} \|P_\Delta^\perp \zeta\|$)

$$\begin{aligned}
\|x_0 + \zeta\|_{l^1} &\geq \|x_0\|_{l^1} + \mathrm{Re}\langle \zeta, \mathrm{sgn}(x_0) + P_\Delta^\perp \overline{\mathrm{sgn}(\zeta)} - \rho\rangle \\
&\geq \|x_0\|_{l^1} + \|P_\Delta^\perp \zeta\|_{l^1} - (|\langle \zeta, \mathrm{sgn}(x_0) - P_\Delta \rho\rangle| + |\langle \zeta, P_\Delta^\perp \rho\rangle|) \\
&\geq \|x_0\|_{l^1} + \|P_\Delta^\perp \zeta\|_{l^1} - \left(\frac{\sqrt{q}}{4}\|P_\Delta \zeta\|_{l^1} + \frac{1}{2}\|P_\Delta^\perp \zeta\|_{l^1}\right) \\
&\geq \|x_0\|_{l^1} + \|P_\Delta^\perp \zeta\|_{l^1} - \left(\frac{\sqrt{2}}{4}\|P_\Delta^\perp \zeta\|_{l^1} + \frac{1}{2}\|P_\Delta^\perp \zeta\|_{l^1}\right).
\end{aligned} \tag{9.8}$$

Thus, if $\zeta \neq 0$ this gives $\|x_0 + \zeta\|_{l^1} > \|x_0\|_{l^1}$ contradicting the fact that $\|\xi\|_{l^1} \leq \|x_0\|_{l^1}$. Hence $\zeta = 0$, and this gives the first part of the proposition. The argument for the second part of the proposition is almost identical. By letting $\zeta = \xi_M - x_0$ we may use exactly the same analysis as previously, except for the transition from the second line in (9.8) to the third line. In that case, since $\zeta \in \mathrm{ran}(P_M)$, we only need the requirement that $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2$. $\square$

10 Stability analysis for infinite-dimensional convex optimization

In the previous section we established conditions that guarantee recovery of $x_0 \in l^1(\mathbb{N})$ by solving

$$\min_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega U x_0, \tag{10.1}$$

and its finite-dimensional approximations

$$\min_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U P_k \eta = P_\Omega U x_0. \tag{10.2}$$

In particular, we gave a proof of Proposition 7.4.

We now consider the issue of stability in such optimization problems. In other words, we consider the effect of replacing x_0 by $x_0 + h$, where h is small in norm, on the minimizers ξ and ξ_k of (10.1) and (10.2) respectively. Note that this is the first step towards a proof of Theorems 7.2 and 7.3 concerning the recovery of compressible signals which are described by the semi/fully infinite-dimensional models §3. However, at this moment we do not consider either sparsity or randomness. This comes in §11, in which the results proved in this and the previous section are applied to the sparse recovery problems (7.1) and (7.4) to yield proofs of Theorems 7.1–7.3.

10.1 Stability

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

Definition 10.1. *Let Ω, Δ be finite subsets of $\mathbb{N}$, $U \in \mathcal{B}(\mathcal{H})$ and let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function such that $\lim_{t\to 0} f(t) = 0$. If $\xi \in \mathcal{H}$, $\mathrm{supp}(\xi) = \Delta$, is the unique minimizer of*

$$\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \xi\}, \tag{10.3}$$

and for any $\epsilon > 0$ and $\zeta \in \mathcal{H}$ such that $\|\xi - \zeta\|_{l_1} \le \epsilon$, we have that

$$\|x - \xi\|_{l_1} \le f(\epsilon),$$

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 ξ. If $f(t) = Ct$ for some constant $C > 0$ then $\{U, \Omega, \Delta\}$ is said to be locally linearly stable at ξ. We say that $\{U, \Omega, \Delta\}$ is globally f-stable (linearly stable) if the above statements hold for all $\xi \in \mathcal{H}$, $\mathrm{supp}(\xi) = \Delta$, such that ξ is the unique minimizer of (10.3).

Proposition 10.2. *Let $U \in \mathcal{B}(\mathcal{H})$ be unitary and let Ω, Δ be finite subsets of $\mathbb{N}$. Suppose that $\{U, \Omega, \Delta\}$ is globally f-stable. Suppose also that there exists $x \in \mathcal{H}$, $\mathrm{supp}(x) = \Delta$, such that x is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U x\}$. Then, if $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta|_{P_\Delta \mathcal{H}}$ is invertible, and $y \in \mathcal{H}$, $\mathrm{supp}(y) = \Delta$, is arbitrary, then y is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U y\}$.*

Proposition 10.3. *Let $U \in \mathcal{B}(\mathcal{H})$ be unitary and let Ω, Δ be finite subsets of $\mathbb{N}$. Suppose that for any $\xi \in \mathcal{H}$, $\mathrm{supp}(\xi) = \Delta$, then ξ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U \eta = P_\Omega U \xi\}$, and also that $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta|_{P_\Delta \mathcal{H}}$ is invertible. Then, $\{U, \Omega, \Delta\}$ is globally linearly stable.*

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 $\mathrm{supp}(y) = \Delta$ 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.

10.2 The key result

The key result of this section, which will later lead to the proofs of Theorems 7.1–7.3, is the following:

Proposition 10.4. *Let $U \in \mathcal{B}(\mathcal{H})$ with $\|U\| \le 1$, and suppose that Δ and Ω 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 $\xi, \xi_M \in \mathcal{H}$ satisfy*

$$\|\xi\|_{l^1} = \inf_{\eta \in \mathcal{H}} \{\|\eta\|_{l^1} : P_\Omega U \eta = P_\Omega U (x_0 + h)\}, \tag{10.4}$$

$$\|\xi_M\|_{l^1} = \inf_{\eta\in\mathcal{H}}\{\|\eta\|_{l^1} : P_\Omega U P_M \eta = P_\Omega U(x_0 + P_M h)\}.$$

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

$$\|\xi - x_0\| \leq \left(\frac{16}{q} + 7\right)\|h\|_{l^1}. \tag{10.5}$$

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

$$\|\xi_M - x_0\| \leq \left(\frac{16}{q} + 7\right)\|P_M h\|_{l^1}. \tag{10.6}$$

Proof. Note that (10.4) and (i) yield

$$\begin{aligned}
& P_\Omega U(x_0 - P_\Delta \xi) = P_\Omega U(P_\Delta^\perp \xi - h) \\
\Rightarrow \quad & P_\Delta U^* P_\Omega U(x_0 - P_\Delta \xi) = P_\Delta U^* P_\Omega U(P_\Delta^\perp \xi - h) \\
& \qquad \Rightarrow \quad x_0 - P_\Delta \xi = (P_\Delta U^* P_\Omega U P_\Delta)^{-1} P_\Delta U^* P_\Omega U(P_\Delta^\perp \xi - h).
\end{aligned} \tag{10.7}$$

(note that (i) implies that $P_\Delta U^* P_\Omega U P_\Delta$ is invertible). Hence, from (i) and (10.7), and by using the fact that $\|U\| \leq 1$ we obtain

$$\|x_0 - P_\Delta \xi\| \leq 2/q \|P_\Delta^\perp \xi - h\|. \tag{10.8}$$

Thus,

$$\|x_0 - \xi\| \leq \frac{2}{q}\|P_\Delta^\perp \xi - h\| + \|P_\Delta^\perp \xi\| \leq \left(\frac{2}{q} + 1\right)\|P_\Delta^\perp \xi\|_{l^1} + \frac{2}{q}\|h\|_{l^1}. \tag{10.9}$$

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

Note that 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

$$\mathrm{Re}(\langle x_0, \rho\rangle) + \mathrm{Re}(\langle h, \rho\rangle) = \mathrm{Re}(\langle \xi, \rho\rangle) \leq \mathrm{Re}(\langle \xi, P_\Delta \rho\rangle) + \frac{1}{2}\sum_{j\in\Delta^c}|\xi(j)|. \tag{10.10}$$

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

$$\begin{aligned}
\mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho\rangle = \mathrm{Re}\,\langle x_0, \rho\rangle - \mathrm{Re}\,\langle \xi, P_\Delta \rho\rangle &\leq -\mathrm{Re}\,\langle h, \rho\rangle + \frac{1}{2}\|P_\Delta^\perp \xi\|_{l^1} \\
&= -\mathrm{Re}\,\langle h, P_\Delta^\perp \rho\rangle + \frac{1}{2}\|P_\Delta^\perp \xi\|_{l^1} \\
&\leq \frac{1}{2}\left(\|h\|_{l^1} + \|P_\Delta^\perp \xi\|_{l^1}\right).
\end{aligned} \tag{10.11}$$

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

$$\mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho\rangle \geq \mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho - \mathrm{sgn}(x_0)\rangle - \|h\|_{l^1} + \|P_\Delta^\perp \xi\|_{l^1}. \tag{10.12}$$

Moreover, using (ii) and (10.8), 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 (10.12) now gives

$$\begin{aligned}\mathrm{Re}\,\langle x_0 - \xi, P_\Delta \rho\rangle &\geq -\frac{1}{4}\|P_\Delta^\perp \xi - h\| - \|h\|_{l^1} + \|P_\Delta^\perp \xi\|_{l^1} \\ &\geq -\frac{5}{4}\|h\|_{l^1} + \frac{3}{4}\|P_\Delta^\perp \xi\|_{l^1}. \end{aligned} \tag{10.13}$$

Combining (10.11) and (10.13) and rearranging now gives

$$\|P_\Delta^\perp \xi\|_{l^1} \leq 7\|h\|_{l^1}.$$

Substituting this into (10.9) now yields (10.5). The proof of (10.6) is almost identical, and we omit the details. $\square$

11 Proofs of the main results

11.1 The Idea

Before we present proofs of Theorems 7.1–7.3, we would like to sketch the key ideas. Our approach is to use Proposition 10.4 to show the existence of some $\rho \in \mathrm{ran}(U^* P_\Omega)$ with the following properties

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

for some $\theta > 0$ (recall the setup in Theorems 7.1 and 7.2).

Throughout the paper we will be concerned with randomly choosing a set $\Omega \subset \{1, \ldots, N\}$. In our models we will choose Ω 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. We will typically take a sequence $\{\delta_1, \ldots \delta_N\}$ of independent identically distributed Bernoulli random variables taking values 0 and 1 with $\mathbb{P}(\delta_j = 1) = q$ for all j and let $\Omega = \{j : \delta_j = 1\}$. We will refer to this type of random selection of Ω as the Bernoulli model and we will denote such a procedure by $\{N, \ldots, 1\} \supset \Omega \sim \mathrm{Ber}(q)$.

We will assume that $\{N, \ldots, 1\} \supset \Omega \sim \mathrm{Ber}(\theta)$, for some finite $N \in \mathbb{N}$. However, we will construct Ω in an equivalent, but slightly different way. Namely, we let

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

where the specific value of μ will be determined later. Note that as long as the q_js are chosen according to θ this is equivalent to letting $\Omega \sim \mathrm{Ber}(\theta)$. Indeed, we have that $\Omega \sim \mathrm{Ber}(\theta)$ is equivalent to $\Omega^c \sim \mathrm{Ber}(1-\theta)$. 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

$$(1-q_1)(1-q_2)\cdots(1-q_\mu) = (1-\theta) \tag{11.1}$$

it is easy to see (by independence) that the two models are equivalent. Note that, obviously, there might be overlaps between the Ω_js. This automatically gives us the following:

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

This fact will be used several times in the arguments that follow and is a very crucial observation. We can now present the Golfing Scheme.

11.2 The Golfing Scheme

Let $U \in \mathcal{B}(\mathcal{H})$ be an isometry and let $\{N, \ldots, 1\} \supset \Omega_j \sim \mathrm{Ber}(q_j)$ for $j = 1, \ldots, \mu$ for some $\mu \in \mathbb{N}$ where the q_js satisfy (11.1) for some $0 < \theta \leq 1$. Suppose also that $x_0 \in \mathcal{H}$. Define the operator

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

The construction of ρ is based on the following idea. Let

$$\begin{aligned} \rho = Y_\mu, \quad Y_i = \sum_{j=1}^{i} q_j^{-1} E_{\Omega_j} Z_{j-1} \\ Z_i = \mathrm{sgn}(x_0) - P_\Delta Y_i, \quad Z_0 = \mathrm{sgn}(x_0), \end{aligned} \tag{11.2}$$

where the specific value of μ will be determined later. The construction suggested in (11.2) will be referred to as the golfing scheme, and is a variant of the extremely useful original golfing scheme introduced in [33] by David 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 theorems we need to establish some results that will be crucial in the construction of ρ.

11.3 The Proofs

We first require the following three results. Proofs are found in the Appendix:

Proposition 11.1. *Let $U \in \mathcal{B}(\mathcal{H}))$ be an isometry. Let $\{N, \ldots, 1\} \supset \Omega \sim \mathrm{Ber}(q)$ for some $0 < q \leq 1$, and $\Delta \subset \mathbb{N}$ with $|\Delta| < \infty$. Also, let $M \in \mathbb{N}$ be so large that $\Delta \subset \{1, \ldots, M\}$ and define $E_\Omega = U^* P_\Omega U$. Then, for $\eta \in \mathcal{H}$ and $t, \gamma > 0$*

$$\mathbb{P}\left(\|q^{-1} P_M P_\Delta^\perp E_\Omega P_\Delta \eta\|_{l^\infty} > (t + \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}})\|\eta\|\right) \leq \gamma \tag{11.3}$$

provided

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

Also,

$$\mathbb{P}\left(\|q^{-1} P_\Delta^\perp E_\Omega P_\Delta \eta\|_{l^\infty} > (t + \|P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}})\|\eta\|\right) \leq \gamma \tag{11.4}$$

whenever

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

where $\omega = \tilde{\omega}_{M,U}(|\Delta|, tq, N)$ (recall $\tilde{\omega}_{M,U}$ from (6.2)). In addition, if $q = 1$,the left-hand sides of (11.3) and (11.4) are equal to zero.

This proposition states that, for a sparse vector η supported in Δ, the vector $U^* P_\Omega U \eta$ cannot be too large outside the support of η.

Proposition 11.2. *Let $U \in \mathcal{B}(\mathcal{H}))$ be an isometry, $\Delta \subset \mathbb{N}$ with $|\Delta| < \infty$ and $\{N, \ldots, 1\} \supset \Omega \sim \mathrm{Ber}(q)$ 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 13.2).

Theorem 11.3. *There exists a constant $C > 0$ with the following property. Suppose that $U \in \mathcal{B}(\mathcal{H})$ is an isometry, Δ 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 that*

$$\mathbb{P}\left(\left\|\theta^{-1}P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\right\| \geq \frac{1}{\gamma} + \|P_\Delta U^* P_N U P_\Delta - P_\Delta\|\right) \leq \epsilon, \tag{11.5}$$

provided that

$$\begin{aligned} &\theta \geq C \cdot \gamma \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log(|\Delta|), \\ &\theta \geq C \cdot \gamma \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log(C\epsilon^{-1}) \cdot \left(\log\left(1 + \frac{1}{4\gamma}\right)\right)^{-1}. \end{aligned} \tag{11.6}$$

If $\theta = 1$ then the left hand side of (11.5) is equal to zero.

With these results presented, we can now embark on the task of proving the main theorems of this paper.

Proof **of Theorem 7.1 and Theorem 7.2**. The set $\Omega \subset \{1, \dots, N\}$ is chosen uniformly at random with $|\Omega| = m$. By Proposition 10.4 it suffices to show that there exists a $\rho \in \mathrm{ran}(U^* P_\Omega)$ such that

$$(i)\ \|\theta^{-1}P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\| \leq 1/2, \quad (ii)\ \|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq \theta/8, \quad (iii)\ \|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2, \tag{11.7}$$

with large probability. Note that we may (without loss of generality) replace this way of choosing Ω with the model that $\{N, \dots, 1\} \supset \Omega \sim \mathrm{Ber}(\theta)$ for $\theta = m/N$ (θ will have this value throughout the proof). Doing so may only change the constant C in (7.2). This trick has almost become standard in the literature and we will thus skip the specifics (see [16, 17] for details). Note that, as discussed in Section 11.1, the model $\{N, \dots, 1\} \supset \Omega \sim \mathrm{Ber}(\theta)$ is equivalent to choosing Ω as

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

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

$$(1 - q_1)(1 - q_2) \cdots (1 - q_\mu) = (1 - \theta). \tag{11.8}$$

The latter model is the one we will use throughout the proof and the specific value of μ will be chosen later. The theorems will follow if we can show that the conditions in (11.7) occur with probability exceeding $1-\epsilon$, and what follows is a setup to ensure this eventually. We will focus on (ii) and (iii) in (11.7) and deal with (i) at the end of the proof. The proof proceeds in a number of steps.

Step I (The construction of ρ): Let ν 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 μ, ν, $\{\alpha_i\}_{i=1}^\mu$ and $\{\beta_i\}_{i=1}^\mu$ will be carefully chosen later in the proof. Consider now the following construction of ρ : 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 define

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

where $E_{\Omega_j} = U^* P_{\Omega_j} U$, and $\{q_1, \dots, q_\mu\}$ stem from (11.8). The precise values of the q_j's will be chosen later. Let also $\Theta_1 = \{1\}$ and $\Theta_2 = \{1, 2\}$. Then define recursively, for $i \geq 3$, the following:

$$\Theta_i = \begin{cases} \Theta_{i-1} \cup \{i\} & \text{if } \left\|\left(P_\Delta - q_i^{-1} P_\Delta E_{\Omega_i} P_\Delta\right) Z_{i-1}\right\| \leq \alpha_i \|Z_{i-1}\|, \\ & \text{and } \left\|q_i^{-1} P_M P_\Delta^\perp E_{\Omega_i} P_\Delta Z_{i-1}\right\|_{l^\infty} \leq \beta_i \|Z_{i-1}\|, \\ \Theta_{i-1} & \text{otherwise,} \end{cases} \tag{11.9}$$

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

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

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

$$
\begin{aligned}
&A_i: &&\left\|\left(P_\Delta - q_i^{-1} P_\Delta E_{\Omega_i} P_\Delta\right) Z_{i-1}\right\| \leq \alpha_i \left\|Z_{i-1}\right\|, \qquad i = 1,2,\\
&B_i: &&\left\|q_i^{-1} P_M P_\Delta^\perp E_{\Omega_i} P_\Delta Z_{i-1}\right\|_{l^\infty} \leq \beta_i \|Z_{i-1}\|, \qquad i = 1,2,\\
&B_3: &&|\Theta_\mu| \geq \nu,\\
&B_4: &&(\cap_{i=1}^2 A_i) \cap (\cap_{i=1}^3 B_i),
\end{aligned}
\tag{11.10}
$$

where $|\Theta_\mu|$ denotes the length of Θ_μ.

Also, let $\tau(j)$ denote the j^{th} element in Θ_μ (e.g. $\tau(1) = 1, \tau(2) = 2$ etc. We also let $\tau(0) = 0$.) and finally define ρ by

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

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

$$
\begin{aligned}
Z_{\tau(i)} &= \mathrm{sgn}(x_0) - P_\Delta \left(q_{\tau(1)}^{-1} E_{\Omega_{\tau(1)}} \mathrm{sgn}(x_0) + q_{\tau(2)}^{-1} E_{\Omega_{\tau(2)}} Z_1 + \ldots + q_{\tau(i)}^{-1} E_{\Omega_{\tau(i)}} Z_{\tau(i-1)} \right)\\
&= Z_{\tau(i-1)} - q_{\tau(i)}^{-1} P_\Delta E_{\Omega_{\tau(i)}} P_\Delta Z_{\tau(i-1)} = (P_\Delta - q_{\tau(i)}^{-1} P_\Delta E_{\Omega_{\tau(i)}} P_\Delta) Z_{\tau(i-1)}.
\end{aligned}
\tag{11.11}
$$

Hence, if the event B_4 occurs, we have

$$
\|P_\Delta \rho - \mathrm{sgn}(x_0)\| = \|Z_{\tau(\nu)}\| \leq \sqrt{|\Delta|} \prod_{i=1}^{\nu} \alpha_{\tau(i)}, \tag{11.12}
$$

$$
\begin{aligned}
\|P_M P_\Delta^\perp \rho\|_{l^\infty} &\leq \sum_{i=1}^{\nu} \|q_{\tau(i)}^{-1} P_M P_\Delta^\perp E_{\Omega_{\tau(i)}} Z_{\tau(i-1)}\|_{l^\infty}\\
&\leq \sum_{i=1}^{\nu} \beta_{\tau(i)} \|Z_{\tau(i-1)}\| \leq \sqrt{|\Delta|} \sum_{i=1}^{\nu} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)},
\end{aligned}
\tag{11.13}
$$

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 ν, $\{\beta_j\}_{j=1}^\mu$ and $\{\alpha_j\}_{j=1}^\mu$ then (ii) and (iii) in (11.7) are satisfied when the event B_4 occurs. We delay specifying a the value for μ until Step IV. Let $L \geq 2$, (we will give a value for L in a moment) and

$$
\alpha_1 = \alpha_2 = \frac{1}{2 \log_2^{1/2}(L)}, \qquad \alpha_i = 1/2, \quad 3 \leq i \leq \mu,
$$

$$
\beta_1 = \beta_2 = \frac{1}{4\sqrt{|\Delta|}}, \qquad \beta_i = \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{4\sqrt{|\Delta|}}, \quad 3 \leq i \leq \mu.
$$

It follows that

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

Hence, if

$$
\nu = \left\lceil \log_2 \left(8\theta^{-1} \sqrt{|\Delta|} \right) \right\rceil, \tag{11.14}
$$

then it follows by (11.12) that

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

(recall that $L \geq 2$) yielding (ii) in (11.7). Also, after inserting the values of ν, $\{\beta_j\}_{j=1}^\mu$ and $\{\alpha_j\}_{j=1}^\mu$ into (11.13) we get:

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

if we let $L = 4\theta^{-1}\sqrt{|\Delta|}$. Thus, by (11.13) we have

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

yielding (iii) in (11.7). In particular, we have showed that, if ν, $\{\beta_j\}_{j=1}^{\mu}$ 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 framework for proving (ii) and (iii) in (11.7) with a certain probability. To do this, we will make a careful choice of μ 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

$$\|P_\Delta U^* P_N U P_\Delta - P_\Delta\| \leq \frac{1}{4\log_2^{1/2}(4\theta^{-1}\sqrt{|\Delta|})}, \tag{11.15}$$

and

$$q_1 = q_2 \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1}\right) + 1\right) \cdot \log\left(\theta^{-1}\sqrt{|\Delta|}\right), \tag{11.16}$$

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 6.2 these requirements are well defined. Now note that Proposition 11.2 gives, for $i = 1, 2$ and $0 < t, \gamma \leq 1$ that

$$\mathbb{P}\left(\left\|\left(q_i^{-1} P_\Delta U^* P_{\Omega_i} U P_\Delta - P_\Delta\right) Z_{i-1}\right\| > \left(t + \|P_\Delta U^* P_N U P_\Delta - P_\Delta\|\right)\|Z_{i-1}\|\right) \leq \gamma, \tag{11.17}$$

if

$$q_i(1-q_i)^{-1} \geq 4t^{-2} \cdot \upsilon^2(U) \cdot |\Delta|, \tag{11.18}$$

and

$$\log\left(1 + \frac{t}{4}\right) \geq \frac{2K}{t}\max\{q^{-1} - 1, 1\} \cdot \upsilon^2(U) \cdot |\Delta| \cdot \log\left(\frac{3}{\gamma}\right), \tag{11.19}$$

where K is the constant in Talagrand's Theorem (Theorem 13.2). Thus, by (11.17), (11.18) and (11.19) (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 (11.17) is zero, and hence the last assertion follows.

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

$$\|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}, \tag{11.20}$$

and

$$q_1 = q_2 \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1} M\right) + 1\right), \tag{11.21}$$

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 11.1 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

$$\mathbb{P}(B_1^c \cup B_2^c) = 0, \qquad q_1 = q_2 = 1,$$

follows automatically from Proposition 11.1.

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

$$\mu = 8\lceil 3\nu + \log(\gamma^{-1})\rceil, \tag{11.22}$$

$$\|P_\Delta U^* P_N U P_\Delta - P_\Delta\| \leq 1/4, \tag{11.23}$$

and

$$\|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{8\sqrt{|\Delta|}}, \tag{11.24}$$

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

$$q \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\frac{\log{(M)}}{\log_2(4\theta^{-1}\sqrt{|\Delta|})} + 1 \right), \tag{11.25}$$

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 (11.22) on μ. 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

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

However, the random variables $X_1, \ldots X_{\mu-2}$ are not independent. Thus, to use standard tools such as Chernoff's inequality we must apply a couple of tricks. Observe that

$$\begin{aligned} &\mathbb{P}(X_1 + \ldots + X_{\mu-2} > \mu - \nu) \\ &\leq \sum_{l=1}^{\binom{\mu-2}{\mu-\nu}} \mathbb{P}(X_{\pi(l)_1} = 1, X_{\pi(l)_2} = 1, \ldots + X_{\pi(l)_{\mu-\nu}} = 1) \\ &= \sum_{l=1}^{\binom{\mu-2}{\mu-\nu}} \mathbb{P}(X_{\pi(l)_{\mu-\nu}} = 1 \,|\, X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-1}} = 1) \mathbb{P}(X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-1}} = 1) \\ &= \sum_{l=1}^{\binom{\mu-2}{\mu-\nu}} \mathbb{P}(X_{\pi(l)_{\mu-\nu}} = 1 \,|\, X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-1}} = 1) \\ &\qquad \times \mathbb{P}(X_{\pi(l)_{\mu-\nu-1}} = 1 \,|\, X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-2}} = 1) \cdots \mathbb{P}(X_{\pi(l)_1} = 1) \end{aligned} \tag{11.26}$$

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

$$\begin{aligned} &P \geq \mathbb{P}(X_{\pi(l)_{\mu-\nu-j}} = 1 \,|\, X_{\pi(l)_1} = 1, \ldots, X_{\pi(l)_{\mu-\nu-(j+1)}} = 1), \\ &P \geq \mathbb{P}(X_{\pi(l)_1} = 1), \end{aligned} \tag{11.27}$$

$$l = 1, \ldots, \binom{\mu-2}{\mu-\nu}, \quad j = 0, \ldots, \mu-\nu-2,$$

then, by (11.26),

$$\mathbb{P}(X_1 + \ldots + X_{\mu-2} > \mu - \nu) \leq \binom{\mu-2}{\mu-\nu} P^{\mu-\nu}. \tag{11.28}$$

Now let $\widetilde{X}_1, \ldots, \widetilde{X}_{\mu-2}$ be independent binary random variables, with $\mathbb{P}(\widetilde{X}_k = 1) = P$ and $\mathbb{P}(\widetilde{X}_k = 0) = 1 - P$ for each k. From Lemma 13.3 and (11.28), we have that

$$\mathbb{P}\left(|\Theta_\mu| < \nu\right) = \mathbb{P}\left(\sum_{i=1}^{\mu-2} X_i \geq \mu - \nu\right) \leq \left(\frac{(\mu-2) \cdot e}{\mu - \nu}\right)^{(\mu-\nu)} \mathbb{P}\left(\sum_{i=1}^{\mu-2} \widetilde{X}_i \geq \mu - \nu\right). \tag{11.29}$$

Note that by the standard Chernoff bound ([43, Theorem 2.1]) it follows that, for $t > 0$,

$$\mathbb{P}\left(\widetilde{X}_1 + \ldots + \widetilde{X}_{\mu-2} \geq (\mu-2)(t+P)\right) \leq e^{-2(\mu-2)t^2}. \tag{11.30}$$

If we let $t = (\mu-\nu)/(\mu-2) - P$, then (11.30) and (11.29) gives that

$$\mathbb{P}\left(\sum_{i=1}^{\mu-2} X_i \geq \mu-\nu\right) \leq \gamma \tag{11.31}$$

whenever

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

Hence, by observing that $\log((\mu-2)/(\mu-\nu))+1 \leq (\mu-2)/(\mu-\nu)$, we have that (11.31) is satisfied whenever

$$\mu \geq x, \qquad (x-2)\left(\frac{x-\nu}{x-2} - P\right)^2 - \log\left(\gamma^{-1/2}\right) - \frac{x-2}{2} = 0 \tag{11.32}$$

where x is the largest root satisfying (11.32). In particular, we have shown that $\mathbb{P}(B_3^c) \leq \gamma$ when (11.32) is satisfied. Choosing $P = 1/4$ will yield $x \leq 8\lceil 3\nu + \log(\gamma^{-1/2})\rceil$. Hence (11.22) yields (11.32).

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 (11.27) is satisfied with $P = 1/4$. Note that $X_k = 1$ if and only if one of the following events occur:

$$\begin{aligned} D_1: &\quad \left\|\left(P_\Delta - q_j^{-1} P_\Delta E_{\Omega_j} P_\Delta\right) Z_{j-1}\right\| > \alpha_j \left\|Z_{j-1}\right\|, \qquad j = k+2, \\ D_2: &\quad \left\|q_j^{-1} P_M P_\Delta^\perp E_{\Omega_j} P_\Delta Z_{j-1}\right\|_{l^\infty} > \beta_j \|Z_{j-1}\|, \qquad j = k+2. \end{aligned} \tag{11.33}$$

Observe that we may argue exactly as in the proof of Step II (via Proposition 11.2) and regardless of the vector Z_{j-1}, we may deduce that $\mathbb{P}(D_1) \leq 1/8$ when N and q_j, are chosen such that

$$\begin{aligned} \|P_\Delta U^* P_N U P_\Delta - P_\Delta\| &\leq \alpha_j/2, \\ q_j \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \alpha_j^{-2} &\cdot (\log(24) + 1), \quad j = k+2, \end{aligned} \tag{11.34}$$

for some universal constant $C > 0$. Observe also that we may argue exactly as in the proof of Step III (via Proposition 11.1) 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

$$\begin{aligned} \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} &\leq \beta_j/2, \\ q_j \geq C \cdot \upsilon^2(U) \cdot \left(\frac{1}{\beta_j^2} + \frac{1}{\beta_j}\sqrt{|\Delta|}\right) &\cdot (\log(32M) + 1), \quad j = k+2, \end{aligned} \tag{11.35}$$

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 $k = \pi(l)_{\mu-\nu-i}$,

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

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

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

whenever (11.34) and (11.35) are satisfied. In particular, (11.34) and (11.35) imply (11.27). But (11.34) and (11.35) follow from (11.23), (11.24) and (11.25) (with a possibly different, however universal, C) and thus, 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 11.2 and 11.1.

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

$$\mathbb{P}\left(\|P_\Delta \rho - \mathrm{sgn}(x_0)\| > \theta/8 \cup \|P_M P_\Delta^\perp \rho\|_{l^\infty} > 1/2\right) \leq 5\gamma, \tag{11.36}$$

when $N \in \mathbb{N}$ and $\theta > 0$ are chosen according to (6.3), (6.4) and

$$\theta \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1}\right) + 1\right) \cdot \log\left(M\theta^{-1}\sqrt{|\Delta|}\right), \tag{11.37}$$

for some universal constant $C > 0$. Also, if $\theta = 1$ then the left hand side of (11.36) 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 $\|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq \theta/8$ and $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2$. It therefore suffices to show that, given the conditions (6.3), (6.4) and (11.37), it holds that

$$\mathbb{P}\left(B_4^c\right) \leq 5\gamma. \tag{11.38}$$

To do this we begin by making some observations. First

$$\mathbb{P}\left(B_4^c\right) \leq \mathbb{P}(A_1^c \cup A_2^c) + \mathbb{P}(B_1^c \cup B_2^c) + \mathbb{P}(B_3^c), \tag{11.39}$$

and second

$$q_1 + q_2 + \ldots + q_\mu \geq \theta. \tag{11.40}$$

Recall from Step II we have that $\mathbb{P}(A_1^c \cup A_2^c) \leq 2\gamma$ whenever (11.15) and (11.16) are satisfied. Also, by Step III, $\mathbb{P}(B_1^c \cup B_2^c) \leq 2\gamma$ whenever (11.20) and (11.21) are fulfilled. Finally, from Step IV we have that $\mathbb{P}(B_3^c) \leq \gamma$ provided

$$\mu = 8\left\lceil \log(\gamma^{-1}) + 3\left\lceil \log_2\left(8\theta^{-1}\sqrt{|\Delta|}\right)\right\rceil\right\rceil, \tag{11.41}$$

and (11.23), (11.24) and (11.25) are satisfied. In particular, using (11.39) we find that (11.38) follows from (11.15), (11.16), (11.20), (11.21), (11.23), (11.24) and (11.25). Thus, we must then show that these equations follow from (6.3), (6.4) and (11.37). Now let $q_1 = q_2 = \theta/4$. Then, by (11.37), we have that (11.16) follows (with a possibly different constant), and similarly (11.21) follows. Let $q = q_3 = \ldots = q_\mu$. By (11.40) and (11.41) 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 (11.25) follows. The only thing left to do is to deal with the requirements on N. In particular, we need to show that (11.15), (11.20), (11.23) and (11.24) follow when (6.3) and (6.4) are satisfied. Note that (11.23) and (11.24) are weaker than (11.15) and (11.20). Thus, we only need to concentrate on (11.15) and (11.20). To see that (6.3) and (6.4) imply (11.15) and (11.20), note that (since $P_M \geq P_\Delta$)

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

and so

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

Hence (11.15) follows from (6.3). The fact that (11.20) follows from (6.4) is clear. Also, the fact that the left-hand side of (11.36) is equal to zero when $\theta = 1$ follows from Steps II - IV and the fact that when $\theta = 1$ we have $q_1 = \ldots = q_\mu = 1$.

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

$$\mathbb{P}(\|\theta^{-1}P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\| > 1/2) \leq \gamma, \tag{11.42}$$

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

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

for some universal constant C. Also, if $\theta = 1$ then the left hand side of (11.42) is equal to zero.

To prove this claim note that, by Theorem 11.3, 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 (11.42) 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 (11.7) follows with probability exceeding $1-\epsilon$. Note that in Step V we have shown that if $\gamma > 0$, then (ii) and (iii) in (11.7) are satisfied with probability exceeding $1 - 5\gamma$, provided (6.3), (6.3) and (11.37) 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 (6.4) and (11.37), and hence (i) in (11.7) follows with probability exceeding $1 - \gamma$. This implies that (i), (ii) and (iii) in (11.7) hold with probability exceeding $1 - 6\gamma$. By choosing γ such that $6\gamma = \epsilon$ we observe that (11.37) follows (with possibly a different C) from the conditions in Theorems 7.1 and 7.2 and we have finally proved the first assertions in Theorem 7.1 and Theorem 7.2. The last assertions follow by the fact that $\theta = 1$ when $m = N$, (and hence also $q_1 = \ldots = q_\mu = 1$) and Step V - VI. □

Proof **of Theorem 7.3**. We will follow the recipe from the of proof of Theorem 7.2 almost word for word, and we will only point out where the differences lie. The first such difference is the set of conditions provided by Proposition 10.4. In particular we must show that there exists a $\rho \in \mathrm{ran}(U^* P_\Omega)$ such that

$$(i)\ \|\theta^{-1} P_\Delta U^* P_\Omega U P_\Delta - P_\Delta\| \leq 1/2, \quad (ii)\ \|P_\Delta \rho - \mathrm{sgn}(x_0)\| \leq \theta/8 \quad (iii)\ \|P_\Delta^\perp \rho\|_{l^\infty} \leq 1/2, \quad (11.43)$$

is true with probability exceeding $1 - \epsilon$. (Note that only condition (iii) is changed from the proof of Theorem 7.2).

Step I: Almost as in the proof of Theorem 7.2, except that (11.9) should read

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

and the events B_1 and B_2 in (11.10) should be

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

Also, (11.13) 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 7.1.

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

$$\|P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{1}{8\sqrt{|\Delta|}}, \tag{11.44}$$

and

$$q_1 = q_2 \geq C \cdot \upsilon^2(U) \cdot |\Delta| \cdot \left(\log\left(\gamma^{-1}\omega_1\right) + 1\right), \tag{11.45}$$

where

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

(recall $\tilde{\omega}_{M,U}$ from (6.2)) 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 7.1 by using the last part of Proposition 11.1.

Step IV: We claim that, for $\gamma > 0$, then $\mathbb{P}(B_3^c) \leq \gamma$, if μ, (recall μ and ν from Step I) N and $\{q_3, \ldots, q_\mu\}$ are chosen according to (11.22), (11.23) and

$$\|P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}} \leq \frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{8\sqrt{|\Delta|}}, \tag{11.46}$$

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

$$q \geq C \cdot v^2(U) \cdot |\Delta| \cdot \left(\frac{\log(\omega_2)}{\log_2(4\theta^{-1}\sqrt{|\Delta|})} + 1 \right), \tag{11.47}$$

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

The proof is almost as in the proof of Theorem 7.2, except that the last part of (11.33) 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 (11.35) should be

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

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

$$\mathbb{P}\left(\|P_\Delta \rho - \mathrm{sgn}(x_0)\| > \theta/8 \cup \|P_\Delta^\perp \rho\|_{l^\infty} > 1/2 \right) \leq 5\gamma, \tag{11.48}$$

when $N \in \mathbb{N}$ and $\theta > 0$ are chosen according to (6.3), (6.5) and

$$\theta \geq C \cdot v^2(U) \cdot |\Delta| \cdot \left(\log(\gamma^{-1}) + 1 \right) \cdot \log\left(\omega\theta^{-1}\sqrt{|\Delta|} \right), \tag{11.49}$$

where

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

and $\tilde{\omega}_{M,U}$ is defined in (6.2), for some universal constant $C > 0$. Also, if $\theta = 1$ then the left hand side of (11.48) is equal to zero.

The strategy is almost as in the proof of Step V in Theorem 7.1. In particular, we argue by using Step II - IV that $\mathbb{P}(B_4^c) \leq 5\gamma$ when (11.15), (11.16), (11.44), (11.45), (11.23), (11.46) and (11.47) are satisfied, and thus (11.48) follows. We then need to show that these equations follow from (6.3), (6.5) and (11.49). To do this, let $q_1 = q_2 = \theta/4$. Then, by (11.49), we have that (11.16) follows (with a possibly different constant). To show that (11.45) is implied by (11.49) it suffices to show that $\omega \geq \omega_1$. This will follow by the definition (6.2) 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

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

To show that (11.47) follows from (11.49) it suffices to show that $\omega \geq \omega_2$. To do this (as argued above) it is sufficient to prove that

$$q\frac{\log_2(4\theta^{-1}\sqrt{|\Delta|})}{8\sqrt{|\Delta|}} \geq s. \tag{11.50}$$

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 μ (from (11.22)) from Step IV and noting 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

$$q\log_2(4\theta^{-1}\sqrt{|\Delta|}) \geq \frac{\theta\log_2(4\theta^{-1}\sqrt{|\Delta|})}{16(\log(\gamma^{-1}) + 3\log_2(8\theta^{-1}\sqrt{|\Delta|}) + 1)} > \frac{\theta}{8\log(e^4\gamma^{-1})}. \tag{11.51}$$

Thus, we have shown (11.50).

We are now left with the task of showing that (11.15), (11.44), (11.45), (11.23) and (11.46) follow from (6.3) and (6.5), and this follows by arguing exactly as in the proof of Step V in the proof of Theorem 7.1

Step VI and **Step VII**: Exactly as in the proof of Theorem 7.1. □

12 Acknowledgments

The authors would like to thank Akram Aldroubi, Emmanuel Candès, Ron DeVore, David Donoho, Karlheinz Gröchenig, Holger Rauhut, Jared Tanner, 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.

13 Appendix

The appendix contains all the proofs that have not been displayed so far. However, before do this, there are two results that are absolutely crucial. The first is a due to Rudelson [49].

Lemma 13.1. *(Rudelson) Let $\eta_1, \ldots, \eta_M \in \mathbb{C}^n$ and let $\varepsilon_1, \ldots \varepsilon_M$ 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 [49] does not apply in this case. Actually, we need the complex version proved in [56]. We will, however, still refer to it as Rudelson's Lemma. The following theorem is also indispensable (note that we deliberately forgo the use of any vector/matrix Bernstein inequalities and prefer Talagrand's result instead. This allows for more flexibility in the infinite-dimensional setting):

Theorem 13.2. *(Talagrand [54, 41]) There exists a number K with the following property. Consider n independent random variables X_i valued in a measurable space Ω. Let $\mathcal{F}$ be a (countable) class of measurable functions on Ω and consider the random variable $Z = \sup_{f \in \mathcal{F}} \sum_{i \leq n} f(X_i)$. Let*

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

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

We also require the following lemma:

Lemma 13.3. *Let $\{\tilde{X}_j\}_{j=1}^N$ be independent binary variables taking values 0 and 1, such that $\tilde{X}_j = 1$ with probability P. Then,*

$$\mathbb{P}\left(\sum_{i=1}^{N} \tilde{X}_i \geq k\right) \geq \left(\frac{N \cdot e}{k}\right)^{-k} \binom{N}{k} P^k. \tag{13.1}$$

Proof. First observe that

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{N} \tilde{X}_i \geq k\right) &= \sum_{i=k}^{N} \binom{N}{i} P^i (1-P)^{N-i} = \sum_{i=0}^{N-k} \binom{N}{i+k} P^{i+k} (1-P)^{N-k-i} \\
&= \binom{N}{k} P^k \sum_{i=0}^{N-k} \frac{(N-k)!k!}{(N-i-k)!(i+k)!} P^i (1-P)^{N-k-i} \\
&= \binom{N}{k} P^k \sum_{i=0}^{N-k} \binom{N-k}{i} P^i (1-P)^{N-k-i} \left[\binom{i+k}{k}\right]^{-1}.
\end{aligned}$$

The result now follows because $\sum_{i=0}^{N-k} \binom{N-k}{i} P^i(1-P)^{N-k-i} = 1$ and for $i = 0, \dots, N-k$, we have that

$$\binom{i+k}{k} \leq \left(\frac{(i+k)\cdot e}{k}\right)^k \leq \left(\frac{N\cdot e}{k}\right)^k,$$

where the first inequality follows from Stirling's approximation (see [21], p. 1186). $\square$

Proof **of Proposition 9.5**. By the assumptions, 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)

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

Thus, by using duality (recall Proposition 9.3), 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

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

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

Proof **of Proposition 10.2**. Let $\alpha = |\Delta|$ and also $\omega = \{\omega_j\}_{j=1}^\alpha$ be a sequence, where $\omega_j \in \mathbb{C}$. Now define

$$V_\omega = I_{\Delta^c} \oplus S_\omega : P_\Delta^\perp \mathcal{H} \oplus P_\Delta\mathcal{H} \to P_\Delta^\perp\mathcal{H} \oplus P_\Delta\mathcal{H}, \tag{13.2}$$

where $S_\omega = \mathrm{diag}(\{\omega_j\}_{j=1}^\alpha)$ on $P_\Delta\mathcal{H}$ and I_{Δ^c} is the identity on $P_\Delta^\perp\mathcal{H}$. Define $U(\omega) = UV_\omega$. Note that to prove our claim 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 ω, where

$$\omega \in \Lambda = \{(e^{i\theta_1}, \dots, e^{i\theta_\alpha}) \in \mathbb{C}^\alpha : \theta_j \in [0, 2\pi), 1 \leq j \leq \alpha\}. \tag{13.3}$$

Indeed, if that is the case then, by Proposition 9.5, for every $\omega \in \Lambda$ there exists $\zeta_\omega \in P_\Omega\mathcal{H}$ such that

$$\pi_\omega = U^* P_\Omega \zeta_\omega, \quad P_\Delta \pi_\omega = \mathrm{sgn}(V_\omega x), \quad \|P_{\Delta^c}\pi_\omega\|_{l^\infty} < 1. \tag{13.4}$$

Thus, for any $y \in \mathcal{H}$ such that $\mathrm{supp}(y) = \Delta$ choose $\omega \in \Lambda$ such that $\mathrm{sgn}(y) = \mathrm{sgn}(V_\omega x)$. Then, since $(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta \restriction_{P_\Delta\mathcal{H}}$ is invertible it follows by 13.4 and Proposition 9.5 that y is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U\eta = P_\Omega U y\}$. Note also that if $\omega \in \Lambda$ then V_ω 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 ζ 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 for that for all $\omega \in \Lambda$. To do that, it suffices, by Proposition 9.5 and the fact that $U(\omega)$ is unitary, to show that there exists a vector $\rho \in \mathcal{H}$ such that

$$P_{\Omega^c} U(\omega)\rho = 0, \quad P_\Delta \rho = \mathrm{sgn}(x), \quad \|P_{\Delta^c}\rho\|_{l^\infty} < 1. \tag{13.5}$$

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

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

where $\iota_\Delta : P_\Delta\mathcal{H} \to \mathcal{H}$ is the inclusion operator. Then (13.5) is satisfied if and only if $\varphi(\omega) < 1$. Thus, to show 13.5 we must show that $\varphi(\omega) < 1$ for all $\omega \in \Lambda$.

Suppose for the moment that ϵ is chosen such that φ is defined on its domain. We will show that φ is continuous. It suffices to show that φ is continuous on $\mathcal{B}(a,\epsilon)$ for $a \in \Lambda$. Note that, by the fact that $\mathcal{B}(a,\epsilon)$ is open it suffices to show that φ is convex. To see that φ is convex, let $\omega_1, \omega_2 \in \mathcal{B}(a,\epsilon)$ and $t \in (0,1)$. Let also $\xi, \eta \in \mathcal{H}$ such that

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

Note that the existence of such vectors is guaranteed by the assumption that φ is defined on its domain. Now, observe that

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

Thus, taking infimum on the right hand side yields $\varphi(t\omega_1 + (1-t)\omega_2) \leq t\varphi(\omega_1) + (1-t)\varphi(\omega_2)$, and we have shown the assertion that φ is convex. Returning to the question on the domain of φ, note that if $(P_\Omega UP_\Delta)^* P_\Omega UP_\Delta \restriction_{P_\Delta\mathcal{H}}$ is invertible, then

$$P_\Omega U(\omega)P_\Delta)^* P_\Omega U(\omega)P_\Delta \restriction_{P_\Delta\mathcal{H}}$$

is invertible if $\|U(\tilde{\omega}) - U(\omega)\|$ is small and $\tilde{\omega} \in \Lambda$. Letting

$$\rho = U(\omega)^* P_\Omega U(\omega)P_\Delta((P_\Omega U(\omega)P_\Delta)^* P_\Omega U(\omega)P_\Delta \restriction_{P_\Delta\mathcal{H}})^{-1}\mathrm{sgn}(x)$$

we get

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

Thus, φ is defined on its domain for small ϵ.

Let Γ 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 Γ 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 $\xi \in \mathcal{H}$ satisfying

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

we have

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

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

Note also that Γ is open. Indeed, for if $\tilde{\omega} \in \Gamma$ then there exist $\rho \in \mathcal{H}$ such that ρ satisfies 13.5 (with ω replaced by $\tilde{\omega}$) e.g. $\varphi(\tilde{\omega}) < 1$. But, by continuity of φ it follows that φ is strictly less than one on a neighborhood of $\tilde{\omega}$. Since $(P_\Omega UP_\Delta)^* P_\Omega UP_\Delta \restriction_{P_\Delta\mathcal{H}}$ is invertible, then it is easy to see that $P_\Omega U(\omega)P_\Delta)^* P_\Omega U(\omega)P_\Delta \restriction_{P_\Delta\mathcal{H}}$ is invertible, for all $\omega \in \Lambda$ thus it follows by Proposition 9.5 that (13.5) is satisfied for all $\omega \in \Lambda$ in a neighborhood of $\tilde{\omega}$ and hence Γ is open.

The fact that Γ 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 10.3**. Let V_ω and Λ be defined as in (13.2) and (13.3) respectively. Suppose that $y \in \mathcal{H}$ such that $\mathrm{supp}(y) = \Delta$. Then, by assumption, $V_\omega y$ is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U\eta = P_\Omega UV_\omega y\}$. Thus, by Proposition 9.5 it follows that there exists a $\rho_\omega \in \mathcal{H}$ such that

$$P_{\Omega^c}U\rho_\omega = 0, \quad P_\Delta\rho_\omega = \mathrm{sgn}(V_\omega y), \quad \|P_{\Delta^c}\rho_\omega\|_{l^\infty} < 1. \tag{13.6}$$

Let $\beta = \sup_{\omega\in\Lambda}\{\|P_{\Delta^c}\rho_\omega\|_{l^\infty}\}$. Note that $\beta < 1$, since Λ is closed. Thus, for every $y \in \mathcal{H}$ with $\mathrm{supp}(y) = \Delta$ there exists $\rho_\omega \in \mathcal{H}$ satisfying (13.6) 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 [18]) there exists a constant $C > 0$ (depending on β) such that, if $\xi \in \mathcal{H}$, $\mathrm{supp}(\xi) = \Delta$, is the unique minimizer of $\inf\{\|\eta\|_{l_1} : P_\Omega U\eta = P_\Omega U\xi\}$, $\zeta \in \mathcal{H}$ 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

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

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

Proof **of Proposition 11.1**. 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, clearly (by using the fact that U is an isometry), we have

$$\begin{aligned} q^{-1}P_M P_\Delta^\perp E_\Omega P_\Delta\eta &= q^{-1}\sum_{j=1}^N P_M P_\Delta^\perp U^*\delta_j(e_j \otimes e_j)UP_\Delta\eta \\ &= q^{-1}\sum_{j=1}^N P_M P_\Delta^\perp U^*(\delta_j - q)(e_j \otimes e_j)UP_\Delta\eta + P_M P_\Delta^\perp U^* P_N^\perp UP_\Delta\eta. \end{aligned} \tag{13.7}$$

Our goal is to eventually use Bernstein's inequality and the following is therefore a setup for that. Define, for $1 \leq j \leq N$ the random variables

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

Thus, by (13.7) it follows that for $s > 0$ we have

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

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 $s = t + \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}}$ it follows that

$$\mathbb{P}\left(\left\|q^{-1} P_M P_\Delta^\perp E_\Omega P_\Delta \eta\right\|_{l^\infty} > t + \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{\mathrm{mr}}\right) \leq \sum_{i \in \Delta^c \cap \{1,\dots,M\}} \mathbb{P}\left(\left|\sum_{j=1}^N X_j^i\right| > t\right). \quad (13.8)$$

To get a bound on the right hand side of (13.8) we will be using Bernstein's inequality, and in order to do that we need a couple of observations. First note that

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

Thus

$$\sum_{j=1}^N \mathbb{E}\left(|X_j^i|^2\right) \leq q^{-1}(1-q)\|\eta\|^2 \upsilon^2(U) = q^{-1}(1-q)\upsilon^2(U), \qquad i \in \Delta^c \cap \{1, \dots, M\}. \quad (13.9)$$

Also, observe that

$$|X_j^i| = q^{-1}|(\delta_j - q)||\langle \eta, P_\Delta U^*(e_j \otimes e_j) U e_i \rangle| \leq \max\{(1-q)/q, 1\}\upsilon^2(U)\sqrt{|\Delta|}, \quad (13.10)$$

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

$$\mathbb{P}\left(\left|\sum_{j=1}^N X_j^i\right| > t\right) \leq 4 \exp\left(-\frac{t^2/4}{q^{-1}(1-q)\upsilon^2(U) + \max\{q^{-1}(1-q), 1\}\upsilon^2(U)\sqrt{|\Delta|}t/3\sqrt{2}}\right), \quad (13.11)$$

for all $i \in \Delta^c \cap \{1, \dots, M\}$. Thus, by invoking (13.11) and (13.8) 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

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

and the first part of the proposition follows. The fact that the left hand side of (11.3) when $q = 1$ is clear from (13.9) and (13.10).

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 $P_M P_\Delta^\perp E_\Omega P_\Delta \eta$) actually 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

$$q^{-1} P_\Delta^\perp E_\Omega P_\Delta \eta = \sum_{j=1}^{N} \widetilde{Y}_j + P_\Delta^\perp U^* P_N^\perp U P_\Delta \eta, \qquad \widetilde{Y}_j = q^{-1} P_\Delta^\perp U^*(\delta_j - q)(e_j \otimes e_j) U P_\Delta \eta. \tag{13.12}$$

Define (as we did above) the random variables

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

Note that we now have infinitely many X_j^is, however, suppose for a moment that for every $t > 0$ there exists a non-empty set $\Lambda_t \subset \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{i \in \Lambda_t} \left| \sum_{j=1}^{N} X_j^i \right| > t \right) = 0 \qquad |\Delta^c \setminus \Lambda_t| < \infty. \tag{13.13}$$

Then, if that was the case, we would immediately get (by arguing as in Step I and using (13.12) 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 (13.11), 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

$$q \geq \left(\frac{4}{t^2} + \frac{2\sqrt{2}}{3t} \sqrt{|\Delta|} \right) \log\left(\frac{4}{\gamma} |\Delta^c \setminus \Lambda_t| \right) \upsilon^2(U). \tag{13.14}$$

Hence, if we could show the existence of Λ_t and provide a bound on $|\Delta^c \setminus \Lambda_t|$ we could appeal to (13.12) and (13.14) and be done. 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 automatically get that $|\Delta^c \setminus \Lambda_t| < \infty$. Note also that (13.13) 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-Schwartz inequality. With the existence of Λ_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 Δ 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

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

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 $\Gamma_1 \subset \{1, \ldots, M\}$ and $\Gamma_2 \subset \{1, \ldots, N\}$ then $\|P_{\Gamma_1} U^* P_{\Gamma_2} U e_i\| \to \infty$ 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

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

and the second part of the proposition follows. The fact that the left hand side of (11.4) is zero when $q = 1$ is clear from (13.9) and (13.10).

□

Proof **of Proposition 11.2**. Without loss of generality we may assume that $\|\eta\| = 1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = q$. Let also, for $k \in \mathbb{N}$, $\xi_k = (UP_\Delta)^* e_k$. Observe that, since U is an isometry,

$$q^{-1}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta = \sum_{k=1}^{N} q^{-1} \delta_k \xi_k \otimes \bar{\xi}_k, \quad P_\Delta = \sum_{k=1}^{\infty} \xi_k \otimes \bar{\xi}_k, \tag{13.15}$$

and

$$\left\| \left(\frac{1}{q}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta \right) \eta \right\| \leq \left\| \left(\sum_{k=1}^{N} (q^{-1}\delta_k - 1) \xi_k \otimes \bar{\xi}_k \right) \eta \right\| + \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\eta\|, \tag{13.16}$$

where the infinite series in (13.15) converges in operator norm. Also, (13.16) follows directly from (13.15). To get the desired result we first focus on getting 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 the following is really 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 $\hat{\zeta}$. Also, let $\mathcal{F}$ be a countable collection of unit vectors such that for any $\xi \in \mathcal{H}$ we have that $\|\xi\| = \sup_{\zeta \in \mathcal{F}} \hat{\zeta}(\xi)$. Now define

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

Observe 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:

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

Note that for V we get the following estimate:

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

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

$$\hat{\zeta}(Z_k) = |\left(q^{-1}\delta_k - 1\right)| \langle \xi_k, \zeta \rangle || \langle \xi_k, \eta \rangle| \leq \max\{q^{-1} - 1, 1\} \upsilon^2(U) |\Delta|, \quad k \leq N, \tag{13.17}$$

thus

$$S \le \max\{q^{-1}-1,1\}\upsilon^2(U)|\Delta|, \tag{13.18}$$

where (13.18) is a direct consequence of (13.17). Finally, we can estimate R as follows

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

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

$$\mathbb{E}\left(\left\|\sum_{k\le N} Z_k\right\|\right) \le \sqrt{\mathbb{E}\left(\left\|\sum_{k\le N} Z_k\right\|^2\right)} \le \sqrt{q^{-1}(1-q)\upsilon^2(U)|\Delta|}. \tag{13.19}$$

With the estimates on V, S and R now established we may appeal to Theorem 13.2 and deduce that there is a constant $K>0$ such that for $\theta>0$ it follows that as long as q is chosen such that the right hand side of (13.19) is bounded by 1 (this is guarantied by the assumptions of the proposition),

$$\begin{aligned}&\mathbb{P}\left(\left\|\left(\sum_{k=1}^{N}(q^{-1}\delta_k-1)\xi_k\otimes\bar{\xi}_k\right)\eta\right\| \ge \theta + \sqrt{q^{-1}(1-q)\upsilon^2(U)|\Delta|}\right)\\ &\le 3\exp\left(-\frac{\theta}{K}(\max\{q^{-1}-1,1\}\upsilon^2(U)|\Delta|)^{-1}\log\left(1+\frac{\theta}{2}\right)\right).\end{aligned} \tag{13.20}$$

But by (13.16) it follows that for any $r>0$, we have

$$\begin{aligned}&\mathbb{P}\left(\left\|\left(\frac{1}{q}(P_\Omega UP_\Delta)^*P_\Omega UP_\Delta - P_\Delta\right)\eta\right\| \ge r\right)\\ &\le \mathbb{P}\left(\left\|\left(\sum_{k=1}^{N}(q^{-1}\delta_k-1)\xi_k\otimes\bar{\xi}_k\right)\eta\right\| \ge r - \|(P_\Delta U^*P_NUP_\Delta - P_\Delta)\|\right).\end{aligned} \tag{13.21}$$

Therefore, by appealing to (13.21) and (13.20) we obtain that for $\theta>0$

$$\begin{aligned}&\mathbb{P}\left(\left\|\left(\frac{1}{q}(P_\Omega UP_\Delta)^*P_\Omega UP_\Delta - P_\Delta\right)\eta\right\| \ge \theta + \sqrt{q^{-1}(1-q)\upsilon^2(U)|\Delta|} + \Xi\right)\\ &\le 3\exp\left(-\frac{\theta}{K}(\max\{q^{-1}-1,1\}\upsilon^2(U)|\Delta|)^{-1}\log\left(1+\frac{\theta}{2}\right)\right),\end{aligned}$$

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

Proof **of Theorem 11.3**. The proof is quite similar to the proof of Proposition 11.2. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \theta$. Note that we may argue as in (13.15) and observe that

$$\left\|\theta^{-1}(P_\Omega UP_\Delta)^*P_\Omega UP_\Delta - P_\Delta\right\| \le \left\|\sum_{k=1}^{N}(\theta^{-1}\delta_k-1)\xi_k\otimes\bar{\xi}_k\right\| + \|(P_\Delta U^*P_NUP_\Delta - P_\Delta)\|, \tag{13.22}$$

where $\xi_k = (UP_\Delta)^*e_k$. 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 11.2 the goal is to use Talagrand's powerful inequality and the first step 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

$$\mathbb{E}(\|Z\|)^2 \le 48\max\{\log(|\Delta|),1\}\theta^{-1}\upsilon^2(U)|\Delta|, \tag{13.23}$$

when

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

To prove the claim we simply rework the techniques used in [49]. This is now standard and has also been used in [16, 56].

We we start by observing that by letting $\tilde{\delta} = \{\tilde{\delta}_k\}_{k=1}^N$ be independent copies of $\delta = \{\delta_k\}_{k=1}^N$. Then

$$\begin{aligned} \mathbb{E}_\delta\left(\|Z\|\right) &= \mathbb{E}_\delta\left(\left\|Z - \mathbb{E}_{\tilde{\delta}}\left(\sum_{k=1}^N \left(\theta^{-1}\tilde{\delta}_k - 1\right)\xi_k \otimes \bar{\xi}_k\right)\right\|\right) \\ &\leq \mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|Z - \sum_{k=1}^N \left(\theta^{-1}\tilde{\delta}_k - 1\right)\xi_k \otimes \bar{\xi}_k\right\|\right)\right), \end{aligned} \tag{13.24}$$

by Jensen's inequality. Let $\varepsilon = \{\varepsilon_j\}_{j=1}^N$ be a sequence of Bernoulli variables taking values ± 1 with probability $1/2$. Then, by (13.24), symmetry, Fubini's Theorem and the triangle inequality, it follows that

$$\begin{aligned} \mathbb{E}_\delta\left(\|Z\|\right) \quad &\leq \mathbb{E}_\varepsilon\left(\mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|\sum_{k=1}^N \varepsilon_k\left(\theta^{-1}\delta_k - \theta^{-1}\tilde{\delta}_k\right)\xi_k \otimes \bar{\xi}_k\right\|\right)\right)\right) \\ &\leq 2\mathbb{E}_\delta\left(\mathbb{E}_\varepsilon\left(\left\|\sum_{k=1}^N \varepsilon_k\theta^{-1}\delta_k\xi_k \otimes \bar{\xi}_k\right\|\right)\right). \end{aligned} \tag{13.25}$$

Now, by Lemma 13.1 we get that

$$\mathbb{E}_\varepsilon\left(\left\|\sum_{k=1}^N \varepsilon_k\theta^{-1}\delta_k\xi_k \otimes \bar{\xi}_k\right\|\right) \leq 3\sqrt{\max\{2\log(|\Delta|), 2\}\theta^{-1}}\max_{1\leq k\leq N}\|\xi_k\|\sqrt{\left\|\sum_{k=1}^N \theta^{-1}\delta_k\xi_k \otimes \bar{\xi}_k\right\|}. \tag{13.26}$$

And hence, by using (13.25) and (13.26), 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 we have that $r \leq c(1+\sqrt{5})/2$, 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 Talagrands formula there are some preparations that have to be done. First write

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

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. Let, for $\eta \in \mathcal{G}$, 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 $\|Z\| = \sup_{\hat{\eta}_i\in\mathcal{F}}\sum_{k\leq N}\hat{\eta}_i(Z_k)$, and the setup for Talagrand's theorem is ready. Then, for $k = 1, \ldots, N$ we have

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

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

$$S = \sup_{\eta_i\in\mathcal{F}}\|\hat{\eta}_i\|_\infty \leq \theta^{-1}\max_{k\leq N}\|\xi_k\|^2 \leq \theta^{-1}\upsilon^2(U)|\Delta|. \tag{13.27}$$

Also, note that

$$\begin{aligned} V = \sup_{\hat{\eta}_i\in\mathcal{F}}\mathbb{E}\left(\sum_{k\leq N}\hat{\eta}(Z_k)^2\right) &\leq \sup_{\hat{\eta}\in\mathcal{F}}\mathbb{E}\left(\sum_{k\leq N}\left(\theta^{-1}\delta_k - 1\right)^2|\langle\xi_k, \eta\rangle|^4\right) \\ \leq \max_{k\leq N}\|\xi_k\|^2\left(\theta^{-1} - 1\right)\sup_{\hat{\eta}\in\mathcal{F}}\sum_{k\leq N}|\langle e_k, UP_\Delta\eta\rangle|^2 & \\ \leq \left(\theta^{-1} - 1\right)\max_{k\leq N}\|\xi_k\|^2 \leq \left(\theta^{-1} - 1\right)\upsilon^2(U)|\Delta|, & \end{aligned} \tag{13.28}$$

where the third inequality follows from the fact that U is an isometry. It follows by Talagrand's inequality (Theorem 13.2), by using the claim (and requiring that the right hand side of (13.23) is bounded by one, which is guarantied by the assumption of the theorem), the first part of the assumed (11.6), (13.27) and (13.28), that there is a constant $K > 0$ such that for $t > 0$

$$\begin{aligned}\mathbb{P}\left(\left\|\sum_{k=1}^{N}(\theta^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\right\| \geq t + 48\log(|\Delta|)\theta^{-1}\upsilon^2(U)|\Delta|\right)\\ \leq 3\exp\left(-\frac{t}{K}(\theta^{-1}\upsilon^2(U)|\Delta|)^{-1}\log\left(1+\frac{t}{2}\right)\right).\end{aligned} \tag{13.29}$$

But by (13.22) it follows that for any $r > 0$, we have

$$\begin{aligned}&\mathbb{P}\left(\left\|\frac{1}{\theta}(P_\Omega U P_\Delta)^* P_\Omega U P_\Delta - P_\Delta\right\| \geq r\right)\\ &\leq \mathbb{P}\left(\left\|\sum_{k=1}^{N}(\theta^{-1}\delta_k - 1)\xi_k \otimes \bar{\xi}_k\right\| \geq r - \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\|\right).\end{aligned} \tag{13.30}$$

Therefore, by appealing to (13.30) and (13.29) we obtain that for $t > 0$

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

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

References

[1] 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.

[2] 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.

[3] 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.

[4] B. Adcock, A. C. Hansen, and C. Poon. On optimal wavelet reconstructions from Fourier samples: linearity and universality of the stable sampling rate. *Technical report NA2012/07, DAMTP, University of Cambridge*, 2012.

[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. (accepted)*, 2013.

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

[7] B. Adcock, A. C. Hansen, and A. Shadrin. A stability barrier for reconstructions from Fourier samples. *Preprint*, 2012.

[8] A. Aldroubi. Oblique projections in atomic spaces. *Proc. Amer. Math. Soc.*, 124(7):2051–2060, 1996.

[9] W. Arveson. C^*-algebras and numerical linear algebra. *J. Funct. Anal.*, 122(2):333–360, 1994.

[10] 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.

[11] A. Böttcher. Infinite matrices and projection methods. In *Lectures on operator theory and its applications (Waterloo, ON, 1994)*, volume 3 of *Fields Inst. Monogr.*, pages 1–72. Amer. Math. Soc., Providence, RI, 1996.

[12] A. Böttcher and B. Silbermann. *Introduction to large truncated Toeplitz matrices*. Universitext. Springer-Verlag, New York, 1999.

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

[14] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math. (to appear)*, 2012.

[15] E. J. Candes and Y. Plan. A Probabilistic and RIPless Theory of Compressed Sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.

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

[17] 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.

[18] 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.

[19] 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.

[20] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.

[21] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001.

[22] 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.

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

[24] 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.

[25] 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.

[26] I. Ekeland and T. Turnbull. *Infinite-dimensional optimization and convexity*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1983.

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

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

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

[30] A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM J. Imaging Sci.*, 5:179–202, 2012.

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

[32] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhauser, 2013.

[33] D. Gross. Recovering Low-Rank Matrices From Few Coefficients in Any Basis. *IEEE Transactions on Information Theory*, 57(3), 2011.

[34] 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.

[35] 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.

[36] A. C. Hansen. On the approximation of spectra of linear operators on Hilbert spaces. *J. Funct. Anal.*, 254(8):2092–2126, 2008.

[37] 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.

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

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

[40] T. W. Körner. *Fourier Analysis*. Cambridge University Press, 1988.

[41] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

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

[43] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

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

[45] 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.

[46] S. K. Mitter. Convex optimization in infinite dimensional spaces. In *Recent advances in learning and control*, volume 371 of *Lecture Notes in Control and Inform. Sci.*, pages 161–179. Springer, London, 2008.

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

[48] H. Rauhut and R. Ward. Sparse Legendre expansions via l1-minimization. *J. Approx. Theory*, 164(5):517–533, 2012.

[49] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999.

[50] G. Strang and T. Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, 1997.

[51] T. Strohmer. Measure what should be measured: progress and challenges in compressive sensing. *IEEE Signal Process. Letters*, 19(12):887–893, 2012.

[52] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[53] 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.

[54] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[55] G. Tang, B. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *Preprint*, 2012.

[56] J. A. Tropp. On the conditioning of random subdictionaries. *Appl. Comput. Harmon. Anal.*, 25(1):1–24, 2008.

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

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

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