Breaking the coherence barrier: A new theory for compressed sensing

B. Adcock — Purdue Univ.
A. C. Hansen — Univ. of Cambridge
C. Poon — Univ. of Cambridge
B. Roman — Univ. of Cambridge

1 Introduction

This paper provides an extension of compressed sensing which bridges a substantial gap between existing theory and its current use in real-world applications.

Compressed sensing (CS), introduced by Candès, Romberg & Tao [14] and Donoho [25], has been one of the major developments in applied mathematics in the last decade [10, 27, 26, 22, 28, 29, 30]. Subject to appropriate conditions, it allows one to circumvent the traditional barriers of sampling theory (e.g. the Nyquist rate), and thereby recover signals from far fewer measurements than is classically considered possible. This has important implications in many practical applications, and for this reason CS has, and continues to be, very intensively researched.

The theory of CS is based on three fundamental concepts: *sparsity*, *incoherence* and *uniform random subsampling*. Whilst there are examples where these apply, in many applications one or more of these principles may be lacking. This includes virtually all of medical imaging – Magnetic Resonance Imaging (MRI), Computerized Tomography (CT) and other versions of tomography such as Thermoacoustic, Photoacoustic or Electrical Impedance Tomography – most of electron microscopy, as well as seismic tomography, fluorescence microscopy, Hadamard spectroscopy and radio interferometry. In many of these problems, it is the principle of incoherence that is lacking, rendering the standard theory inapplicable. Despite this issue, compressed sensing has been, and continues to be, used with great success in many of these areas. Yet, to do so it is typically implemented with sampling patterns that differ substantially from the uniform subsampling strategies suggested by the theory. In fact, in many cases uniform random subsampling yields highly suboptimal numerical results.

The standard mathematical theory of CS has now reached a mature state. However, as this discussion attests, there is a substantial, and arguably widening gap between the theoretical and applied sides of the field. New developments and sampling strategies are increasingly based on empirical evidence lacking mathematical justification. Furthermore, in the above applications one also witnesses a number of intriguing phenomena that are not explained by the standard theory. For example, in such problems, the optimal sampling strategy depends not just on the overall sparsity of the signal, but also on its structure, as will be documented thoroughly in this paper. This phenomenon is in direct contradiction with the usual sparsity-based theory of CS. Theorems that explain this observation – i.e. that reflect how the optimal subsampling strategy depends on the structure of the signal – do not currently exist.

The purpose of this paper is to provide a bridge across this divide. It does so by generalizing the three traditional pillars of CS to three new concepts: *asymptotic sparsity*, *asymptotic incoherence* and *multilevel random subsampling*. This new theory shows that CS is also possible, and reveals several advantages, under these substantially more general conditions. Critically, it also addresses the important issue raised above: the dependence of the subsampling strategy on the structure of the signal.

The importance of this generalization is threefold. First, as will be explained, real-world inverse problems are typically not incoherent and sparse, but asymptotically incoherent and asymptotically sparse. This paper provides the first comprehensive mathematical explanation for a range of empirical usages of CS in applications such as those listed above. Second, in showing that incoherence is not a requirement for CS, but instead that asymptotic incoherence suffices, the new theory offers markedly greater flexibility in the design of sensing mechanisms. In the future, sensors need only satisfy this significantly more relaxed condition. Third, by using asymptotic incoherence and multilevel sampling to exploit not just sparsity, but also structure, i.e. asymptotic sparsity, the new theory paves the way for an improved CS paradigm that achieve better reconstructions in practice from fewer measurements.

A critical aspect of many practical problems such as those listed above is that they do not offer the freedom to design or choose the sensing operator, but instead impose it (e.g. Fourier sampling in MRI). As such, much of the existing CS work, which relies on random or custom-designed sensing matrices, typically to provide universality, is not applicable. This paper shows that in many such applications the imposed sensing operators are highly non-universal and coherent with popular sparsifying bases. Yet they are asymptotically incoherent, and thus fall within the remit of the new theory. Spurred by this observation, this paper also raises the question of whether universality is actually desirable in practice, even in applications where there is flexibility to design sensing operators with this property (e.g. in compressive imaging). The new theory shows that asymptotically incoherent sensing and multilevel sampling allow one to exploit structure, not just sparsity. Doing so leads to notable advantages over universal operators, even for problems where the latter are applicable. Moreover, and crucially, this can be done in a computationally efficient manner using fast Fourier or Hadamard transforms (see §6.1).

This aside, another outcome of this work is that the Restricted Isometry Property (RIP), although a popular tool in CS theory, is of little relevance in many practical inverse problems. As confirmed later via the so-called *flip test*, the RIP does not hold in such applications.

Before we commence with the remainder of this paper, let us make several further remarks. First, many of the problems listed above are analog, i.e. they are modelled with continuous transforms, such as the Fourier or Radon transforms. Conversely, the standard theory of CS is based on a finite-dimensional model. Such *mismatch* can lead to critical errors when applied to real data arising from continuous models, or inverse crimes when the data is inappropriately simulated [16, 34]. To overcome this issue, a theory of CS in infinite dimensions was recently introduced in [1]. This paper fundamentally extends [1] by presenting new theory in both the finite- and infinite-dimensional settings, the infinite-dimensional analysis also being instrumental for obtaining the Fourier and wavelets estimates in §6.

Second, this is primarily a mathematical paper. However, as one may expect in light of the above discussion, there are a range of practical implications. We therefore encourage the reader to consult the paper [53] for further discussions on the practical aspects and more extensive numerical experiments.

2 The need for a new theory

Let us ask the following question: does the standard theory of CS explain its empirical success in the aforementioned applications? We now argue that the answer is no. Specifically, even in well-known applications such as MRI (recall that MRI was one of the first applications of CS, due to the pioneering work of Lustig et al. [42, 44, 45, 46]), there is a significant gap between theory and practice.

2.1 Compressed sensing

Let us commence with a short review of finite-dimensional CS theory – infinite-dimensional CS will be considered in §5. A typical setup, and one which we shall follow in part of this paper, is as follows. Let $\{\psi_j\}_{j=1}^N$ and $\{\varphi_j\}_{j=1}^N$ be two orthonormal bases of $\mathbb{C}^N$, the *sampling* and *sparsity* bases respectively, and write $U = (u_{ij})_{i,j=1}^N \in \mathbb{C}^{N\times N}$, $u_{ij} = \langle \varphi_j, \psi_i \rangle$. Note that U is an isometry, i.e. $U^*U = I$.

Definition 2.1. *Let $U = (u_{ij})_{i,j=1}^N \in \mathbb{C}^{N\times N}$ be an isometry. The coherence of U is precisely*

$$\mu(U) = \max_{i,j=1,\dots,N} |u_{ij}|^2 \in [N^{-1}, 1]. \tag{2.1}$$

We say that U is perfectly incoherent if $\mu(U) = N^{-1}$.

A signal $f \in \mathbb{C}^N$ is said to be s-sparse in the orthonormal basis $\{\varphi_j\}_{j=1}^N$ if at most s of its coefficients in this basis are nonzero. In other words, $f = \sum_{j=1}^N x_j \varphi_j$, and the vector $x \in \mathbb{C}^N$ satisfies $|\mathrm{supp}(x)| \leq s$, where $\mathrm{supp}(x) = \{j : x_j \neq 0\}$. Let $f \in \mathbb{C}^N$ be s-sparse in $\{\varphi_j\}_{j=1}^N$, and suppose we have access to the samples $\hat{f}_j = \langle f, \psi_j \rangle$, $j = 1, \dots, N$. Let $\Omega \subseteq \{1, \dots, N\}$ be of cardinality m and chosen uniformly at random. According to a result of Candès & Plan [12] and Adcock & Hansen [1], f can be recovered exactly with probability exceeding $1 - \epsilon$ from the subset of measurements $\{\hat{f}_j : j \in \Omega\}$, provided

$$m \gtrsim \mu(U) \cdot N \cdot s \cdot \left(1 + \log(\epsilon^{-1})\right) \cdot \log(N), \tag{2.2}$$

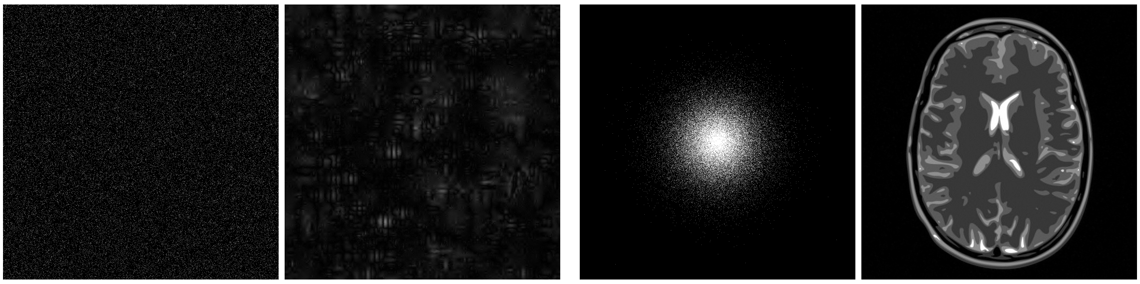


Figure 1: Left to right: (i) 5% uniform random subsampling scheme, (ii) CS reconstruction from uniform subsampling, (iii) 5% multilevel subsampling scheme, (iv) CS reconstruction from multilevel subsampling.

(here and elsewhere in this paper we shall use the notation $a \gtrsim b$ to mean that there exists a constant $C > 0$ independent of all relevant parameters such that $a \geq Cb$). In practice, recovery is achieved by solving the following convex optimization problem:

$$\min_{\eta \in \mathbb{C}^N} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \hat{f}, \tag{2.3}$$

where $\hat{f} = (\hat{f}_1, \ldots, \hat{f}_N)^\top$ and $P_\Omega \in \mathbb{C}^{N \times N}$ is the diagonal projection matrix with j^{th} entry 1 if $j \in \Omega$ and zero otherwise. The key estimate (2.2) shows that the number of measurements m required is, up to a log factor, on the order of the sparsity s, provided the coherence $\mu(U) = \mathcal{O}\left(N^{-1}\right)$. This is the case, for example, when U is the DFT matrix; a problem which was studied in some of the first papers on CS [14].

2.2 Incoherence is rare in practice

To test the practicality of the incoherence condition, let us consider a typical CS problem. In a number of important applications, not least MRI, the sampling is carried out in the Fourier domain. Since images are sparse in wavelets, the usual CS setup is to form the a matrix $U_N = U_{\text{df}} V_{\text{dw}}^{-1} \in \mathbb{C}^{N \times N}$, where U_{df} and V_{dw} represent the discrete Fourier and wavelet transforms respectively. However, in the case the coherence satisfies $\mu(U_N) = \mathcal{O}(1)$ as $N \to \infty$, for any wavelet basis. Thus, this problem has the worst possible coherence, and the standard CS estimate (2.2) states that $m = N$ samples are needed in this case (i.e. full sampling), even though the object to recover is typically highly sparse. Note that this is not an insufficiency of the theory. If uniform random subsampling is employed, then the lack of incoherence does indeed lead to a very poor reconstruction. This can be seen in Figure 1.

The underlying reason for this lack of incoherence can be traced to the fact that this finite-dimensional problem is a discretization of an infinite-dimensional problem. Specifically,

$$\underset{N \to \infty}{\text{WOT-lim}}\; U_{\text{df}} V_{\text{dw}}^{-1} = U, \tag{2.4}$$

where $U : l^2(\mathbb{N}) \to l^2(\mathbb{N})$ is the operator represented as the infinite matrix

$$U = \begin{pmatrix} \langle \varphi_1, \psi_1 \rangle & \langle \varphi_2, \psi_1 \rangle & \cdots \\ \langle \varphi_1, \psi_2 \rangle & \langle \varphi_2, \psi_2 \rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \tag{2.5}$$

and the functions φ_j are the wavelets used, the ψ_j's are the standard complex exponentials and WOT denotes the weak operator topology. Since the coherence of the infinite matrix U – i.e. the supremum of its entries in absolute value – is a fixed number independent of N, we cannot expect incoherence of the discretization U_N for large N. At some point, one will always encounter the so-called *coherence barrier*. Such an issue is not isolated to this example. Heuristically, any problem that arises as a discretization of an infinite-dimensional problem will suffer from the same phenomenon. The list of applications of this type is long, and includes for example, MRI, CT, microscopy and seismology.

To mitigate this problem, one may naturally try to change $\{\varphi_j\}$ or $\{\psi_j\}$. However, this will deliver only marginal benefits, since (2.4) demonstrates that the coherence barrier will always occur for large enough N.

In view of this, one may wonder how it is possible that CS is applied so successfully to many such problems. The key is so-called *asymptotic* incoherence (see §3.1) and the use of a variable density/multilevel subsampling strategy. The success of such subsampling is confirmed numerically in Figure 1. However, it is important to note that this is an empirical solution to the problem. None of the usual theory explains the effectiveness of CS when implemented in this way.

2.3 Sparsity and the flip test

The previous discussion demonstrates that we must dispense with the principles of incoherence and uniform random subsampling in order to develop a new theory of CS. We now claim that sparsity must also be replaced with a more general concept. This may come as a surprise to the reader, since sparsity is a central pillar of not just CS, but much of modern signal processing. However, this can be confirmed by a simple experiment we refer to as the *flip test*.

Sparsity asserts that an unknown vector x has s important coefficients, where the locations can be arbitrary. CS establishes that all s-sparse vectors can be recovered from the same sampling strategy. In particular, the sampling strategy is completely independent of the location of these coefficients. The flip test, described next, allows one to evaluate whether this holds in a given application. Let $x \in \mathbb{C}^N$ and $U \in \mathbb{C}^{N\times N}$. Next we take samples according to some appropriate subset $\Omega \subseteq \{1, \dots, N\}$ with $|\Omega| = m$, and solve:

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \text{ subject to } P_\Omega U z = P_\Omega U x. \tag{2.6}$$

This gives a reconstruction $z = z_1$. Now we flip x through the operation $x \mapsto x^{\text{fp}} \in \mathbb{C}^N$, $x^{\text{fp}}_1 = x_N$, $x^{\text{fp}}_2 = x_{N-1}, \dots, x^{\text{fp}}_N = x_1$, giving a new vector x^{fp} with reversed entries. We next apply the same CS reconstruction to x^{fp}, using the same matrix U and the same subset Ω. That is we solve

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \text{ subject to } P_\Omega U z = P_\Omega U x^{\text{fp}}. \tag{2.7}$$

Let z be a solution of (2.7). In order to get a reconstruction of the original vector x, we perform the flipping operation once more and form the final reconstruction $z_2 = z^{\text{fp}}$.

Suppose now that Ω is a good sampling pattern for recovering x using the solution z_1 of (2.6). If sparsity is the key structure that determines such reconstruction quality, then we expect exactly the same quality in the approximation z_2 obtained via (2.7), since x^{fp} is merely a permutation of x. To investigate whether or not this is true, we consider several examples arising from the following applications: fluorescence microscopy, compressive imaging, MRI, CT, electron microscopy and radio interferometry. These examples are based on the matrix $U = U_{\text{dft}} V^{-1}_{\text{dwt}}$ or $U = U_{\text{Had}} V^{-1}_{\text{dwt}}$, where U_{dft} is the discrete Fourier transform, U_{Had} is a Hadamard matrix and V_{dwt} is the discrete wavelet transform.

The results of this experiment are shown in Figure 2. As is evident, in all cases the flipped reconstructions z_2 are substantially worse than their unflipped counterparts z_1. Hence, we conclude that sparsity alone does not govern the reconstruction quality, and consequently the success in the unflipped case must also be due in part to the structure of the signal. In other words:

The optimal subsampling strategy depends on the signal structure.

Note that the flip test reveals another interesting phenomenon:

There is no Restricted Isometry Property (RIP).

Suppose the matrix $P_\Omega U$ satisfied an RIP for realistic parameter values (i.e. problem size N, subsampling percentage m, and sparsity s) found in applications. Then this would imply recovery of all approximately sparse vectors with the same error. This is in direct contradiction with the results of the flip test.

Note that in all the examples in Figure 2, uniform random subsampling would have given nonsensical results, analogously to what was shown in Figure 1.

2.4 Signals and images are asymptotically sparse in -lets

Given that structure is key, we now ask the question: what, if any, structure is characteristic of such applications? Let us consider a wavelet basis $\{\varphi_n\}_{n\in\mathbb{N}}$. Recall that associated to such a basis, there is a natural

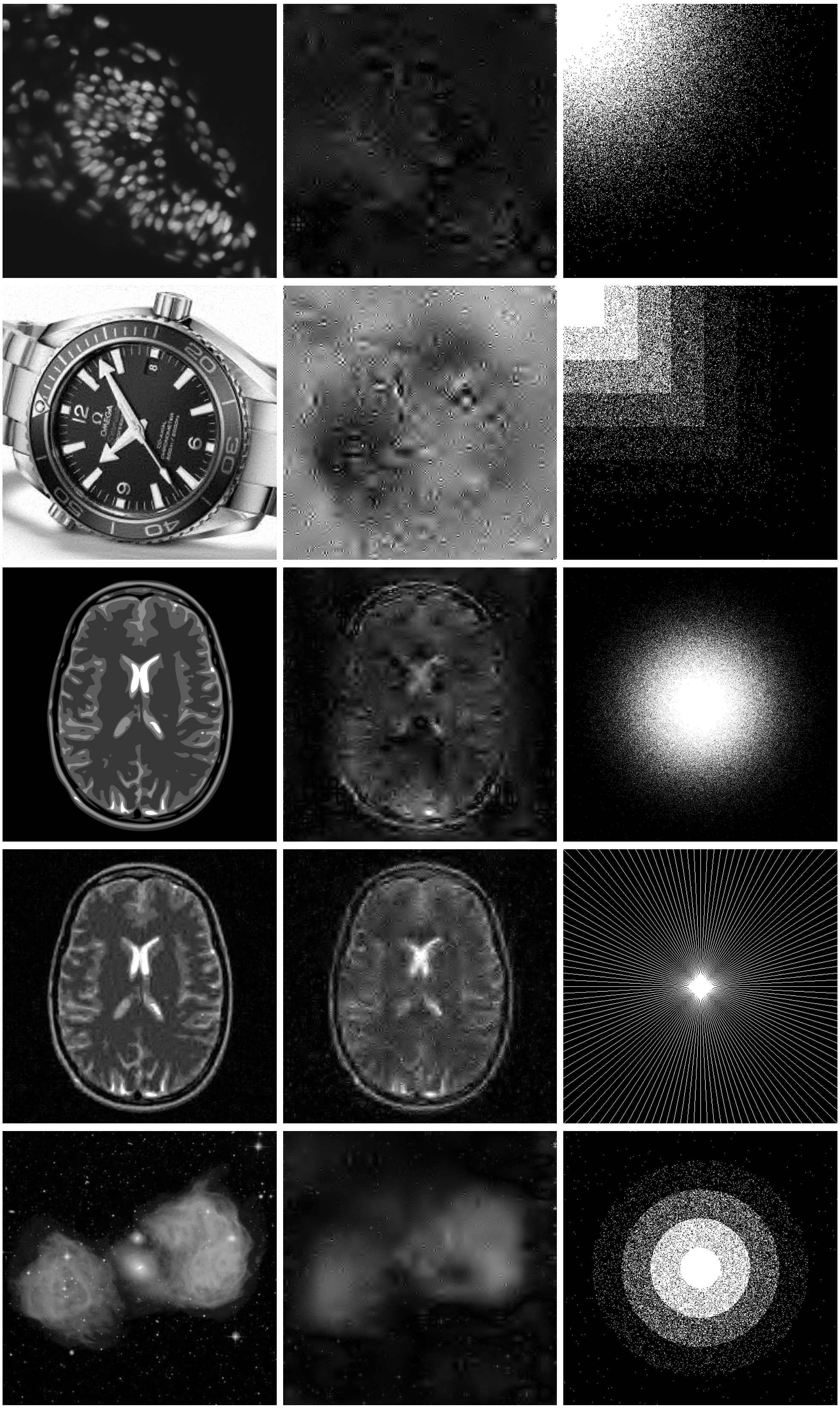


Figure 2: Reconstructions via CS (left column) and the flipped wavelet coefficients (middle column). The right column shows the subsampling map used. The percentage shown is the fraction of Fourier or Hadamard coefficients that were sampled. The reconstruction basis was DB4 for the Fluorescence microscopy example, and DB6 for the rest.

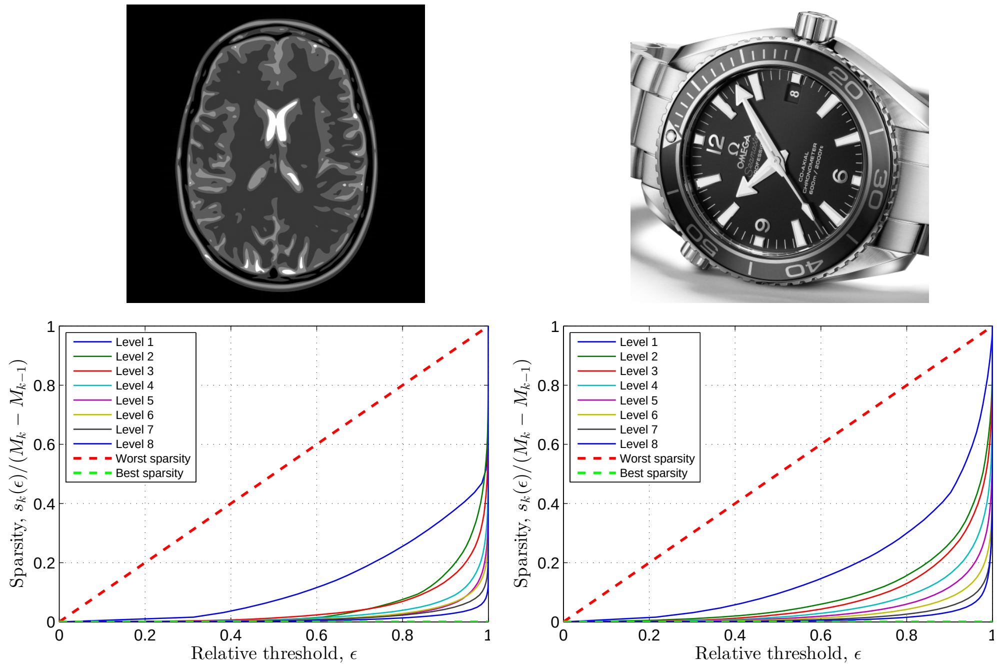


Figure 3: Relative sparsity of the Daubechies-8 wavelet coefficients of two images. Here the levels correspond to wavelet scales and $s_k(\epsilon)$ is given by (2.8). Each curve shows the relative sparsity at level k as a function of ϵ. The decreasing nature of the curves for increasing k confirms (2.9).

decomposition of $\mathbb{N}$ into finite subsets according to different scales, i.e. $\mathbb{N} = \bigcup_{k \in \mathbb{N}} \{M_{k-1}+1, \ldots, M_k\}$, where $0 = M_0 < M_1 < M_2 < \ldots$ and $\{M_{k-1}+1, \ldots, M_k\}$ is the set of indices corresponding to the k^{th} scale. Let $x \in l^2(\mathbb{N})$ be the coefficients of a function f in this basis. Suppose that $\epsilon \in (0, 1]$ is given, and define

$$s_k = s_k(\epsilon) = \min \left\{ K : \Big\| \sum_{i=1}^{K} x_{\pi(i)} \varphi_{\pi(i)} \Big\| \geq \epsilon \Big\| \sum_{i=M_{k-1}+1}^{M_k} x_j \varphi_j \Big\| \right\}, \tag{2.8}$$

where $\pi : \{1, \ldots, M_k - M_{k-1}\} \to \{M_{k-1}+1, \ldots, M_k\}$ is a bijection such that $|x_{\pi(i)}| \geq |x_{\pi(i+1)}|$ for $i = 1, \ldots, M_k - M_{k-1} - 1$. In order words, the quantity s_k is the effective sparsity of the wavelet coefficients of f at the k^{th} scale.

Sparsity of f in a wavelet basis means that for a given maximal scale $r \in \mathbb{N}$, the ratio $s/M_r \ll 1$, where $M = M_r$ and $s = s_1 + \ldots + s_r$ is the total effective sparsity of f. The observation that typical images and signals are approximately sparse in wavelet bases is one of the key results in nonlinear approximation [23, 47]. However, such objects exhibit far more than sparsity alone. In fact, the ratios

$$s_k/(M_k - M_{k-1}) \to 0, \tag{2.9}$$

rapidly as $k \to \infty$, for every fixed $\epsilon \in (0, 1]$. Thus typical signals and images have a distinct sparsity *structure*. They are much more sparse at fine scales (large k) than at coarse scales (small k). This is confirmed in Figure 3. Note that this conclusion does not change if one replaces wavelets by other related approximation systems, such as curvelets [9, 11], contourlets [24, 49] or shearlets [18, 19, 41].

3 New principles

Having argued for their need, we now introduce the main new concepts of the paper: namely, asymptotic incoherence, asymptotic sparsity and multilevel sampling.

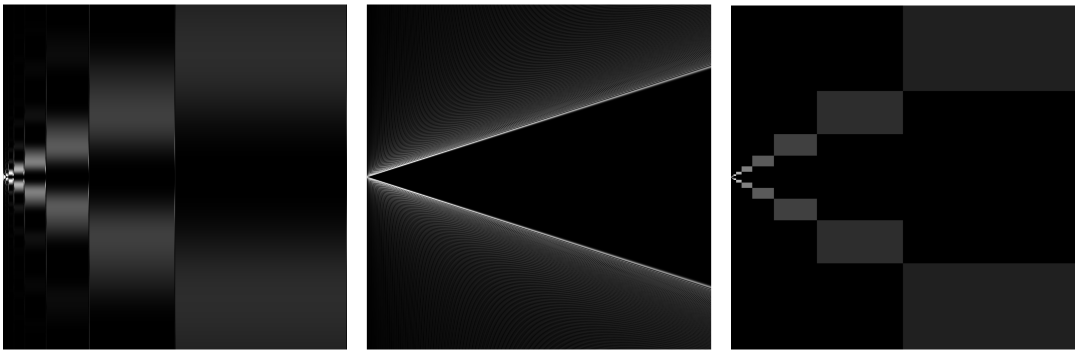


Figure 4: The absolute values of the matrix U in (2.5): (*left*): DB2 wavelets with Fourier sampling. (*middle*): Legendre polynomials with Fourier sampling. (*right*): The absolute values of $U_{\mathrm{Had}} V_{\mathrm{dwt}}^{-1}$, where U_{Had} is a Hadamard matrix and V_{dwt}^{-1} is the discrete Haar transform. Light regions correspond to large values and dark regions to small values.

3.1 Asymptotic incoherence

Recall from §2.2 that the case of Fourier sampling with wavelets as the sparsity basis is a standard example of a coherent problem. Similarly, Fourier sampling with Legendre polynomials is also coherent, as is the case of Hadamard sampling with wavelets. In Figure 4 we plot the absolute values of the entries of the matrix U for these three examples. As is evident, whilst U does indeed have large entries in all three case (since it is coherent), these are isolated to a leading submatrix (note that we enumerate over $\mathbb{Z}$ for the Fourier sampling basis and $\mathbb{N}$ for the wavelet/Legendre sparsity bases). As one moves away from this region the values get progressively smaller. That is, the matrix U is incoherent aside from a leading coherent submatrix. This motivates the following definition:

Definition 3.1 (Asymptotic incoherence)**.** *Let be $\{U_N\}$ be a sequence of isometries with $U_N \in \mathbb{C}^N$ or let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then*

(i) $\{U_N\}$ is asymptotically incoherent if $\mu(P_K^{\perp} U_N),\ \mu(U_N P_K^{\perp}) \to 0$, when $K \to \infty$, with $N/K = c$, for all $c \geq 1$.

(ii) U is asymptotically incoherent if $\mu(P_K^{\perp} U),\ \mu(U P_K^{\perp}) \to 0$, when $K \to \infty$.

Here P_K is the projection onto $\mathrm{span}\{e_j : j = 1, ..., K\}$*, where $\{e_j\}$ is the canonical basis of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$, and $P_K^{\perp}$ is its orthogonal complement.*

In other words, U is asymptotically incoherent if the coherences of the matrices formed by replacing either the first K rows or columns of U are small. As it transpires, the Fourier/wavelets, Fourier/Legendre and Hadamard/wavelets problems are asymptotically incoherent. In particular, $\mu(P_K^{\perp} U),\ \mu(U P_K^{\perp}) = \mathcal{O}\left(K^{-1}\right)$ as $K \to \infty$ for the former (see §6).

3.2 Multi-level sampling

Asymptotic incoherence suggests a different subsampling strategy should be used instead of uniform random sampling. High coherence in the first few rows of U means that important information about the signal to be recovered may well be contained in its corresponding measurements. Hence to ensure good recovery we should fully sample these rows. Conversely, once outside of this region, when the coherence starts to decrease, we can begin to subsample. Let $N_1, N, m \in \mathbb{N}$ be given. This now leads us to consider an index set Ω of the form $\Omega = \Omega_1 \cup \Omega_2$, where $\Omega_1 = \{1, \ldots, N_1\}$, and $\Omega_2 \subseteq \{N_1 + 1, \ldots, N\}$ is chosen uniformly at random with $|\Omega_2| = m$. We refer to this as a *two-level* sampling scheme. As we shall prove later, the amount of subsampling possible (i.e. the parameter m) in the region corresponding to Ω_2 will depend solely on the sparsity of the signal and coherence $\mu(P_{N_1}^{\perp} U)$.

The two-level scheme represents the simplest type of nonuniform density sampling. There is no reason, however, to restrict our attention to just two levels, full and subsampled. In general, we shall consider *multilevel* schemes, defined as follows:

Definition 3.2 (Multilevel random sampling). *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$, $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \ldots, r$, and suppose that $\Omega_k \subseteq \{N_{k-1}+1, \ldots, N_k\}$, $\quad |\Omega_k| = m_k, \quad k = 1, \ldots, r$, are chosen uniformly at random, where $N_0 = 0$. We refer to the set $\Omega = \Omega_{\mathbf{N},\mathbf{m}} = \Omega_1 \cup \ldots \cup \Omega_r$. as an $(\mathbf{N}, \mathbf{m})$-multilevel sampling scheme.*

Note that the idea of sampling the low-order coefficients of an image differently goes back to the early days of CS. In particular, Donoho considers a two-level approach for recovering wavelet coefficients in his seminal paper [25], based on acquiring the coarse scale coefficients directly. This was later extended by Tsaig & Donoho to so-called 'multiscale CS' in [60], where distinct subbands were sensed separately. See also Romberg's work [54], and as well as Candès & Romberg [13].

We also remark that, although motivated by wavelets, our definition is completely general, as are the theorems we present in §4 and §5. Moreover, and critically, we do not assume separation of the coefficients into distinct levels before sampling (as done above), which is often infeasible in practice (in particular, any application based on Fourier or Hadamard sampling). Note also that in MRI similar sampling strategies to what we introduce here are found in most implementations of CS [45, 46, 51, 52]. Additionally, a so-called "half-half" scheme (an example of a two-level strategy) was used by [57] in application of CS in fluorescence microscopy, albeit without theoretical recovery guarantees.

3.3 Asymptotic sparsity in levels

The flip test, the discussion in §2.4 and Figure 3 suggest that we need a different concept to sparsity. Given the structure of modern function systems such as wavelets and their generalizations, we propose the notion of sparsity in levels:

Definition 3.3 (Sparsity in levels). *Let x be an element of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$. For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \leq M_1 < \ldots < M_r$ and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \ldots, r$, where $M_0 = 0$. We say that x is $(\mathbf{s}, \mathbf{M})$-sparse if, for each $k = 1, \ldots, r$, $\Delta_k := \mathrm{supp}(x) \cap \{M_{k-1}+1, \ldots, M_k\}$, satisfies $|\Delta_k| \leq s_k$. We denote the set of $(\mathbf{s}, \mathbf{M})$-sparse vectors by $\Sigma_{\mathbf{s},\mathbf{M}}$.*

Definition 3.4 ($(\mathbf{s}, \mathbf{M})$-term approximation). *Let $f = \sum_j x_j \varphi_j$, where $\{\varphi_j\}$ is some orthonormal basis of a Hilbert space and $x = (x_j)$ is an element of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$. We define the $(\mathbf{s}, \mathbf{M})$-term approximation*

$$\sigma_{\mathbf{s},\mathbf{M}}(f) = \min_{\eta \in \Sigma_{\mathbf{s},\mathbf{M}}} \|x - \eta\|_{l^1}. \tag{3.1}$$

Typically, it is the case that $s_k/(M_k - M_{k-1}) \to 0$ as $k \to \infty$, in which case we say that x is *asymptotically sparse in levels*.

4 Main theorems I: the finite-dimensional case

We now present the main theorems in the finite-dimensional setting. In §5 we address the infinite-dimensional case. To avoid pathological examples we will assume throughout that the total sparsity $s = s_1 + \ldots + s_r \geq 3$. This is simply to ensure that $\log(s) \geq 1$, which is convenient in the proofs.

4.1 Two-level sampling schemes

We commence with the case of two-level sampling schemes. Recall that in practice, signals are never exactly sparse (or sparse in levels), and their measurements are always contaminated by noise. Let $f = \sum_j x_j \varphi_j$ be a fixed signal, and write $y = P_\Omega \hat{f} + z = P_\Omega U x + z$, for its noisy measurements, where $z \in \mathrm{ran}(P_\Omega)$ is a noise vector satisfying $\|z\| \leq \delta$ for some $\delta \geq 0$. If δ is known, we now consider the following problem:

$$\min_{\eta \in \mathbb{C}^N} \|\eta\|_{l^1} \text{ subject to } \|P_\Omega U \eta - y\| \leq \delta. \tag{4.1}$$

Our aim now is to recover x up to an error proportional to δ and the best approximation error $\sigma_{\mathbf{s},\mathbf{M}}(f)$.

Before stating our theorem, it is useful to make the following definition. For $K \in \mathbb{N}$, we write $\mu_K = \mu(P_K^\perp U)$. We now have the following:

Theorem 4.1. *Let $U \in \mathbb{C}^{N\times N}$ be an isometry and $x \in \mathbb{C}^N$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a two-level sampling scheme, where $\mathbf{N} = (N_1, N_2)$, $N_2 = N$, and $\mathbf{m} = (N_1, m_2)$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, M_2) \in \mathbb{N}^2$, $M_1 < M_2$, $M_2 = N$, and $\mathbf{s} = (M_1, s_2) \in \mathbb{N}^2$, $s_2 \leq M_2 - M_1$, be any pair such that the following holds:*

(i) we have

$$\|P_{N_1}^{\perp} U P_{M_1}\| \leq \frac{\gamma}{\sqrt{M_1}} \tag{4.2}$$

and $\gamma \leq s_2\sqrt{\mu_{N_1}}$ for some $\gamma \in (0, 2/5]$;

(ii) for $\epsilon \in (0, e^{-1}]$, let

$$m_2 \gtrsim (N - N_1) \cdot \log(\epsilon^{-1}) \cdot \mu_{N_1} \cdot s_2 \cdot \log(N).$$

Suppose that $\xi \in \mathbb{C}^N$ is a minimizer of (4.1) with $\delta = \tilde{\delta}\sqrt{K^{-1}}$ and $K = (N_2 - N_1)/m_2$. Then, with probability exceeding $1 - s\epsilon$, we have

$$\|\xi - x\| \leq C \cdot \left(\tilde{\delta} \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f)\right), \tag{4.3}$$

for some constant C, where $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is as in (3.1), $L = 1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KM\sqrt{s})}$. If $m_2 = N - N_1$ then this holds with probability 1.

To interpret Theorem 4.1, and in particular, show how it overcomes the coherence barrier, we note the following:

(i) The condition $\|P_{N_1}^{\perp} U P_{M_1}\| \leq \frac{2}{5\sqrt{M_1}}$ (which is always satisfied for some N_1) implies that fully sampling the first N_1 measurements allows one to recover the first M_1 coefficients of f.

(ii) To recover the remaining s_2 coefficients we require, up to log factors, an additional $m_2 \gtrsim (N - N_1) \cdot \mu_{N_1} \cdot s_2$, measurements, taken randomly from the range $M_1 + 1, \dots, M_2$. In particular, if N_1 is a fixed fraction of N, and if $\mu_{N_1} = \mathcal{O}\left(N_1^{-1}\right)$, such as for wavelets with Fourier measurements (Theorem 6.1), then one requires only $m_2 \gtrsim s_2$ additional measurements to recover the sparse part of the signal.

Thus, in the case where x is asymptotically sparse, we require a fixed number N_1 measurements to recover the nonsparse part of x, and then a number m_2 depending on s_2 and the asymptotic coherence μ_{N_1} to recover the sparse part.

Remark 4.1 It is not necessary to know the sparsity structure, i.e. the values $\mathbf{s}$ and $\mathbf{M}$, of the signal f in order to implement the two-level sampling technique (the same also applies to the multilevel technique discussed in the next section). Given a two-level scheme $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$, Theorem 4.1 demonstrates that f will be recovered exactly up to an error on the order of $\sigma_{\mathbf{s},\mathbf{M}}(f)$, where $\mathbf{s}$ and $\mathbf{M}$ are determined implicitly by $\mathbf{N}$, $\mathbf{m}$ and the conditions (i) and (ii) of the theorem. Of course, some *a priori* knowledge of $\mathbf{s}$ and $\mathbf{M}$ will greatly assist in selecting the parameters $\mathbf{N}$ and $\mathbf{m}$ so as to get the best recovery results. However, this is not strictly necessary for implementation.

4.2 Multilevel sampling schemes

We now consider the case of multilevel sampling schemes. Before presenting this case, we need several definitions. The first is key concept in this paper: namely, *the local coherence*.

Definition 4.2 (Local coherence)**.** *Let U be an isometry of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$. If $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ and $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots N_r$ and $1 \leq M_1 < \dots < M_r$ the $(k,l)^{\text{th}}$ local coherence of U with respect to $\mathbf{N}$ and $\mathbf{M}$ is given by*

$$\mu_{\mathbf{N},\mathbf{M}}(k,l) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}) \cdot \mu(P_{N_k}^{N_{k-1}} U)}, \quad k,l = 1,\dots,r,$$

where $N_0 = M_0 = 0$ and P_b^a denotes the projection matrix corresponding to indices $\{a+1, \dots, b\}$. In the case where $U \in \mathcal{B}(l^2(\mathbb{N}))$ (i.e. U belongs to the space of bounded operators on $l^2(\mathbb{N})$), we also define

$$\mu_{\mathbf{N},\mathbf{M}}(k,\infty) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_{r-1}}^{\perp}) \cdot \mu(P_{N_k}^{N_{k-1}} U)}, \quad k = 1,\dots,r.$$

Besides the local sparsities s_k, we shall also require the notion of a relative sparsity:

Definition 4.3 (Relative sparsity). *Let U be an isometry of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$. For $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$, $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$ and $1 \leq M_1 < \ldots < M_r$, $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$ and $1 \leq k \leq r$, the k^{th} relative sparsity is given by $S_k = S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}) = \max_{\eta \in \Theta} \|P_{N_k}^{N_{k-1}} U\eta\|^2$, where $N_0 = M_0 = 0$ and Θ is the set*

$$\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, |\mathrm{supp}(P_{M_l}^{M_{l-1}}\eta)| = s_l,\ l = 1, \ldots, r\}.$$

We can now present our main theorem:

Theorem 4.4. *Let $U \in \mathbb{C}^{N \times N}$ be an isometry and $x \in \mathbb{C}^N$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$, $N_r = N$, and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_r = N$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds: for $\epsilon \in (0, e^{-1}]$ and $1 \leq k \leq r$,*

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r} \mu_{\mathbf{N},\mathbf{M}}(k, l) \cdot s_l \right) \cdot \log(N), \tag{4.4}$$

where $m_k \gtrsim \hat{m}_k \cdot \log(\epsilon^{-1}) \cdot \log(N)$, and $\hat{m}_k$ is such that

$$1 \gtrsim \sum_{k=1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1 \right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k, l) \cdot \tilde{s}_k, \tag{4.5}$$

for all $l = 1, \ldots, r$ and all $\tilde{s}_1, \ldots, \tilde{s}_r \in (0, \infty)$ satisfying

$$\tilde{s}_1 + \ldots + \tilde{s}_r \leq s_1 + \ldots + s_r, \qquad \tilde{s}_k \leq S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}).$$

Suppose that $\xi \in \mathbb{C}^N$ is a minimizer of (4.1) with $\delta = \tilde{\delta}\sqrt{K^{-1}}$ and $K = \max_{1 \leq k \leq r}\{(N_k - N_{k-1})/m_k\}$. Then, with probability exceeding $1 - s\epsilon$, where $s = s_1 + \ldots + s_r$, we have that

$$\|\xi - x\| \leq C \cdot \left(\tilde{\delta} \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f) \right),$$

for some constant C, where $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is as in (3.1), $L = 1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KM\sqrt{s})}$. If $m_k = N_k - N_{k-1}$, $1 \leq k \leq r$, then this holds with probability 1.

The key component of this theorem is the bounds (4.4) and (4.5). Whereas the standard CS estimate (2.2) relates the total number of samples m to the global coherence and the global sparsity, these bounds now relate the local sampling m_k to the local coherences $\mu_{\mathbf{N},\mathbf{M}}(k, l)$ and local and relative sparsities s_k and S_k. In particular, by relating these local quantities this theorem conforms with the conclusions of the flip test in §2.3: namely, that the optimal sampling strategy must depend on the signal structure. This is exactly what is described in (4.4) and (4.5).

On the face of it, the bounds (4.4) and (4.5) may appear somewhat complicated, not least because they involve the relative sparsities S_k. As we next show, however, they are indeed sharp in the sense that they reduce to the correct information-theoretic limits in several important cases. Furthermore, in the important case of wavelet sparsity with Fourier sampling, they can be used to provide near-optimal recovery guarantees. We discuss this in §6. Note, however, that to do this it is first necessary to generalize Theorem 4.4 to the infinite-dimensional setting, which we do in §5.

4.2.1 Sharpness of the estimates – the block-diagonal case

Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, and suppose for simplicity that $\mathbf{M} = \mathbf{N}$. Consider the block-diagonal matrix

$$A = A_1 \oplus \ldots \oplus A_r \in \mathbb{C}^{N \times N}, \quad A_k \in \mathbb{C}^{(N_k - N_{k-1}) \times (N_k - N_{k-1})}, \quad A_k^* A_k = I,$$

where $N_0 = 0$. Note that in this setting we have $S_k = s_k$, $\mu_{\mathbf{N},\mathbf{M}}(k,l) = 0$, $k \neq l$. Also, since $\mu(\mathbf{N},\mathbf{M})(k,k) = \mu(A_k)$, equations (4.4) and (4.5) reduce to

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \log(\epsilon^{-1}) \cdot \mu(A_k) \cdot s_k \cdot \log(N), \quad 1 \gtrsim \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \mu(A_k) \cdot s_k.$$

In particular, it suffices to take

$$m_k \gtrsim (N_k - N_{k-1}) \cdot \log(\epsilon^{-1}) \cdot \mu(A_k) \cdot s_k \cdot \log(N), \quad 1 \leq k \leq r. \tag{4.6}$$

This is exactly as one expects: the number of measurements in the k^{th} level depends on the size of the level multiplied by the local coherence and the sparsity in that level. Note that this result recovers the standard one-level results in finite dimensions [1, 12] up to a slight deterioration in the probability bound to $1 - s\epsilon$. Specifically, the usual bound would be $1 - \epsilon$. The question as to whether or not this s can be removed in the multilevel setting is open, although such a result would be more of a cosmetic improvement.

4.2.2 Sharpness of the estimates – the non-block diagonal case

The previous argument demonstrated that Theorem 4.4 is sharp, up to the probability term, in the sense that it reduces to the usual estimate (4.6) for block-diagonal matrices, i.e. $S_k = s_k$. This is not true in the general setting. Clearly, $S_k \leq s = s_1 + \ldots + s_r$. However in general there is usually *interference* between different sparsity levels, which means that S_k need not have anything to do with s_k, or can indeed be proportional to the total sparsity s. This may seem an undesirable aspect of the theorems, since S_k may be significantly larger than s_k, and thus the estimate on the number of measurements m_k required in the k^{th} level may also be much larger than the corresponding sparsity s_k. Could it therefore be that the S_ks are an unfortunate artefact of the proof? As we now show by example, this is not the case.

Let $N = rn$ for some $n \in \mathbb{N}$ and $\mathbf{N} = \mathbf{M} = (n, 2n, \ldots, rn)$. Let $W \in \mathbb{C}^{n\times n}$ and $V \in \mathbb{C}^{r\times r}$ be isometries and consider the matrix

$$A = V \otimes W,$$

where $\otimes$ is the usual Kronecker product. Note that $A \in \mathbb{C}^{N\times N}$ is also an isometry. Now suppose that $x = (x_1, \ldots, x_r) \in \mathbb{C}^N$ is an $(\mathbf{s},\mathbf{M})$-sparse vector, where each $x_k \in \mathbb{C}^n$ is s_k-sparse. Then $Ax = y$, $\quad y = (y_1, \ldots, y_r)$, $y_k = W z_k$, $z_k = \sum_{l=1}^{r} v_{kl} x_l$. Hence the problem of recovering x from measurements y with an $(\mathbf{N},\mathbf{m})$-multilevel strategy decouples into r problems of recovering the vector z_k from the measurements $y_k = W z_k$, $k = 1, \ldots, r$. Let $\hat{s}_k$ denote the sparsity of z_k. Since the coherence provides an information-theoretic limit [12], one requires at least

$$m_k \gtrsim n \cdot \mu(W) \cdot \hat{s}_k \cdot \log(n), \quad 1 \leq k \leq r. \tag{4.7}$$

measurements at level k in order to recover each z_k, and therefore recover x, regardless of the reconstruction method used. We now consider two examples of this setup:

Example 4.1 Let $\pi : \{1, \ldots, r\} \to \{1, \ldots, r\}$ be a permutation and let V be the matrix with entries $v_{kl} = \delta_{l,\pi(k)}$. Since $z_k = x_{\pi(k)}$ in this case, the lower bound (4.7) reads

$$m_k \gtrsim n \cdot \mu(W) \cdot s_{\pi(k)} \cdot \log(n), \quad 1 \leq k \leq r. \tag{4.8}$$

Now consider Theorem 4.4 for this matrix. First, we note that $S_k = s_{\pi(k)}$. In particular, S_k is completely unrelated to s_k. Substituting this into Theorem 4.4 and noting that $\mu_{\mathbf{N},\mathbf{M}}(k,l) = \mu(W)\delta_{l,\pi(k)}$ in this case, we arrive at the condition $m_k \gtrsim n \cdot \mu(W) \cdot s_{\pi(k)} \cdot \big(\log(\epsilon^{-1}) + 1\big) \cdot \log(nr)$, which is equivalent to (4.8) provided $r \lesssim n$.

Example 4.2 Now suppose that V is the $r \times r$ DFT matrix. Suppose also that $s \leq n/r$ and that the x_k's have disjoint support sets, i.e. $\mathrm{supp}(x_k) \cap \mathrm{supp}(x_l) = \emptyset$, $k \neq l$. Then by construction, each z_k is s-sparse, and therefore the lower bound (4.7) reads $m_k \gtrsim n \cdot \mu(W) \cdot s \cdot \log n$, for $1 \leq k \leq r$. After a short argument, one finds that $s/r \leq S_k \leq s$ in this case. Hence, S_k is typically much larger than s_k. Moreover, after noting that $\mu_{\mathbf{N},\mathbf{M}}(k,l) = \frac{1}{r}\mu(W)$, we find that Theorem 4.4 gives the condition $m_k \gtrsim n \cdot \mu(W) \cdot s \cdot \big(\log(\epsilon^{-1}) + 1\big) \cdot \log(nr)$. Thus, Theorem 4.4 obtains the lower bound in this case as well.

4.2.3 Sparsity leads to pessimistic reconstruction guarantees

The flip test demonstrates that any sparsity-based theory of CS cannot describe the quality of the reconstructions seen in practice. To conclude this section, we now use the block-diagonal case to further emphasize the need for theorems that go beyond sparsity, such as Theorems 4.1 and 4.4. To see this, consider the block-diagonal matrix

$$U = U_1 \oplus \ldots \oplus U_r, \qquad U_k \in \mathbb{C}^{(N_k - N_{k-1})\times(N_k - N_{k-1})},$$

where each U_k is perfectly incoherent, i.e. $\mu(U_k) = (N_k - N_{k-1})^{-1}$, and suppose we take m_k measurements within each block U_k. Let $x \in \mathbb{C}^N$ be the signal we wish to recover, where $N = N_r$. The question is, how many samples $m = m_1 + \ldots + m_r$ do we require?

Suppose we assume that x is s-sparse, where $s \leq \min_{k=1,\ldots,r}\{N_k - N_{k-1}\}$. Given no further information about the sparsity structure, it is necessary to take $m_k \gtrsim s\log(N)$ measurements in each block, giving $m \gtrsim rs\log(N)$ in total. However, suppose now that x is known to be s_k-sparse within each level, i.e. $|\mathrm{supp}(x) \cap \{N_{k-1}+1,\ldots,N_k\}| = s_k$. Then we now require only $m_k \gtrsim s_k\log(N)$, and therefore $m \gtrsim s\log(N)$ total measurements. Thus, structured sparsity leads to a significant saving by a factor of r in the total number of measurements required.

5 Main theorems II: the infinite-dimensional case

Finite-dimensional CS is suitable in many cases. However, there are some important problems where it can lead to significant problems, since the underlying problem is continuous/analog. Discretization of the problem in order to produce a finite-dimensional, vector-space model can lead to substantial errors [1, 7, 16, 56], due to the phenomenon of model mismatch.

To address this issue, a theory of infinite-dimensional CS was introduced by Adcock & Hansen in [1], based on a new approach to classical sampling known as *generalized sampling* [2, 3, 4, 5, 6, 38]. We describe this theory next. Note that this infinite-dimensional CS model has also been advocated for and implemented in MRI by Guerquin–Kern, Häberlin, Pruessmann & Unser [33]. Note also that sampling theories such as generalized sampling and finite rate of innovation [61] are infinite-dimensional, and hence it is most natural that CS has an infinite-dimensional theory as well.

5.1 Infinite-dimensional CS

Suppose that $\mathcal{H}$ is a separable Hilbert space over $\mathbb{C}$, and let $\{\psi_j\}_{j\in\mathbb{N}}$ be an orthonormal basis on $\mathcal{H}$ (the sampling basis). Let $\{\varphi_j\}_{j\in\mathbb{N}}$ be an orthonormal system in $\mathcal{H}$ (the sparsity system), and suppose that

$$U = (u_{ij})_{i,j\in\mathbb{N}}, \qquad u_{ij} = \langle \varphi_j, \psi_i \rangle, \tag{5.1}$$

is an infinite matrix. We may consider U as an element of $\mathcal{B}(l^2(\mathbb{N}))$; the space of bounded operators on $l^2(\mathbb{N})$. As in the finite-dimensional case, U is an isometry, and we may define its coherence $\mu(U) \in (0,1]$ analogously to (2.1). We want to recover $f = \sum_{j\in\mathbb{N}} x_j\varphi_j \in \mathcal{H}$ from a small number of the measurements $\hat{f} = \{\hat{f}_j\}_{j\in\mathbb{N}}$, where $\hat{f}_j = \langle f, \psi_j\rangle$. To do this, we introduce a second parameter $N \in \mathbb{N}$, and let Ω be a randomly-chosen subset of indices $1,\ldots,N$ of size m. Unlike in finite dimensions, we now consider two cases. Suppose first that $P_M^{\perp}x = 0$, i.e. x has no tail. Then we solve

$$\inf_{\eta\in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } \|P_\Omega U P_M \eta - y\| \leq \delta, \tag{5.2}$$

where $y = P_\Omega \hat{f} + z$ and $z \in \mathrm{ran}(P_\Omega)$ is a noise vector satisfying $\|z\| \leq \delta$, and P_Ω is the projection operator corresponding to the index set Ω. In [1] it was proved that any solution to (5.2) recovers f exactly up to an error determined by $\sigma_{s,M}(f)$, provided N and m satisfy the so-called *weak balancing property* with respect to M and s (see Definition 5.1, as well as Remark 5.1 for a discussion), and provided

$$m \gtrsim \mu(U)\cdot N \cdot s \cdot \left(1 + \log(\epsilon^{-1})\right)\cdot \log\left(m^{-1}MN\sqrt{s}\right). \tag{5.3}$$

As in the finite-dimensional case, which turns out to be a corollary of this result, we find that m is on the order of the sparsity s whenever $\mu(U)$ is sufficiently small.

In practice, the condition $P_M^\perp x = 0$ is unrealistic. In the more general case, $P_M^\perp x \neq 0$, we solve the following problem:

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } \|P_\Omega U \eta - y\| \leq \delta. \tag{5.4}$$

In [1] it was shown that any solution of (5.4) recovers f exactly up to an error determined by $\sigma_{s,M}(f)$, provided N and m satisfy the so-called *strong balancing property* with respect to M and s (see Definition 5.1), and provided a bound similar to (5.3) holds, where the M is replaced by a slightly larger constant (we give the details in the next section in the more general setting of multilevel sampling). Note that (5.4) cannot be solved numerically, since it is infinite-dimensional. Therefore in practice we replace (5.4) by

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } \|P_\Omega U P_R \eta - y\| \leq \delta, \tag{5.5}$$

where R is taken sufficiently large. See [1] for more information.

5.2 Main theorems

We first require the definition of the so-called *balancing property* [1]:

Definition 5.1 (Balancing property). *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then $N \in \mathbb{N}$ and $K \geq 1$ satisfy the weak balancing property with respect to U, $M \in \mathbb{N}$ and $s \in \mathbb{N}$ if*

$$\|P_M U^* P_N U P_M - P_M\|_{l^\infty \to l^\infty} \leq \frac{1}{8}\left(\log_2^{1/2}\left(4\sqrt{s}KM\right)\right)^{-1}, \tag{5.6}$$

where $\|\cdot\|_{l^\infty \to l^\infty}$ is the norm on $\mathcal{B}(l^\infty(\mathbb{N}))$. We say that N and K satisfy the strong balancing property with respect to U, M and s if (5.6) holds, as well as

$$\|P_M^\perp U^* P_N U P_M\|_{l^\infty \to l^\infty} \leq \frac{1}{8}. \tag{5.7}$$

As in the previous section, we commence with the two-level case. Furthermore, to illustrate the differences between the weak/strong balancing property, we first consider the setting of (5.2):

Theorem 5.2. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a two-level sampling scheme, where $\mathbf{N} = (N_1, N_2)$ and $\mathbf{m} = (N_1, m_2)$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, M_2) \in \mathbb{N}^2$, $M_1 < M_2$, and $\mathbf{s} = (M_1, s_2) \in \mathbb{N}^2$, be any pair such that the following holds:*

(i) we have $\|P_{N_1}^\perp U P_{M_1}\| \leq \frac{\gamma}{\sqrt{M_1}}$ and $\gamma \leq s_2\sqrt{\mu_{N_1}}$ for some $\gamma \in (0, 2/5]$;

(ii) the parameters $N = N_2$, $K = (N_2 - N_1)/m_2$ satisfy the weak balancing property with respect to U, $M := M_2$ and $s := M_1 + s_2$;

(iii) for $\epsilon \in (0, e^{-1}]$, let

$$m_2 \gtrsim (N - N_1) \cdot \log(\epsilon^{-1}) \cdot \mu_{N_1} \cdot s_2 \cdot \log\left(KM\sqrt{s}\right).$$

Suppose that $P_{M_2}^\perp x = 0$ and let $\xi \in l^1(\mathbb{N})$ be a minimizer of (5.2) with $\delta = \tilde{\delta}\sqrt{K^{-1}}$. Then, with probability exceeding $1 - s\epsilon$, we have

$$\|\xi - x\| \leq C \cdot \left(\tilde{\delta} \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f)\right), \tag{5.8}$$

for some constant C, where $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is as in (3.1), and $L = 1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KM\sqrt{s})}$. If $m_2 = N - N_1$ then this holds with probability 1.

We next state a result for multilevel sampling in the more general setting of (5.4). For this, we require the following notation: $\tilde{M} = \min\{i \in \mathbb{N} : \max_{k \geq i} \|P_N U e_k\| \leq 1/(32K\sqrt{s})\}$, where N, s and K are as defined below.

Theorem 5.3. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds:*

(i) the parameters $N = N_r, K = \max_{k=1,\dots,r}\left\{\frac{N_k - N_{k-1}}{m_k}\right\}$*, satisfy the strong balancing property with respect to* U*,* $M := M_r$ *and* $s := s_1 + \ldots + s_r$*;*

(ii) for $\epsilon \in (0, e^{-1}]$ *and* $1 \le k \le r$*,*

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r} \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l\right) \cdot \log\left(K\tilde{M}\sqrt{s}\right),$$

(with $\mu_{\mathbf{N},\mathbf{M}}(k,r)$ *replaced by* $\mu_{\mathbf{N},\mathbf{M}}(k,\infty)$*) and* $m_k \gtrsim \hat{m}_k \cdot \log(\epsilon^{-1}) \cdot \log\left(K\tilde{M}\sqrt{s}\right)$*, where* $\hat{m}_k$ *satisfies (4.5).*

Suppose that $\xi \in l^1(\mathbb{N})$ *is a minimizer of (5.4) with* $\delta = \tilde{\delta}\sqrt{K^{-1}}$*. Then, with probability exceeding* $1 - s\epsilon$*,*

$$\|\xi - x\| \le C \cdot \left(\tilde{\delta} \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f)\right),$$

for some constant C*, where* $\sigma_{\mathbf{s},\mathbf{M}}(f)$ *is as in (3.1), and* $L = C \cdot \left(1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KM\sqrt{s})}\right)$*. If* $m_k = N_k - N_{k-1}$ *for* $1 \le k \le r$ *then this holds with probability* 1*.*

This theorem removes the condition in Theorem 5.2 that x has zero tail. Note that the price to pay is the $\tilde{M}$ in the logarithmic term rather than M ($\tilde{M} \ge M$ because of the balancing property). Observe that $\tilde{M}$ is finite, and in the case of Fourier sampling with wavelets, we have that $\tilde{M} = \mathcal{O}(KN)$ (see §6). Note that Theorem 5.2 has a strong form analogous to Theorem 5.3 which removes the tail condition. The only difference is the requirement of the strong, as opposed to the weak, balancing property, and the replacement of M by $\tilde{M}$ in the log factor. Similarly, Theorem 5.3 has a weak form involving a tail condition. For succinctness we do not state these.

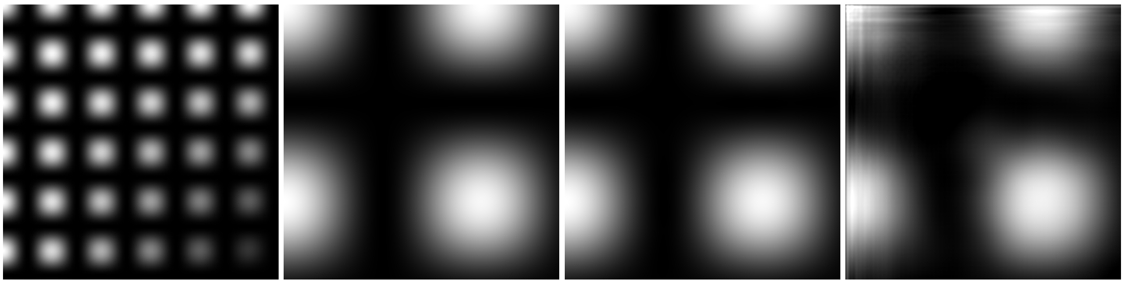
Remark 5.1 The balancing property is the main difference between the finite- and infinite-dimensional theorems. Its role is to ensure that the truncated matrix $P_N U P_M$ is close to an isometry. In reconstruction problems, the presence of an isometry ensures stability in the mapping between measurements and coefficients [2], which explains the need for a such a property in our theorems. As explained in [1], without the balancing property the lack of stability in this mapping leads to numerically useless reconstructions. Note that the balancing property is usually not satisfied for $N = M$. In general, one requires $N > M$ for the balancing property to hold. However, there is always a finite N for which it is satisfied, since the infinite matrix U is an isometry. For details we refer to [1]. We will provide specific estimates in §6 for the required magnitude of N in the case of Fourier sampling with wavelet sparsity.

5.3 The need for infinite-dimensional CS

As mentioned, infinite-dimensional CS is necessary to avoid the artefacts that are introduced when one applies finite-dimensional CS techniques to analog problems. To illustrate this, we consider the problem of recovering a smooth phantom, i.e. a C^∞ bivariate function, from its Fourier data. Note that this arises in both electron microscopy and spectroscopy. The test function is $f(x,y) = \cos^2(17\pi x/2)\cos^2(17\pi y/2)\exp(-x-y)$. In Figure 5, we compare finite-dimensional CS, based on solving (4.1) with $U = U_{\mathrm{dft}} V_{\mathrm{dwt}}^{-1}$ (discrete Fourier and wavelet transform respectively) with infinite-dimensional CS, which solves (5.5) with the Fourier basis $\{\psi_j\}_{j\in\mathbb{N}}$ and boundary wavelet basis $\{\varphi_j\}_{j\in\mathbb{N}}$. The improvement one gets is due to that fact that that the error in infinite-dimensional case is dominated by the wavelet approximation error, whereas in the finite-dimensional case (due mismatch between the continuous Fourier samples and the discrete Fourier transform) the error is dominated by the Fourier approximation error. As is well known [47], wavelet approximation is superior to Fourier approximation and depends on the number of vanishing moments of the wavelet used (DB4 in this case).

6 Recovery of wavelet coefficients from Fourier samples

As noted, Fourier sampling with wavelet sparsity is a important reconstruction problem in CS, with numerous applications ranging from medical imaging to seismology and interferometry. Here we consider the Fourier sampling basis $\{\psi_j\}_{j\in\mathbb{N}}$ and wavelet reconstruction basis $\{\varphi_j\}_{j\in\mathbb{N}}$ (see §7.4.1 for a formal definition) with the infinite matrix U as in (5.1). The incoherence properties can be described as follows.



Original | Original (zoomed) | Infinite-dim. CS (zoomed) Err 0.6% | Finite-dim. CS (zoomed) Err 12.7%

Figure 5: Subsampling 6.15%. Both reconstructions are based on identical sampling information.

Theorem 6.1. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be the matrix from (7.107) corresponding to the Fourier/wavelets system described in §7.4. Then $\mu(U) \geq \omega$, where ω is the sampling density, and $\mu(P_N^{\perp}U), \mu(UP_N^{\perp}) = \mathcal{O}\left(N^{-1}\right)$.*

Thus, Fourier sampling with wavelet sparsity is indeed globally coherent, yet asymptotically incoherent. This result holds for essentially any wavelet basis in one dimension (see [39] for the multidimensional case). To recover wavelet coefficients, we seek to appl a multilevel sampling strategy, which raises the question: how do we design this strategy, and how many measurements are required? If the levels $\mathbf{M} = (M_1, \ldots, M_r)$ correspond to the wavelet scales, and $\mathbf{s} = (s_1, \ldots, s_r)$ to the sparsities within them, then the best one could hope to achieve is that the number of measurements m_k in the k^{th} sampling level is proportional to the sparsity s_k in the corresponding sparsity level. Our main theorem below shows that multilevel sampling can achieve this, up to an exponentially-localized factor and the usual log terms.

Theorem 6.2. *Consider an orthonormal basis of compactly supported wavelets with a multiresolution analysis (MRA). Let Φ and Ψ denote the scaling function and mother wavelet respectively satisfying (7.100) with $\alpha \geq 1$. Suppose that Ψ has $v \geq 1$ vanishing moments, that the Fourier sampling density ω satisfies (7.105) and that the wavelets $\{\varphi_j\}$ are ordered according to (7.102). Let $f = \sum_{j=1}^{\infty} x_j \varphi_j$. Suppose that $\mathbf{M} = (M_1, \ldots, M_r)$ corresponds to wavelet scales with $M_k = \mathcal{O}\left(2^{R_k}\right)$ with $R_k \in \mathbb{N}$, $R_{k+1} = a + R_k$, $a \geq 1$, $k = 1, \ldots, r$ and $\mathbf{s} = (s_1, \ldots, s_r)$ corresponds to the sparsities within them. Let $\epsilon \in (0, e^{-1}]$ and let $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ be a multilevel sampling scheme such that the following holds:*

(i) The parameters $N = N_r$, $K = \max_{k=1,\ldots,r}, \{(N_k - N_{k-1})/m_k\}$, $M = M_r$, $s = s_1 + \ldots + s_r$ satisfy $N \gtrsim M^{1+1/(2\alpha-1)} \cdot (\log_2(4MK\sqrt{s}))^{1/(2\alpha-1)}$. Alternatively, if Φ and Ψ satisfy the slightly stronger Fourier decay property (7.101), then $N \gtrsim M \cdot (\log_2(4KM\sqrt{s}))^{1/(4\alpha-2)}$.

(ii) For each $k = 1, \ldots, r-1$, $N_k = 2^{R_k}\omega^{-1}$ and for each $k = 1, \ldots, r$,

$$m_k \gtrsim \log(\epsilon^{-1}) \cdot \log(\tilde{N}) \cdot \frac{N_k - N_{k-1}}{N_{k-1}} \cdot \left(\hat{s}_k + \sum_{l=1}^{k-2} s_l \cdot 2^{-(\alpha-1/2)A_{k,l}} + \sum_{l=k+2}^{r} s_l \cdot 2^{-vB_{k,l}} \right), \quad (6.1)$$

where $A_{k,l} = R_{k-1} - R_l$, $B_{k,l} = R_{l-1} - R_k$, $\tilde{N} = (K\sqrt{s})^{1+1/v}N$ and $\hat{s}_k = \max\{s_{k-1}, s_k, s_{k+1}\}$ (see Remark 6.1).

Then, with probability exceeding $1 - s\epsilon$, any minimizer $\xi \in l^1(\mathbb{N})$ of (5.4) with $\delta = \tilde{\delta}\sqrt{K^{-1}}$ satisfies

$$\|\xi - x\| \leq C \cdot \left(\tilde{\delta} \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f) \right),$$

for some constant C, where $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is as in (3.1), and $L = C \cdot \left(1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KM\sqrt{s})}\right)$. If $m_k = N_k - N_{k-1}$ for $1 \leq k \leq r$ then this holds with probability 1.

Remark 6.1 To avoid cluttered notation we have abused notation slightly in (ii) of Theorem 6.2. In particular, we interpret $s_0 = 0$, $\frac{N_k - N_{k-1}}{N_{k-1}} = N_1$ for $k = 1$, and $\sum_{l=1}^{k-2} s_l \cdot 2^{-(\alpha-1/2)A_{k,l}} = 0$ when $k \leq 2$.

Original image | Random Bernoulli Err = 15.7% | Multilevel Hadamard Err = 9.6% | Multilevel Fourier Err 8.7%

Figure 6: 12.5% subsampling at 256×256 resolution using DB4 wavelets and various different measurements.

This theorem provides the first comprehensive explanation for the observed success of CS in applications based on the Fourier/wavelets model. To see why, note that the key estimate (6.1) shows that m_k need only scale as a linear combination of the local sparsities s_l, $1 \leq l \leq r$, and critically, the dependence of the sparsities s_l for $l \neq k$ is exponentially diminishing in $|k-l|$. Note that the presence of the off-diagonal terms is due to the previously-discussed phenomenon of interference, which occurs since the Fourier/wavelets system is not exactly block diagonal. Nonetheless, the system is nearly block-diagonal, and this results in the near-optimality seen in (6.1).

Observe that (6.1) is in agreement with the flip test: if the local sparsities s_k change, then the subsampling factors m_k must also change to ensure the same quality reconstruction. Having said that, it is straightforward to deduce from (6.1) the following global sparsity bound:

$$m \gtrsim s \cdot \log(\epsilon^{-1}) \cdot \log(\tilde{N}),$$

where $m = m_1 + \ldots + m_r$ is the total number of measurements and $s = s_1 + \ldots + s_r$ is the total sparsity. Note in particular the optimal exponent in the log factor.

Remark 6.2 The Fourier/wavelets recovery problem was studied by Candès & Romberg in [13]. Their result shows that if, in an ideal setting, an image can be first separated into separate wavelet subbands before sampling, then it can be recovered using approximately s_k measurements (up to a log factor) in each sampling band. Unfortunately, such separation into separate wavelet subbands before sampling is infeasible in most practical situations. Theorem 6.2 improves on this result by removing this substantial restriction, with the sole penalty being the slightly worse bound (6.1).

Note also that a recovery result for bivariate Haar wavelets, as well as the related technique of TV minimization, was given in [40]. Similarly [8] analyzes block sampling strategies with application to MRI. However, these results are based on sparsity, and therefore they do not explain how the sampling strategy will depend on the signal structure.

6.1 Universality and RIP or structure?

Theorem 6.2 explains the success of CS when one is constrained to acquire Fourier measurements. Yet, due primarily to the their high global coherence with wavelets, Fourier measurements are often viewed as suboptimal for CS. If one had complete freedom to choose the measurements, and no physical constraints (such as are always present in MRI, for example), then standard CS intuition would suggest random Gaussian or Bernoulli measurements, since they are universal and satisfy the RIP.

However, in reality such measurements are actually highly suboptimal in the presence of structured sparsity. This is demonstrated in Figure 6, where an image is recovered from $m = 8192$ measurements taken either as random Bernoulli or multilevel Hadamard or Fourier. As is evident, the latter gives an error that is almost 50% smaller. The reason for this improvement is that whilst Fourier or Hadamard measurements are highly coherent with wavelets, they are asymptotically incoherent, and this can be exploited through multilevel random subsampling to recover asymptotically sparse wavelet coefficients. Random Gaussian/Bernoulli measurements on the other hand cannot take advantage of this structure since they satisfy an RIP.

This observation is an important consequence of our theory. In conclusion, whenever structured sparsity is present (such is the case in the majority of imaging applications, for example) there are substantial improvements to be gained by designing the measurements according to not just the sparsity, but also the additional structure. For a more comprehensive discussion see [53], see also [15, 62].

7 Proofs

The proofs rely on some key propositions from which one can deduce the main theorems. The main work is to prove these proposition, and that will be done subsequently.

7.1 Key results

Proposition 7.1. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ and suppose that Δ and $\Omega = \Omega_1 \cup \ldots \cup \Omega_r$ (where the union is disjoint) are subsets of $\mathbb{N}$. Let $x_0 \in \mathcal{H}$ and $z \in \mathrm{ran}(P_\Omega U)$ be such that $\|z\| \leq \delta$ for $\delta \geq 0$. Let $M \in \mathbb{N}$ and $y = P_\Omega U x_0 + z$ and $y_M = P_\Omega U P_M x_0 + z$. Suppose that $\xi \in \mathcal{H}$ and $\xi_M \in \mathcal{H}$ satisfiy*

$$\|\xi\|_{l^1} = \inf_{\eta \in \mathcal{H}} \{\|\eta\|_{l^1} : \|P_\Omega U \eta - y\| \leq \delta\}. \tag{7.1}$$

$$\|\xi_M\|_{l^1} = \inf_{\eta \in \mathbb{C}^M} \{\|\eta\|_{l^1} : \|P_\Omega U P_M \eta - y_M\| \leq \delta\}. \tag{7.2}$$

If there exists a vector $\rho = U^* P_\Omega w$ such that

(i) $\|P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right) U P_\Delta - I_\Delta\| \leq \frac{1}{4}$

(ii) $\max_{i \in \Delta^c} \| \left(q_1^{-1/2} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1/2} P_{\Omega_r}\right) U e_i\| \leq \sqrt{\frac{5}{4}}$

(iii) $\|P_\Delta \rho - \mathrm{sgn}(P_\Delta x_0)\| \leq \frac{q}{8}$.

(iv) $\|P_\Delta^\perp \rho\|_{l^\infty} \leq \frac{1}{2}$

(v) $\|w\| \leq L \cdot \sqrt{|\Delta|}$

for some $L > 0$ and $0 < q_k \leq 1$, $k = 1, \ldots, r$, then we have that

$$\|\xi - x_0\| \leq C \cdot \left(\delta \cdot \left(\frac{1}{\sqrt{q}} + L\sqrt{s}\right) + \|P_\Delta^\perp x_0\|_{l^1}\right),$$

for some constant C, where $s = |\Delta|$ and $q = \min\{q_k\}_{k=1}^r$. Also, if (ii) is replaced by

$$\max_{i \in \{1,\ldots,M\} \cap \Delta^c} \| \left(q_1^{-1/2} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1/2} P_{\Omega_r}\right) U e_i\| \leq \sqrt{\frac{5}{4}}$$

and (iv) is replaced by $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq \frac{1}{2}$ then

$$\|\xi_M - x_0\| \leq C \cdot \left(\delta \cdot \left(\frac{1}{\sqrt{q}} + L\sqrt{s}\right) + \|P_M P_\Delta^\perp x_0\|_{l^1}\right). \tag{7.3}$$

Proof. First observe that (i) implies that $(P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right) U P_\Delta|_{P_\Delta(\mathcal{H})})^{-1}$ exists and

$$\|(P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right) U P_\Delta|_{P_\Delta(\mathcal{H})})^{-1}\| \leq \frac{4}{3}. \tag{7.4}$$

Also, (i) implies that

$$\| \left(q_1^{-1/2} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1/2} P_{\Omega_r}\right) U P_\Delta\|^2 = \|P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right) U P_\Delta\| \leq \frac{5}{4}, \tag{7.5}$$

and

$$
\begin{aligned}
&\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)\|^2 = \|\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta\|^2\\
&= \sup_{\|\eta\|=1}\|\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta\eta\|^2\\
&= \sup_{\|\eta\|=1}\sum_{k=1}^{r}\|q_k^{-1}P_{\Omega_k}UP_\Delta\eta\|^2 \leq \frac{1}{q}\sup_{\|\eta\|=1}\sum_{k=1}^{r}q_k^{-1}\|P_{\Omega_k}UP_\Delta\eta\|^2, \quad \frac{1}{q}=\max_{1\leq k\leq r}\{\frac{1}{q_k}\}\\
&= \frac{1}{q}\sup_{\|\eta\|=1}\langle P_\Delta U^*\left(\sum_{k=1}^{r}q_k^{-1}P_{\Omega_k}\right)UP_\Delta\eta,\eta\rangle \leq \frac{1}{q}\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta\|.
\end{aligned}
\tag{7.6}
$$

Thus, (7.5) and (7.6) imply

$$
\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)\| \leq \sqrt{\frac{5}{4q}}. \tag{7.7}
$$

Suppose that there exists a vector ρ, constructed with $y_0 = P_\Delta x_0$, satisfying (iii)-(v). Let ξ be a solution to (7.1) and let $h = \xi - x_0$. Let $A_\Delta = P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta|_{P_\Delta(\mathcal{H})}$. Then, it follows from (ii) and observations (7.4), (7.5), (7.7) that

$$
\begin{aligned}
&\|P_\Delta h\| = \|A_\Delta^{-1}A_\Delta P_\Delta h\|\\
&\leq \|A_\Delta^{-1}\|\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)U(I-P_\Delta^\perp)h\|\\
&\leq \frac{4}{3}\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)\|\|P_\Omega Uh\|\\
&+\frac{4}{3}\max_{i\in\Delta^c}\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)Ue_i\|\|P_\Delta^\perp h\|_{l^1}\\
&\leq \frac{4}{3}\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r}\right)\|\|P_\Omega Uh\|\\
&+\frac{4}{3}\left\|P_\Delta U^*\left(q_1^{-1/2}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1/2}\right)\right\|\max_{i\in\Delta^c}\left\|\left(q_1^{-1/2}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1/2}P_{\Omega_r}\right)Ue_i\right\|\|P_\Delta^\perp h\|_{l^1}\\
&\leq \frac{4\sqrt{5}}{3\sqrt{q}}\delta+\frac{5}{3}\|P_\Delta^\perp h\|_{l^1},
\end{aligned}
\tag{7.8}
$$

where in the final step we use $\|P_\Omega Uh\| \leq \|P_\Omega U\zeta - y\| + \|z\| \leq 2\delta$. We will now obtain a bound for $\|P_\Delta^\perp h\|_{l^1}$. First note that

$$
\begin{aligned}
\|h+x_0\|_{l^1} &= \|P_\Delta h + P_\Delta x_0\|_{l^1} + \|P_\Delta^\perp(h+x_0)\|_{l^1}\\
&\geq \mathrm{Re}\,\langle P_\Delta h, \mathrm{sgn}(P_\Delta x_0)\rangle + \|P_\Delta x_0\|_{l^1} + \|P_\Delta^\perp h\|_{l^1} - \|P_\Delta^\perp x_0\|_{l^1}\\
&\geq \mathrm{Re}\,\langle P_\Delta h, \mathrm{sgn}(P_\Delta x_0)\rangle + \|x_0\|_{l^1} + \|P_\Delta^\perp h\|_{l^1} - 2\|P_\Delta^\perp x_0\|_{l^1}.
\end{aligned}
\tag{7.9}
$$

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

$$
\|P_\Delta^\perp h\|_{l^1} \leq |\langle P_\Delta h, \mathrm{sgn}(P_\Delta x_0)\rangle| + 2\|P_\Delta^\perp x_0\|_{l^1}. \tag{7.10}
$$

We will use this equation later on in the proof, but before we do that observe that some basic adding and subtracting yields

$$
\begin{aligned}
|\langle P_\Delta h, \mathrm{sgn}(x_0)\rangle| &\leq |\langle P_\Delta h, \mathrm{sgn}(P_\Delta x_0) - P_\Delta\rho\rangle| + |\langle h,\rho\rangle| + \left|\langle P_\Delta^\perp h, P_\Delta^\perp\rho\rangle\right|\\
&\leq \|P_\Delta h\|\|\mathrm{sgn}(P_\Delta x_0) - P_\Delta\rho\| + |\langle P_\Omega Uh, w\rangle| + \|P_\Delta^\perp h\|_{l^1}\|P_\Delta^\perp\rho\|_{l^\infty}\\
&\leq \frac{q}{8}\|P_\Delta h\| + 2L\delta\sqrt{s} + \frac{1}{2}\|P_\Delta^\perp h\|_{l^1}\\
&\leq \frac{\sqrt{5q}}{6}\delta + \frac{5q}{24}\|P_\Delta^\perp h\|_{l^1} + 2L\delta\sqrt{s} + \frac{1}{2}\|P_\Delta^\perp h\|_{l^1}
\end{aligned}
\tag{7.11}
$$

where the last inequality utilises (7.8) and the penultimate inequality follows from properties (iii), (iv) and (v) of the dual vector ρ. Combining this with (7.10) and the fact that $q \leq 1$ gives that

$$
\|P_\Delta^\perp h\|_{l^1} \leq \delta\left(\frac{4\sqrt{5q}}{3} + 8L\sqrt{s}\right) + 8\|P_\Delta^\perp x_0\|_{l^1}. \tag{7.12}
$$

Thus, (7.8) and (7.12) yields:

$$\|h\| \leq \|P_\Delta h\| + \left\|P_\Delta^\perp h\right\| \leq \frac{8}{3}\|P_\Delta^\perp h\|_{l^1} + \frac{4\sqrt{5}}{3\sqrt{q}}\delta \leq \left(8\sqrt{q} + 22L\sqrt{s} + \frac{3}{\sqrt{q}}\right) \cdot \delta + 22\left\|P_\Delta^\perp x_0\right\|_{l^1}. \quad (7.13)$$

The proof of the second part of this proposition follows the proof as outlined above and we omit the details. $\square$

The next two propositions give sufficient conditions for Proposition 7.1 to be true. But before we state them we need to define the following.

Definition 7.2. *Let U be an isometry of either $\mathbb{C}^{N\times N}$ or $\mathcal{B}(l^2(\mathbb{N}))$. For $\mathbf{N} = (N_1,\ldots,N_r) \in \mathbb{N}^r$, $\mathbf{M} = (M_1,\ldots,M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$ and $1 \leq M_1 < \ldots < M_r$, $\mathbf{s} = (s_1,\ldots,s_r) \in \mathbb{N}^r$ and $1 \leq k \leq r$, let*

$$\kappa_{\mathbf{N},\mathbf{M}}(k,l) = \max_{\eta\in\Theta} \|P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}} \eta\|_{l^\infty} \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U)}.$$

where

$$\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, |\mathrm{supp}(P_{M_l}^{M_{l-1}}\eta)| = s_l,\ l = 1,\ldots,r-1,\ |\mathrm{supp}(P_{M_{r-1}}^{\perp}\eta)| = s_r, \},$$

and $N_0 = M_0 = 0$. We also define

$$\kappa_{\mathbf{N},\mathbf{M}}(k,\infty) = \max_{\eta\in\Theta} \|P_{N_k}^{N_{k-1}} U P_{M_{r-1}}^{\perp} \eta\|_{l^\infty} \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U)}.$$

Proposition 7.3. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1,\ldots,N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1,\ldots,m_r) \in \mathbb{N}^r$. Let $(\mathbf{s},\mathbf{M})$, where $\mathbf{M} = (M_1,\ldots,M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1,\ldots,s_r) \in \mathbb{N}^r$, be any pair such that the following holds:*

(i) The parameters $N := N_r$, and $K := \max_{k=1,\ldots,r}(N_k - N_{k-1})/m_k$, satisfy the weak balancing property with respect to U, $M := M_r$ and $s := s_1 + \ldots + s_r$;

(ii) for $\epsilon > 0$ and $1 \leq k \leq r$,

$$1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N_k - N_{k-1}}{m_k} \cdot \left(\sum_{l=1}^{r} \kappa_{\mathbf{N},\mathbf{M}}(k,l)\right) \cdot \log\left(KM\sqrt{s}\right), \quad (7.14)$$

(iii)

$$m_k \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_k \cdot \log\left(KM\sqrt{s}\right), \quad (7.15)$$

where $\hat{m}_k$ satisfies

$$1 \gtrsim \sum_{k=1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k, \qquad \forall\, l = 1,\ldots,r,$$

where $\tilde{s}_1 + \ldots + \tilde{s}_r \leq s_1 + \ldots + s_r$, $\tilde{s}_k \leq S_k(s_1,\ldots,s_r)$ and S_k is defined in (4.3).

Then (i)-(v) in Proposition 7.1 follow with probability exceeding $1-\epsilon$, with (ii) replaced by

$$\max_{i\in\{1,\ldots,M\}\cap\Delta^c} \|\left(q_1^{-1/2} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1/2} P_{\Omega_r}\right) U e_i\| \leq \sqrt{\frac{5}{4}}, \quad (7.16)$$

(iv) replaced by $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq \frac{1}{2}$ and L in (v) is given by

$$L = C \cdot \sqrt{K} \cdot \left(1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KM\sqrt{s})}\right). \quad (7.17)$$

If $m_k = N_k - N_{k-1}$ for all $1 \leq k \leq r$ then (i)-(v) follow with probability one (with the alterations suggested above).

Proposition 7.4. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds:*

(i) The parameters N and K (as in Proposition 7.3) satisfy the strong balancing property with respect to U, $M = M_r$ and $s := s_1 + \ldots + s_r$;

(ii) for $\epsilon > 0$ and $1 \leq k \leq r$,

$$1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N_k - N_{k-1}}{m_k} \cdot \left(\kappa_{\mathbf{N},\mathbf{M}}(k, \infty) + \sum_{l=1}^{r-1} \kappa_{\mathbf{N},\mathbf{M}}(k, l) \right) \cdot \log\left(K\tilde{M}\sqrt{s} \right), \quad (7.18)$$

(iii)

$$m_k \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_k \cdot \log\left(K\tilde{M}\sqrt{s} \right), \quad (7.19)$$

where $\tilde{M} = \min\{i \in \mathbb{N} : \| \max_{j \geq i} P_N U P_{\{j\}} \| \leq 1/(K32\sqrt{s})\}$, and $\hat{m}_k$ is as in Proposition 7.3.

Then (i)-(v) in Proposition 7.1 follow with probability exceeding $1-\epsilon$ with L as in (7.17). If $m_k = N_k - N_{k-1}$ for all $1 \leq k \leq r$ then (i)-(v) follow with probability one.

Lemma 7.5 (Bounds for $\kappa_{\mathbf{N},\mathbf{M}}(k,l)$)**.** *For $k, l = 1, \ldots, r$*

$$\kappa_{\mathbf{N},\mathbf{M}}(k, l) \leq \min \left\{ \mu_{\mathbf{N},\mathbf{M}}(k, l) \cdot s_l, \ \sqrt{s_l \cdot \mu(P_{N_k}^{N_{k-1}} U)} \cdot \left\| P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}} \right\| \right\}. \quad (7.20)$$

Also, for $k = 1, \ldots, r$

$$\kappa_{\mathbf{N},\mathbf{M}}(k, \infty) \leq \min \left\{ \mu_{\mathbf{N},\mathbf{M}}(k, \infty) \cdot s_r, \ \sqrt{s_r \cdot \mu(P_{N_k}^{N_{k-1}} U)} \cdot \left\| P_{N_k}^{N_{k-1}} U P_{M_{r-1}}^{\perp} \right\| \right\}. \quad (7.21)$$

Proof. For $k, l = 1, \ldots, r$

$$\begin{aligned}
\kappa_{\mathbf{N},\mathbf{M}}(k, l) &= \max_{\eta \in \Theta} \| P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}} \eta \|_{l^\infty} \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U)} \\
&= \max_{\eta \in \Theta} \max_{N_{k-1} < i \leq N_k} \left| \sum_{M_{l-1} < j \leq M_l} \eta_j u_{ij} \right| \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U)} \\
&\leq s_l \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}})} \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U)} \leq s_l \cdot \mu_{\mathbf{N},\mathbf{M}}(k, l)
\end{aligned}$$

since $|u_{ij}| \leq 1$, and similarly,

$$\begin{aligned}
\kappa_{\mathbf{N},\mathbf{M}}(k, \infty) &= \max_{\eta \in \Theta} \| P_{N_k}^{N_{k-1}} U P_{M_{r-1}}^{\perp} \eta \|_{l^\infty} \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U)} \\
&= \max_{\eta \in \Theta} \max_{N_{k-1} < i \leq N_k} \left| \sum_{M_{r-1} < j} \eta_j u_{ij} \right| \cdot \sqrt{\mu(P_{N_k}^{N_{k-1}} U)} \leq s_r \cdot \mu_{\mathbf{N},\mathbf{M}}(k, \infty).
\end{aligned}$$

Finally, it is straightforward to show that for $k, l = 1, \ldots, r$,

$$\kappa_{\mathbf{N},\mathbf{M}}(k, l) \leq \sqrt{s_l} \cdot \left\| P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}} \right\| \sqrt{\mu(P_{N_k}^{N_{k-1}} U)}$$

and

$$\kappa_{\mathbf{N},\mathbf{M}}(k, \infty) \leq \sqrt{s_r} \cdot \left\| P_{N_k}^{N_{k-1}} U P_{M_{r-1}}^{\perp} \right\| \sqrt{\mu(P_{N_k}^{N_{k-1}} U)}.$$

□

We are now ready to prove the main theorems.

Proof of Theorems 4.1 and 5.2. It is clear that Theorem 4.1 follows from Theorem 5.2, thus it remains to prove the latter. We will apply Proposition 7.3 to a two-level sampling scheme $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$, where $\mathbf{N} = (N_1, N_2)$ and $\mathbf{m} = (m_1, m_2)$ with $m_1 = N_1$ and $m_2 = m$. Also, consider $(\mathbf{s}, \mathbf{M})$, where $\mathbf{s} = (M_1, s_2)$, $\mathbf{M} = (M_1, M_2)$. Thus, if $N_1, N_2, m_1, m_2 \in \mathbb{N}$ are such that

$$N = N_2, \quad K = \max\left\{\frac{N_2 - N_1}{m_2}, \frac{N_1}{m_1}\right\}$$

satisfy the weak balancing property with respect to U, $M = M_2$ and $s = M_1 + s_2$, we have that (i) - (v) in Proposition 7.1 follow with probability exceeding $1 - s\epsilon$, with (ii) replaced by

$$\max_{i \in \{1,\ldots,M\} \cap \Delta^c} \| \left(P_{N_1} \oplus \frac{N_2 - N_1}{m_2} P_{\Omega_2}\right) U e_i \| \leq \sqrt{\frac{5}{4}},$$

(iv) replaced by $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq \frac{1}{2}$ and L in (v) is given by (7.17), if

$$1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N - N_1}{m_2} \cdot (\kappa_{\mathbf{N},\mathbf{M}}(2,1) + \kappa_{\mathbf{N},\mathbf{M}}(2,2)) \cdot \log\left(KM\sqrt{s}\right), \tag{7.22}$$

$$m_2 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_2 \cdot \log\left(KM\sqrt{s}\right), \tag{7.23}$$

where $\hat{m}_2$ satisfies $1 \gtrsim ((N_2 - N_1)/\hat{m}_2 - 1) \cdot \mu_{N_1} \cdot \tilde{s}_2$, and $\tilde{s}_2 \leq S_2$ (recall S_2 from Definition 4.3). Recall from (7.20) that

$$\kappa_{\mathbf{N},\mathbf{M}}(2,1) \leq \sqrt{s_1 \cdot \mu_{N_1}} \cdot \left\|P_{N_1}^\perp U P_{M_1}\right\|, \quad \kappa_{\mathbf{N},\mathbf{M}}(2,2) \leq s_2 \cdot \mu_{N_1}.$$

Also, it follows directly from Definition 4.3 that

$$S_2 \leq \left(\left\|P_{N_1}^\perp U P_{M_1}\right\| \cdot \sqrt{M_1} + \sqrt{s_2}\right)^2.$$

Thus, provided that $\left\|P_{N_1}^\perp U P_{M_1}\right\| \leq \gamma/\sqrt{M_1}$ where γ is as in (i) of Theorem 5.2, we observe that (iii) of Theorem 5.2 implies (7.22) and (7.23). Thus, the theorem now follows from Proposition 7.1. $\square$

Proof of Theorem 4.4 and Theorem 5.3. It is straightforward that Theorem 4.4 follows from Theorem 5.3. Now, recall from Lemma 7.20 that

$$\kappa_{\mathbf{N},\mathbf{M}}(k,l) \leq s_l \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l), \quad \kappa_{\mathbf{N},\mathbf{M}}(k,\infty) \leq s_r \cdot \mu_{\mathbf{N},\mathbf{M}}(k,\infty), \qquad k,l = 1,\ldots,r.$$

Thus, a direct application of Proposition 7.4 and Proposition 7.1 completes the proof.

$\square$

It remains now to prove Propositions 7.3 and 7.4. This is the content of the next sections.

7.2 Preliminaries

Before we commence on the rather length proof of these propositions, let us recall one of the monumental results in probability theory that will be of greater use later on.

Theorem 7.6. *(Talagrand [58, 43]) There exists a number K with the following property. Consider n independent random variables X_i valued in a measurable space Ω and let $\mathcal{F}$ be a (countable) class of measurable functions on Ω. Let Z be the random variable $Z = \sup_{f\in\mathcal{F}} \sum_{i\leq n} f(X_i)$ and define*

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

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

Note that this version of Talagrand's theorem is found in [43, Cor. 7.8]. We next present a theorem and several technical propositions that will serve as the main tools in our proofs of Propositions 7.3 and 7.4. A crucial tool herein is the Bernoulli sampling model. We will use the notation $\{a, \dots, b\} \supset \Omega \sim \mathrm{Ber}(q)$, where $a < b$ $a, b \in \mathbb{N}$, when Ω is given by $\Omega = \{k : \delta_k = 1\}$ and $\{\delta_k\}_{k=1}^N$ is a sequence of Bernoulli variables with $\mathbb{P}(\delta_k = 1) = q$.

Definition 7.7. *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots < N_r$, $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \dots, r$, and suppose that*

$$\Omega_k \subseteq \{N_{k-1}+1, \dots, N_k\}, \quad \Omega_k \sim \mathrm{Ber}\left(\frac{m_k}{N_k - N_{k-1}}\right), \quad k = 1, \dots, r,$$

where $N_0 = 0$. We refer to the set $\Omega = \Omega_{\mathbf{N},\mathbf{m}} := \Omega_1 \cup \dots \cup \Omega_r$. as an $(\mathbf{N}, \mathbf{m})$-multilevel Bernoulli sampling scheme.

Theorem 7.8. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel Bernoulli sampling scheme, where $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$. Consider $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$, $M_1 < \dots < M_r$, and $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$, and let*

$$\Delta = \Delta_1 \cup \dots \cup \Delta_r, \qquad \Delta_k \subset \{M_{k-1}+1, \dots, M_k\}, \qquad |\Delta_k| = s_k$$

where $M_0 = 0$. If $\|P_{M_r} U^* P_{N_r} U P_{M_r} - P_{M_r}\| \leq 1/8$ then, for $\gamma \in (0, 1)$,

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1} P_{\Omega_1} \oplus \dots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta - P_\Delta\| \geq 1/4) \leq \gamma, \tag{7.24}$$

where $q_k = m_k/(N_k - N_{k-1})$, provided that

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \left(\sum_{l=1}^r \kappa_{\mathbf{N},\mathbf{M}}(k, l)\right) \cdot \left(\log\left(\gamma^{-1} s\right) + 1\right). \tag{7.25}$$

In addition, if $q = \min\{q_k\}_{k=1}^r = 1$ then

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1} P_{\Omega_1} \oplus \dots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta - P_\Delta\| \geq 1/4) = 0.$$

In proving this theorem we deliberately avoid the use of the Matrix Bernstein inequality [32], as Talagrand's theorem is more convenient for our infinite-dimensional setting. Before we can prove this theorem, we need the following technical lemma.

Lemma 7.9. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ with $\|U\| \leq 1$, and consider the setup in Theorem 7.8. Let $N = N_r$ and let $\{\delta_j\}_{j=1}^N$ be independent random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \tilde{q}_j$, $\tilde{q}_j = m_k/(N_k - N_{k-1})$ and $j \in \{N_{k-1}+1, \dots, N_k\}$, and define $Z = \sum_{j=1}^N Z_j$, $Z_j = \left(\tilde{q}_j^{-1}\delta_j - 1\right) \eta_j \otimes \bar{\eta}_j$ and $\eta_j = P_\Delta U^* e_j$. Then*

$$\mathbb{E}\left(\|Z\|\right)^2 \leq 48 \max\{\log(|\Delta|), 1\} \max_{1 \leq j \leq N} \left\{\tilde{q}_j^{-1} \|\eta_j\|^2\right\},$$

when $(\max\{\log(|\Delta|), 1\})^{-1} \geq 18 \max_{1 \leq j \leq N} \left\{\tilde{q}_j^{-1} \|\eta_j\|^2\right\}$.

The proof of this lemma involves essentially reworking an argument due to Rudelson [55], and is similar to arguments given previously in [1] (see also [13]). We include it here for completeness as the setup deviates slightly. We shall also require the following result:

Lemma 7.10. *(Rudelson) Let $\eta_1, \dots, \eta_M \in \mathbb{C}^n$ and let $\varepsilon_1, \dots \varepsilon_M$ be independent Bernoulli variables taking values $1, -1$ with probability $1/2$. Then*

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

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

Lemma 7.10 is often referred to as Rudelson's Lemma [55]. However, we use the above complex version that was proven by Tropp [59, Lem. 22].

Proof of Lemma 7.9. We commence by letting $\tilde{\delta} = \{\tilde{\delta}_j\}_{j=1}^N$ be independent copies of $\delta = \{\delta_j\}_{j=1}^N$. Then, since $\mathbb{E}(Z) = 0$,

$$\begin{aligned}\mathbb{E}_\delta(\|Z\|) &= \mathbb{E}_\delta\left(\left\|Z - \mathbb{E}_{\tilde{\delta}}\left(\sum_{j=1}^N \left(\tilde{q}_j^{-1}\tilde{\delta}_j - 1\right)\eta_j \otimes \bar{\eta}_j\right)\right\|\right) \\ &\leq \mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|Z - \sum_{j=1}^N \left(\tilde{q}_j^{-1}\tilde{\delta}_j - 1\right)\eta_j \otimes \bar{\eta}_j\right\|\right)\right),\end{aligned} \tag{7.26}$$

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 (7.26), symmetry, Fubini's Theorem and the triangle inequality, it follows that

$$\begin{aligned}\mathbb{E}_\delta(\|Z\|) &\leq \mathbb{E}_\varepsilon\left(\mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|\sum_{j=1}^N \varepsilon_j\left(\tilde{q}_j^{-1}\delta_j - \tilde{q}_j^{-1}\tilde{\delta}_j\right)\eta_j \otimes \bar{\eta}_j\right\|\right)\right)\right) \\ &\leq 2\mathbb{E}_\delta\left(\mathbb{E}_\varepsilon\left(\left\|\sum_{j=1}^N \varepsilon_j \tilde{q}_j^{-1}\delta_j\eta_j \otimes \bar{\eta}_j\right\|\right)\right).\end{aligned} \tag{7.27}$$

We are now able to apply Rudelson's Lemma (Lemma 7.10). However, as specified before, it is the complex version that is crucial here. By Lemma 7.10 we get that

$$\mathbb{E}_\varepsilon\left(\left\|\sum_{j=1}^N \varepsilon_j\tilde{q}_j^{-1}\delta_j\eta_j \otimes \bar{\eta}_j\right\|\right) \leq \frac{3}{2}\sqrt{\max\{2\log(s), 2\}}\max_{1\leq j\leq N}\tilde{q}_j^{-1/2}\|\eta_j\|\sqrt{\left\|\sum_{j=1}^N q_j^{-1}\tilde{q}_j^{-1}\delta_j\eta_j \otimes \bar{\eta}_j\right\|}, \tag{7.28}$$

where $s = |\Delta|$. And hence, by using (7.27) and (7.28), it follows that

$$\mathbb{E}_\delta(\|Z\|) \leq 3\sqrt{\max\{2\log(s), 2\}}\max_{1\leq j\leq N}\tilde{q}_j^{-1/2}\|\eta_j\|\sqrt{\mathbb{E}_\delta\left(\left\|Z + \sum_{j=1}^N \eta_j \otimes \bar{\eta}_j\right\|\right)}.$$

Note that $\|\sum_{j=1}^N \eta_j \otimes \bar{\eta}_j\| \leq 1$, since U is an isometry. The result now follows from the straightforward calculus fact that if $r > 0$, $c \leq 1$ and $r \leq c\sqrt{r+1}$ then we have that $r \leq c(1+\sqrt{5})/2$. $\square$

Proof of Theorem 7.8. Let $N = N_r$ just to be clear here. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables as defined in Lemma 7.9 and define $Z = \sum_{j=1}^N Z_j$, $Z_j = \left(\tilde{q}_j^{-1}\delta_j - 1\right)\eta_j \otimes \bar{\eta}_j$ with $\eta_j = P_\Delta U^* e_j$. Now observe that

$$P_\Delta U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta = \sum_{j=1}^N \tilde{q}_j^{-1}\delta_j\eta_j \otimes \bar{\eta}_j, \quad P_\Delta U^* P_N U P_\Delta = \sum_{j=1}^N \eta_j \otimes \bar{\eta}_j. \tag{7.29}$$

Thus, it follows that

$$\|P_\Delta U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta - P_\Delta\| \leq \|Z\| + \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\| \leq \|Z\| + \frac{1}{8}, \tag{7.30}$$

by the assumption that $\|P_{M_r}U^*P_{N_r}UP_{M_r} - P_{M_r}\| \leq 1/8$. Thus, to prove the assertion we need to estimate $\|Z\|$, and Talagrand's Theorem (Theorem 7.6) will be our main tool. Note that clearly, since Z is self-adjoint, we have that $\|Z\| = \sup_{\zeta\in\mathcal{G}}|\langle Z\zeta, \zeta\rangle|$, where $\mathcal{G}$ is a countable set of vectors in the unit ball of $P_\Delta(\mathcal{H})$. For $\zeta \in \mathcal{G}$ define the mappings

$$\hat{\zeta}_1(T) = \langle T\zeta, \zeta\rangle, \quad \hat{\zeta}_2(T) = -\langle T\zeta, \zeta\rangle, \qquad T \in \mathcal{B}(\mathcal{H}).$$

In order to use Talagrand's Theorem 7.6 we restrict the domain $\mathcal{D}$ of the mappings ζ_i to

$$\mathcal{D} = \{T \in \mathcal{B}(\mathcal{H}) : \|T\| \leq \max_{1\leq j\leq N}\{\tilde{q}_j^{-1}\|\eta_j\|^2\}\}.$$

Let $\mathcal{F}$ denote the family of mappings $\hat{\zeta}_1, \hat{\zeta}_2$ for $\zeta \in \mathcal{G}$. Then $\|Z\| = \sup_{\hat{\zeta}\in\mathcal{F}} \hat{\zeta}(Z)$, and for $i = 1, 2$ we have

$$|\hat{\zeta}_i(Z_j)| = \left|\left(\tilde{q}_j^{-1}\delta_j - 1\right)\right| |\langle(\eta_j \otimes \bar{\eta}_j)\,\zeta, \zeta\rangle| \leq \max_{1\leq j\leq N}\{\tilde{q}_j^{-1}\|\eta_j\|^2\}.$$

Thus, $Z_j \in \mathcal{D}$ for $1 \leq j \leq N$ and $S := \sup_{\zeta\in\mathcal{F}} \|\hat{\zeta}\|_\infty = \max_{1\leq j\leq N}\{\tilde{q}_j^{-1}\|\eta_j\|^2\}$. Note that

$$\|\eta_j\|^2 = \langle P_\Delta U^* e_j, P_\Delta U^* e_j\rangle = \sum_{k=1}^{r} \langle P_{\Delta_k} U^* e_j, P_{\Delta_k} U^* e_j\rangle.$$

Also, note that an easy application of Holder's inequality gives the following (note that the l^1 and l^∞ bounds are finite because all the projections have finite rank),

$$\begin{aligned}
&|\langle P_{\Delta_k} U^* e_j, P_{\Delta_k} U^* e_j\rangle| \leq \|P_{\Delta_k} U^* e_j\|_{l^1} \|P_{\Delta_k} U^* e_j\|_{l^\infty} \\
&\leq \|P_{\Delta_k} U^* P_{N_l}^{N_{l-1}}\|_{l^1\to l^1} \|P_{\Delta_k} U^* e_j\|_{l^\infty} \leq \|P_{N_l}^{N_{l-1}} U P_{\Delta_k}\|_{l^\infty\to l^\infty} \cdot \sqrt{\mu(P_{N_l}^{N_{l-1}} U)} \leq \kappa_{\mathbf{N},\mathbf{M}}(l,k),
\end{aligned}$$

for $j \in \{N_{l-1}+1, \ldots, N_l\}$ and $l \in \{1, \ldots, r\}$. Hence, it follows that

$$\|\eta_j\|^2 \leq \max_{1\leq k\leq r} (\kappa_{\mathbf{N},\mathbf{M}}(k,1) + \ldots + \kappa_{\mathbf{N},\mathbf{M}}(k,r)), \tag{7.31}$$

and therefore $S \leq \max_{1\leq k\leq r}\left(q_k^{-1}\sum_{j=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,j)\right)$. Finally, note that by (7.31) and the reasoning above, it follows that

$$\begin{aligned}
V &:= \sup_{\hat{\zeta}_i\in\mathcal{F}} \mathbb{E}\left(\sum_{j=1}^{N} \hat{\zeta}_i(Z_j)^2\right) = \sup_{\zeta\in\mathcal{G}} \mathbb{E}\left(\sum_{j=1}^{N} \left(\tilde{q}_j^{-1}\delta_j - 1\right)^2 |\langle P_\Delta U^* e_j, \zeta\rangle|^4\right) \\
&\leq \max_{1\leq k\leq r} \|\eta_k\|^2 \left(\frac{N_k - N_{k-1}}{m_k} - 1\right) \sup_{\zeta\in\mathcal{G}} \sum_{j=1}^{N} |\langle e_j, U P_\Delta \zeta\rangle|^2, \qquad (7.32) \\
&\leq \max_{1\leq k\leq r} \frac{N_k - N_{k-1}}{m_k}\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right) \sup_{\zeta\in\mathcal{G}} \|U\zeta\|^2 = \max_{1\leq k\leq r} \frac{N_k - N_{k-1}}{m_k}\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right),
\end{aligned}$$

where we used the fact that U is an isometry to deduce that $\|U\| = 1$. Also, by Lemma 7.9 and (7.31) , it follows that

$$\mathbb{E}\left(\|Z\|\right)^2 \leq 48 \max_{1\leq k\leq r} \frac{N_k - N_{k-1}}{m_k}\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right) \cdot \log(s) \tag{7.33}$$

when

$$1 \geq 18 \max_{1\leq k\leq r} \frac{N_k - N_{k-1}}{m_k}\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right) \cdot \log(s), \tag{7.34}$$

(recall that we have assumed $s \geq 3$). Thus, by (7.30) and Talagrand's Theorem 7.6, it follows that

$$\begin{aligned}
&\mathbb{P}\left(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}) U P_\Delta - P_\Delta\| \geq 1/4\right) \\
&\leq \mathbb{P}\left(\|Z\| \geq \frac{1}{16} + \sqrt{24 \max_{1\leq k\leq r} \frac{N_k - N_{k-1}}{m_k}\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right) \cdot \log(s)}\right) \\
&\leq 3\exp\left(-\frac{1}{16K}\left(\max_{1\leq k\leq r} \frac{N_k - N_{k-1}}{m_k}\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right)\right)^{-1} \log\left(1 + 1/32\right)\right), \qquad (7.35)
\end{aligned}$$

when m_k's are chosen such that the right hand side of (7.33) is less than or equal to 1. Thus, by (7.30) and Talagrand's Theorem 7.6, it follows that

$$\begin{aligned}
&\mathbb{P}\left(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}) U P_\Delta - P_\Delta\| \geq 1/4\right) \\
&\leq \mathbb{P}\left(\|Z\| \geq 1/8\right) \leq \mathbb{P}\left(\|Z\| \geq \frac{1}{16} + \mathbb{E}\|Z\|\right) \leq \mathbb{P}\left(\left|\|Z\| - \mathbb{E}\|Z\|\right| \geq \frac{1}{16}\right) \\
&\leq 3\exp\left(-\frac{1}{16K}\left(\max_{1\leq k\leq r} \frac{N_k - N_{k-1}}{m_k}\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right)\right)^{-1} \log\left(1 + 1/32\right)\right), \qquad (7.36)
\end{aligned}$$

when m_k's are chosen such that the right hand side of (7.33) is less than or equal to $1/16^2$. Note that this condition is implied by the assumptions of the theorem as is (7.34). This yields the first part of the theorem. The second claim of this theorem follows from the assumption that $\|P_{M_r}U^*P_{N_r}UP_{M_r} - P_{M_r}\| \leq 1/8$. $\square$

Proposition 7.11. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel Bernoulli sampling scheme, where $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$. Consider $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$, $M_1 < \dots < M_r$, and $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$, and let $\Delta = \Delta_1 \cup \dots \cup \Delta_r$, $\Delta_k \subset \{M_{k-1}, \dots, M_k\}$, $|\Delta_k| = s_k$, where $M_0 = 0$. Let $\beta \geq 1/4$.*

(i) If

$$N := N_r, \quad K := \max_{k=1,\dots,r} \left\{ \frac{N_k - N_{k-1}}{m_k} \right\},$$

satisfy the weak balancing property with respect to U, $M := M_r$ and $s := s_1 + \dots + s_r$, then, for $\xi \in \mathcal{H}$ and $\beta, \gamma > 0$, we have that

$$\mathbb{P}\left(\|P_M P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1} \oplus \dots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta \xi\|_{l^\infty} > \beta \|\xi\|_{l^\infty}\right) \leq \gamma, \tag{7.37}$$

provided that

$$\frac{\beta}{\log\left(\frac{4}{\gamma}(M-s)\right)} \geq C\,\Lambda, \qquad \frac{\beta^2}{\log\left(\frac{4}{\gamma}(M-s)\right)} \geq C\,\Upsilon, \tag{7.38}$$

for some constant $C > 0$, where $q_k = m_k/(N_k - N_{k-1})$ for $k = 1, \dots, r$,

$$\Lambda = \max_{1 \leq k \leq r} \left\{ \frac{N_k - N_{k-1}}{m_k} \cdot \left(\sum_{l=1}^r \kappa_{\mathbf{N},\mathbf{M}}(k,l) \right) \right\}, \tag{7.39}$$

$$\Upsilon = \max_{1 \leq l \leq r} \sum_{k=1}^r \left(\frac{N_k - N_{k-1}}{m_k} - 1 \right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k, \tag{7.40}$$

for all $\{\tilde{s}_k\}_{k=1}^r$ such that $\tilde{s}_1 + \dots + \tilde{s}_r \leq s_1 + \dots + s_r$ and $\tilde{s}_k \leq S_k(s_1, \dots, s_r)$. Moreover, if $q_k = 1$ for all $k = 1, \dots, r$, then (7.38) is trivially satisfied for any $\gamma > 0$ and the left-hand side of (7.37) is equal to zero.

(ii) If N satisfies the strong Balancing Property with respect to U, M and s, then, for $\xi \in \mathcal{H}$ and $\beta, \gamma > 0$, we have that

$$\mathbb{P}\left(\|P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1} \oplus \dots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta \xi\|_{l^\infty} > \beta \|\xi\|_{l^\infty}\right) \leq \gamma, \tag{7.41}$$

provided that

$$\frac{\beta}{\log\left(\frac{4}{\gamma}(\tilde{\theta}-s)\right)} \geq C\,\Lambda, \qquad \frac{\beta^2}{\log\left(\frac{4}{\gamma}(\tilde{\theta}-s)\right)} \geq C\,\Upsilon, \tag{7.42}$$

for some constant $C > 0$, $\tilde{\theta} = \tilde{\theta}(\{q_k\}_{k=1}^r, 1/8, \{N_k\}_{k=1}^r, s, M)$ and Υ, Λ as defined in (i) and

$$\tilde{\theta}(\{q_k\}_{k=1}^r, t, \{N_k\}_{k=1}^r, s, M) = \left| \left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\dots,M\},\ |\Gamma_1| = s \\ \Gamma_{2,j} \subset \{N_{j-1}+1,\dots,N_j\},\ j=1,\dots,r}} \|P_{\Gamma_1}U^*(q_1^{-1}P_{\Gamma_{2,1}} \oplus \dots \oplus q_r^{-1}P_{\Gamma_{2,r}})Ue_i\| > \frac{t}{\sqrt{s}} \right\} \right|.$$

Moreover, if $q_k = 1$ for all $k = 1, \dots, r$, then (7.42) is trivially satisfied for any $\gamma > 0$ and the left-hand side of (7.41) is equal to zero.

Proof. To prove (i) we note that, without loss of generality, we can assume that $\|\xi\|_{l^\infty} = 1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \tilde{q}_j = q_k$, for $j \in \{N_{k-1}+1, \dots, N_k\}$ and $1 \leq k \leq r$. A key observation that will be crucial below is that

$$\begin{aligned} P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1} \oplus \dots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta \xi &= \sum_{j=1}^N P_\Delta^\perp U^* \tilde{q}_j^{-1}\delta_j (e_j \otimes e_j)UP_\Delta \xi \\ &= \sum_{j=1}^N P_\Delta^\perp U^*(\tilde{q}_j^{-1}\delta_j - 1)(e_j \otimes e_j)UP_\Delta \xi + P_\Delta^\perp U^* P_N U P_\Delta \xi. \end{aligned} \tag{7.43}$$

We will use this equation at the end of the argument, but first we will estimate the size of the individual components of $\sum_{j=1}^{N} P_\Delta^\perp U^*(\tilde{q}_j^{-1}\delta_j - 1)(e_j \otimes e_j)UP_\Delta\xi$. To do that define, for $1 \leq j \leq N$, the random variables

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

We will show using Bernstein's inequality that, for each $i \in \Delta^c$ and $t > 0$,

$$\mathbb{P}\left(\left|\sum_{j=1}^{N} X_j^i\right| > t\right) \leq 4\exp\left(-\frac{t^2/4}{\Upsilon + \Lambda t/3}\right). \tag{7.44}$$

To prove the claim, we need to estimate $\mathbb{E}\left(|X_j^i|^2\right)$ and $|X_j^i|$. First note that,

$$\mathbb{E}\left(|X_j^i|^2\right) = (\tilde{q}_j^{-1} - 1)|\langle e_j, UP_\Delta\xi\rangle|^2|\langle e_j, Ue_i\rangle|^2,$$

and note that $|\langle e_j, Ue_i\rangle|^2 \leq \mu_{\mathbf{N},\mathbf{M}}(k,l)$ for $j \in \{N_{k-1}+1, \ldots, N_k\}$ and $i \in \{M_{l-1}+1, \ldots, M_l\}$. Hence

$$\begin{aligned}\sum_{j=1}^{N} \mathbb{E}\left(|X_j^i|^2\right) &\leq \sum_{k=1}^{r}(q_k^{-1} - 1)\mu_{\mathbf{N},\mathbf{M}}(k,l)\|P_{N_k}^{N_{k-1}}UP_\Delta\xi\|^2 \\ &\leq \sup_{\zeta\in\Theta}\left\{\sum_{k=1}^{r}(q_k^{-1} - 1)\mu_{\mathbf{N},\mathbf{M}}(k,l)\|P_{N_k}^{N_{k-1}}U\zeta\|^2\right\},\end{aligned}$$

where

$$\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, |\mathrm{supp}(P_{M_l}^{M_{l-1}}\eta)| = s_l,\, l = 1, \ldots, r\}.$$

The supremum in the above bound is attained for some $\tilde{\zeta} \in \Theta$. If $\tilde{s}_k = \|P_{N_k}^{N_{k-1}}U\tilde{\zeta}\|^2$, then we have

$$\sum_{j=1}^{N} \mathbb{E}\left(|X_j^i|^2\right) \leq \sum_{k=1}^{r}(q_k^{-1} - 1)\mu_{\mathbf{N},\mathbf{M}}(k,l)\tilde{s}_k. \tag{7.45}$$

Note that it is clear from the definition that $s_k \leq S_k(s_1, \ldots, s_r)$ for $1 \leq k \leq r$. Also, using the fact that $\|U\| \leq 1$ and the definition of Θ, we note that

$$\tilde{s}_1 + \ldots + \tilde{s}_r = \sum_{k=1}^{r}\|P_{N_k}^{N_{k-1}}UP_\Delta\zeta\|^2 \leq \|UP_\Delta\zeta\|^2 = \|\zeta\|^2 \leq s_1 + \ldots + s_r.$$

To estimate $|X_j^i|$ we start by observing that, by the triangle inequality, the fact that $\|\xi\|_{l^\infty} = 1$ and Holder's inequality, it follows that $|\langle \xi, P_\Delta U^* e_j\rangle| \leq \sum_{k=1}^{r}|\langle P_{M_k}^{M_{k-1}}\xi, P_\Delta U^* e_j\rangle|$, and

$$|\langle P_{M_k}^{M_{k-1}}\xi, P_\Delta U^* e_j\rangle| \leq \|P_{N_l}^{N_{l-1}}UP_{\Delta_k}\|_{l^\infty\to l^\infty}, \quad j \in \{N_{l-1}+1, \ldots, N_l\}, \quad l \in \{1, \ldots, r\}.$$

Hence, it follows that for $1 \leq j \leq N$ and $i \in \Delta^c$,

$$\begin{aligned}|X_j^i| &= \tilde{q}_j^{-1}|(\delta_j - \tilde{q}_j)||\langle \xi, P_\Delta U^* e_j\rangle||\langle e_j, Ue_i\rangle|, \\ &\leq \max_{1\leq k\leq r}\left\{\frac{N_k - N_{k-1}}{m_k}\cdot(\kappa_{\mathbf{N},\mathbf{M}}(k,1) + \ldots + \kappa_{\mathbf{N},\mathbf{M}}(k,r))\right\}.\end{aligned} \tag{7.46}$$

Now, clearly $\mathbb{E}(X_j^i) = 0$ for $1 \leq j \leq N$ and $i \in \Delta^c$. Thus, by applying Bernstein's inequality to $\mathrm{Re}(X_j^i)$ and $\mathrm{Im}(X_j^i)$ for $j = 1, \ldots, N$, via (7.45) and (7.46), the claim (7.44) follows.

Now, by (7.44), (7.43) and the assumed weak Balancing property (wBP), it follows that

$$\begin{aligned}&\mathbb{P}\left(\|P_M P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta\xi\|_{l^\infty} > \beta\right) \\ &\leq \sum_{i\in\Delta^c\cap\{1,\ldots,M\}}\mathbb{P}\left(\left|\sum_{j=1}^{N} X_j^i + \langle P_M P_\Delta^\perp U^* P_N^\perp UP_\Delta\xi, e_i\rangle\right| > \beta\right) \\ &\leq \sum_{i\in\Delta^c\cap\{1,\ldots,M\}}\mathbb{P}\left(\left|\sum_{j=1}^{N} X_j^i\right| > \beta - \|P_M P_\Delta^\perp U^* P_N UP_\Delta\|_{l^\infty}\right) \\ &\leq 4(M - s)\exp\left(-\frac{t^2/4}{\Upsilon + \Lambda t/3}\right), \quad t = \frac{1}{2}\beta, \qquad \text{by (7.44), (wBP)},\end{aligned}$$

Also,

$$4(M-s)\exp\left(-\frac{t^2/4}{\Upsilon+\Lambda t/3}\right)\leq\gamma$$

when

$$\log\left(\frac{4}{\gamma}(M-s)\right)^{-1}\geq\left(\frac{4\Upsilon}{t^2}+\frac{4\Lambda}{3t}\right).$$

And this concludes the proof of (i). To prove (ii), for $t>0$, suppose that there is a set $\Lambda_t\subset\mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{i\in\Lambda_t}|\langle P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta\eta,e_i\rangle|>t\right)=0,\qquad|\Lambda_t^c|<\infty.$$

Then, as before, by (7.44), (7.43) and the assumed strong Balancing property (sBP), it follows that

$$\begin{aligned}&\mathbb{P}\left(\|P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta\xi\|_{l^\infty}>\beta\right)\\&\leq\sum_{i\in\Delta^c\cap\Lambda_t^c}\mathbb{P}\left(\left|\sum_{j=1}^N X_j^i+\langle P_\Delta^\perp U^*P_N^\perp UP_\Delta\xi,e_i\rangle\right|>\beta\right),\end{aligned}$$

yielding

$$\begin{aligned}&\mathbb{P}\left(\|P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta\xi\|_{l^\infty}>\beta\right)\\&\leq\sum_{i\in\Delta^c\cap\Lambda_t^c}\mathbb{P}\left(\left|\sum_{j=1}^N X_j^i\right|>\beta-\|P_\Delta^\perp U^*P_NUP_\Delta\|_{l^\infty}\right)\\&\leq 4(|\Lambda_t^c|-s)\exp\left(-\frac{t^2/4}{\Upsilon+\Lambda t/3}\right)<\gamma,\quad t=\frac{1}{2}\beta,\qquad\text{by (7.44), (sBP)},\end{aligned}$$

whenever

$$\log\left(\frac{4}{\gamma}(|\Lambda_t^c|-s)\right)^{-1}\geq\left(\frac{4\Upsilon}{t^2}+\frac{4\Lambda}{3t}\right).$$

Hence, it remains to obtain a bound on $|\Lambda_t^c|$. Let

$$\theta(q_1,\ldots,q_r,t,s)=\left\{i\in\mathbb{N}:\max_{\substack{\Gamma_1\subset\{1,\ldots,M\},\quad|\Gamma_1|=s\\\Gamma_{2,j}\subset\{N_{j-1}+1,\ldots,N_j\},\quad j=1,\ldots,r}}\|P_{\Gamma_1}U^*(q_1^{-1}P_{\Gamma_{2,1}}\oplus\ldots\oplus q_r^{-1}P_{\Gamma_{2,r}})Ue_i\|>\frac{t}{\sqrt{s}}\right\}.$$

Clearly, $\Delta_t^c\subset\theta(q_1,\ldots,q_r,t,s)$ and

$$\|P_{\Gamma_1}U^*(q_1^{-1}P_{\Gamma_{2,1}}\oplus\ldots\oplus q_r^{-1}P_{\Gamma_{2,r}})Ue_i\|\leq\max_{1\leq j\leq r}q_j^{-1}\|P_NUP_{i-1}^\perp\|\to0$$

as $i\to\infty$. So, $|\theta(q_1,\ldots,q_r,t,s)|<\infty$. Furthermore, since $\tilde{\theta}(\{q_k\}_{k=1}^r,t,\{N_k\}_{k=1}^r,s,M)$ is a decreasing function in t, for all $t\geq\frac{1}{8}$,

$$|\theta(q_1,\ldots,q_r,t,s)|<\tilde{\theta}(\{q_k\}_{k=1}^r,1/8,\{N_k\}_{k=1}^r,s,M)$$

thus, we have proved (ii). The statements at the end of (i) and (ii) are clear from the reasoning above. $\square$

Proposition 7.12. *Consider the same setup as in Proposition 7.11. If N and K satisfy the weak Balancing Property with respect to U, M and s, then, for $\xi\in\mathcal{H}$ and $\gamma>0$, we have*

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta-P_\Delta)\xi\|_{l^\infty}>\tilde{\alpha}\|\xi\|_{l^\infty})\leq\gamma,\tag{7.47}$$

with $\tilde{\alpha}=(2\log_2^{1/2}(4\sqrt{s}KM))^{-1}$, provided that

$$\begin{aligned}1&\gtrsim\Lambda\cdot\left(\log\left(s\gamma^{-1}\right)+1\right)\cdot\log\left(\sqrt{s}KM\right),\\1&\gtrsim\Upsilon\cdot\left(\log\left(s\gamma^{-1}\right)+1\right)\cdot\log\left(\sqrt{s}KM\right),\end{aligned}$$

where Λ *and* Υ *are defined in (7.39) and (7.40). Also,*

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta - P_\Delta)\xi\|_{l^\infty} > \frac{1}{2}\|\xi\|_{l^\infty}) \leq \gamma \tag{7.48}$$

provided that

$$1 \gtrsim \Lambda\cdot\left(\log\left(s\gamma^{-1}\right)+1\right), \quad 1 \gtrsim \Upsilon\cdot\left(\log\left(s\gamma^{-1}\right)+1\right).$$

Moreover, if $q_k = 1$ *for all* $k = 1,\ldots,r$*, then the left-hand sides of (7.47) and (7.48) are equal to zero.*

Proof. Without loss of generality we may assume that $\|\xi\|_{l^\infty} = 1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \tilde{q}_j := q_k$, with $j \in \{N_{k-1}+1,\ldots,N_k\}$ and $1 \leq k \leq r$. Let also, for $j \in \mathbb{N}$, $\eta_j = (UP_\Delta)^* e_j$. Then, after observing that

$$P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta = \sum_{j=1}^N q_j^{-1}\delta_j\eta_j\otimes\bar{\eta}_j, \quad P_\Delta U^* P_N U P_\Delta = \sum_{j=1}^N \eta_j\otimes\bar{\eta}_j,$$

it follows immediately that

$$P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta - P_\Delta = \sum_{j=1}^N(\tilde{q}_j^{-1}\delta_j - 1)\eta_j\otimes\bar{\eta}_j - (P_\Delta U^*P_NUP_\Delta - P_\Delta). \tag{7.49}$$

As in the proof of Proposition 7.11 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 $Z_j^i = \langle(\tilde{q}_j^{-1}\delta_j - 1)(\eta_j\otimes\bar{\eta}_j)\xi, e_i\rangle$, for $i \in \Delta$. We claim that, for $t > 0$,

$$\mathbb{P}\left(\left|\sum_{j=1}^N Z_j^i\right| > t\right) \leq 4\exp\left(-\frac{t^2/4}{\Upsilon + \Lambda t/3}\right), \qquad i \in \Delta. \tag{7.50}$$

Now, clearly $\mathbb{E}(Z_j^i) = 0$, so we may use Bernstein's inequality. Thus, we need to estimate $\mathbb{E}\left(|Z_j^i|^2\right)$ and $|Z_j^i|$. We will start with $\mathbb{E}\left(|Z_j^i|^2\right)$. Note that

$$\mathbb{E}\left(|Z_j^i|^2\right) = (\tilde{q}_j^{-1} - 1)|\langle e_j, UP_\Delta\xi\rangle|^2|\langle e_j, Ue_i\rangle|^2. \tag{7.51}$$

Thus, we can argue exactly as in the proof of Proposition 7.11 and deduce that

$$\sum_{j=1}^N \mathbb{E}\left(|Z_j^i|^2\right) \leq \sum_{k=1}^r (q_k^{-1} - 1)\mu_{N_{k-1}}\tilde{s}_k, \tag{7.52}$$

where $s_k \leq S_k(s_1,\ldots,s_r)$ for $1 \leq k \leq r$ and $\tilde{s}_1+\ldots+\tilde{s}_r \leq s_1+\ldots+s_r$. To estimate $|Z_j^i|$ we argue as in the proof of Proposition 7.11 and obtain

$$|Z_j^i| \leq \max_{1\leq k\leq r}\left\{\frac{N_k - N_{k-1}}{m_k}\cdot(\kappa_{\mathbf{N},\mathbf{M}}(k,1)+\ldots+\kappa_{\mathbf{N},\mathbf{M}}(k,r))\right\}. \tag{7.53}$$

Thus, by applying Bernstein's inequality to $\mathrm{Re}(Z_1^i),\ldots,\mathrm{Re}(Z_N^i)$ and $\mathrm{Im}(Z_1^i),\ldots,\mathrm{Im}(Z_N^i)$ we obtain, via (7.52) and (7.53) the estimate (7.50), and we have proved the claim.

Now armed with (7.50) we can deduce that , by (7.43) and the assumed weak Balancing property (wBP), it follows that

$$\begin{aligned}
&\mathbb{P}\left(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta - P_\Delta)\xi\|_{l^\infty} > \tilde{\alpha}\right)\\
&\leq \sum_{i\in\Delta}\mathbb{P}\left(\left|\sum_{j=1}^N Z_j^i + \langle(P_\Delta U^*P_NUP_\Delta - P_\Delta)\xi, e_i\rangle\right| > \tilde{\alpha}\right)\\
&\leq \sum_{i\in\Delta}\mathbb{P}\left(\left|\sum_{j=1}^N Z_j^i\right| > \tilde{\alpha} - \|P_MU^*P_NUP_M - P_M\|_{l^1}\right),\\
&\leq 4\,s\exp\left(-\frac{t^2/4}{\Upsilon+\Lambda t/3}\right), \quad t = \tilde{\alpha}, \qquad \text{by (7.50), (wBP).}
\end{aligned} \tag{7.54}$$

Also,

$$4s\exp\left(-\frac{t^2/4}{\Upsilon+\Lambda t/3}\right)\leq\gamma, \tag{7.55}$$

when

$$1\geq\left(\frac{4\Upsilon}{t^2}+\frac{4}{3t}\Lambda\right)\cdot\log\left(\frac{4s}{\gamma}\right).$$

And this gives the first part of the proposition. Also, the fact that the left hand side of (7.47) is zero when $q_k=1$ for $1\leq k\leq r$ is clear from (7.55). Note that (ii) follows by arguing exactly as above and replacing $\tilde{\alpha}$ by $\frac{1}{4}$.

$\square$

Proposition 7.13. *Let $U\in\mathcal{B}(l^2(\mathbb{N}))$ such that $\|U\|\leq 1$. Suppose that $\Omega=\Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel Bernoulli sampling scheme, where $\mathbf{N}=(N_1,\ldots,N_r)\in\mathbb{N}^r$ and $\mathbf{m}=(m_1,\ldots,m_r)\in\mathbb{N}^r$. Consider $(\mathbf{s},\mathbf{M})$, where $\mathbf{M}=(M_1,\ldots,M_r)\in\mathbb{N}^r$, $M_1<\ldots<M_r$, and $\mathbf{s}=(s_1,\ldots,s_r)\in\mathbb{N}^r$, and let $\Delta=\Delta_1\cup\ldots\cup\Delta_r$, where $\Delta_k\subset\{M_{k-1}+1,\ldots,M_k\}$, $|\Delta_k|=s_k$, and $M_0=0$. Then, for any $t\in(0,1)$ and $\gamma\in(0,1)$,*

$$\mathbb{P}\left(\max_{i\in\{1,\ldots,M\}\cap\Delta^c}\|P_{\{i\}}U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_{\{i\}}\|\geq 1+t\right)\leq\gamma$$

provided that

$$\frac{t^2}{4}\geq\log\left(\frac{2M}{\gamma}\right)\cdot\max_{1\leq k\leq r}\left\{\left(\frac{N_k-N_{k-1}}{m_k}-1\right)\cdot\mu_{\mathbf{N},\mathbf{M}}(k,l)\right\} \tag{7.56}$$

for all $l=1,\ldots,r$ when $M=M_r$ and for all $l=1,\ldots,r-1,\infty$ when $M>M_r$. In addition, if $m_k=N_k-N_{k-1}$ for each $k=1,\ldots r$, then

$$\mathbb{P}(\|P_{\{i\}}U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_{\{i\}}\|\geq 1+t)=0,\quad\forall i\in\mathbb{N}. \tag{7.57}$$

Proof. Fix $i\in\{1,\ldots,M\}$. Let $\{\delta_j\}_{j=1}^N$ be random independent Bernoulli variables with $\mathbb{P}(\delta_j=1)=\tilde{q}_j:=q_k$ for $j\in\{N_{k-1}+1,\ldots,N_k\}$. Define $Z=\sum_{j=1}^N Z_j$ and $Z_j=\left(\tilde{q}_j^{-1}\delta_j-1\right)|u_{ji}|^2$. Now observe that

$$P_{\{i\}}U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_{\{i\}}=\sum_{j=1}^N\tilde{q}_j^{-1}\delta_j\,|u_{ji}|^2=\sum_{j=1}^N Z_j+\sum_{j=1}^N|u_{ji}|^2,$$

where we interpret U as the infinite matrix $U=\{u_{ij}\}_{i,j\in\mathbb{N}}$. Thus, since $\|U\|\leq 1$,

$$\|P_{\{i\}}U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_{\{i\}}\|\leq\left|\sum_{j=1}^N Z_j\right|+1 \tag{7.58}$$

and it is clear that (7.57) is true. For the case where $q_k<1$ for some $k\in\{1,\ldots,r\}$, observe that for $i\in\{M_{l-1}+1,\ldots,M_l\}$ (recall that Z_j depend on i), we have that $\mathbb{E}(Z_j)=0$. Also,

$$|Z_j|\leq\begin{cases}\max_{1\leq k\leq r}\{\max\{q_k^{-1}-1,1\}\cdot\mu_{\mathbf{N},\mathbf{M}}(k,l)\}:=B_i & i\in\{M_{l-1}+1,\ldots,M_l\}\\ \max_{1\leq k\leq r}\{\max\{q_k^{-1}-1,1\}\cdot\mu_{\mathbf{N},\mathbf{M}}(k,\infty)\}:=B_\infty & i>M_r,\end{cases}$$

and, by again using the assumption that $\|U\|\leq 1$,

$$\begin{aligned}\sum_{j=1}^N\mathbb{E}(|Z_j|^2)&=\sum_{j=1}^N(\tilde{q}_j^{-1}-1)\,|u_{ji}|^4\\ &\leq\begin{cases}\max_{1\leq k\leq r}\{(q_k^{-1}-1)\,\mu_{\mathbf{N},\mathbf{M}}(k,l)\}=:\sigma_i^2 & i\in\{M_{l-1}+1,\ldots,M_l\}\\ \max_{1\leq k\leq r}\{(q_k^{-1}-1)\,\mu_{\mathbf{N},\mathbf{M}}(k,\infty)\}=:\sigma_\infty^2 & i>M_r.\end{cases}\end{aligned}$$

Thus, by Bernstein's inequality and (7.58),

$$
\begin{aligned}
&\mathbb{P}(\|P_{\{i\}}U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r})UP_{\{i\}}\| \geq 1+t) \\
&\qquad \leq \mathbb{P}\left(\left|\sum_{j=1}^{N} Z_j\right| \geq t\right) \leq 2\exp\left(-\frac{t^2/2}{\sigma^2 + Bt/3}\right),
\end{aligned}
$$

$$
B = \begin{cases} \max_{1\leq i\leq r} B_i & M = M_r, \\ \max_{i\in\{1,\ldots,r-1,\infty\}} B_i & M > M_r \end{cases}, \quad \sigma^2 = \begin{cases} \max_{1\leq i\leq r} \sigma_i^2 & M = M_r, \\ \max_{i\in\{1,\ldots,r-1,\infty\}} \sigma_1^2 & M > M_r. \end{cases}
$$

Applying the union bound yields

$$
\mathbb{P}\left(\max_{i\in\{1,\ldots,M\}} \|P_{\{i\}}U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r})UP_{\{i\}}\| \geq 1+t\right) \leq \gamma
$$

whenever (7.56) holds. □

7.3 Proofs of Propositions 7.3 and 7.4

The proof of the propositions relies on an idea that originated in a paper by D. Gross [32], namely, the golfing scheme. The variant we are using here is based on an idea from [1] as well as uneven section techniques from [36, 35], see also [31]. However, the informed reader will recognise that the setup here differs substantially from both [32] and [1]. See also [12] for other examples of the use of the golfing scheme. Before we embark on the proof, we will state and prove a useful lemma.

Lemma 7.14. *Let $\tilde{X}_k$ be independent binary variables taking values 0 and 1, such that $\tilde{X}_k = 1$ with probability P. Then,*

$$
\mathbb{P}\left(\sum_{i=1}^{N} \tilde{X}_i \geq k\right) \geq \left(\frac{N\cdot e}{k}\right)^{-k} \binom{N}{k} P^k. \tag{7.59}
$$

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 $i = 0,\ldots,N-k$, 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 [17], p. 1186). □

Proof of Proposition 7.3. We start by mentioning that converting from the Bernoulli sampling model and uniform sampling model has become standard in the literature. In particular, one can do this by showing that the Bernoulli model implies (up to a constant) the uniform sampling model in each of the conditions in Proposition 7.1. This is straightforward and the reader may consult [14, 13, 30] for details. We will therefore consider (without loss of generality) only the multilevel Bernoulli sampling scheme.

Recall that we are using the following Bernoulli sampling model: Given $N_0 = 0, N_1, \ldots, N_r \in \mathbb{N}$ we let

$$
\{N_{k-1}+1,\ldots,N_k\} \supset \Omega_k \sim \mathrm{Ber}(q_k), \quad q_k = \frac{m_k}{N_k - N_{k-1}}.
$$

Note that we may replace this Bernoulli sampling model with the following equivalent sampling model (see [1]):

$$\Omega_k = \Omega_k^1 \cup \Omega_k^2 \cup \cdots \cup \Omega_k^u, \qquad \Omega_k^j \sim \mathrm{Ber}(q_k^j), \qquad 1 \le k \le r,$$

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

$$(1-q_k^1)(1-q_k^2)\cdots(1-q_k^u) = (1-q_k). \tag{7.60}$$

The latter model is the one we will use throughout the proof and the specific value of u will be chosen later. Note also that because of overlaps we will have

$$q_k^1 + q_k^2 + \ldots + q_k^u \ge q_k, \qquad 1 \le k \le r. \tag{7.61}$$

The strategy of the proof is to show the validity of (i) and (ii), and the existence of a $\rho \in \mathrm{ran}(U^*(P_{\Omega_1} \oplus \ldots \oplus P_{\Omega_r}))$ that satisfies (iii)-(v) in Proposition 7.1 with probability exceeding $1-\epsilon$, where (iii) is replaced by (7.16), (iv) is replaced by $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \le \frac{1}{2}$ and L in (v) is given by (7.17).

Step I: The construction of ρ: We start by defining $\gamma = \epsilon/6$ (the reason for this particular choice will become clear later). We also define a number of quantities (and the reason for these choices will become clear later in the proof):

$$u = 8\lceil 3v + \log(\gamma^{-1})\rceil, \qquad v = \lceil \log_2(8KM\sqrt{s})\rceil, \tag{7.62}$$

as well as

$$\{q_k^i : 1 \le k \le r, 1 \le i \le u\}, \quad \{\alpha_i\}_{i=1}^u, \quad \{\beta_i\}_{i=1}^u$$

by

$$q_k^1 = q_k^2 = \frac{1}{4}q_k, \qquad \tilde{q}_k = q_k^3 = \ldots = q_k^u, \qquad q_k = (N_k - N_{k-1})m_k^{-1}, \quad 1 \le k \le r, \tag{7.63}$$

with

$$(1-q_k^1)(1-q_k^2)\cdots(1-q_k^u) = (1-q_k)$$

and

$$\alpha_1 = \alpha_2 = (2\log_2^{1/2}(4KM\sqrt{s}))^{-1}, \qquad \alpha_i = 1/2, \quad 3 \le i \le u, \tag{7.64}$$

as well as

$$\beta_1 = \beta_2 = \frac{1}{4}, \qquad \beta_i = \frac{1}{4}\log_2(4KM\sqrt{s}), \quad 3 \le i \le u. \tag{7.65}$$

Consider now the following construction of ρ. We will define recursively the sequences $\{Z_i\}_{i=0}^u \subset \mathcal{H}$, $\{Y_i\}_{i=1}^u \subset \mathcal{H}$ and $\{\omega_i\}_{i=0}^u \subset \mathbb{N}$ as follows: first let $\omega_0 = \{0\}$, $\omega_1 = \{0,1\}$ and $\omega_2 = \{0,1,2\}$. Then define recursively, for $i \ge 3$, the following:

$$\omega_i = \begin{cases} \omega_{i-1} \cup \{i\} & \text{if } \|(P_\Delta - P_\Delta U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta)Z_{i-1}\|_{l^\infty} \le \alpha_i \|P_{\Delta_k} Z_{i-1}\|_{l^\infty}, \\ & \text{and } \|P_M P_\Delta^\perp U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta Z_{i-1}\|_{l^\infty} \le \beta_i \|Z_{i-1}\|_{l^\infty}, \\ \omega_{i-1} & \text{otherwise}, \end{cases}$$

$$Y_i = \begin{cases} \sum_{j \in \omega_i} U^*(\frac{1}{q_1^j}P_{\Omega_1^j} \oplus \ldots \oplus \frac{1}{q_r^j}P_{\Omega_r^j})UZ_{j-1} & \text{if } i \in \omega_i, \\ Y_{i-1} & \text{otherwise}, \end{cases} \qquad i \ge 1, \tag{7.66}$$

$$Z_i = \begin{cases} \mathrm{sgn}(x_0) - P_\Delta Y_i & \text{if } i \in \omega_i, \\ Z_{i-1} & \text{otherwise}, \end{cases} \qquad i \ge 1, \qquad Z_0 = \mathrm{sgn}(x_0).$$

Now, let $\{A_i\}_{i=1}^2$ and $\{B_i\}_{i=1}^5$ denote the following events

$$\begin{aligned}
A_i: &\quad \|(P_\Delta - U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta)Z_{i-1}\|_{l^\infty} \le \alpha_i \|Z_{i-1}\|_{l^\infty}, \qquad i = 1,2, \\
B_i: &\quad \|P_M P_\Delta^\perp U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta Z_{i-1}\|_{l^\infty} \le \beta_i \|Z_{i-1}\|_{l^\infty}, \qquad i = 1,2, \\
B_3: &\quad \|P_\Delta U^*(\frac{1}{q_1}P_{\Omega_1} \oplus \ldots \oplus \frac{1}{q_r}P_{\Omega_r})UP_\Delta - P_\Delta\| \le 1/4, \\
&\quad \max_{i \in \Delta^c \cap \{1,\ldots,M\}} \|\left(q_1^{-1/2}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1/2}P_{\Omega_r}\right)Ue_i\| \le \sqrt{5/4} \\
B_4: &\quad |\omega_u| \ge v, \\
B_5: &\quad (\cap_{i=1}^2 A_i) \cap (\cap_{i=1}^4 B_i).
\end{aligned} \tag{7.67}$$

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

$$\rho = \begin{cases} Y_{\tau(v)} & \text{if } B_5 \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that, clearly, $\rho \in \operatorname{ran}(U^* P_\Omega)$, and we just need to show that when the event B_5 occurs, then (i)-(v) in Proposition 7.1 will follow.

Step II: $B_5 \Rightarrow$ **(i)**, **(ii)**. To see that the assertion is true, note that if B_5 occurs then B_3 occurs, which immediately (i) and (ii).

Step III: $B_5 \Rightarrow$ **(iii)**, **(iv)**. To show the assertion, we start by making the following observations: By the construction of $Z_{\tau(i)}$ and the fact that $Z_0 = \operatorname{sgn}(x_0)$, it follows that

$$\begin{aligned} Z_{\tau(i)} &= Z_0 - (P_\Delta U^* (\frac{1}{q_1^{\tau(1)}} P_{\Omega_1^{\tau(1)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(1)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_0 \\ &\quad + \ldots + P_\Delta U^* (\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_{\tau(i-1)}) \\ &= Z_{\tau(i-1)} - P_\Delta U^* (\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_{\tau(i-1)} \qquad i \leq |\omega_u|, \end{aligned}$$

so we immediately get that

$$Z_{\tau(i)} = (P_\Delta - P_\Delta U^* (\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_{\tau(i-1)}, \qquad i \leq |\omega_u|.$$

Hence, if the event B_5 occurs, we have, by the choices in (7.64) and (7.65)

$$\|\rho - \operatorname{sgn}(x_0)\| = \|Z_{\tau(v)}\| \leq \sqrt{s} \|Z_{\tau(v)}\|_{l^\infty} \leq \sqrt{s} \prod_{i=1}^{v} \alpha_{\tau(i)} \leq \frac{\sqrt{s}}{2^v} \leq \frac{1}{8K}, \tag{7.68}$$

since we have chosen $v = \lceil \log_2(8KM\sqrt{s}) \rceil$. Also,

$$\begin{aligned} \|P_M P_\Delta^\perp \rho\|_{l^\infty} &\leq \sum_{i=1}^{v} \|P_M P_\Delta^\perp U^* (\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta Z_{\tau(i-1)}\|_{l^\infty} \\ &\leq \sum_{i=1}^{v} \beta_{\tau(i)} \|Z_{\tau(i-1)}\|_{l^\infty} \leq \sum_{i=1}^{v} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)} \\ &\leq \frac{1}{4}(1 + \frac{1}{2\log_2^{1/2}(a)} + \frac{\log_2(a)}{2^3 \log_2(a)} + \ldots + \frac{1}{2^{v-1}}) \leq \frac{1}{2}, \qquad a = 4KM\sqrt{s}. \end{aligned} \tag{7.69}$$

In particular, (7.68) and (7.69) imply (iii) and (iv) in Proposition 7.1.

Step IV: $B_5 \Rightarrow$ **(v)**. To show that, note that we may write the already constructed ρ as $\rho = U^* P_\Omega w$ where

$$w = \sum_{i=1}^{v} w_i, \quad w_i = \left(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r} \right) U P_\Delta Z_{\tau(i-1)}.$$

To estimate $\|w\|$ we simply compute

$$\begin{aligned} \|w_i\|^2 &= \left\langle \left(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}} \right) U P_\Delta Z_{\tau(i-1)}, \left(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}} \right) U P_\Delta Z_{\tau(i-1)} \right\rangle \\ &= \sum_{k=1}^{r} \left(\frac{1}{q_k^{\tau(i)}} \right)^2 \|P_{\Omega_k^{\tau(i)}} U Z_{\tau(i-1)}\|^2, \end{aligned}$$

and then use the assumption that the event B_5 holds to deduce that

$$\sum_{k=1}^{r}\left(\frac{1}{q_k^{\tau(i)}}\right)^2 \|P_{\Omega_k^{\tau(i)}}UZ_{\tau(i-1)}\|^2 \leq \max_{1\leq k\leq r}\left\{\frac{1}{q_k^{\tau(i)}}\right\}\langle\sum_{k=1}^{r}\frac{1}{q_k^{\tau(i)}}P_\Delta U^* P_{\Omega_k^{\tau(i)}}UZ_{\tau(i-1)}, Z_{\tau(i-1)}\rangle$$
$$= \max_{1\leq k\leq r}\left\{\frac{1}{q_k^{\tau(i)}}\right\}\langle\left(\sum_{k=1}^{r}\frac{1}{q_k^{\tau(i)}}P_\Delta U^* P_{\Omega_k^{\tau(i)}}U - P_\Delta\right)Z_{\tau(i-1)}, Z_{\tau(i-1)}\rangle + \|Z_{\tau(i-1)}\|^2$$
$$\leq \max_{1\leq k\leq r}\left\{\frac{1}{q_k^{\tau(i)}}\right\}\left(\|Z_{\tau(i-1)}\|\|Z_{\tau(i)}\| + \|Z_{\tau(i-1)}\|^2\right)$$
$$\leq \max_{1\leq k\leq r}\left\{\frac{1}{q_k^{\tau(i)}}\right\} s\left(\|Z_{\tau(i-1)}\|_{l^\infty}\|Z_{\tau(i)}\|_{l^\infty} + \|Z_{\tau(i-1)}\|_{l^\infty}^2\right) \leq \max_{1\leq k\leq r}\left\{\frac{1}{q_k^{\tau(i)}}\right\} s(\alpha_i+1)\left(\prod_{j=1}^{i-1}\alpha_j\right)^2,$$

where the last inequality follows from the assumption that the event B_5 holds. Hence

$$\|w\| \leq \sqrt{s}\sum_{i=1}^{v}\left(\max_{1\leq k\leq r}\left\{\frac{1}{\sqrt{q_k^{\tau(i)}}}\right\}\sqrt{\alpha_i+1}\prod_{j=1}^{i-1}\alpha_j\right) \tag{7.70}$$

Note that, due to the fact that $q_k^1+\ldots+q_k^u \geq q_k$, we have that

$$\tilde{q}_k \geq \frac{m_k}{2(N_k-N_{k-1})}\frac{1}{8\lceil\log(\gamma^{-1})+3\lceil\log_2(8KM\sqrt{s})\rceil\rceil-2}.$$

This gives, in combination with the chosen values of $\{\alpha_j\}$ and (7.70) that

$$\begin{aligned}
\|w\| &\leq 2\sqrt{s}\max_{1\leq k\leq r}\sqrt{\frac{N_k-N_{k-1}}{m_k}}\left(1+\frac{1}{2\log_2^{1/2}(4KM\sqrt{s})}\right)^{3/2} \\
&\quad + \sqrt{s}\max_{1\leq k\leq r}\sqrt{\frac{N_k-N_{k-1}}{m_k}}\cdot\frac{\sqrt{3}}{2}\cdot\frac{\sqrt{8\lceil\log(\gamma^{-1})+3\lceil\log_2(8KM\sqrt{s})\rceil\rceil-2}}{\log_2(4KM\sqrt{s})}\cdot\sum_{i=3}^{v}\frac{1}{2^{i-3}} \\
&\leq 2\sqrt{s}\max_{1\leq k\leq r}\sqrt{\frac{N_k-N_{k-1}}{m_k}}\left(\left(\frac{3}{2}\right)^{3/2}+\sqrt{\frac{6}{\log_2(4KM\sqrt{s})}}\sqrt{1+\frac{\log_2(\gamma^{-1})+6}{\log_2(4KM\sqrt{s})}}\right) \\
&\leq \sqrt{s}\max_{1\leq k\leq r}\sqrt{\frac{N_k-N_{k-1}}{m_k}}\left(\frac{3\sqrt{3}}{\sqrt{2}}+\frac{2\sqrt{6}}{\sqrt{\log_2(4KM\sqrt{s})}}\sqrt{1+\frac{\log_2(\gamma^{-1})+6}{\log_2(4KM\sqrt{s})}}\right).
\end{aligned} \tag{7.71}$$

Step V: The weak balancing property, (7.14) and (7.15) $\Rightarrow \mathbb{P}(A_1^c\cup A_2^c\cup B_1^c\cup B_2^c\cup B_3^c)\leq 5\gamma$.

To see this, note that by Proposition 7.12 we immediately get (recall that $q_k^1=q_k^2=1/4q_k$) that $\mathbb{P}(A_1^c)\leq\gamma$ and $\mathbb{P}(A_2^c)\leq\gamma$ as long as the weak balancing property and

$$1\gtrsim\Lambda\cdot\left(\log\left(s\gamma^{-1}\right)+1\right)\cdot\log\left(\sqrt{s}KM\right),\quad 1\gtrsim\Upsilon\cdot\left(\log\left(s\gamma^{-1}\right)+1\right)\cdot\log\left(\sqrt{s}KM\right), \tag{7.72}$$

are satisfied, where $K=\max_{1\leq k\leq r}(N_k-N_{k-1})/m_k$,

$$\Lambda=\max_{1\leq k\leq r}\left\{\frac{N_k-N_{k-1}}{m_k}\cdot\left(\sum_{l=1}^{r}\kappa_{\mathbf{N},\mathbf{M}}(k,l)\right)\right\}, \tag{7.73}$$

$$\Upsilon=\max_{1\leq l\leq r}\sum_{k=1}^{r}\left(\frac{N_k-N_{k-1}}{m_k}-1\right)\cdot\mu_{\mathbf{N},\mathbf{M}}(k,l)\cdot\tilde{s}_k, \tag{7.74}$$

and where $\tilde{s}_1+\ldots+\tilde{s}_r\leq s_1+\ldots+s_r$ and $\tilde{s}_k\leq S_k(s_1,\ldots,s_r)$. However, clearly, (7.14) and (7.15) imply (7.72). Also, Proposition 7.11 yields that $\mathbb{P}(B_1^c)\leq\gamma$ and $\mathbb{P}(B_2^c)\leq\gamma$ as long as the weak balancing property and

$$1\gtrsim\Lambda\cdot\log\left(\frac{4}{\gamma}(M-s)\right),\qquad 1\gtrsim\Upsilon\cdot\log\left(\frac{4}{\gamma}(M-s)\right), \tag{7.75}$$

are satisfied. However, again, (7.14) and (7.15) imply (7.75). Finally, it remains to bound $\mathbb{P}(B_3^c)$. First note that by Theorem 7.8, we may deduce that

$$\mathbb{P}\left(\|P_\Delta U^*(\frac{1}{q_1}P_{\Omega_1}\oplus\ldots\oplus\frac{1}{q_r}P_{\Omega_r})UP_\Delta - P_\Delta\| > 1/4,\right) \leq \gamma/2,$$

when the weak balancing property and

$$1 \gtrsim \Lambda\cdot\left(\log\left(\gamma^{-1}\,s\right)+1\right) \tag{7.76}$$

holds and (7.14) implies (7.76).

For the second part of B_3, we may deduce from Proposition 7.13 that

$$\mathbb{P}\left(\max_{i\in\Delta^c\cap\{1,\ldots,M\}}\|\left(q_1^{-1/2}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1/2}P_{\Omega_r}\right)Ue_i\| > \sqrt{5/4}\right) \leq \frac{\gamma}{2},$$

whenever

$$1 \gtrsim \log\left(\frac{2M}{\gamma}\right)\cdot\max_{1\leq k\leq r}\left\{\left(\frac{N_k-N_{k-1}}{m_k}-1\right)\cdot\mu_{\mathbf{N},\mathbf{M}}(k,l)\right\},\qquad l=1,\ldots,r. \tag{7.77}$$

which is true whenever (7.14) holds. Indeed, recalling the definition of $\kappa_{\mathbf{N},\mathbf{M}}(k,j)$ and Θ in Definition 7.2, observe that

$$\max_{\eta\in\Theta,\|\eta\|_\infty=1}\sum_{l=1}^{r}\left\|P_{N_k}^{N_{k-1}}UP_{M_l}^{M_{l-1}}\eta\right\|_\infty \geq \max_{\eta\in\Theta,\|\eta\|_\infty=1}\left\|P_{N_k}^{N_{k-1}}U\eta\right\|_\infty \geq \sqrt{\mu(P_{N_k}^{N_{k-1}}UP_{M_l}^{M_{l-1}})} \tag{7.78}$$

for each $l=1,\ldots,r$ which implies that $\sum_{j=1}^r\kappa_{\mathbf{N},\mathbf{M}}(k,j)\geq\mu_{\mathbf{N},\mathbf{M}}(k,l)$, for $l=1,\ldots,r$. Consequently, (7.77) follows from (7.14). Thus, $\mathbb{P}(B_3^c)\leq\gamma$.

Step VI: The weak balancing property, (7.14) and (7.15) $\Rightarrow$ $\mathbb{P}(B_4^c)\leq\gamma$. To see this, define the random variables $X_1,\ldots X_{u-2}$ by

$$X_j = \begin{cases} 0 & \omega_{j+2}\neq\omega_{j+1}, \\ 1 & \omega_{j+2}=\omega_{j+1}. \end{cases} \tag{7.79}$$

We immediately observe that

$$\mathbb{P}(B_4^c)=\mathbb{P}(|\omega_u|<v)=\mathbb{P}(X_1+\ldots+X_{u-2}>u-v). \tag{7.80}$$

However, the random variables $X_1,\ldots X_{u-2}$ are not independent, and we therefore cannot directly apply the standard Chernoff bound. In particular, we must adapt the setup slightly. Note that

$$\begin{aligned}
&\mathbb{P}(X_1+\ldots+X_{u-2}>u-v)\\
&\leq \sum_{l=1}^{\binom{u-2}{u-v}}\mathbb{P}(X_{\pi(l)_1}=1,X_{\pi(l)_2}=1,\ldots,X_{\pi(l)_{u-v}}=1)\\
&= \sum_{l=1}^{\binom{u-2}{u-v}}\mathbb{P}(X_{\pi(l)_{u-v}}=1\,|\,X_{\pi(l)_1}=1,\ldots,X_{\pi(l)_{u-v-1}}=1)\mathbb{P}(X_{\pi(l)_1}=1,\ldots,X_{\pi(l)_{u-v-1}}=1)\\
&= \sum_{l=1}^{\binom{u-2}{u-v}}\mathbb{P}(X_{\pi(l)_{u-v}}=1\,|\,X_{\pi(l)_1}=1,\ldots,X_{\pi(l)_{u-v-1}}=1)\\
&\qquad\times\mathbb{P}(X_{\pi(l)_{u-v-1}}=1\,|\,X_{\pi(l)_1}=1,\ldots,X_{\pi(l)_{u-v-2}}=1)\cdots\mathbb{P}(X_{\pi(l)_1}=1)
\end{aligned} \tag{7.81}$$

where $\pi:\{1,\ldots,\binom{u-2}{u-v}\}\to\mathbb{N}^{u-v}$ ranges over all $\binom{u-2}{u-v}$ ordered subsets of $\{1,\ldots,u-2\}$ of size $u-v$. Thus, if we can provide a bound P such that

$$\begin{aligned}
&P\geq\mathbb{P}(X_{\pi(l)_{u-v-j}}=1\,|\,X_{\pi(l)_1}=1,\ldots,X_{\pi(l)_{u-v-(j+1)}}=1),\\
&P\geq\mathbb{P}(X_{\pi(l)_1}=1)
\end{aligned} \tag{7.82}$$

$$l = 1, \ldots, \binom{u-2}{u-v}, \quad j = 0, \ldots, u-v-2,$$

then, by (7.81),

$$\mathbb{P}(X_1 + \ldots + X_{u-2} > u-v) \leq \binom{u-2}{u-v} P^{u-v}. \tag{7.83}$$

We will continue assuming that (7.82) is true, and then return to this inequality below.

Let $\{\tilde{X}_k\}_{k=1}^{u-2}$ be independent binary variables taking values 0 and 1, such that $\tilde{X}_k = 1$ with probability P. Then, by Lemma 7.14, (7.83) and (7.80) it follows that

$$\mathbb{P}(B_4^c) \leq \mathbb{P}\left(\tilde{X}_1 + \ldots + \tilde{X}_{u-2} \geq u-v\right) \left(\frac{(u-2) \cdot e}{u-v}\right)^{u-v}. \tag{7.84}$$

Then, by the standard Chernoff bound ([48, Theorem 2.1, equation 2]), it follows that, for $t > 0$,

$$\mathbb{P}\left(\tilde{X}_1 + \ldots + \tilde{X}_{u-2} \geq (u-2)(t+P)\right) \leq e^{-2(u-2)t^2}. \tag{7.85}$$

Hence, if we let $t = (u-v)/(u-2) - P$, it follows from (7.84) and (7.85) that

$$\mathbb{P}(B_4^c) \leq e^{-2(u-2)t^2 + (u-v)(\log(\frac{u-2}{u-v})+1)} \leq e^{-2(u-2)t^2 + u - 2}.$$

Thus, by choosing $P = 1/4$ we get that $\mathbb{P}(B_4^c) \leq \gamma$ whenever $u \geq x$ and x is the largest root satisfying

$$(x-u)\left(\frac{x-v}{u-2} - \frac{1}{4}\right) - \log(\gamma^{-1/2}) - \frac{x-2}{2} = 0,$$

and this yields $u \geq 8\lceil 3v + \log(\gamma^{-1/2})\rceil$ which is satisfied by the choice of u in (7.62). Thus, we would have been done with Step VI if we could verify (7.82) with $P = 1/4$, and this is the theme in the following claim.

Claim: The weak balancing property, (7.14) and (7.15) $\Rightarrow$ (7.82) with $P = 1/4$. To prove the claim we first observe that $X_j = 0$ when

$$\|(P_\Delta - P_\Delta U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta) Z_{i-1}\|_{l^\infty} \leq \frac{1}{2}\|Z_{i-1}\|_{l^\infty}$$

$$\|P_M P_\Delta^\perp U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta Z_{i-1}\|_{l^\infty} \leq \frac{1}{4}\log_2(4KM\sqrt{s})\|Z_{i-1}\|_{l^\infty}, \qquad i = j+2,$$

where we recall from (7.63) that

$$q_k^3 = q_k^4 = \ldots = q_k^u = \tilde{q}_k, \qquad 1 \leq k \leq r.$$

Thus, by choosing $\gamma = 1/8$ in (7.48) in Proposition 7.12 and $\gamma = 1/8$ in (i) in Proposition 7.11, it follows that $\frac{1}{4} \geq \mathbb{P}(X_j = 1)$, for $j = 1, \ldots, u-2$, when the weak balancing property is satisfied and

$$(\log(8s) + 1)^{-1} \gtrsim \tilde{q}_k^{-1} \cdot \sum_{l=1}^r \kappa_{\mathbf{N},\mathbf{M}}(k,l), \quad 1 \leq k \leq r \tag{7.86}$$

$$(\log(8s) + 1)^{-1} \gtrsim \left(\sum_{k=1}^r \left(\tilde{q}_k^{-1} - 1\right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k\right), \quad 1 \leq l \leq r, \tag{7.87}$$

as well as

$$\frac{\log_2(4KM\sqrt{s})}{\log(32(M-s))} \gtrsim \tilde{q}_k^{-1} \cdot \sum_{l=1}^r \kappa_{\mathbf{N},\mathbf{M}}(k,l), \quad 1 \leq k \leq r \tag{7.88}$$

$$\frac{\log_2(4KM\sqrt{s})}{\log(32(M-s))} \gtrsim \left(\sum_{k=1}^r \left(\tilde{q}_k^{-1} - 1\right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k\right), \quad 1 \leq l \leq r, \tag{7.89}$$

with $K = \max_{1 \leq k \leq r}(N_k - N_{k-1})/m_k$. Thus, to prove the claim we must demonstrate that (7.14) and (7.15) $\Rightarrow$ (7.86), (7.87), (7.88) and (7.89). We split this into two stages:

Stage 1: (7.15) $\Rightarrow$ (7.89) and (7.87). To show the assertion we must demonstrate that if, for $1 \leq k \leq r$,

$$m_k \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_k \cdot \log\left(KM\sqrt{s}\right), \tag{7.90}$$

where $\hat{m}_k$ satisfies

$$1 \gtrsim \sum_{k=1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k, \qquad l = 1, \ldots, r, \tag{7.91}$$

we get (7.89) and (7.87). To see this, note that by (7.61) we have that

$$q_k^1 + q_k^2 + (u-2)\tilde{q}_k \geq q_k, \qquad 1 \leq k \leq r, \tag{7.92}$$

so since $q_k^1 = q_k^2 = \frac{1}{4}q_k$, and by (7.92), (7.90) and the choice of u in (7.62), it follows that

$$\begin{aligned} 2(8(\lceil\log(\gamma^{-1}) + 3\lceil\log_2(8KM\sqrt{s})\rceil\rceil) - 2)\tilde{q}_k \geq q_k &= \frac{m_k}{N_k - N_{k-1}} \\ &\geq C\frac{\hat{m}_k}{N_k - N_{k-1}}(\log(s\epsilon^{-1}) + 1)\log\left(KM\sqrt{s}\right) \\ &\geq C\frac{\hat{m}_k}{N_k - N_{k-1}}(\log(s) + 1)(\log\left(KM\sqrt{s}\right) + \log(\epsilon^{-1})), \end{aligned}$$

for some constant C (recall that we have assumed that $\log(s) \geq 1$). And this gives (by recalling that $\gamma = \epsilon/6$) that $\tilde{q}_k \geq \hat{C}\frac{\hat{m}_k}{N_k - N_{k-1}}(\log(s) + 1)$, for some constant $\hat{C}$. Thus, (7.15) implies that for $1 \leq l \leq r$,

$$\begin{aligned} 1 &\gtrsim (\log(s) + 1)\left(\sum_{k=1}^{r}\left(\frac{N_k - N_{k-1}}{m_k(\log(s) + 1)} - \frac{1}{\log(s) + 1}\right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k\right) \\ &\gtrsim (\log(s) + 1)\left(\sum_{k=1}^{r}\left(\tilde{q}_k^{-1} - 1\right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k\right), \end{aligned}$$

and this implies (7.89) and (7.87), given an appropriate choice of the constant C.

Stage 2: (7.14) $\Rightarrow$ (7.88) and (7.86). To show the assertion we must demonstrate that if, for $1 \leq k \leq r$,

$$1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N_k - N_{k-1}}{m_k} \cdot (\sum_{l=1}^{r} \kappa_{\mathbf{N},\mathbf{M}}(k,l)) \cdot \log\left(KM\sqrt{s}\right), \tag{7.93}$$

we obtain (7.88) and (7.86). To see this, note that by arguing as above via the fact that $q_k^1 = q_k^2 = \frac{1}{4}q_k$, and by (7.92), (7.93) and the choice of u in (7.62) we have that

$$\begin{aligned} 2(8(\lceil\log(\gamma^{-1}) + 3\lceil\log_2(8KM\sqrt{s})\rceil\rceil) - 2)\tilde{q}_k \geq q_k &= \frac{m_k}{N_k - N_{k-1}} \\ &\geq C \cdot (\sum_{l=1}^{r} \kappa_{\mathbf{N},\mathbf{M}}(k,l)) \cdot (\log(s\epsilon^{-1}) + 1) \cdot \log\left(KM\sqrt{s}\right) \\ &\geq C \cdot (\sum_{l=1}^{r} \kappa_{\mathbf{N},\mathbf{M}}(k,l)) \cdot (\log(s) + 1)\left(\log(\epsilon^{-1}) + \log\left(KM\sqrt{s}\right)\right), \end{aligned}$$

for some constant C. Thus, we have that for some appropriately chosen constant $\hat{C}$, $\tilde{q}_k \geq \hat{C} \cdot (\log(s) + 1) \cdot \sum_{l=1}^{r} \kappa_{\mathbf{N},\mathbf{M}}(k,l)$. So, (7.88) and (7.86) holds given an appropriately chosen C. This yields the last puzzle of the proof, and we are done. $\square$

Proof of Proposition 7.4. The proof is very close to the proof of Proposition 7.3 and we will simply point out the differences. The strategy of the proof is to show the validity of (i) and (ii), and the existence of a $\rho \in \mathrm{ran}(U^*(P_{\Omega_1} \oplus \ldots \oplus P_{\Omega_r}))$ that satisfies (iii)-(v) in Proposition 7.1 with probability exceeding $1 - \epsilon$.

Step I: The construction of ρ: The construction is almost identical to the construction in the proof of Proposition 7.3, except that

$$u = 8\lceil\log(\gamma^{-1}) + 3v\rceil, \qquad v = \lceil\log_2(8K\tilde{M}\sqrt{s})\rceil, \tag{7.94}$$

$$\alpha_1 = \alpha_2 = (2\log_2^{1/2}(4K\tilde{M}\sqrt{s}))^{-1}, \qquad \alpha_i = 1/2, \quad 3 \leq i \leq u,$$

as well as

$$\beta_1 = \beta_2 = \frac{1}{4}, \qquad \beta_i = \frac{1}{4}\log_2(4K\tilde{M}\sqrt{s}), \quad 3 \leq i \leq u,$$

and (7.66) gets changed to

$$\omega_i = \begin{cases} \omega_{i-1} \cup \{i\} & \text{if } \|(P_\Delta - P_\Delta U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta)Z_{i-1}\|_{l^\infty} \leq \alpha_i \|P_{\Delta_k} Z_{i-1}\|_{l^\infty}, \\ & \text{and } \|P_\Delta^\perp U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta Z_{i-1}\|_{l^\infty} \leq \beta_i \|Z_{i-1}\|_{l^\infty}, \\ \omega_{i-1} & \text{otherwise,} \end{cases}$$

the events B_i, $i = 1, 2$ in (7.67) get replaced by

$$\widetilde{B}_i : \qquad \|P_\Delta^\perp U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta Z_{i-1}\|_{l^\infty} \leq \beta_i \|Z_{i-1}\|_{l^\infty}, \qquad i = 1, 2.$$

and the second part of B_3 becomes

$$\max_{i \in \Delta^c} \| \left(q_1^{-1/2} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1/2} P_{\Omega_r} \right) U e_i \| \leq \sqrt{5/4}.$$

Step II: $B_5 \Rightarrow$ (**i**), (**ii**). This step is identical to Step II in the proof of Proposition 7.3.

Step III: $B_5 \Rightarrow$ (**iii**), (**iv**). Equation (7.69) gets changed to

$$\begin{aligned} \|P_\Delta^\perp \rho\|_{l^\infty} &\leq \sum_{i=1}^{v} \|P_\Delta^\perp U^*(\frac{1}{q_1^{\tau(i)}}P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}}P_{\Omega_r^{\tau(i)}})UP_\Delta Z_{\tau(i-1)}\|_{l^\infty} \\ &\leq \sum_{i=1}^{v} \beta_{\tau(i)} \|Z_{\tau(i-1)}\|_{l^\infty} \leq \sum_{i=1}^{v} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)} \\ &\leq \frac{1}{4}(1 + \frac{1}{2\log_2^{1/2}(a)} + \frac{\log_2(a)}{2^3\log_2(a)} + \ldots + \frac{1}{2^{v-1}}) \leq \frac{1}{2}, \qquad a = 4\tilde{M}K\sqrt{s}. \end{aligned}$$

Step IV: $B_5 \Rightarrow$ (**v**). This step is identical to Step IV in the proof of Proposition 7.3.

Step V: The strong balancing property, (7.18) and (7.19) $\Rightarrow \mathbb{P}(A_1^c \cup A_2^c \cup \widetilde{B}_1^c \cup \widetilde{B}_2^c \cup B_3^c) \leq 5\gamma$. We will start by bounding $\mathbb{P}(\widetilde{B}_1^c)$ and $\mathbb{P}(\widetilde{B}_2^c)$. Note that by Proposition 7.11 (ii) it follows that $\mathbb{P}(\widetilde{B}_1^c) \leq \gamma$ and $\mathbb{P}(\widetilde{B}_2^c) \leq \gamma$ as long as the strong balancing property is satisfied and

$$1 \gtrsim \Lambda \cdot \log\left(\frac{4}{\gamma}(\tilde{\theta} - s)\right), \qquad 1 \gtrsim \Upsilon \cdot \log\left(\frac{4}{\gamma}(\tilde{\theta} - s)\right) \tag{7.95}$$

where $\tilde{\theta} = \tilde{\theta}(\{q_k^i\}_{k=1}^r, 1/8, \{N_k\}_{k=1}^r, s, M)$ for $i = 1, 2$ and where $\tilde{\theta}$ is defined in Proposition 7.11 (ii) and Λ and Υ are defined in (7.73) and (7.74). Note that it is easy to see that we have

$$\left|\left\{ j \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\ldots,M\}, \quad |\Gamma_1| = s \\ \Gamma_{2,j} \subset \{N_{j-1}+1,\ldots,N_j\}, \quad j=1,\ldots,r}} \|P_{\Gamma_1} U^*((q_1^i)^{-1}P_{\Gamma_{2,1}} \oplus \ldots \oplus (q_r^i)^{-1}P_{\Gamma_{2,r}})Ue_j\| > \frac{1}{8\sqrt{s}} \right\}\right| \leq \tilde{M},$$

where

$$\tilde{M} = \min\{i \in \mathbb{N} : \max_{j \geq i} \|P_N U P_{\{j\}}\| \leq 1/(K32\sqrt{s})\},$$

and this follows from the choice in (7.63) where $q_k^1 = q_k^2 = \frac{1}{4}q_k$ for $1 \leq k \leq r$. Thus, it immediately follows that (7.18) and (7.19) imply (7.95). To bound $\mathbb{P}(B_3^c)$, we first deduce as in Step V of the proof of Proposition 7.3 that

$$\mathbb{P}\left(\|P_\Delta U^*(\frac{1}{q_1}P_{\Omega_1} \oplus \ldots \oplus \frac{1}{q_r}P_{\Omega_r})UP_\Delta - P_\Delta\| > 1/4, \right) \leq \gamma/2$$

when the strong balancing property and (7.18) holds. For the second part of B_3, we know from the choice of $\tilde{M}$ that

$$\max_{i \geq \tilde{M}} \| \left(q_1^{-1/2} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1/2} P_{\Omega_r} \right) U e_i \| \leq \sqrt{\frac{5}{4}}$$

and we may deduce from Proposition 7.13 that

$$\mathbb{P}\left(\max_{i\in\Delta^c\cap\{1,\dots,\tilde{M}\}}\|\left(q_1^{-1/2}P_{\Omega_1}\oplus\dots\oplus q_r^{-1/2}P_{\Omega_r}\right)Ue_i\|>\sqrt{5/4}\right)\leq\frac{\gamma}{2},$$

whenever

$$1\gtrsim\log\left(\frac{2\tilde{M}}{\gamma}\right)\cdot\max_{1\leq k\leq r}\left\{\left(\frac{N_k-N_{k-1}}{m_k}-1\right)\mu_{\mathbf{N},\mathbf{M}}(k,l)\right\},\quad l=1,\dots,r-1,\infty,$$

which is true whenever (7.18) holds, since by a similar argument to (7.78),

$$\kappa_{\mathbf{N},\mathbf{M}}(k,\infty)+\sum_{j=1}^{r-1}\kappa_{\mathbf{N},\mathbf{M}}(k,j)\geq\mu_{\mathbf{N},\mathbf{M}}(k,l),\qquad l=1,\dots,r-1,\infty.$$

Thus, $\mathbb{P}(B_3^c)\leq\gamma$. As for bounding $\mathbb{P}(A_1^c)$ and $\mathbb{P}(A_2^c)$, observe that by the strong balancing property $\tilde{M}\geq M$, thus this is done exactly as in Step V of the proof of Proposition 7.3.

Step VI: The strong balancing property, (7.18) and (7.19) $\Rightarrow$ $\mathbb{P}(B_4^c)\leq\gamma$. To see this, define the random variables $X_1,\dots X_{u-2}$ as in (7.79). Let π be defined as in Step VI of the proof of Proposition 7.3. Then it suffices to show that (7.18) and (7.19) imply that for $l=1,\dots\binom{u-2}{u-v}$ and $j=0,\dots,u-v-2$, we have

$$\begin{aligned}\frac{1}{4}&\geq\mathbb{P}(X_{\pi(l)_{u-v-j}}=1\,|\,X_{\pi(l)_1}=1,\dots,X_{\pi(l)_{u-v-(j+1)}}=1),\\ \frac{1}{4}&\geq\mathbb{P}(X_{\pi(l)_1}=1).\end{aligned}\tag{7.96}$$

Claim: The strong balancing property, (7.18) and (7.19) $\Rightarrow$ (7.96). To prove the claim we first observe that $X_j=0$ when

$$\|(P_\Delta-P_\Delta U^*(\frac{1}{q_1^i}P_{\Omega_1^i}\oplus\dots\oplus\frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta)Z_{i-1}\|_{l^\infty}\leq\frac{1}{2}\|Z_{i-1}\|_{l^\infty}$$

$$\|P_\Delta^\perp U^*(\frac{1}{q_1^i}P_{\Omega_1^i}\oplus\dots\oplus\frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta Z_{i-1}\|_{l^\infty}\leq\frac{1}{4}\log_2(4K\tilde{M}\sqrt{s})\|Z_{i-1}\|_{l^\infty},\qquad i=j+2.$$

Thus, by again recalling from (7.63) that $q_k^3=q_k^4=\dots=q_k^u=\tilde{q}_k$, $1\leq k\leq r$, and by choosing $\tilde{\gamma}=1/4$ in (7.48) in Proposition 7.12 and $\tilde{\gamma}=1/4$ in (ii) in Proposition 7.11, we conclude that (7.96) follows when the strong balancing property is satisfied as well as (7.86) and (7.87). and

$$\frac{\log_2(4K\tilde{M}\sqrt{s})}{\log\left(16(\tilde{M}-s)\right)}\geq C_2\cdot\tilde{q}_k^{-1}\cdot\left(\sum_{l=1}^{r-1}\kappa_{\mathbf{N},\mathbf{M}}(k,l)+\kappa_{\mathbf{N},\mathbf{M}}(k,\infty)\right),\quad k=1,\dots,r\tag{7.97}$$

$$\frac{\log_2(4K\tilde{M}\sqrt{s})}{\log\left(16(\tilde{M}-s)\right)}\geq C_2\cdot\left(\sum_{l=1}^{r}\left(\tilde{q}_k^{-1}-1\right)\cdot\mu_{\mathbf{N},\mathbf{M}}(k,l)\cdot\tilde{s}_k\right),\quad l=1,\dots,r-1,\infty\tag{7.98}$$

for $K=\max_{1\leq k\leq r}(N_k-N_{k-1})/m_k$. for some constants C_1 and C_2. Thus, to prove the claim we must demonstrate that (7.18) and (7.19) $\Rightarrow$ (7.86), (7.87), (7.97) and (7.98). This is done by repeating Stage 1 and Stage 2 in Step VI of the proof of Proposition 7.3 almost verbatim, except replacing M by $\tilde{M}$. $\square$

7.4 Proof of Theorem 6.2

Throughout this section, we use the notation

$$\hat{f}(\xi)=\int_{\mathbb{R}}f(x)e^{-ix\xi}\mathrm{d}x,\tag{7.99}$$

to denote the Fourier transform of a function $f\in L^1(\mathbb{R})$.

7.4.1 Setup

We first introduce the wavelet sparsity and Fourier sampling bases that we consider, and in particular, their orderings. Consider an orthonormal basis of compactly supported wavelets with an MRA [20, 21]. For simplicity, suppose that $\mathrm{supp}(\Psi) = \mathrm{supp}(\Phi) = [0, a]$ for some $a \geq 1$, where Ψ and Φ are the mother wavelet and scaling function respectively. For later use, we recall the following three properties of any such wavelet basis:

1. There exist $\alpha \geq 1$, C_Ψ and $C_\Phi > 0$, such that
$$\left|\hat{\Phi}(\xi)\right| \leq \frac{C_\Phi}{(1+|\xi|)^\alpha}, \quad \left|\hat{\Psi}(\xi)\right| \leq \frac{C_\Psi}{(1+|\xi|)^\alpha}. \tag{7.100}$$
See [21, Eqn. (7.1.4)]. We will denote $\max\{C_\Psi, C_\Phi\}$ by $C_{\Phi,\Psi}$.

2. Ψ has $v \geq 1$ vanishing moments and $\hat{\Psi}(z) = (-iz)^v \theta_\Psi(z)$ for some bounded function θ_Ψ (see [47, p.208 & p.284]).

3. $\|\hat{\Phi}\|_{L^\infty}, \|\hat{\Psi}\|_{L^\infty} \leq 1$.

Remark 7.1 The three properties above are based on the standard setup for an MRA, however, we also consider a stronger assumption on the decay of the Fourier transform of derivatives of the scaling function and the mother wavelet. In particular, in addition, we sometimes assume that for $C > 0$ and $\alpha \geq 1.5$,
$$\left|\hat{\Phi}^{(k)}(\xi)\right| \leq \frac{C}{(1+|\xi|)^\alpha}, \quad \left|\hat{\Psi}^{(k)}(\xi)\right| \leq \frac{C}{(1+|\xi|)^\alpha}, \quad \xi \in \mathbb{R}, \quad k = 0, 1, 2, \tag{7.101}$$
where $\hat{\Phi}^{(k)}$ and $\hat{\Psi}^{(k)}$ denotes the k^{th} derivative of the Fourier transform of Φ and Ψ respectively. As is evident from Theorem 6.2, the faster decay, the closer the relationship between N and M in the balancing property gets to linear. Also, faster decay and more vanishing moments yield a closer to block-diagonal structure of the matrix U.

We now wish to construct a wavelet basis for the compact interval $[0, a]$. The most standard approach is to consider the following collection of functions
$$\Lambda_a = \{\Phi_k, \Psi_{j,k} : \mathrm{supp}(\Phi_k)^o \cap [0,a] \neq \emptyset,\ \mathrm{supp}(\Psi_{j,k})^o \cap [0,a] \neq \emptyset, j \in \mathbb{Z}_+, k \in \mathbb{Z}, \},$$
where $\Phi_k = \Phi(\cdot - k)$, and $\Psi_{j,k} = 2^{\frac{j}{2}}\Psi(2^j \cdot -k)$. (the notation K^o denotes the interior of a set $K \subseteq \mathbb{R}$). This gives
$$\left\{f \in \mathrm{L}^2(\mathbb{R}) : \mathrm{supp}(f) \subseteq [0,a]\right\} \subseteq \overline{\mathrm{span}\{\varphi : \varphi \in \Lambda_a\}} \subseteq \left\{f \in \mathrm{L}^2(\mathbb{R}) : \mathrm{supp}(f) \subseteq [-T_1, T_2]\right\},$$
where $T_1, T_2 > 0$ are such that $[-T_1, T_2]$ contains the support of all functions in Λ_a. Note that the inclusions may be proper (but not always, as is the case with the Haar wavelet). It is easy to see that
$$\Psi_{j,k} \notin \Lambda_a \Longleftrightarrow \frac{a+k}{2^j} \leq 0, \quad a \leq \frac{k}{2^j},$$
$$\Phi_k \notin \Lambda_a \Longleftrightarrow a + k \leq 0, \quad a \leq k,$$
and therefore
$$\Lambda_a = \{\Phi_k : |k| = 0, \ldots, \lceil a \rceil - 1\} \cup \{\Psi_{j,k} : j \in \mathbb{Z}_+, k \in \mathbb{Z}, -\lceil a \rceil < k < 2^j \lceil a \rceil\}.$$
We order Λ_a in increasing order of wavelet resolution as follows:
$$\begin{aligned}&\{\Phi_{-\lceil a \rceil+1}, \ldots, \Phi_{-1}, \Phi_0, \Phi_1, \ldots, \Phi_{\lceil a \rceil-1},\\ &\Psi_{0,-\lceil a \rceil+1}, \ldots, \Psi_{0,-1}, \Psi_{0,0}, \Psi_{0,1}, \ldots, \Psi_{0,\lceil a \rceil-1}, \Psi_{1,-\lceil a \rceil+1}, \ldots\},\end{aligned} \tag{7.102}$$
and then we finally denote the functions according to this ordering by $\{\varphi_j\}_{j \in \mathbb{N}}$. By the definition of Λ_a, we let $T_1 = \lceil a \rceil - 1$ and $T_2 = 2\lceil a \rceil - 1$. Finally, for $R \in \mathbb{N}$, let $\Lambda_{R,a}$ contain all wavelets in Λ_a with resolution less than R, so that
$$\Lambda_{R,a} = \{\varphi \in \Lambda_a : \varphi = \Psi_{j,k},\ 0 \leq j < R,\ \text{or } \varphi = \Phi_k\}. \tag{7.103}$$

We also denote the size of $\Lambda_{R,a}$ by W_R. It is easy to verify that

$$W_R = 2^R \lceil a \rceil + (R+1)(\lceil a \rceil - 1). \tag{7.104}$$

Having constructed an orthonormal wavelet system for $[0, a]$, we now introduce the appropriate Fourier sampling basis. We must sample at a rate that is at least that of the Nyquist rate. Hence we let $\omega \leq 1/(T_1 + T_2)$ be the *sampling density* (note that $1/(T_1 + T_2)$ is the Nyquist criterion for functions supported on $[-T_1, T_2]$). For simplicity, we assume throughout that

$$\omega \in (0, 1/(T_1 + T_2)), \quad \omega^{-1} \in \mathbb{N}, \tag{7.105}$$

and remark that this assumption is an artefact of our proofs and is not necessary in practice. The Fourier sampling vectors are now defined as follows.

$$\psi_j(x) = \sqrt{\omega} e^{-2\pi i j \omega x} \chi_{[-T_1/(\omega(T_1+T_2)), T_2/(\omega(T_1+T_2))]}(x), \qquad j \in \mathbb{Z}. \tag{7.106}$$

This gives an orthonormal sampling basis for the space $\{f \in \mathrm{L}^2(\mathbb{R}) : \mathrm{supp}(f) \subseteq [-T_1, T_2]\}$. Since Λ_a is an orthonormal system in for this space, it follows that the infinite matrix

$$U = \begin{pmatrix} u_{11} & u_{12} & u_{13} & \cdots \\ u_{21} & u_{22} & u_{23} & \cdots \\ u_{31} & u_{32} & u_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \qquad u_{ij} = \langle \varphi_j, \tilde{\psi}_i \rangle, \tag{7.107}$$

is an isometry, where $\{\varphi_j\}_{j\in\mathbb{N}}$ represents the wavelets ordered according to (7.102) and $\{\tilde{\psi}_j\}_{j\in\mathbb{N}}$ is the standard ordering of the Fourier basis (7.106) over $\mathbb{N}$ ($\tilde{\psi}_1 = \psi_0$, $\tilde{\psi}_{2n} = \psi_n$ and $\tilde{\psi}_{2n+1} = \psi_{-n}$). With slight abuse of notation it is this ordering that we are using in Theorem 6.2.

7.4.2 Some preliminary estimates

Throughout this section, we assume the setup and notation introduced above.

Theorem 7.15. *Let U be the matrix of the Fourier/wavelets pair introduced in (7.107) with sampling density ω as in (7.105) . Suppose that Φ and Ψ satisfy the decay estimate (7.100) with $\alpha \geq 1$ and that Ψ has $v \geq 1$ vanishing moments. Then the following holds.*

(i) We have $\mu(U) \geq \omega$.

(ii) We have that

$$\mu(P_N^{\perp} U) \leq \frac{C_{\Phi,\Psi}^2}{\pi N (2\alpha - 1)(1 + 1/(2\alpha - 1))^{2\alpha}}, \qquad N \in \mathbb{N},$$
$$\mu(U P_N^{\perp}) \leq \|\Psi\|_{L^\infty}^2 \frac{4\omega \lceil a \rceil}{N}, \qquad N \geq 2\lceil a \rceil + 2(\lceil a \rceil - 1),$$

and consequently $\mu(P_N^{\perp} U), \mu(U P_N^{\perp}) = \mathcal{O}\left(N^{-1}\right)$.

(iii) If the wavelet and scaling function satisfy the decay estimate (7.100) with $\alpha > 1/2$, then, for R and N such that $\omega^{-1} 2^R \leq N$ and $M = |\Lambda_{R,a}|$ (recall the definition of $\Lambda_{R,a}$ from (7.103)),

$$\mu(P_N^{\perp} U P_M) \leq \frac{C_{\Phi,\Psi}^2}{\pi^{2\alpha} \omega^{2\alpha - 1}} (2^{R-1} N^{-1})^{2\alpha - 1} N^{-1}.$$

(iv) If the wavelet has $v \geq 1$ vanishing moments, $\omega^{-1} 2^R \geq N$ and $M = |\Lambda_{R,a}|$ with $R \geq 1$, then

$$\mu(P_N U P_M^{\perp}) \leq \frac{\omega}{2^R} \cdot \left(\frac{\pi \omega N}{2^R}\right)^{2v} \cdot \|\theta_\Psi\|_{L^\infty}^2,$$

where θ_Ψ is the function such that $\hat{\Psi}(z) = (-iz)^v \theta_\Psi(z)$ (see above).

Proof. Note that $\mu(U) \geq |\langle \Phi, \psi_0 \rangle|^2 = \omega \left|\hat{\Phi}(0)\right|^2$, moreover, it is known that $\hat{\Phi}(0) = 1$ [37, Ch. 2, Thm. 1.7]. Thus, (i) follows.

To show (ii), let $R \in \mathbb{N}$, $-\lceil a \rceil < j < 2^R \lceil a \rceil$ and $k \in \mathbb{Z}$. Then, by the choice of j, we have that $\Psi_{R,j}$ is supported on $[-T_1, T_2]$. Also, $\psi_k(x) = \sqrt{\omega} e^{-2\pi i k \omega x} \chi_{[-T_1/(\omega(T_1+T_2)), T_2/(\omega(T_1+T_2))]}(x)$. Thus, since by (7.105) we have $\omega \in (0, 1/(T_1+T_2))$, it follows that

$$\begin{aligned} \langle \Psi_{R,j}, \psi_k \rangle &= \sqrt{\omega} \int_{-\frac{T_1}{\omega(T_1+T_2)}}^{\frac{T_2}{\omega(T_1+T_2)}} \Psi_{R,j}(x) e^{2\pi i \omega k x} \mathrm{d}x \\ &= \sqrt{\omega} \hat{\Psi}_{R,j}(-2\pi\omega k) = \sqrt{\frac{\omega}{2^R}} \hat{\Psi}\left(\frac{-2\pi k \omega}{2^R}\right) e^{2\pi i \omega k j / 2^R}. \end{aligned} \tag{7.108}$$

Also, similarly, it follows that

$$\langle \Phi_j, \psi_k \rangle = \sqrt{\omega} \int_{-\frac{T_1}{\omega(T_1+T_2)}}^{\frac{T_2}{\omega(T_1+T_2)}} \Phi_j(x) e^{2\pi i \omega k x} \mathrm{d}x = \sqrt{\omega} \hat{\Phi}_j\left(-2\pi k \omega\right) = \sqrt{\omega} \hat{\Phi}\left(-2\pi k \omega\right) e^{2\pi i \omega k j}. \tag{7.109}$$

Thus, the decay estimate in (7.100) yields

$$\begin{aligned} \mu(P_N^\perp U) &\leq \sup_{|k| \geq \frac{N}{2}} \max_{\varphi \in \Lambda_a} |\langle \varphi, \psi_k \rangle|^2 \\ &= \max \left\{ \sup_{|k| \geq \frac{N}{2}} \max_{R \in \mathbb{Z}_+} \frac{\omega}{2^R} \left| \hat{\Psi}\left(\frac{-2\pi\omega k}{2^R}\right) \right|^2, \omega \sup_{|k| \geq \frac{N}{2}} \left| \hat{\Phi}\left(-2\pi\omega k\right) \right|^2 \right\} \\ &\leq \max_{|k| \geq \frac{N}{2}} \max_{R \in \mathbb{Z}_+} \frac{\omega}{2^R} \frac{C_{\Phi,\Psi}^2}{(1 + |2\pi\omega k 2^{-R}|)^{2\alpha}} \leq \max_{R \in \mathbb{Z}_+} \frac{\omega}{2^R} \frac{C_{\Phi,\Psi}^2}{(1 + |\pi\omega N 2^{-R}|)^{2\alpha}}. \end{aligned}$$

The function $f(x) = x^{-1}(1 + \pi\omega N/x)^{-2\alpha}$ on $[1, \infty)$ satisfies $f'(\pi\omega N(2\alpha - 1)) = 0$. Hence

$$\mu(P_N^\perp U) \leq \frac{C_{\Phi,\Psi}^2}{\pi N (2\alpha - 1)(1 + 1/(2\alpha - 1))^{2\alpha}},$$

which gives the first part of (ii). For the second part, we first recall the definition of W_R for $R \in \mathbb{N}$ from (7.104). Then, given any $N \in \mathbb{N}$ such that $N \geq W_1 = 2\lceil a \rceil + 2(\lceil a \rceil - 1)$, let R be such that $W_R \leq N < W_{R+1}$. Then, for each $n \geq N$, there exists some $j \geq R$ and $l \in \mathbb{Z}$ such that the n^{th} element via the ordering (7.102) is $\varphi_n = \Psi_{j,l}$ (note that we only need $\Psi_{j,l}$ here and not Φ_j as we have chosen $N \geq W_1$). Hence, by using (7.108),

$$\begin{aligned} \mu(U P_N^\perp) &= \max_{n \geq N} \max_{k \in \mathbb{Z}} |\langle \varphi_n, \psi_k \rangle|^2 = \max_{j \geq R} \max_{k \in \mathbb{Z}} \frac{\omega}{2^j} \left| \hat{\Psi}\left(\frac{-2\pi\omega k}{2^j}\right) \right|^2 \\ &\leq \|\hat{\Psi}\|_{L^\infty}^2 \frac{\omega}{2^R} \leq 4 \|\hat{\Psi}\|_{L^\infty}^2 \frac{\omega \lceil a \rceil}{N}, \end{aligned}$$

where the last line follows because $N < W_{R+1} = 2^{R+1}\lceil a \rceil + (R+2)(\lceil a \rceil - 1)$ implies that

$$2^{-R} < \frac{1}{N}\left(2\lceil a \rceil + (R+2)(\lceil a \rceil - 1)2^{-R}\right) \leq \frac{4\lceil a \rceil}{N}.$$

This concludes the proof of (ii).

To show (iii), let R and N be such that $\omega^{-1}2^R \leq N$ and $M = |\Lambda_{R,a}|$. Observe that (7.108) and (7.109) together with the decay estimate in (7.100) yield

$$\begin{aligned} \mu(P_N^\perp U P_{W_R}) &\leq \max_{|k| \geq \frac{N}{2}} \max_{\varphi \in \Lambda_{R,a}} |\langle \varphi, \psi_k \rangle|^2 \\ &= \max \left\{ \max_{|k| \geq \frac{N}{2}} \max_{j < R} \frac{\omega}{2^j} \left| \hat{\Psi}\left(\frac{-2\pi\omega k}{2^j}\right) \right|^2, \max_{|k| \geq \frac{N}{2}} \left| \hat{\Phi}\left(-2\pi\omega k\right) \right|^2 \right\} \\ &\leq \max_{|k| \geq \frac{N}{2}} \max_{j < R} \frac{\omega}{2^j} \frac{C_{\Phi,\Psi}^2}{(1 + |2\pi\omega k 2^{-j}|)^{2\alpha}} \leq \max_{k \geq \frac{N}{2}} \max_{j < R} \frac{C_{\Phi,\Psi}^2}{\pi^{2\alpha}\omega^{2\alpha-1}} \frac{2^{j(2\alpha-1)}}{(2k)^{2\alpha}} \\ &= \frac{C_{\Phi,\Psi}^2}{\pi^{2\alpha}\omega^{2\alpha-1}} (2^{R-1} N^{-1})^{2\alpha-1} N^{-1}, \end{aligned}$$

and this colludes the proof of (iii).

To show (iv), first note that because $R \geq 1$, for all $n > W_R$, $\varphi_n = \Psi_{j,k}$ for some $j \geq 0$ and $k \in \mathbb{Z}$. Then, recalling the properties of Daubechies wavelets with v vanishing moments, and by using (7.108) we get that

$$\begin{aligned}\mu(P_N U P_{W_R}^{\perp}) &= \max_{n > W_R} \max_{|k| \leq \frac{N}{2}} |\langle \varphi_n, \psi_k \rangle|^2 = \max_{j \geq R} \max_{|k| \leq \frac{N}{2}} \frac{\omega}{2^j} \left| \hat{\Psi}\left(\frac{-2\pi\omega k}{2^j} \right) \right|^2 \\ &\leq \frac{\omega}{2^R} \cdot \left(\frac{\pi \omega N}{2^R} \right)^{2v} \cdot \|\theta_\Psi\|_{L^\infty}^2,\end{aligned}$$

as required. □

Corollary 7.16. *Let* $\mathbf{N}$ *and* $\mathbf{M}$ *be as in Theorem 6.2 and recall the definition of* $\mu_{\mathbf{N},\mathbf{M}}(k,j)$ *in (4.2). Suppose that* Φ *and* Ψ *satisfy the decay estimate (7.100) with* $\alpha \geq 1$ *and that* Ψ *has* $v \geq 1$ *vanishing moments. Then,*

$$\text{for } k \geq 2, \quad \mu_{\mathbf{N},\mathbf{M}}(k,j) \leq B_{\Phi,\Psi} \cdot \begin{cases} \frac{\sqrt{\omega}}{\sqrt{N_{k-1} 2^{R_{j-1}}}} \cdot \left(\frac{\omega N_k}{2^{R_{j-1}}} \right)^v & j \geq k+1 \\ \frac{1}{N_{k-1}} \left(\frac{2^{R_{j-1}}}{\omega N_{k-1}} \right)^{\alpha - 1/2} & j \leq k-1 \\ \frac{1}{N_{k-1}} & j = k, \end{cases} \tag{7.110}$$

$$\text{for } k \geq 2, \quad \mu_{\mathbf{N},\mathbf{M}}(k,\infty) \leq B_{\Phi,\Psi} \cdot \begin{cases} \frac{\sqrt{\omega}}{\sqrt{N_{k-1} 2^{R_{r-1}}}} \cdot \left(\frac{\omega N_k}{2^{R_{r-1}}} \right)^v & k \leq r-1 \\ \frac{1}{N_{r-1}} & k = r, \end{cases} \tag{7.111}$$

$$\mu_{\mathbf{N},\mathbf{M}}(1,j) \leq B_{\Phi,\Psi} \cdot \begin{cases} \frac{\sqrt{\omega}}{\sqrt{2^{R_{j-1}}}} \cdot \left(\frac{\omega N_1}{2^{R_{j-1}}} \right)^v & j \geq 2 \\ 1 & j = 1, \end{cases} \tag{7.112}$$

$$\mu_{\mathbf{N},\mathbf{M}}(1,\infty) \leq B_{\Phi,\Psi} \cdot \frac{\sqrt{\omega}}{\sqrt{2^{R_{r-1}}}} \cdot \left(\frac{\omega N_1}{2^{R_{r-1}}} \right)^v, \tag{7.113}$$

where $B_{\Phi,\Psi}$ *is a constant which depends only on* Φ *and* Ψ *and* $R_0 = 0$.

Proof. Throughout this proof, $B_{\Phi,\Psi}$ is a constant which depends only on Φ and Ψ, although its value may change from instance to instance. Note that

$$\begin{aligned}\mu_{\mathbf{N},\mathbf{M}}(k,j) &= \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_j}^{M_{j-1}}) \cdot \mu(P_{N_k}^{N_{k-1}} U)} \\ &\leq B_{\Phi,\Psi} N_{k-1}^{-1/2} \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_j}^{M_{j-1}})}, \qquad k \geq 2,\ j \in \{1,\ldots,r\},\end{aligned} \tag{7.114}$$

since we have $\mu(P_{N_{k-1}}^{\perp} U) \leq B_{\Phi,\Psi} N_{k-1}^{-1}$ by (ii) of Theorem 7.15. Also, clearly

$$\mu_{\mathbf{N},\mathbf{M}}(1,j) = \sqrt{\mu(P_{N_1}^{N_0} U P_{M_j}^{M_{j-1}}) \cdot \mu(P_{N_1}^{N_0} U)} \leq B_{\Phi,\Psi} \sqrt{\mu(P_{N_1}^{N_0} U P_{M_j}^{M_{j-1}})}, \tag{7.115}$$

for $j \in \{1,\ldots,r\}$. Thus, for $k \geq 2$, it follows that $\mu_{\mathbf{N},\mathbf{M}}(k,k) \leq \mu(P_{N_{k-1}}^{\perp} U) \leq B_{\Phi,\Psi} \frac{1}{N_{k-1}}$, yielding the last part of (7.110). Also, the last part of (7.112) is clear from (7.115).

As for the middle part of (7.110), note that for $k \geq 2$, and with $j \leq k-1$, we may use (iii) of Theorem 7.15 to obtain

$$\sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_j}^{M_{j-1}})} \leq \sqrt{\mu(P_{N_{k-1}}^{\perp} U P_{M_j})} \leq B_{\Phi,\Psi} \cdot \frac{1}{\sqrt{N_{k-1}}} \left(\frac{2^{R_{j-1}}}{\omega N_{k-1}} \right)^{\alpha - 1/2},$$

and thus, in combination with (7.114), we obtain the $j \leq k-1$ part of (7.110). Observe that if $k \in \{1,\ldots,r\}$ and $j \geq k+1$, then by applying (iv) of Theorem 7.15, we obtain

$$\sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_j}^{M_{j-1}})} \leq \sqrt{\mu(P_{N_k} U P_{M_{j-1}}^{\perp})} \leq B_{\Phi,\Psi} \cdot \frac{\sqrt{\omega}}{\sqrt{2^{R_{j-1}}}} \cdot \left(\frac{\omega N_k}{2^{R_{j-1}}} \right)^v. \tag{7.116}$$

Thus, by combining (7.116) with (7.114), we obtain the $j \geq k+1$ part of (7.110). Also, by combining (7.116) with (7.114) we get the $j \geq 2$ part of (7.112). Finally, recall that

$$\mu_{\mathbf{N},\mathbf{M}}(k,\infty) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_{r-1}}^{\perp}) \cdot \mu(P_{N_{k-1}}^{\perp} U)}$$

and similarly to the above, (7.111) and (7.113) are direct consequences of parts (ii) and (iv) of Theorem 7.15. ☐

The following lemmas inform us of the range of Fourier samples required for accurate reconstruction of wavelet coefficients. Specifically, Lemma 7.17 will provide a quantitative understanding of the balancing property, whilst Lemma 7.18 and Lemma 7.19 will be used in bounding the relative sparsity terms.

Lemma 7.17 ([50, Corollary 5.4]). *Consider the setup in §7.4.1. Let the sampling density ω be such that $\omega^{-1} \in \mathbb{N}$ and suppose that there exists $C_\Phi, C_\Psi > 0$ and $\alpha \geq 1.5$ such that*

$$\left|\hat{\Phi}^{(k)}(\xi)\right| \leq \frac{C_\Phi}{(1+|\xi|)^\alpha}, \quad \left|\hat{\Psi}^{(k)}(\xi)\right| \leq \frac{C_\Psi}{(1+|\xi|)^\alpha}, \quad \xi \in \mathbb{R}, \qquad k = 0,1,2.$$

Then given $\gamma \in (0,1)$, we have that $\|P_M U^* P_N U P_M - P_M\|_{l^\infty \to l^\infty} \leq \gamma$ wherever $N \geq C\gamma^{-1/(2\alpha-1)} M$ and $\left\|P_M^{\perp} U^* P_N U P_M\right\|_{l^\infty \to l^\infty} \leq \gamma$ wherever $N \geq C\gamma^{-1/(\alpha-1)} M$ where C is some constant independent of N but dependent on C_Φ, C_Ψ and ω.

Lemma 7.18 ([50, Lemma 5.1]). *Let φ_k denote the k^{th} wavelet via the ordering in (7.102). Let $R \in \mathbb{N}$ and $M \leq W_R$ be such that $\{\varphi_j : j \leq M\} \subset \Lambda_{R,a}$, where W_R and $\Lambda_{R,a}$ are defined in (7.104) and (7.103) respectively. Also, let the sampling density ω be such that $\omega^{-1} \in \mathbb{N}$. Then for any $\gamma \in (0,1)$, we have that $\left\|P_N^{\perp} U P_M\right\| \leq \gamma$, whenever N is such that*

$$N \geq \omega^{-1} \left(\frac{4C_\Phi^2}{(2\pi)^{2\alpha} \cdot (2\alpha-1)}\right)^{\frac{1}{2\alpha-1}} \cdot 2^{R+1} \cdot \gamma^{-\frac{2}{2\alpha-1}}$$

and C_Φ is a constant depending on Φ.

Lemma 7.19. *Let φ_k denote the k^{th} wavelet the ordering in (7.102). Let $R_1, R_2 \in \mathbb{N}$ with $R_2 > R_1$, and $M_1, M_2 \in \mathbb{N}$ with $M_2 > M_1$ be such that*

$$\{\varphi_j : M_2 \geq j > M_1\} \subset \Lambda_{R_2,a} \setminus \Lambda_{R_1,a},$$

where $\Lambda_{R_i,a}$ is defined in (7.103). Then for any $\gamma \in (0,1)$

$$\left\|P_N U P_{M_2}^{M_1}\right\| \leq \frac{\pi^2}{4} \|\theta_\Psi\|_{L^\infty} \cdot (2\pi\gamma)^v \cdot \sqrt{\frac{1-2^{2v(R_1-R_2)}}{1-2^{-2v}}}$$

whenever N is such that $N \leq \gamma\omega^{-1} 2^{R_1}$.

Proof. Let $\eta \in l^2(\mathbb{N})$ be such that $\|\eta\| = 1$. Note that, by the definition of U in (7.107), it follows that

$$\|P_N U P_{M_2}^{M_1} \eta\|^2 \leq \sum_{|k| \leq N/2} \left| \langle \psi_k, \sum_{j=M_1+1}^{M_2} \eta_j \varphi_j \rangle \right|^2 \leq \sum_{|k| \leq N/2} \left| \langle \psi_k, \sum_{l=R_1}^{R_2-1} \sum_{j \in \Delta_l} \eta_{\rho(l,j)} \Psi_{l,j} \rangle \right|^2,$$

where we have defined

$$\Delta_l = \{j \in \mathbb{Z} : \Psi_{l,j} \in \Lambda_{l+1,a} \setminus \Lambda_{l,a}\}, \quad \rho : \{(l,\Delta_l)\}_{l \in \mathbb{N}} \to \mathbb{N} \setminus \{1,\ldots,|\Lambda_{1,a}|\}$$

to be the bijection such that $\varphi_{\rho(l,j)} = \Psi_{l,j}$. Now, observe that we may argue as in the proof of Theorem 7.15 and use (7.108) to deduce that given $l \in \mathbb{N}$, $-\lceil a \rceil < j < 2^l \lceil a \rceil$ and $k \in \mathbb{Z}$, we have that $\langle \Psi_{l,j}, \psi_k \rangle = \sqrt{\frac{\omega}{2^l}} \hat{\Psi}\left(-\frac{2\pi\omega k}{2^l}\right) e^{2\pi i \omega j k}$. Hence, it follows that

$$\sum_{|k| \leq N/2} \left| \langle \psi_k, \sum_{l=R_1}^{R_2-1} \sum_{j \in \Delta_l} \eta_{\rho(l,j)} \Psi_{l,j} \rangle \right|^2 = \sum_{|k| \leq N/2} \left| \sum_{l=R_1}^{R_2-1} \frac{\sqrt{\omega}}{\sqrt{2^l}} \sum_{j \in \Delta_l} \eta_{\rho(l,j)} \hat{\Psi}\left(-\frac{2\pi\omega k}{2^l}\right) e^{2\pi i \omega j k/2^l} \right|^2,$$

which again gives us that

$$
\begin{aligned}
\|P_N U P_{M_2}^{M_1}\eta\|^2 &\leq \sum_{|k|\leq N/2} \left| \sum_{l=R_1}^{R_2-1} \frac{\sqrt{\omega}}{\sqrt{2^l}} \hat{\Psi}\left(-\frac{2\pi\omega k}{2^l}\right) f^{[l]}\left(\frac{\omega k}{2^l}\right) \right|^2 \\
&\leq \sum_{|k|\leq N/2} \sum_{l=R_1}^{R_2-1} \left|\hat{\Psi}\left(-\frac{2\pi\omega k}{2^l}\right)\right|^2 \cdot \sum_{l=R_1}^{R_2-1} \left| \frac{\sqrt{\omega}}{\sqrt{2^l}} f^{[l]}\left(\frac{\omega k}{2^l}\right)\right|^2 \\
&\leq \sum_{l=R_1}^{R_2-1} \max_{|k|\leq N/2} \left|\hat{\Psi}\left(-\frac{2\pi\omega k}{2^l}\right)\right|^2 \cdot \sum_{l=R_1}^{R_2-1} \sum_{|k|\leq N/2} \frac{\omega}{2^l} \left| f^{[l]}\left(\frac{\omega k}{2^l}\right)\right|^2,
\end{aligned}
\tag{7.117}
$$

where $f^{[l]}(z) = \sum_{j\in\Delta_l} \eta_{\rho(l,j)} e^{2\pi i z j}$. Let $H = \chi_{[0,1)}$ and, for $l \in \mathbb{N}$, $-\lceil a\rceil < j < 2^j\lceil a\rceil$, define $H_{l,j} = 2^{\frac{l}{2}} H(2^l\cdot - j)$. By the choice of j, we have that $H_{l,j}$ is supported on $[-T_1, T_2]$. Also, since by (7.105) we have $\omega \in (0, 1/(T_1+T_2))$, we may argue as in (7.108) and find that $\langle H_{l,j}, \psi_k\rangle = \sqrt{\frac{\omega}{2^l}}\hat{H}\left(\frac{-2\pi k\omega}{2^l}\right) e^{2\pi i\omega k j/2^l}$. Thus,

$$
\langle \sum_{j\in\Delta_l} \eta_{\rho(l,j)} H_{l,j}, \psi_k\rangle = \sqrt{\frac{\omega}{2^l}} \sum_{j\in\Delta_l} \eta_{\rho(l,j)} \hat{H}\left(\frac{-2\pi k\omega}{2^l}\right) e^{2\pi i\omega k j/2^l}. \tag{7.118}
$$

It is straightforward to show that $\inf_{|x|\leq\pi}\left|\hat{H}(x)\right| \geq 2/\pi$, and since $N \leq 2^{R_1}/\omega$, for each $l \geq R_1$, it follows directly from (7.118) and the definition of $f^{[l]}$ that

$$
\begin{aligned}
\sum_{|k|\leq N/2} \frac{\omega}{2^l}\left|f^{[l]}\left(\frac{\omega k}{2^l}\right)\right|^2 &\leq \left(\inf_{|x|\leq\pi}\left|\hat{H}(x)\right|^2\right)^{-1} \sum_{|k|\leq N/2} \left|\langle \sum_{j\in\Delta_l} \eta_{\rho(l,j)} H_{l,j}, \psi_k\rangle\right|^2 \\
&\leq \frac{\pi^2}{4}\left\|\sum_{j\in\Delta_l} \eta_{\rho(l,j)} H_{l,j}\right\|^2 \leq \frac{\pi^2}{4}\|P_{\Delta_l}\eta\|^2.
\end{aligned}
$$

Hence, we immediately get that

$$
\sum_{l=R_1}^{R_2-1} \sum_{|k|\leq N/2} \frac{\omega}{2^l}\left|f^{[l]}\left(\frac{\omega k}{2^l}\right)\right|^2 \leq \frac{\pi^2}{4}\sum_{l=R_1}^{R_2-1}\|P_{\Delta_l}\eta\|^2 \leq \frac{\pi^2}{4}\|\eta\|^2 \leq \frac{\pi^2}{4}. \tag{7.119}
$$

Also, since Ψ has v vanishing moments, we have that $\hat{\Psi}(z) = (-iz)^v\theta_\Psi(z)$ for some bounded L^∞ function θ_Ψ. Thus, since $N \leq \gamma\cdot 2^{R_1}/\omega$, we have

$$
\begin{aligned}
\sum_{l=R_1}^{R_2-1} \max_{|k|\leq N/2}\left|\hat{\Psi}\left(\frac{2\pi\omega k}{2^l}\right)\right|^2 &\leq \frac{\pi^2}{4}\|\theta_\Psi\|_{L^\infty}^2 \sum_{l=R_1}^{R_2-1}\left(2\pi\gamma 2^{R_1-l}\right)^{2v} \\
&\leq \frac{\pi^2}{4}(2\pi\gamma)^{2v}\|\theta_\Psi\|_{L^\infty}^2 \frac{1-2^{2v(R_1-R_2)}}{1-2^{-2v}}.
\end{aligned}
$$

Thus, by applying (7.117), (7.118) and (7.119), it follows that

$$
\|P_N U P_{M_2}^{M_1}\eta\|^2 \leq \frac{\pi^2}{4}\|\theta_\Psi\|_{L^\infty}^2\cdot(2\pi\gamma)^{2v}\frac{1-2^{2v(R_1-R_2)}}{1-2^{-2v}},
$$

and we have proved the desired estimate. □

7.4.3 The proof

Proof of Theorem 6.2. In this proof, we will let $B_{\Phi,\Psi}$ be some constant which depends only on Φ and Ψ, although its value may change from instance to instance. The assertions of the theorem will follow if we can show that the conditions in Theorem 5.3 are satisfied. We will begin with condition (i). First observe that since U is an isometry we have that $\|P_M U^* P_N U P_M - P_M\|_{l^\infty} = \|P_M U^* P_N^\perp U P_M\|_{l^\infty\to l^\infty} \leq$

$\sqrt{M}\left\|P_N^{\perp}UP_M\right\|$ and $\|P_M^{\perp}U^*P_NUP_M\|_{l^\infty\to l^\infty} = \|P_M^{\perp}U^*P_N^{\perp}UP_M\|_{l^\infty\to l^\infty} \leq \sqrt{M}\left\|P_N^{\perp}UP_M\right\|$. So N, K satisfy the strong balancing property with respect to U, M and s if

$$\left\|P_N^{\perp}UP_M\right\| \leq \frac{1}{8}\left(M\log_2(4KM\sqrt{s})\right)^{-1/2}.$$

In the case of $\alpha \geq 1$, by applying Lemma 7.18 with $\gamma = \frac{1}{8}\left(M\log_2(4KM\sqrt{s})\right)^{-1/2}$, it follows that N, K satisfy the strong balancing property with respect to U, M, s whenever

$$N \geq C_{\omega,\Phi}\cdot 2^{R+1}\cdot\left(\frac{1}{8}\left(M\log_2(4KM\sqrt{s})\right)^{-1/2}\right)^{-\frac{2}{2\alpha-1}},$$

where R is the smallest integer such that $M \leq W_R$ (where W_R is defined in (7.104)) and $C_{\omega,\Phi}$ is a constant which depends only on the Fourier decay of Φ and ω. By the choice of R, we have that $M = \mathcal{O}\left(2^R\right)$ since $W_R = \mathcal{O}\left(2^R\right)$ by (7.104). Thus, the strong balancing property holds provided that

$$N \gtrsim M^{1+1/(2\alpha-1)}\cdot\left(\log_2(4MK\sqrt{s})\right)^{1/(2\alpha-1)}$$

where the constant involved depends only on ω and the Fourier decay of Φ. Furthermore, if (7.101) holds, then a direct application of Lemma 7.17 gives that N, K satisfy the strong balancing property with respect to U, M, s whenever $N \gtrsim M\cdot(\log_2(4KM\sqrt{s}))^{1/(4\alpha-2)}$. So, condition (i) of Theorem 6.2 implies condition (i) of Theorem 5.3.

To show that (ii) in Theorem 5.3 is satisfied, we need to demonstrate that

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k}\cdot\log(\epsilon^{-1})\cdot\left(\sum_{l=1}^{r}\mu_{\mathbf{N},\mathbf{M}}(k,l)\cdot s_l\right)\cdot\log\left(K\tilde{M}\sqrt{s}\right), \tag{7.120}$$

(with $\mu_{\mathbf{N},\mathbf{M}}(k,r)$ replaced by $\mu_{\mathbf{N},\mathbf{M}}(k,\infty)$, and also recall that $N_0 = 0$) and

$$\begin{aligned} m_k &\gtrsim \hat{m}_k\cdot\log(\epsilon^{-1})\cdot\log\left(K\tilde{M}\sqrt{s}\right), \\ 1 &\gtrsim \sum_{k=1}^{r}\left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right)\cdot\mu_{\mathbf{N},\mathbf{M}}(k,l)\cdot\tilde{s}_k, \qquad \forall\, l = 1,\ldots,r, \end{aligned} \tag{7.121}$$

where

$$\tilde{M} = \min\{i\in\mathbb{N} : \max_{k\geq i}\|P_NUe_k\| \leq 1/(32K\sqrt{s})\}. \tag{7.122}$$

We will first consider (7.120). By applying the bounds (7.110) and (7.111) on the local coherences derived in Corollary 7.16, we have that (7.120) is implied by

$$\begin{aligned} \frac{m_k}{(N_k - N_{k-1})} &\gtrsim B_{\Phi,\Psi}\cdot\left(\sum_{j=1}^{k-1}\frac{s_j}{N_{k-1}}\left(\frac{2^{R_{j-1}}}{\omega N_{k-1}}\right)^{\alpha-1/2} + \frac{s_k}{N_{k-1}}\right. \\ &\left. + \sum_{j=k+1}^{r}\frac{s_j\cdot\sqrt{\omega}}{\sqrt{N_{k-1}2^{R_{j-1}}}}\cdot\left(\frac{\omega N_k}{2^{R_{j-1}}}\right)^{v}\right)\cdot\log(\epsilon^{-1})\cdot\log\left(K\tilde{M}\sqrt{s}\right), \qquad k = 2,\ldots,r \end{aligned} \tag{7.123}$$

$$\frac{m_1}{N_1} \gtrsim B_{\Phi,\Psi}\cdot\left(s_1 + \sum_{j=2}^{r}\frac{s_j\cdot\sqrt{\omega}}{\sqrt{2^{R_{j-1}}}}\cdot\left(\frac{\omega N_1}{2^{R_{j-1}}}\right)^{v}\right)\cdot\log(\epsilon^{-1})\cdot\log\left(K\tilde{M}\sqrt{s}\right). \tag{7.124}$$

To obtain a bound on the value of $\tilde{M}$ in (7.122), observe that by Lemma 7.19, $\left\|P_NUP_{\{j\}}\right\| \leq 1/(32K\sqrt{s})$ whenever $j = 2^J$ such that $2^J \geq (32K\sqrt{s})^{1/v}\cdot N\cdot\omega$. Thus, $\tilde{M} \leq \lceil(32K\sqrt{s})^{1/v}\cdot N\cdot\omega\rceil$, and by recalling that $N_k = 2^{R_k}\omega^{-1}$, we have that (7.123) is implied by

$$\begin{aligned} \frac{m_k\cdot N_{k-1}}{N_k - N_{k-1}} &\gtrsim B_{\Phi,\Psi}\cdot\log(\epsilon^{-1})\cdot\log\left((K\sqrt{s})^{1+1/v}N\right) \\ &\cdot\left(\sum_{j=1}^{k-1}s_j\cdot\left(2^{\alpha-1/2}\right)^{-(R_{k-1}-R_{j-1})} + s_k + s_{k+1}\cdot 2^{-(R_k-R_{k-1})/2}\right. \\ &\left. + \sum_{j=k+2}^{r}s_j\cdot 2^{-(R_{j-1}-R_{k-1})/2}\cdot 2^{-v(R_{j-1}-R_k)}\right), \quad k\geq 2, \end{aligned} \tag{7.125}$$

and when $k = 1$, (7.124) is implied by

$$\frac{m_1}{N_1} \gtrsim B_{\Phi,\Psi} \cdot \log(\epsilon^{-1}) \cdot \log\left((K\sqrt{s})^{1+1/v} N\right) \cdot \left(s_1 + s_2 \cdot 2^{-R_1/2} + \sum_{j=k+2}^{r} s_j \cdot 2^{-(R_{j-1}-R_{k-1})/2} \cdot 2^{-v(R_{j-1}-R_k)}\right). \tag{7.126}$$

However, the condition (6.1) obviously implies (7.125) and (7.124), hence we have established that condition (6.1) implies (7.120). As for condition (7.121), we will first derive upper bounds for the $\tilde{s}_k$ values. Recall that according to Theorem 5.3 we have

$$\tilde{s}_k \le S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}) = \max\{\|P_{N_k}^{N_{k-1}} U\eta\|^2 : \|\eta\|_{l^\infty} \le 1, |\mathrm{supp}(P_{M_l}^{M_{l-1}}\eta)| = s_l,\, l = 1, \dots, r\},$$

where $N_0 = M_0 = 0$. Thus, we will concentrate on bounding S_k. First note that by a direct rearrangement of terms in Lemma 7.18, for any $\gamma \in (0,1)$ and $R \in \mathbb{N}$ such that $M \le W_R$, we have that $\left\|P_N^\perp U P_M\right\| \le \gamma$ whenever N is such that

$$\gamma \ge \left(\frac{2^R}{\omega N}\right)^{\frac{2\alpha-1}{2}} \cdot \sqrt{\frac{2}{2\alpha-1}} \cdot \frac{C_\Phi}{\pi^\alpha}.$$

So for any $L > 0$, by letting $\gamma = \sqrt{\frac{2}{2\alpha-1}} \cdot \frac{C_\Phi}{\pi^\alpha} \cdot L^{-\frac{2\alpha-1}{2}}$, if $\gamma \in (0,1)$, then $\left\|P_N^\perp U P_M\right\| \le \gamma$ provided that $N \ge \omega^{-1} \cdot L \cdot 2^R$. Also, if $\gamma > 1$, then $\left\|P_N^\perp U P_M\right\| \le \gamma$ is trivially true since $\|U\| = 1$. Therefore, for $k \ge 2$ we have that

$$\|P_{N_{k-1}}^\perp U P_{M_l}\| < \sqrt{\frac{2}{2\alpha-1}} \cdot \frac{C_\Phi}{\pi^\alpha} \cdot \left(\frac{2^{R_l}}{2^{R_{k-1}}}\right)^{\alpha-1/2}, \qquad l \le k-1.$$

Also, by Lemma 7.19, it follows that

$$\|P_{N_k} U P_{M_l}^{M_{l-1}}\eta\| < (2\pi)^v \cdot \|\theta_\Psi\|_{L^\infty} \cdot \left(\frac{2^{R_k}}{2^{R_{l-1}}}\right)^v, \qquad l \ge k+1.$$

Consequently, for $k = 3, \dots, r$

$$\begin{aligned}\sqrt{\tilde{s}_k} \le \sqrt{S_k} &= \max_{\eta\in\Theta} \|P_{N_k}^{N_{k-1}} U\eta\| \le \sum_{l=1}^{r} \|P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}\| \sqrt{s_l} \\ &\le B_{\Phi,\Psi}\left(\sum_{l=1}^{k-2} \sqrt{s_l} \cdot \left(\frac{2^{R_l}}{2^{R_{k-1}}}\right)^{\alpha-1/2} + \sqrt{s_{k-1}} + \sqrt{s_k} + \sqrt{s_{k+1}} + \sum_{l=k+2}^{r} \sqrt{s_l} \cdot \left(\frac{2^{R_k}}{2^{R_{l-1}}}\right)^v\right),\end{aligned}$$

where

$$\Theta = \{\eta : \|\eta\|_{l^\infty} \le 1, |\mathrm{supp}(P_{M_l}^{M_{l-1}}\eta)| = s_l,\, l = 1, \dots, r\},$$

and for $k = 1, 2$ we have

$$\sqrt{\tilde{s}_k} \le B_{\Phi,\Psi}\left(\sqrt{s_{k-1}} + \sqrt{s_k} + \sqrt{s_{k+1}} + \sum_{l=k+2}^{r} \sqrt{s_l} \cdot \left(\frac{2^{R_k}}{2^{R_{l-1}}}\right)^v\right),$$

where we let $s_0 = 0$. Hence, for $k = 3, \dots, r$, $A_\alpha = 2^{\alpha-1/2}$ and $A_v = 2^v$

$$\tilde{s}_k \le B_{\Phi,\Psi}\left(\sqrt{\hat{s}_k} + \sum_{l=1}^{k-2} \sqrt{s_l} \cdot A_\alpha^{-(R_{k-1}-R_l)} + \sum_{l=k+2}^{r} \sqrt{s_l} \cdot A_v^{-(R_{l-1}-R_k)}\right)^2,$$

where $\hat{s}_k = \max\{s_{k-1}, s_k, s_{k+1}\}$. So, by using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\tilde{s}_k &\le B_{\Phi,\Psi}\left(1 + \sum_{l=1}^{k-2} A_\alpha^{-(R_{k-1}-R_l)} + \sum_{l=k+2}^{r} A_v^{-(R_{l-1}-R_k)}\right) \\ &\qquad \cdot \left(\hat{s}_k + \sum_{l=1}^{k-2} s_l \cdot A_\alpha^{-(R_{k-1}-R_l)} + \sum_{l=k+2}^{r} s_l \cdot A_v^{-(R_{l-1}-R_k)}\right) \\ &\le B_{\Phi,\Psi}\left(\hat{s}_k + \sum_{l=1}^{k-2} s_l \cdot A_\alpha^{-(R_{k-1}-R_l)} + \sum_{l=k+2}^{r} s_l \cdot A_v^{-(R_{l-1}-R_k)}\right),\end{aligned}$$

and similarly, for $k = 1, 2$, it follows that $\tilde{s}_k \leq B_{\Phi,\Psi}(\hat{s}_k + \sum_{l=k+2}^{r} s_l \cdot A_v^{-(R_{l-1}-R_k)})$. Finally, we will use the above results to show that condition (6.1) implies (7.121): By our coherence estimates in (7.110), (7.112), (7.111) and (7.113), we see that (7.121) holds if $m_k \gtrsim \hat{m}_k \cdot (\log(\epsilon^{-1}) + 1) \cdot \log\left((K\sqrt{s})^{1+1/v} N\right)$ and for each $l = 2, \ldots, r$,

$$\begin{aligned}
1 \gtrsim B_{\Phi,\Psi} \Bigg(&\left(\frac{N_1}{\hat{m}_1} - 1\right) \cdot \tilde{s}_1 \cdot \sqrt{\frac{\omega}{2^{R_{l-1}}}} \cdot \left(\frac{\omega N_1}{2^{R_{l-1}}}\right)^v \\
&+ \sum_{k=2}^{l-1} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \tilde{s}_k \cdot \sqrt{\frac{\omega}{N_{k-1} 2^{R_{l-1}}}} \cdot \left(\frac{\omega N_k}{2^{R_{l-1}}}\right)^v \\
&+ \left(\frac{N_l - N_{l-1}}{\hat{m}_l} - 1\right) \cdot \tilde{s}_l \cdot \frac{1}{N_{l-1}} + \sum_{k=l+1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \tilde{s}_k \cdot \frac{1}{N_{k-1}} \left(\frac{2^{R_{l-1}}}{\omega N_{k-1}}\right)^{\alpha - 1/2} \Bigg),
\end{aligned} \tag{7.127}$$

(where we with slight abuse of notation define $\sum_{k=2}^{l-1}(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1)\tilde{s}_k \sqrt{\frac{\omega}{N_{k-1}2^{R_{l-1}}}}(\frac{\omega N_k}{2^{R_{l-1}}})^v = 0$ when $l = 2$), and for $l = 1$

$$1 \gtrsim B_{\Phi,\Psi} \left(\left(\frac{N_1}{\hat{m}_1} - 1\right) \cdot \tilde{s}_1 + \sum_{k=2}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \tilde{s}_k \cdot \frac{1}{N_{k-1}} \left(\frac{1}{\omega N_{k-1}}\right)^{\alpha - 1/2} \right). \tag{7.128}$$

Recalling that $N_k = \omega^{-1} 2^{R_k}$, (7.127) becomes, for $l = 2, \ldots, r$,

$$\begin{aligned}
1 \gtrsim B_{\Phi,\Psi} \cdot \Bigg(&\left(\frac{N_1}{\hat{m}_1} - 1\right) \cdot \frac{\tilde{s}_k}{N_{k-1}} \cdot 2^{-v(R_{l-1}-R_k)} + \sum_{k=1}^{l-1} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \frac{\tilde{s}_k}{N_{k-1}} \cdot 2^{-v(R_{l-1}-R_k)} \\
&+ \left(\frac{N_l - N_{l-1}}{\hat{m}_l} - 1\right) \cdot \frac{\tilde{s}_l}{N_{l-1}} + \sum_{k=l+1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \frac{\tilde{s}_k}{N_{k-1}} \cdot \left(2^{\alpha - 1/2}\right)^{-(R_{k-1} - R_{l-1})} \Bigg),
\end{aligned}$$

and (7.128) becomes

$$1 \gtrsim B_{\Phi,\Psi} \cdot \left(\left(\frac{N_1}{\hat{m}_1} - 1\right) \cdot \tilde{s}_1 + \sum_{k=l+1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1\right) \cdot \frac{\tilde{s}_k}{N_{k-1}} \cdot \left(2^{\alpha - 1/2}\right)^{-R_{k-1}} \right).$$

Observe that for $l = 2, \ldots, r$

$$1 + \sum_{k=1}^{l-1} 2^{-v(R_{l-1}-R_k)} + \sum_{k=l+1}^{r} \left(2^{\alpha-1/2}\right)^{-(R_{k-1}-R_{l-1})} \leq B_{\Phi,\Psi},$$

and that $1 + \sum_{k=l+1}^{r} \left(2^{\alpha-1/2}\right)^{-(R_{k-1})} \leq B_{\Phi,\Psi}$. Thus, (7.121) holds provided that for each $k = 2, \ldots, r$,

$$\hat{m}_k \geq B_{\Phi,\Psi} \cdot \frac{N_k - N_{k-1}}{N_{k-1}} \cdot \tilde{s}_k, \qquad \hat{m}_1 \geq B_{\Phi,\Psi} \cdot N_1 \cdot \tilde{s}_1,$$

and combining with our estimates of $\tilde{s}_k$, we may deduce that (6.1) implies (7.121). □

Acknowledgements

The authors would like to thank Akram Aldroubi, Emmanuel Candès, Massimo Fornasier, Karlheinz Gröchenig, Felix Krahmer, Gitta Kutyniok, Thomas Strohmer, Gerd Teschke, Michael Unser, Martin Vetterli and Rachel Ward for useful discussions and comments. The authors also thank Stuart Marcelle and Homerton College, University of Cambridge for the provision of computing hardware used in some of the experiments. BA acknowledges support from the NSF DMS grant 1318894. ACH acknowledges support from a Royal Society University Research Fellowship as well as the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L003457/1. CP acknowledges support from the EPSRC grant EP/H023348/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis.

References

[1] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Technical report NA2011/02, DAMTP, University of Cambridge*, 2011.

[2] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *J. Fourier Anal. Appl.*, 18(4):685–716, 2012.

[3] B. Adcock and A. C. Hansen. Stable reconstructions in Hilbert spaces and the resolution of the Gibbs phenomenon. *Appl. Comput. Harmon. Anal.*, 32(3):357–388, 2012.

[4] B. Adcock, A. C. Hansen, E. Herrholz, and G. Teschke. Generalized sampling: extension to frames and inverse and ill-posed problems. *Inverse Problems*, 29(1):015008, 2013.

[5] B. Adcock, A. C. Hansen, and C. Poon. Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem. *SIAM J. Math. Anal.*, 45(5):3114–3131, 2013.

[6] B. Adcock, A. C. Hansen, and C. Poon. On optimal wavelet reconstructions from Fourier samples: linearity and universality of the stable sampling rate. *Appl. Comput. Harmon. Anal.*, 36(3):387–415, 2014.

[7] B. Adcock, A. C. Hansen, B. Roman, and G. Teschke. Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. *Advances in Imaging and Electron Physics*, 182:187–279, 2014.

[8] J. Bigot, C. Boyer, and P. Weiss. An analysis of block sampling strategies in compressed sensing. *Arxiv:1305.4446*, 2013.

[9] E. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: optimality of curvelet frames. *Ann. Statist.*, 30(3):784–842, 2002.

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

[11] E. J. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

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

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

[14] 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.

[15] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin. Communications-inspired projection design with application to compressive sensing. *SIAM J. Imaging Sci.*, 5(4):1185–1212, 2012.

[16] Y. Chi, L. L. Scharf, A. Pezeshki, and R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.*, 59(5):2182–2195, 2011.

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

[18] S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, H.-G. Stark, and G. Teschke. The uncertainty principle associated with the continuous shearlet transform. *Int. J. Wavelets Multiresolut. Inf. Process.*, 6(2):157–181, 2008.

[19] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke. Shearlet coorbit spaces and associated Banach frames. *Appl. Comput. Harmon. Anal.*, 27(2):195–214, 2009.

[20] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41(7):909—996, 1988.

[21] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992.

[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] R. A. DeVore. Nonlinear approximation. *Acta Numer.*, 7:51–150, 1998.

[24] M. N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Trans. Image Proc.*, 14(12):2091–2106, 2005.

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

[26] D. L. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. *Proc. Natl Acad. Sci. USA*, 102(27):9452–9457, 2005.

[27] 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.

[28] Y. C. Eldar and G. Kutyniok, editors. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

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

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

[31] K. Gröchenig, Z. Rzeszotnik, and T. Strohmer. Quantitative estimates for the finite section method. *Integral Equations Operator Theory*, to appear.

[32] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theor.*, 57(3):1548–1566, Mar. 2011.

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

[34] 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.

[35] A. C. Hansen. On the approximation of spectra of linear operators on hilbert spaces. *Journal of Functional Analysis*, 254(8):2092 – 2126, 2008.

[36] 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.

[37] E. Hernández and G. Weiss. *A First Course on Wavelets*. Studies in Advanced Mathematics. CRC Press, 1996.

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

[39] A. D. Jones, B. Adcock, and A. C. Hansen. On the asymptotic incoherence of wavelets and polynomials with the fourier basis, and its implications for infinite-dimensional compressed sensing. *Preprint*, 2013.

[40] F. Krahmer and R. Ward. Stable and robust recovery from variable density frequency samples. *IEEE Trans. Image Proc. (to appear)*, 2014.

[41] G. Kutyniok, J. Lemvig, and W.-Q. Lim. Compactly supported shearlets. In M. Neamtu and L. Schumaker, editors, *Approximation Theory XIII: San Antonio 2010*, volume 13 of *Springer Proceedings in Mathematics*, pages 163–186. Springer New York, 2012.

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

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

[44] M. Lustig. *Sparse MRI*. PhD thesis, Stanford University, 2008.

[45] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: the application of compressed sensing for rapid MRI imaging. *Magn. Reson. Imaging*, 58(6):1182–1195, 2007.

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

[47] S. G. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3 edition, 2009.

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

[49] D. D.-Y. Po and M. N. Do. Directional multiscale modeling of images using the contourlet transform. *IEEE Trans. Image Proc.*, 15(6):1610–1620, June 2006.

[50] C. Poon. A stable and consistent approach to generalized sampling. *Preprint*, 2013.

[51] G. Puy, J. P. Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux. Spread spectrum Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging*, 31(3):586–598, 2012.

[52] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *IEEE Signal Process. Letters*, 18:595–598, 2011.

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

[54] J. Romberg. Imaging via compressive sampling. *IEEE Signal Process. Mag.*, 25(2):14–20, 2008.

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

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

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

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

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

[60] Y. Tsaig and D. L. Donoho. Extensions of compressed sensing. *Signal Process.*, 86(3):549–571, 2006.

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

[62] L. Wang, D. Carlson, M. R. D. Rodrigues, D. Wilcox, R. Calderbank, and L. Carin. Designed measurements for vector count data. In *Advances in Neural Information Processing Systems*, pages 1142–1150, 2013.