Structure dependent sampling in compressed sensing:
theoretical guarantees for tight frames

Clarice Poon*
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
May 21, 2015

Abstract

Many of the applications of compressed sensing have been based on variable density sampling, where certain sections of the sampling coefficients are sampled more densely. Furthermore, it has been observed that these sampling schemes are dependent not only on sparsity but also on the sparsity structure of the underlying signal. This paper extends the recent result of (Adcock, Hansen, Poon and Roman, arXiv:1302.0561, 2013) to the case where the sparsifying system forms a tight frame. By dividing the sampling coefficients into levels, our main result will describe how the amount of subsampling in each level is determined by the local coherences between the sampling and sparsifying operators and the localized level sparsities – the sparsity structure in each level under the sparsifying operator.

1 Introduction

Over the past decades, much of the research in signal processing has been based on the assumption that natural signals can be sparsely represented. One of the achievements resulting from this realization was compressed sensing, which made it possible to recover a sparse signal from very few non-adaptive linear measurements. Compressed sensing is typically modelled as follows. Given an unknown vector $x \in \mathbb{C}^N$ and a measurement device represented by a matrix $V$, one aims to recover $x$ from a highly incomplete set of measurements by solving

$$ R(x, \Omega) \in \arg\min_{z \in \mathbb{C}^N} \|Dz\|_1 \text{ subject to } P_{\Omega}Vz = P_{\Omega}Vx, \quad (1.1) $$

where $\Omega$ indexes the given measurements, $P_{\Omega}$ is a projection matrix which restricts a vector to its coefficients indexed by $\Omega$ and $D$ a sparsifying matrix under which $Dx$ is assumed to be sparse. Typical results in compressed sensing describe how under certain conditions, one can guarantee recovery when the number of measurements $|\Omega|$ scales up to a log factor linearly with sparsity $[9, 8, 7]$.

Although a large part of the theoretical development of compressed sensing has revolved around the construction of random sampling operators (such as sample matrices constructed from random Gaussian ensembles) where the choice of the samples is completely independent of the sparsifying system $[16, 32, 34, 40]$, in the majority of applications where compressed sensing has been of interest, one is concerned with the recovery of a signal from structured measurements. For example, the measurements in magnetic resonance imaging (MRI) are modelled via the Fourier transform, while the measurements in radio interferometry are modelled via the Radon transform. In these cases, how one can achieve subsampling is highly dependent on the sparsifying transform. To explain this statement, we recall some results of compressed sensing on recovery of a vector of length $N$ from its discrete Fourier coefficients under various sparsifying transforms.

1. If the underlying vector is $s$-sparse in its canonical basis, then one can guarantee perfect recovery from $O(s \log N)$ Fourier coefficients drawn uniformly at random $[3]$. 

---

*cmhsp2@cam.ac.uk
(2) If the underlying vector is \( s \)-sparse with respect to its total variation \([8]\), then \( O(s \log N) \) Fourier coefficients drawn uniformly at random will again guarantee perfect recovery, however, in the presence of noise and approximate sparsity, then one can obtain far superior error bounds with sampling strategies which sample more densely at low frequency coefficients instead \([30]\).

(3) If the underlying vector is \( s \)-sparse with respect to some wavelet basis, then it is impossible to guarantee recovery from \( O(s \log N) \) samples from sampling uniformly at random. This is a phenomenon which has been observed since the early days of compressed sensing and there has been extensive investigations into how subsampling is still achievable by sampling more densely at low frequencies \([28, 20, 35, 20, 39, 41]\). These approaches were often referred to as variable density sampling and theoretical guarantees for these approaches were recently derived in \([38]\) and \([1]\).

More generally, whether one can sample uniformly at random depends on whether the sampling and sparsifying matrices are sufficiently incoherent. In the absence of incoherence (as is the case in (3) above), how one should choose \( \Omega \) in \((1.1)\) becomes a far more delicate issue. To explain the use of compressed sensing in this case, a theoretical framework was developed in \([1]\) on the basis of three new principles: multilevel sampling, asymptotic incoherence and asymptotic sparsity. By modelling a nonuniform sampling strategies via multilevel sampling, the need for dense sampling at low frequencies in (3) is due to the following two reasons.

(i) The high correspondence between Fourier and wavelet bases at low Fourier frequencies and wavelet scales, but the low correspondence at high Fourier frequencies and wavelet scales (asymptotic incoherence).

(ii) Typical signals or images exhibit distinctive sparsity patterns in their wavelet coefficients, and become increasingly sparse at higher wavelet scales (asymptotic sparsity).

In contrast to the large body of results in compressed sensing where the strategy is based on sparsity alone, the results of \([1]\) demonstrated that one of the driving forces behind the success of variable density sampling strategies is their correspondence to the sparsity structure of the underlying signals of interest. These new principles provided a framework under which one can understand how to exploit, both the sparsity structure of the underlying signal, and the correspondences between the sampling and sparsifying systems to devise optimal subsampling strategies \([37, 33]\).

1.1 Contribution and overview

The paper \([1]\) is concerned only with the case where the sparsifying system is an orthonormal basis. On the other hand, many of the sparsifying transforms in applications tend to be constructed from overcomplete dictionaries, such as contourlets \([14]\), curvelets \([4, 5]\), shearlets \([12, 25]\) and wavelet frames \([13, 15]\). Therefore, the purpose of this paper is to develop a theory on how to structure one’s samples based on the sparsity structure with respect to a tight frame. Although the recent work of \([24]\) derives theoretical guarantee for nonuniform sampling strategies, their theory do not reveal any dependence between the sampling strategy and any sparsity structure; \([2]\) presents a numerical example to explain why an understanding of this dependence is crucial to achieving subsampling. \([3]\) recalls the key principles from \([1]\) and a result on solutions of \((1.1)\) in the case where \( D \) is constructed from an orthonormal basis. The main result of this paper are presented in \([4]\) where we reveal how the main result of \([1]\) can be extended in the case where \( D \) is constructed from a tight frame. The remainder of this paper will be devoted to proving the result of \([4]\).

Notation  Giving Banach spaces \( X \) and \( Y \), let \( \mathcal{B}(X, Y) \) denote the space of bounded linear operators from \( X \) to \( Y \) and let \( \mathcal{B}(X) \) denote the space of bounded linear operators from \( X \) to \( X \). Let \( H \) be a Hilbert space and given any subspace \( S \subseteq H \), \( Q_S \) denotes the orthogonal projection onto \( S \). We say that \( \{ \varphi_j : j \in \mathbb{N} \} \) is a frame for \( H \) if there exists \( c, C > 0 \) such that

\[
c\|g\|_H^2 \leq \sum_{j \in \mathbb{N}} |\langle g, \varphi_j \rangle|^2 \leq C\|g\|_H^2, \quad \forall g \in H.
\]
We say that \{\varphi_j : j \in \mathbb{N}\} is a tight frame if \(c = C\). If \(c = C = 1\), then \{\varphi_j : j \in \mathbb{N}\} is said to be a Parseval frame. Given any linear operator \(U\), let \(\mathcal{R}(U)\) denote its range and let \(\mathcal{N}(U)\) denote its null space.

We will also consider the sequence spaces \(\ell^p(\mathbb{N})\) for \(p \in [1, \infty]\). Let \{\varepsilon_j : j \in \mathbb{N}\} denote the canonical basis for the \(\ell^p(\mathbb{N})\) space under consideration. Given any \(\Delta \subset \mathbb{N}\), \(P_\Delta\) denotes the orthogonal projection onto \(\text{span}\{\varepsilon_j : j \in \Delta\}\). Given \(M \in \mathbb{N}\), let \([M] := \{1, \ldots, M\}\). Given \(z \in \ell^2(\mathbb{N})\), let \(\text{sgn}(z) \in \ell^\infty(\mathbb{N})\) be such that for each \(j \in \mathbb{N}\),

\[
\text{sgn}(z)_j = \begin{cases} 
\frac{z_j}{|z_j|} & z_j \neq 0 \\
0 & \text{otherwise}
\end{cases}.
\]

Given \(q \in (0, \infty]\), the \(\ell^q\) norm (or quasi-norm if \(q \in (0, 1)\)) is defined for \(z = (z_j)_{j \in \mathbb{N}}\) as

\[
\|z\|_{\ell^q} = \sum_j |z_j|^q, \quad q \in (0, \infty), \quad \|z\|_{\ell^\infty} = \sup_j |z_j|.
\]

Let \(\|\cdot\|_{p \to q}\) denote the operator norm of \(\mathcal{B}(\ell^p(\mathbb{N}), \ell^q(\mathbb{N}))\) for \(p, q \in [1, \infty]\). If \(X\) and \(Y\) are Hilbert spaces, we will simply denote the operator norm of \(\mathcal{B}(X, Y)\) by \(\|\cdot\|\). Given \(a, b \in \mathbb{R}, a \preceq b\) denotes \(a \leq Cb\) where \(C\) is a constant which is independent of all variables under consideration. The identity operator is denoted by \(I\), and the space on which this is defined will be clear from context.

### 2 The need for structure dependent sampling

To illustrate the need to account for sparsity structure when devising subsampling strategies, let us consider the case of recovering finite dimensional vectors, where we are given access to a subset of their Fourier coefficients and the sparsifying system is the two redundant discrete Haar wavelet frame. The Haar frame is defined in detail in the appendix \ref{appendixA}. In the following example, \(A\) will denote the discrete Fourier transform, and \(D\) will denote the discrete Haar wavelet transform.

**A numerical example.** Let \(N = 1024\) and consider the recovery of the two signals \(x_1\) and \(x_2\) shown in Figure \ref{fig:signals} from subsampling their discrete Fourier coefficients by solving (\ref{eq:optimization}). These signals are constructed such that \(\|Dx_1\|_0 = \|Dx_2\|_0 = 100\), where we define the sparsity measure of a signal by \(\|z\|_0 := |\{j : |z_j| \neq 0\}|\) for any \(z \in \mathbb{C}^M\) with \(M \in \mathbb{N}\). The sparsity patterns of \(Dx_1\) and \(Dx_2\) are shown in Figure \ref{fig:sparsity}. Observe that compared to \(Dx_2\), \(Dx_1\) has a higher proportion of large coefficients with respect to the higher scale frame elements. Let \(\Omega_V\) index 130 of the rows of \(A\) (12.7\% subsampling), so that the indices correspond to the first 41 Fourier coefficients of lowest frequencies plus 89 of the remaining coefficients drawn uniformly at random. The reconstruction of \(R(x_1, \Omega_V)\) and \(R(x_2, \Omega_V)\) from their partial Fourier samples are shown in the top row of Figure \ref{fig:reconstructions}. Note that although the same sampling pattern is used for both reconstructions, and both signals have the same sparsity with respect to \(D\), \(R(x_2, \Omega_V)\) is an exact reconstruction of \(x_2\) whilst \(R(x_1, \Omega_V)\) incurs a relative error of 34.85\%. This simple example suggests that to subsample efficiently, it is not sufficient to consider sparsity alone. We remark also that unlike sampling with unstructured operators such as random Gaussian matrices, uniform random sampling will yield poor reconstructions for both signals. The second row of Figure \ref{fig:reconstructions} shows the reconstruction \(R(x_1, \Omega_U)\) and \(R(x_2, \Omega_U)\), where \(\Omega_U\) indexes 130 of the available coefficients uniformly at random. Finally, it is interesting to note that the high frequency samples indexed by \(\Omega_V\) are required for an exact reconstruction of \(x_2\) as an error is incurred when one simply samples the Fourier coefficients of lowest frequency (see the bottom row of Figure \ref{fig:reconstructions}).

**Remark 2.1** In the context of sampling the Fourier transform of a signal, which is sparse with respect to some multiscale transform (such as wavelets, curvelets or shearlets), it is now commonly observed that uniform random sampling yields highly inferior results, when compared with variable density sampling patterns which focus on low frequencies. The numerical example in this section simply highlights this observation, and reminds us that the performance of these variable density sampling patterns are highly dependent on the sparsity structure of the underlying signal, and not just the sparsity level alone. Thus, there is a need for a theory which describes how the sparsity structure of the underlying signal should impact the choice of the sampling pattern.
Figure 1: Top row: Two test signals. Only a zoom of $x_1$ is shown since it is supported only on the indices ranging between 100 and 158. Both signals have equal sparsity – for each $i = 1, 2$, $\|Dx_i\|_0 = 100$. The second to the bottom rows show the reconstructions from different sampling maps. Bottom row: the sparsity structure of $Dx_1$ and $Dx_2$. The graph for $|Dx_2|$ has been capped off at 20 to allow for a clear comparison with $|Dx_1|$.

3 Structured sampling with orthonormal systems

The main result of this paper will be an extension of the abstract result of [1] to the case where the sparsifying transform is a tight frame. This section recalls the key concepts introduced in [1] to analyse the use of variable density sampling schemes for orthonormal sparsifying bases. We first remark that although compressed sensing originally considered only finite dimensional vector spaces, the applications in which variable density sampling tends to be of interest arise are more naturally modelled on infinite dimensional Hilbert spaces. For this purpose, a Hilbert space framework for compressed sensing was introduced in [2] and [3].

For a Hilbert space $\mathcal{H}$, and given orthonormal bases $\{\psi_j\}_{j \in \mathbb{N}}$ (the sampling vectors) and $\{\varphi_j\}_{j \in \mathbb{N}}$ (the sparsifying vectors), define the operators

$$V : \mathcal{H} \to \ell^2(\mathbb{N}), \quad f \mapsto ((f, \psi_j))_{j \in \mathbb{N}}, \quad D : \mathcal{H} \to \ell^2(\mathbb{N}), \quad f \mapsto ((f, \varphi_j))_{j \in \mathbb{N}}.$$  

Suppose we wish to recover some $f \in \mathcal{H}$ from samples of the form $y = ((f, \psi_j))_{j \in \Omega} + \eta = P_\Omega Vf + \eta$ for some $\Omega \subset \mathbb{N}$ and noise vector $\eta$ of $\ell^2$-norm at most $\delta$. A key question in compressed sensing is how solutions to the following minimization problem allows one to exploit the sparsity of some $f \in \mathcal{H}$ with respect to $D$ to obtain accurate recovery from a minimal number of samples.

$$\inf_{g \in \mathcal{H}, Dg \in \ell^2(\mathbb{N})} \|Dg\|_{\ell^1} \text{ subject to } \|y - P_\Omega Vg\|_{\ell^2} \leq \delta.$$  

The property of coherence (defined below) of the operator $VD^*$ has been recognized to be an important factor in determining the minimal cardinality of the sampling set $\Omega$. Note that this can be seen as a
measure of the correlation between the sampling system associated with $V$ and the sparsifying system associated with $D$.

**Definition 3.1** (coherence). Let $U$ be a bounded linear operator on $\ell^2(N)$ (or let $U \in C^{N \times N}$ for some $N \in \mathbb{N}$) be such that $\|Ue_j\|_2 = 1$ for all $j \in \mathbb{N}$ (or $j = 1, \ldots, N$). Let $\{e_j : j \in \mathbb{N}\}$ be the canonical basis of $\ell^2(N)$ (or $C^N$). The coherence of $U$ is defined as $\mu(U) = \sup_{k,j} |\langle U e_j, e_k \rangle|$.

For the case where $VD^* \in C^{N \times N}$ is finite dimensional isometry, the main result of [1] showed that $\Omega \subset \{1, \ldots, N\}$ consists of $O \left( s \mu^2(VD^*) N \log N \right)$ samples drawn uniformly at random, where $f$ is $s$-sparse, then any solution $\hat{f}$ to (3.2) satisfies $\|f - \hat{f}\| \leq C\delta$ for some universal constant $C > 0$. Furthermore, one cannot improve upon the estimate of $O \left( s \mu^2(VD^*) N \log N \right)$. Thus, for the recovery of sparse signals, the minimal sampling cardinality was completely determined by this coherence quantity.

Unfortunately, when $\mu(VD^*) \approx 1$, this result merely concludes that $\Omega$ must index all available samples. This is especially problematic because when $VD^*$ is a bounded linear operator defined on the infinite dimensional Hilbert space $\ell^2(N)$ – it is necessarily the case that $\mu(VD^*) \geq c > 0$ for some constant $c$ and one cannot expect the coherence of any finite dimensional discretization of $VD^*$ to be of order $O \left( N^{-1/2} \right)$ (see [1] for a detailed explanation of this phenomenon). In the case where $V$ is associated with a Fourier basis and $D$ is associated with a wavelet basis, it is necessarily the case that...
\[ \mu(VD^+) = 1. \]

The key idea of \( \mathcal{P} \) is to recognize that by placing additional assumptions on the sparsity or compressibility structure of the underlying signal, one can make non trivial statements on how \( \Omega \) can be chosen in accordance to the underlying sparsity. Thus, to consider how one should draw samples from the first \( M \) samples in order to accurately recover \( f \in \mathcal{H} \), with \( \|P_{\Delta}Df\|_{\Omega} \ll \|f\|_{\mathcal{H}} \) for some \( \Delta \subset \{1, \ldots, N\} \) with \( |\Delta| = s \), one approach is to divide the sampling and sparsifying vectors into levels then analyse the correspondence between the different sampling and sparsifying levels. The main theoretical result from \( \mathcal{F} \) is based on three principles.

- Multi-level sampling - instead of considering sampling uniformly at random across all available samples, partition the samples into levels and consider sampling uniformly at random with different densities at each level. This model was introduced to analyse the effects of nonuniform sampling patterns.
- Local coherence - the coherence of partial sections of \( VD^* \).
- Sparsity in levels - instead of considering sparsity across all available coefficients, partition the coefficients into levels and consider the sparsity within each level.

We define each of these concepts below.

**Definition 3.2** (Multi-level sampling). Let \( r \in \mathbb{N} \), \( \mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r \) with \( 0 = M_0 < M_1 < \ldots < M_r \), \( \mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r \), with \( m_k \leq M_k - M_{k-1} \), \( k = 1, \ldots, r \), and suppose that

\[
\Omega_k \subseteq \{M_{k-1} + 1, \ldots, M_k\}, \quad |\Omega_k| = m_k, \quad k = 1, \ldots, r,
\]

are chosen uniformly at random. We refer to the set \( \Omega = \Omega_{\mathbf{M},\mathbf{m}} = \Omega_1 \cup \cdots \cup \Omega_r \) as an \((\mathbf{M}, \mathbf{m})\)-multilevel sampling scheme.

**3.1 Sparsity in levels**

The notion of sparsity in levels is defined as follows. As explained below, this notion is particularly important when considering the wavelet sparsity for imaging purposes.

**Definition 3.3** (Sparsity in levels). Let \( x \) be an element of either \( \mathbb{C}^N \) or \( \ell^2(\mathbb{N}) \). For \( r \in \mathbb{N} \) let \( \mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r \) with \( 0 = N_0 < N_1 < \ldots < N_r \) and \( \mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r \), with \( s_k \leq N_k - N_{k-1} \), \( k = 1, \ldots, r \). We say that \( x \) is \((\mathbf{s}, \mathbf{N})\)-sparse if, for each \( k = 1, \ldots, r \), \( \Delta_k := \text{supp}(x) \cap \{N_{k-1} + 1, \ldots, N_k\} \), satisfies \( |\Delta_k| \leq s_k \). We denote the set of \((\mathbf{s}, \mathbf{N})\)-sparse vectors by \( \Sigma_{\mathbf{s}, \mathbf{N}} \).

**Definition 3.4** \((\mathbf{s}, \mathbf{N})\)-term approximation). Let \( x = (x_j) \) be an element of either \( \mathbb{C}^N \) or \( \ell^2(\mathbb{N}) \). We define the \((\mathbf{s}, \mathbf{N})\)-term approximation

\[
\sigma_{\mathbf{s}, \mathbf{N}}(x) = \min_{\eta \in \Sigma_{\mathbf{s}, \mathbf{N}}} \|x - \eta\|_{\ell^1}.
\]  

(3.3)

As well as the level sparsities \( s_k \) defined in Definition 3.3 we shall also require the notion of a relative sparsity, which takes into account the sampling operator \( V \) and will account for how different levels interfere with each other.

**Definition 3.5** (Relative sparsity). Let \( V, D \in \mathcal{B}(\mathcal{H}, \mathcal{H}') \) where \( \mathcal{H} \) is a Hilbert space and \( \mathcal{H}' \) is either \( \mathbb{C}^N \) or \( \ell^2(\mathbb{N}) \). Let \( \mathbf{s} = (s_j)_{j=1}^r \in \mathbb{N}^r \), \( \mathbf{N} = (N_j)_{j=1}^r \in \mathbb{N}^r \) and \( \mathbf{M} = (M_j)_{j=1}^r \in \mathbb{N}^r \) with \( 0 = N_0 < N_1 < \cdots < N_r \) and \( 0 = M_0 < M_1 < \cdots < M_r \). For \( 1 \leq k \leq r \), the \( k^{\text{th}} \) relative sparsity is given by

\[
\hat{\kappa}_k = \hat{\kappa}_k(\mathbf{N}, \mathbf{M}, \mathbf{s}) = \max_{\gamma \in \Theta} \|P_{\Gamma_k}Vg\|^2,
\]

where \( \Gamma_k = (M_{k-1}, M_k] \cap \mathbb{N} \) and \( \Theta \) is the set

\[
\Theta = \{g \in \mathcal{H} : \|Dg\|_{\ell^\infty} \leq 1, |\text{supp}(P_{\Lambda_l}Dg)| = s_l, l = 1, \ldots, r\}.
\]

where \( \Lambda_l = (N_{k-1}, N_k] \cap \mathbb{N} \).
The Fourier/wavelets case

On level sparsities. It has been established that natural images are not simply sparse in their wavelet coefficients, but exhibit a distinctive ‘tree-structure’ in their coefficient [1]. Given a wavelet basis \( \{ \varphi_j \}_{j \in \mathbb{N}} \), it is often the case that a typical image with sparse approximation \( \sum_{j \in \Delta} \alpha_j \varphi_j \) will actually not be sparse with respect to the wavelets of low scales, but will become increasingly sparse with respect to the wavelets of higher scaling factors. In particular, if \( \{ N_k \}_{k \in \mathbb{N}} \) corresponds to the wavelet scales such that \( \{ \varphi_j \}_{j \leq N_k} \) consists of all wavelets up to the \( k \)th scale, and \( s_k = |\Delta \cap (N_{k-1}, N_k]| \) is the sparsity at the \( k \)th wavelet scale, then one typically observes that although \( s_1/N_1 \approx 1 \), one has asymptotic sparsity with \( s_k/(N_k - N_{k-1}) \rightarrow 0 \) as \( k \) increases. This phenomenon is illustrated in Figure 3.

Thus, for the purpose of reconstructing natural images, it is perhaps too general to consider the recovery of all sparse wavelet coefficients and suffices to consider the recovery of images whose sparse representations exhibit asymptotic sparsity. This is the motivation behind the concept of sparsity in levels.

On relative sparsities In the case where \( V \) is the Fourier sampling operator and \( D \) is the analysis operator associated with an orthonormal basis, one can in fact show that the change of basis matrix \( VD^* \in \mathcal{B}(\ell^2(\mathbb{N})) \) is near block diagonal and by letting \( M \) and \( N \) correspond to wavelet scales,

\[
\hat{\kappa}_k(N, M, s) \lesssim \sum_{j=1}^r s_j A^{-|j-k|},
\]

for some \( A > 1 \) which depends only on the given wavelet basis. So, the dependence of the \( k \)th relative sparsity on each \( s_j \) decays exponentially in \( |j - k| \) and moreover, it follows that \( \sum_{k} \hat{\kappa}_k \lesssim \sum_{k} s_k \). The reader is referred to [1] for a proof of this.

![Figure 3](image.png)

Figure 3: Left: reconstruction from the largest 6% of the Daubechies-4 wavelet coefficients of a 1024 x 1024 image. Centre: location of coefficients in the sparse representation - coefficients are ordered in increasing wavelet scales away from the top left corner. Right: fraction of coefficients at each wavelet scale \( k \) which contribute to the sparse representation.

3.2 Local coherence

Although the coherence between the sampling and sparsifying systems is a crucial concept in the understanding of the minimal sampling cardinality required for the recovery of sparse signals, there are important systems of interest in applications where it is simply too crude to consider coherence alone. Instead, we require the more refined notion of local coherence.

**Definition 3.6** (Local coherence). Let \( V, D \in \mathcal{B}(\mathcal{H}, \mathcal{H}') \) where \( \mathcal{H} \) is a Hilbert space and \( \mathcal{H}' \) is either \( \mathbb{C}^N \) or \( \ell^2(\mathbb{N}) \). If \( N = (N_1, \ldots, N_r) \in \mathbb{N}^r \) and \( M = (M_1, \ldots, M_r) \in \mathbb{N}^r \) with \( 0 = N_0 < N_1 < \cdots < N_r \) and \( 0 = M_0 < M_1 < \cdots < M_r \). For \( k = 1, \ldots, r \), let \( \Gamma_k = \{ M_{k-1} + 1, \ldots, M_k \} \). For \( k = 1, \ldots, r - 1 \), let \( \Lambda_k = \{ N_{k-1} + 1, \ldots, N_k \} \) and let \( \Lambda_r = \{ n \in \mathbb{N} : n > N_r \} \). The \((k,l)\)th local coherence between \( V \) and \( D \)
with respect to $N$ and $M$ is given by
\[ \mu_{N,M}(k,l) = \sqrt{\mu(P_T^* V D^* P_{\Lambda_k}) \mu(P_T^* V D^*)}, \quad k = 1, \ldots, r, \quad l = 1, \ldots, r. \]

The Fourier/wavelets case

If $VD^*$ is constructed from any orthonormal wavelet basis with Fourier sampling, then it is necessarily the case that $\mu(V D^*) = 1$. However, it is only the initial section of $VD^*$ associated with low Fourier frequencies and low wavelet scales that has high coherence. In particular, one can show that
\[ \mu(P_{\Lambda}^* V D^*), \mu(V D^* P_{\Lambda}^*) = O\left(N^{-1/2}\right). \]

Finally, we remark that this property of asymptotic incoherence (decay in the coherence away from initial finite sections) is not unique to the Fourier/wavelets case, but can also be observed for other representation systems such as Fourier/Legendre polynomial systems. In the Fourier/wavelets case, it is this decay in the local coherences that makes it possible to exploit sparsity to subsample the Fourier coefficients.

3.3 Recovery guarantees in the case of orthonormal sparsifying transforms

When we are considering the recovery of an infinite dimensional object by drawing finitely many samples, there is a question on what the range of the samples $M$ should be in order to recover a sparse representation with respect to the first $N$ sparsifying elements. This question is addressed by the balancing property.

Definition 3.7 (Balancing property [1]). Let $VD^* \in B(\ell^2(N))$ be an isometry. Then $M \in \mathbb{N}$ and $K \geq 1$ satisfy the balancing property with respect to $V$, $D$, $N \in \mathbb{N}$ and $s \in \mathbb{N}$ if
\[ \|P_{\Lambda}^* V D^* P_{\Lambda} V D^* P_{\Lambda}\|_{\ell^\infty \to \ell^\infty} \leq \frac{1}{8} \left(\log_2^2 (4\sqrt{\pi N} M)\right)^{-1}, \]
and
\[ \|P_{\Lambda}^* V D^* P_{\Lambda} V D^* P_{\Lambda}\|_{\ell^\infty \to \ell^\infty} \leq \frac{1}{8}, \]
where $\|\cdot\|_{\ell^\infty \to \ell^\infty}$ is the norm on $B(\ell^\infty(N))$.

We now recall the main result of [1] which informs on how multilevel sampling will depend on local coherences and the underlying sparsity structure. For this, we require the following notation:
\[ \tilde{M} = \min\{i \in \mathbb{N} : \max_{k \geq 1} \|P_{M[i]} U e_k\| \leq 1/(32q^{-1} \sqrt{s})\}, \]
where $M$, $s$ and $q$ are as defined below.

Theorem 3.8. [1] Let $VD^*$ be an isometry either on $\ell^2(N)$ or $\mathbb{C}^N$. Let $f \in \mathcal{H}$. Suppose that $\Omega = \Omega_{M,m}$ is a multilevel sampling scheme, where $M = (M_1, \ldots, M_r) \in \mathbb{N}^r$ and $m = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(s, N)$, where $N = (N_1, \ldots, N_r) \in \mathbb{N}^r$, $N_1 < \ldots < N_r$, and $s = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds:

(i) the parameters $M = M_r, q^{-1} = \max_{k=1,\ldots,r} \left\{ \frac{M_k-M_{k-1}}{m_k} \right\}$, satisfy the balancing property with respect to $V$, $D$, $N := N_r$ and $s := s_1 + \ldots + s_r$;

(ii) for $\epsilon \in (0, e^{-1})$,
\[ 1 \geq \frac{M_k - M_{k-1}}{m_k} \log(s^{-1}) \log\left(q^{-1} \tilde{M} \sqrt{s}\right) \left(\sum_{l=1}^r \mu_{N,M}(k,l) s_l\right), \quad k = 1, \ldots, r, \]
and $m_k \geq \hat{m}_k \log(s^{-1}) \log\left(q^{-1} \tilde{M} \sqrt{s}\right)$, where $\hat{m}_k$ is such that
\[ 1 \geq \sum_{k=1}^r \left( \frac{M_k - M_{k-1}}{\hat{m}_k} - 1 \right) \mu_{N,M}(k,l) s_l, \quad l = 1, \ldots, r, \]
for all $(\hat{s}_1, \ldots, \hat{s}_r) \in \mathbb{R}_+^r$ with $\hat{s}_1 + \ldots + \hat{s}_r \leq s_1 + \ldots + s_r$, and $\hat{s}_k \leq \hat{s}_k(N, M, s)$ for each $k = 1, \ldots, r$. 


Suppose that \( \hat{f} \) is a minimizer of (3.2) with \( y = P_\Omega Vf + \eta \) and \( \| \eta \|_\ell^2 \leq \delta \). Then, with probability exceeding 1 - \( \epsilon \),

\[
\| \hat{f} - f \| \leq C \left( q^{-1/2} \delta \left( 1 + L \sqrt{s} \right) + \sigma_{s,N}(Df) \right),
\]

for some constant \( C \), where \( \sigma_{s,M} \) is as in (3.3), and \( L = C \left( 1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4\epsilon^{-1})} \right) \). If \( m_k = M_k - M_{k-1} \) for 1 \( \leq k \leq r \) then this holds with probability 1.

Notice that the number of samples at each level is dependent on the local coherences between \( V \) and \( D \), the level sparsities \( \{s_k\} \) and the relative level sparsities \( \{\hat{s}_k\} \). As discussed in [11], the relative level sparsities account for the interference between the different sampling and sparsifying levels and cannot be removed from the estimates. However, recall that in the case of Fourier sampling with wavelet sparsity where the levels correspond to the wavelet scales, one can essentially show that the dependence of \( \hat{s}_k \) on each \( s_j \) becomes exponentially small as \( |k - j| \) increases.

This result firstly suggests that even in cases where incoherence is missing, subsampling in accordance to sparsity is still possible provided that the sampling and sparsifying bases are not uniformly coherent – subsampling is possible when local coherence is small. Note also that this result suggests that a change in the sparsity structure, i.e. the distribution of \( \{s_k\} \) and \( \{\hat{s}_k\} \), should result in a change in the sampling strategy.

## 4 Main result

The work of [11] provides an initial understanding on how one can structure sampling in accordance to underlying sparsity structures so that the number of samples require is (up to log factors) linear with sparsity. A natural extension of this work would be to consider this question when \( D \) is an analysis operators associated with a tight frame instead of an orthonormal basis. This is of particular interest due to the recent development of sparse representations with respect to multiscale systems such as wavelet, curvelet and shearlet frames. In this paper, we will consider the case where \( V : \mathcal{H} \rightarrow \ell^2(\mathbb{N}) \) and \( D : \mathcal{H} \rightarrow \ell^2(\mathbb{N}) \) are isometries. This assumption simply states that \( V \) and \( D \) are the analysis operators of Parseval frames, i.e. \( \{\psi_j\}_{j \in \mathbb{N}} \) and \( \{\varphi_j\}_{j \in \mathbb{N}} \) are both Parseval frames of \( \mathcal{H} \) in (3.1).

Note that if \( D \) is associated with an orthonormal basis instead of a Parseval frame (i.e. \( D \) is unitary), then (3.2) is equivalent to

\[
\inf_{x \in \ell^1(\mathbb{N})} \| x \|_{\ell^1} \text{ subject to } \| y - P_\Omega V D^* x \|_{\ell^2} \leq \delta.
\]

This minimization problem is referred to as synthesis regularization. On the other hand, in the case of non-orthonormal systems, (3.2) (often referred to as analysis regularization) and (3.1) are no longer equivalent. Some of the differences between synthesis and analysis regularization were investigated in [17] and while the majority of theoretical works in compressed sensing has focussed on synthesis regularization, the theory behind the solutions of the analysis regularization problem (3.2) is less comprehensive.

### 4.1 Sparsity

In this section, we introduce concepts for describing sparsity under an analysis operator. In considering the solutions of (3.2), it is intuitive that this minimization problem will favour signals \( f \) for which the entries in \( Df \) has fast decay or are mainly zero entries. Note also that if there exists an index set \( \Delta \) such that \( P_{\Delta}^\perp Df = 0 \), then \( f \in \mathcal{N}(P_{\Delta}^\perp D) \subset \mathcal{R}(D^* P_{\Delta}) \) whenever \( D^* D = I \). In the works of [6] [24], the signal space considered is, for each sparsity level \( s \), the union of subspaces spanned by \( s \) columns of \( D^* \), \( W = \cup_{|\Delta| = s} \mathcal{R}(D^* P_{\Delta}) \).

As discussed in [24], to understand the impact of sparsity on the recovery of such a model, it is natural to consider the effects of the analysis operator \( D \) on any given \( f \in W \) and in particular, the approximate sparsity of \( Df \). For this purpose, the localization factor \( \eta \) was introduced in [24] as

\[
\eta = \sup \left\{ \frac{\|Dg\|_{\ell^2}}{\sqrt{s}} : |\Delta| = s, g \in \mathcal{R}(D^* P_{\Delta}), \|g\|_{\ell^2} = 1 \right\}
\]

(4.2)
and recovery estimates were given in terms of $\eta^2 s$. Moreover, as observed in [5], a standard measure of sparsity or compression in a vector is the quasi $\ell^p$ norm with $p \leq 1$. With this in mind, we introduce that concept of localized sparsity below.

**Definition 4.1.** Let $r \in \mathbb{N}$ and let $\mathbf{N} = (N_j)_{j=1}^r \in \mathbb{N}^r$, $s = (s_j)_{j=1}^r \in \mathbb{N}^r$. Assume that $N_1 < N_2 < \cdots < N_r = N$. Let $\Lambda_j = \mathbb{N} \cap (N_{j-1}, N_j]$ for $j = 1, \ldots, r - 1$ and $\Lambda_r = \mathbb{N} \cap (N_{r-1}, \infty)$. Let $p = 2^{-J}$ for some $J \in \mathbb{N} \cup \{0\}$. Let $\kappa > 0$ be the smallest number such that

$$k^{1-p/q} \geq \sup \left\{ \|Dg\|_p^q : g = D^*x, \|Dg\|_{\ell^q} = 1, x \in \Sigma_{s,N} \right\}, \quad q \in \{2, \infty\},$$

where we let $p/\infty = 0$. Then, $\kappa(\mathbf{N}, s, p) = \kappa$ is said to be the localized sparsity with respect to $p$, $\mathbf{N}$ and $s$. For each $j = 1, \ldots, r$, let $\kappa_j > 0$ be the smallest number such that

$$k_j^{1-p/q} \geq \sup \left\{ \|P_{\Lambda_j}Dg\|_p^q : g = D^*x, \|P_{\Lambda_j}Dg\|_{\ell^q} = 1, x \in \Sigma_{s,N} \right\}, \quad q \in \{2, \infty\},$$

Then $\kappa_j(\mathbf{N}, s, p) = \kappa_j$ is said to be the $j$th localized level sparsity with respect to $p$, $\mathbf{N}$ and $s$.

**Remark 4.1.** Observe that the localized sparsity is related to the localization factor in (4.2): if $\kappa = 2$ in (4.3), then it suffices to let $\kappa = \eta^2 s$.

One can consider $\kappa(s, \mathbf{N})$ to be a measure of the analysis sparsity of an element $f$ (i.e. sparsity of $Df$) given that it is synthesis sparse with respect to the frame $\{f_j\}$ associated with $D$ (i.e. $f = \sum_{j=\Delta} x_j f_j$ with $|\Delta| = s$ and $x \in \mathbb{C}^2$). Note that if $D$ is associated with an orthonormal basis, then $DD^*$ is the identity and it suffices to let $\kappa = \eta_1 + \cdots + \eta_s$.

The localized level sparsities $\kappa_j(s, \mathbf{N})$ describe the sparsity structure of $Df$ given that $f$ is synthesis sparse with a $(s, \mathbf{N})$-sparsity pattern. Again, if $D$ is associated with an orthonormal basis, then these localized level sparsities are simply the level sparsities $\{s_j\}_{j=1}^r$.

We also require the definition of relative sparsity, note that the only difference to Definition 3.5 is that the set $\Theta$ is defined in terms of $\|\cdot\|_\ell^2$ instead of $\|\cdot\|_\infty$.

**Definition 4.2** (Relative sparsity). Let $V, D \in B(\mathcal{H}, \mathcal{H'})$ where $\mathcal{H}$ is a Hilbert space and $\mathcal{H'}$ is either $\mathbb{C}^N$ or $\ell^2(\mathbb{N})$. Let $\kappa = (\kappa_j)_{j=1}^r \in \mathbb{N}^r$, $\mathbf{N} = (N_j)_{j=1}^r \in \mathbb{N}^r$ and $\mathbf{M} = (M_j)_{j=1}^r \in \mathbb{N}^r$ with $0 = N_0 < N_1 < \cdots < N_r$ and $0 = M_0 < M_1 < \cdots < M_r$. For $1 \leq k \leq r$, the $k$th relative sparsity is given by

$$\hat{\kappa}_k = \frac{\kappa_k(\mathbf{N}, \mathbf{M}, \kappa)}{\kappa_k(\mathbf{N}, \mathbf{M}, \kappa)} = \max_{g \in \mathbf{G}} \|P_{\Gamma_k}Vg\|^2,$$

where $\Gamma_k = (M_{k-1}, M_k) \cap \mathbb{N}$ and $\Theta$ is the set

$$\Theta = \{g \in \mathcal{H} : g = D^*s, \|P_{\Lambda_l}Dg\|^2_{\ell^2} \leq \kappa_l, l = 1, \ldots, r\},$$

where $\Lambda_1 = (N_{k-1}, N_k) \cap \mathbb{N}$.

### 4.2 Main result

The main result of this paper describes how the reconstruction error of any solution of (3.2) depends on the choice of samples. Note that the problem of considering the minimizers of (3.2) is well posed since minimizers necessarily exist (see Proposition 3.1).

In the case of an orthonormal systems, the balancing property provided an indication of the range that one should sample from when recovering a sparse support set $\Delta \subset [N]$ for some $N \in \mathbb{N}$. This condition essentially described how large $M$ must be such that $P_{\lfloor M \rfloor}V$ be close to an isometry on $\mathcal{R}(D^*P_{\Delta})$ for all $\Delta \subset [N]$. In the case where $D$ is no longer constructed from an orthonormal basis, we define the balancing property as follows.

**Definition 4.3.** Let $V, D \in B(\mathcal{H}, \ell^2(\mathbb{N}))$ be isometries. Then $M \in \mathbb{N}$ and $K \geq 1$ satisfy the balancing property with respect to $V$, $D$, $N$, $s \in \mathbb{N}$ and $s_k \geq s_\kappa > 0$ if for all $W = \mathcal{R}(D^*P_{\Delta})$ where $\Delta \subset [N]$ is such that $|\Delta| = s$,

$$\|DQW^*P_{\lfloor M \rfloor}^{1/2}VQW^*D^*\|_{\ell^2 \rightarrow \ell^2} \leq \frac{\sqrt{s_1/s_2}}{8} \left( \log_2(4\sqrt{s_2}KM) \right)^{-1},$$

(4.4)
Theorem 4.4.

\[ \|DQ_{W}V^*P_MVQW^*\|_{\ell^2\to\ell^\infty} \leq \frac{1}{8\sqrt{2}}, \quad (4.5) \]

Although this balancing property conceptually enforces the same isometry properties as the balancing property presented in the case of orthonormal system, note that the conditions are stated in terms of the $\ell^2$ norm instead. This difference is due to a slightly different dual certificate construction in the proof of our main result, and this slightly stronger balancing property will allow us to derive sharper bounds on the number of samples required. We remark also that this balancing property is in fact analogous to the original balancing property introduced in [2].

In the following theorem, for $r \in \mathbb{N}$, let $M = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $m = (m_1, \ldots, m_r) \in \mathbb{N}^r$, $N = (N_1, \ldots, N_r) \in \mathbb{N}^r$, $N_1 < \ldots < N_r$, and $s = (s_1, \ldots, s_r) \in \mathbb{N}^r$. For $p \in (0, 1]$, let $\kappa = \{\kappa_j\}_{j=1}^r$ with $\kappa_j = \kappa_j(N, s, p)$ and let $\kappa_j = \kappa_j(N, M, \kappa)$. Let

\[
\hat{M} = \|DD^*\|_{\ell^\infty \to \ell^\infty} \min \left\{ i \in \mathbb{N} : \max_{j \geq 1} \|P_MVD^*e_j\|_2 \leq \frac{q}{8\sqrt{\kappa_{\max}}}, \max_{j \geq 1} \|Q_{R(D^*P_M)}D^*e_j\| \leq \frac{\sqrt{2}}{4} \right\},
\]

and

\[
B(s, N) = \sup \left\{ \tilde{B}(\Delta) : \Delta \text{ is } (s, N)\text{-sparse} \right\},
\]

where, given any $\Delta \subset \mathbb{N}$ and $W_\Delta = \mathcal{R}(D^*P_\Delta)$,

\[
\tilde{B}(\Delta) = \max \left\{ \|DQ_{W_\Delta}D^*\|_{\ell^\infty \to \ell^\infty}, \|DQ_{W_\Delta}D^*\|_{\ell^\infty \to \ell^\infty} \cdot \max_{l=1}^r \left\{ \sum_{i=1}^r |P_{\Lambda_i}DQ_{W_\Delta}D^*P_{\Lambda_i}|_{\ell^\infty \to \ell^\infty} \right\} \right\}.
\]

(4.6)

**Theorem 4.4.** Let $\mathcal{H}$ be a Hilbert space and let $V, D \in \mathcal{B}(\mathcal{H}, \ell^2(\mathbb{N}))$ be isometric linear operators. Let $f \in \mathcal{H}$. Suppose that $\Omega = \Omega_{M, m}$ is a multilevel sampling scheme. Let $(s, N)$ be such that the following holds:

(i) the parameters $M = M_r, q^{-1} = \max_{k=1, \ldots, r} \left\{ \frac{M_k - M_{k-1}}{m_k} \right\}$, satisfy the balancing property with respect to $V, D, N := N_r$, $\kappa_{\min} = r \min \{ \kappa_j \}$ and $\kappa_{\max} = r \max \{ \kappa_j \}$;

(ii) For $\epsilon \in (0, \epsilon^{-1}],$

\[
1 \geq \sqrt{r} \log(\epsilon^{-1}) \log \left( q^{-1} \hat{M} \sqrt{\kappa_{\max}} \right) B(s, N) \frac{M_k - M_{k-1}}{m_k} \left( \sum_{l=1}^r \mu_{N, M}^2(k, l) \kappa_l \right), \quad k = 1, \ldots, r,
\]

and $m_k \geq r \hat{m}_k B(s, N)^2 \log(\epsilon^{-1}) \log \left( q^{-1} \hat{M} \sqrt{\kappa_{\max}} \right)$, where $\hat{m}_k$ is such that

\[
1 \geq \sum_{k=1}^r \left( \frac{M_k - M_{k-1}}{\hat{m}_k} - 1 \right) \mu_{N, M}^2(k, l) \kappa_l, \quad l = 1, \ldots, r.
\]

Suppose that $\hat{f}$ is a minimizer of $\min_{\|f\|_{\ell^2} \leq \delta}$ with $y = P_NVf + \eta$ and $\|\eta\|_{\ell^2} \leq \delta$. Then, with probability exceeding $1 - \epsilon$,

\[
\|\hat{f} - f\| \leq C \left( q^{-1/2} \delta \left( 1 + L \sqrt{\kappa_{\max}} \right) + \sigma_{s, N}(Df) \right),
\]

for some constant $C$, where $\sigma_{s, N}$ is as in (3.3), and $L = 1 + \frac{\sqrt{\log(\epsilon^{-1})}}{m_k \hat{m}_k \sqrt{\kappa_{\max}}}$. If $m_k = M_k - M_{k-1}$ for $1 \leq k \leq r$ then this holds with probability 1.
4.2.1 The unconstrained minimization problem

Instead of solving the constrained minimization problem in Theorem 4.4 for computational reasons, it is often of interest to solve instead an unconstrained minimization problem for some \( \alpha > 0 \),

\[
\inf_{g \in \mathcal{H}} \alpha \|Dg\|_{L^1} + \|P_{\Omega} V g - y\|_{L^2}^2 .
\]

(4.7)

The following result presents a recovery guarantee for this unconstrained problem.

**Corollary 4.5.** Consider the setting of Theorem 4.4, and let \( \alpha = \sqrt{\delta} \). Then, with probability exceeding \( (1 - \epsilon) \), any minimizer \( \hat{f} \) of \( \{4.7\} \) satisfies

\[
\|f - \hat{f}\|_{\mathcal{H}} \leq \delta \left( q^{-1/2} + L\sqrt{\kappa_{\max}} + L^2\kappa_{\max}\right) + \sigma_{N,s}(Df).
\]

**Remark 4.2** Note that by choosing \( \alpha = \sqrt{\delta} \), the guaranteed error bound is, up to \( \sqrt{q}L^2s \), the same as the guaranteed error bound of solutions to the constrained problem

\[
\inf_{g \in \mathcal{H}} \|Dg\|_{L^1} \text{ subject to } \|P_{\Omega} V g - y\|_{L^2} \leq \delta.
\]

This affirms the finding in [3, Figure 7], which numerically demonstrates that there exists a linear relation between the regularization parameter \( \alpha \) and noise level of the measurements \( \delta \). Moreover, this linear scaling increases as \( q \) increases.

4.2.2 Remarks on the main result

In the case where \( D \) is associated with an orthonormal basis, the key difference between our main result and Theorem 3.8 is that bounds on the number of samples in Theorem has a factor of \( \log(q) \) while the bounds in Theorem 4.4 have a factor of \( r \) (the number of levels) instead. In general, the sparsity \( s \) may grow as the ambient dimension \( N_r \) increases, whilst the number of levels \( r \) can be thought of as simply a constant (for example, \( r = 2 \) in the case of the half-half schemes presented in [35]). Therefore, Theorem 4.4 may be considered to provide slightly sharper bounds than Theorem 3.8 and is in fact optimal in the case where \( r = 1 \) (since the optimal sampling cardinality is \( O(s \log N) \) [7]).

It was shown in [11] how this abstract result can be used to derive sharp bounds on how one may exploit sparsity to subsample in the case of recovering wavelet coefficients from Fourier samples. More precisely, [11] considered the case where

\[
D : L^2[0,1) \rightarrow \ell^2(\mathbb{N}), \quad f \mapsto (\langle f, \varphi_j \rangle)_{j \in \mathbb{N}}
\]

for some orthonormal wavelet basis \( \{\varphi_j\} \) associated with a scaling function \( \Phi \) and a mother wavelet \( \Psi \) satisfying for all \( \xi \in \mathbb{R} \),

\[
\Phi(\xi) \lesssim (1 + |\xi|)^{-\beta}, \quad \Psi(\xi) \lesssim (1 + |\xi|)^{-\beta}.
\]

(4.8)

where \( \Phi \) and \( \Psi \) denote the Fourier transforms of \( \Phi \) and \( \Psi \) respectively, and the Fourier sampling operator is

\[
V : L^2[0,1) \rightarrow \ell^2(\mathbb{N}), \quad f \mapsto ((f, e^{i2\pi\omega k}))_{k \in \mathbb{Z}}
\]

for some appropriate Fourier sampling density \( \omega \in (0,1] \). Then, if we let \( (M_k) \) and \( (N_k) \) correspond to wavelet scales, condition (i) is implied by

\[
M_r \geq \omega^{-1}N_r^{1+1/(2\beta-1)} \log_2(\epsilon^{-1}N_r\sqrt{q}^{-1})^{1/(2\beta-1)},
\]

and condition (iii) is implied by

\[
m_k \geq L \cdot \frac{M_k - M_{k-1}}{M_{k-1}}, \left( s_k + \sum_{j \leq k-1} A^{R_j-R_{k-1}} s_j + \sum_{j \geq k+1} B^{R_k-R_{j-1}} s_j \right),
\]

with \( L = C\log(s\epsilon^{-1}) \log(M), A = 2^\beta, B = 2^\nu \), where \( \beta \geq 1 \) and \( \nu \geq 1 \) are respectively the Fourier decay exponent and the number of vanishing moments of the wavelet transform.

Under Theorem 4.4, we obtain the same result but with \( L \) being replaced with \( C\epsilon \log(\epsilon^{-1}) \log(M) \), note that this would be a slight improvement when considering sampling schemes with a fixed number of levels.
On localized sparsity. It is likely that by carrying out a similar analysis in [1], one can apply Theorem 4.4 to derive sharp recovery results for the recovery of coefficients with other multiscale systems, such as curvelets and shearlets from Fourier samples. This work is beyond the scope of this paper, however, we simply highlight two aspects of any such analysis. The first is that a key aspect of the analysis in [1] were precise estimates on the local coherence terms, and these estimates were essentially due to the Fourier decay (4.8). Since similar Fourier decay estimates exist for multiscale systems such as wavelet frames, shearlets and curvelets, it would be possible to derive estimates on the decay of the local coherences. Secondly, the key difference between Theorems 3.8 and 4.4 are the localized sparsity \( \kappa(s, N) \) and the localized level sparsities with respect to \( D \). As mentioned, these terms are equal to the sparsity and level sparsity terms when \( DD^* \) is the identity, furthermore, it is known that multiscale systems such as wavelet frames, shearlets and curvelets are intrinsically localized with near-diagonal Gram matrices. It is therefore conceivable that this property can be exploited to show that localized sparsity \( \kappa \) is close to the true sparsity \( s \). This idea is further discussed in §5.1.

On \( B(s, N) \). In the case where \( D \) is associated with an orthonormal basis, \( B(s, N) = 1 \). Further analysis of this quantity will be left as future work, however, we simply remark here that initial computations of this quantity suggest that the impact of \( B(s, N) \) will not be significant: To test the behaviour of \( B(s, N) \), consider the following experiment where we test the behaviour of this quantity when considering the support of piecewise constant vectors, under the redundant Haar transform \( D \). Given \( p \in \mathbb{N} \), let \( N = 2^p \), \( D \) be the discrete Haar wavelet frame transform and compute

\[
E(p) = \max \left\{ \tilde{B}(\Delta) : \Delta = \text{supp}(Dx), x \in \Lambda_N \right\}
\]

where \( \tilde{B} \) is as defined in (4.6), and \( \Lambda_N \) is a collection of 1000 randomly generated piecewise constant vectors, each of length \( N \). A plot of \( E \) for \( p = 4, \ldots, 10 \) is shown in Figure 4.

![Figure 4: Plot of \( E \) from (4.9).](image)

5 Localized sparsity

Given a partition \( \{M_j\} \) of the sampling coefficients and a partition \( \{N_j\} \) of the sparsifying coefficients, for some \( p \in (0, 1] \), the signal dependence in the estimates of Theorem 4.4 arise in the localized level sparsities \( \kappa = \{\kappa_j(N, s, p)\}_j \) and the relative sparsities \( \{\hat{\kappa}_j(N, M, \kappa)\}_j \). The following lemma shows that provided that \( VD^* \) satisfies some “block diagonal” structure (so that \( \omega(j, k) \) in Lemma 5.1 decays sufficiently as \( |j - k| \) increases), \( \hat{\kappa}_j \) can be controlled in terms of \( \{\kappa_l\}_{l=1} \) and the dependence on each \( \kappa_l \) decays as \( |j - l| \) increases.

This block diagonal property can be shown to exist when \( V \) is a Fourier sampling transform and \( D \) is a wavelet transform [1], and a consequence of this is that the estimates in Theorem 4.4 can be reduced to depend on \( \{\kappa_j\} \) only. In this section, we present some basic properties of the localized sparsity defined in Definition 4.1. The key findings which would be of use in the proof of Theorem 4.4 are summarized in Corollary 5.3.
Lemma 5.1. Let $V$ and $D$ be isometries. Let $\{N_j\}_{j=1}^r, \{M_j\}_{j=1}^r \in \mathbb{N}^r$ with $\Gamma_j = (M_{j-1}, \ldots, M_j) \cap \mathbb{N}$ and $A_j = (N_{j-1}, \ldots, N_j) \cap \mathbb{N}$ where $M_0 = N_0 = 0$. Suppose that $\|P_{X^*}DV^*P_{\Lambda}\| \leq \omega(k,l)$ and
\[
\sum_{l=1}^r \omega(k,l) \leq C, \quad k = 1, \ldots, r.
\]
Then
\[
\tilde{\kappa}_j \leq C \sum_{l=1}^r \omega(j,l) \kappa_l.
\]

Proof. Since $D^*D = I$,
\[
\tilde{\kappa}_j = \max_{g \in \Theta} \|P_{\Gamma_j \Lambda} V D^* D g\|^2 \leq \max_{g \in \Theta} \left( \sum_{l=1}^r \|P_{\Gamma_j \Lambda} V D^* P_{\Lambda} \| \|D g\| \right)^2 
\]
\[
\leq \max_{g \in \Theta} \left( \sum_{l=1}^r \|P_{\Gamma_j \Lambda} V D^* P_{\Lambda} \| \|D g\| \right)^2 \leq \left( \sum_{l=1}^r \omega(j,l) \sqrt{\kappa_j} \right)^2
\]
\[
\leq \left( \sum_{l=1}^r \omega(j,l) \right) \left( \sum_{l=1}^r \omega(j,l) \kappa_l \right) \leq C \sum_{l=1}^r \omega(j,l) \kappa_l.
\]
\]

Lemma 5.2. Let $p \leq 1$. Suppose that $x$ has at most $s$ non-zero entries and $\|x\|_{\ell^q} = 1$ for some $q \geq 1$. Then, $\|x\|^p_{\ell^p} \leq s^{1-p/q}$.

Proof. Let $\Delta$ denote the support of $x$. By Hölder’s inequality,
\[
\sum_{j \in \Delta} |x_j|^p \leq \left( \sum_{j \in \Delta} |x_j|^q \right)^{p/q} (|\Delta|)^{1-p/q} = s^{1-p/q}.
\]

Lemma 5.3. Let $p \in (0,1]$ and let $q \in [1, \infty]$. Suppose $\kappa > 0$ is such that $\|x\|_{\ell^1} \leq 1$ implies that $\|x\|^p_{\ell^p} \leq \kappa^{1-p/q}$.

(i) $\|x\|_{\ell^1} \leq \kappa \|x\|_{\ell^\infty}$.

(ii) $\|x\|^2_{\ell^2} \leq \kappa \|x\|^2_{\ell^2}$.

(iii) If $p = 2^{-L}$ for some $L \in \mathbb{N} \cup \{0\}$, then $\|x\|^2_{\ell^p} \leq \kappa \|x\|^2_{\ell^2}$.

Proof. Without loss of generality, first assume that $nm \ell^\infty = 1$. Then, $\|x\|_{\ell^p} \leq \kappa$. So, (i) follows because
\[
\sum_{j} |x_j| \leq \sum_{j} |x_j|^p |x_j|^{1-p} \leq \sum_{j} |x_j|^p \leq \kappa,
\]
and (ii) follows because
\[
\sum_{j} |x_j|^2 \leq \sum_{j} |x_j|^p |x_j|^{2-p} \leq \sum_{j} |x_j|^p \leq \kappa.
\]

To show (iii), assume (without loss of generality) that $\|x\|_{\ell^2} = 1$ and recall that $p = 2^{-L}$. Then, by repeatedly applying the Cauchy-Schwarz inequality,
\[
\|x\|^2_{\ell^2} = \left( \sum_{j} |x_j| \right)^2 \leq \sum_{j} |x_j|^p \sum_{j} |x_j|^{2(1-p/2)} \leq \kappa^{1-p/2} \|x\|_{\ell^2} \left( \sum_{j} |x_j|^{2-2p} \right)^{1/2}
\]
\[
\leq \kappa^{1-p/2} \|x\|^{1+1/2}_{\ell^2} \left( \sum_{j} |x_j|^{2-4p} \right)^{1/4} \leq \kappa^{1-p/2} \|x\|^{1+1/2}_{\ell^2} \left( \sum_{j} |x_j|^{2-2^L p} \right)^{1/2^L} = \kappa^{1-p/2} \|x\|^p_{\ell^1},
\]
and by dividing both sides of the inequality by $\|x\|_{p'}$, we obtain

$$\|x\|_{q}^{2(1-p'/2)} \leq \kappa^{1-p'/2} \implies \|x\|_{q}^2 \leq \kappa.$$  

Corollary 5.4. In the notation of Definition 4.1, a direct application of the two lemmas presented above (with $x := P_{\Lambda_j}DD^*y$, $l = 1, \ldots, r$) would yield the following results:

1. Lemma 5.2 : if $D$ was the analysis operator of an orthonormal basis (so that $DD^* = I$) in Definition 4.1, then $\kappa(s, N) = s_1 + \cdots + s_r$.

2. Lemma 5.3 (i) : if $y \in \Sigma_{s, N}$ and $\|P_{\Lambda_j}DD^*y\|_{\infty} = 1$, then

$$\|P_{\Lambda_j}DD^*y\|_{q}^{2} \leq \kappa_j(p, N, s), \quad j = 1, \ldots, r.$$  

3. Lemma 5.3 (ii) : if $y \in \Sigma_{s, N}$ and $\|P_{\Lambda_j}DD^*y\|_{\infty} = 1$, then

$$\|P_{\Lambda_j}DD^*y\|_{q}^{2} \leq \kappa_j(N, s, p), \quad j = 1, \ldots, r.$$  

4. Lemma 5.3 (ii) : if $y \in \Sigma_{s, N}$ and $\|P_{\Lambda_j}DD^*y\|_{\infty} = 1$, then

$$\|P_{\Lambda_j}DD^*y\|_{q}^{2} \leq \kappa_j(N, s, p), \quad j = 1, \ldots, r.$$  

Numerical example: the Haar frame

In the case where $D$ is associated with a Haar frame on $\mathbb{C}^N$, it can be shown that if $|\text{supp}(x)| = s$ then $DD^*x$ has at most $O(s \log N)$ nonzero entries (see [24]). Therefore, from Lemma 5.3, $\kappa(N, s) \lesssim s \log(N)$ where $s = (s_j)_{j=1}^{r}$ and $s = s_1 + \cdots + s_r$. In the case of a Haar frame, experimental results suggest that the localized level sparsities $\{\kappa_j\}$ tend to follow a similar pattern to the level sparsities $\{s_j\}$: Let $D$ be the discrete Haar frame, and let $f \in \mathbb{R}^{1024}$ be as shown on the right of Figure 5. Let $\Delta$ be the support of $Df$. Let $\mathcal{S}$ consist of 1000 randomly generated vectors, each supported on $\Delta$. For each $j = 0, \ldots, 10$, let $\Lambda_j$ index all Haar framelet coefficients in the $j$th scale and let

$$\tilde{\kappa}_j = \sup \left\{ \|P_{\Lambda_j}\eta\|_{q^2} : \eta \in \mathcal{S} \right\}.$$  

We also let $s_j = |\Delta \cap \Lambda_j|$. The bar charts in Figure 5 show for each $j = 0, \ldots, 10$, $\{s_j/|\Lambda_j|\}$ (centre plot) and $\{|\Delta_j|/|\Lambda_j|\}$ (left plot). Note that $\{\tilde{\kappa}_j\}$ merely approximate the localized level sparsities $\{\kappa_j\}$, because otherwise, we would need to consider all $(s, N)$-sparse support sets instead of just one support set $\Delta$ and we would also need to maximize over all vectors supported on $\Delta$, instead of just 1000 randomly generated vectors. Nonetheless, Figure 5 provides some indication of the behaviour of the localized level sparsities.

5.1 Intrinsic localization

As mentioned previously, many of the popular frames such as curvelets, shearlets and wavelet frames are intrinsically localized so that their Gram matrices are near diagonal. This property has been studied for wavelet frames in [23][10][18] and more recently for anisotropic systems such as shearlets and curvelets in [20][21]. In this section, we will show how the property of intrinsic localization can yield estimates on the localized sparsity term, $\kappa$, and the localized level sparsity terms, $\kappa_j$’s. We first define the notion of intrinsic localization. Note that the precise bound on the matrix entries of the Gram matrices of anisotropic systems is more complex than the version presented below. However, the key idea of how to exploit this property to obtain bounds on the localized sparsity values should still be applicable.

Definition 5.5. Let $\mathcal{H}$ be a Hilbert space and let $\Psi = \{\psi_j\}_{j \in \mathbb{N}}$ be a frame for $\mathcal{H}$. $\Psi$ is said to be intrinsically localized with respect to $c > 0$ and $L \geq 1$ if

$$|\langle \psi_j, \psi_k \rangle| \leq \frac{c}{(1 + |j - k|)^{L}}, \quad \forall j, k \in \mathbb{N}.$$
Given $\Delta, \Lambda \subseteq \mathbb{N}$ and $p \in (L^{-1}, 1]$, let

$$I_p(\Delta, \Lambda) = \max_{j \in \Delta} \sum_{k \in \Lambda} |\langle \psi_k, \psi_j \rangle|^p.$$

**Remark 5.1** Note that given any $\Delta, \Lambda \subseteq \mathbb{N}$ and $p \in (L^{-1}, 1]$, note that $Lp > 1$ and

$$I_p(\Delta, \Lambda) \leq \max_{j \in \Delta} \sum_{k \in \Lambda} \frac{c}{(1 + |j - k|)^{Lp}} \leq 1 + \int_1^\infty \frac{c}{|x|^{Lp}} dx \leq 1 + \frac{c}{Lp - 1}.$$  

So $I_p(\Delta, \Lambda)$ is finite. Moreover, if we let

$$d = \operatorname{dist}(\Delta, \Lambda) := \min_{k \in \Delta, j \in \Lambda} |k - j| \geq 1,$$

then

$$I_p(\Delta, \Lambda) \leq \int_{-d}^d \frac{c}{|x|^{Lp}} dx = \frac{c}{(Lp - 1) \cdot d^{Lp-1}}.$$

The main result of this section is as follows.

**Proposition 5.6.** Let $\Psi$ a Parseval frame which is intrinsically localized with respect to $c = 1$ (to simplify the amount of notation only) and $L \geq 1$ and let $D$ be the associated analysis operator. Given any $N = (N_k)_{k=1}^r \in \mathbb{N}^r$ and $s = (s_k)_{k=1}^r \in \mathbb{N}^r$, let $s = s_1 + \cdots + s_r$, and $N_0 = 0$,

$$B = \sup \left\{ \| (P_\Delta D)^j \|_{\ell^\infty \rightarrow \ell^\infty} : \Delta \subset [N_r], |\Delta \cap (N_k - N_{k-1})| = s_k \right\} < \infty,$$

and

$$B' = \sup \left\{ \| (P_\Delta DD^*P_\Delta)^j \|_{\ell^\infty \rightarrow \ell^\infty} : \Delta \subset [N_r], |\Delta \cap (N_k - N_{k-1})| = s_k \right\} < \infty.$$  

Let $d_{j,k} = \operatorname{dist}(\Lambda_j, \Delta_k)$, and recall the definition of localized sparsity and localized level sparsities from Definition 4.1. Let $p \in (L^{-1}, 1]$. Then,

(i) 

$$\kappa(N, s, p) \leq s \cdot \max \left\{ (B^p(1 + 1/(Lp - 1)))^{1/(1-p/2)}, B^p(1 + 1/(Lp - 1)) \right\}.$$

(ii) For $j = 1, \ldots, r$,

$$\kappa_j(N, s, p) \leq s_j \cdot \max \left\{ B^{p/(1-p/2)} \left( \sum_{k=1}^r \frac{s_k/s_j}{d_{j,k}^{Lp-1}} \right)^{1/(1-p/2)}, B^p \sum_{k=1}^r \frac{s_k/s_j}{d_{j,k}^{Lp-1}} \right\}.$$

Note that larger values of $L$ would allow us to select smaller values of $p$, so the larger $L$ is, the closer $\kappa$ and $\kappa_k$ are to $s$ and $s_k$ respectively.
Proof. Note that $B$ and $B'$ are both finite, since there are finitely many subsets of $[N_r]$ and for each subset $\Delta$, $\{\psi_j\}_{j \in \Delta}$ is necessarily a frame for its span with a strictly positive lower frame bound. (i) follows from taking the maximum of (i) and (ii) of Proposition 5.7 over all $\Delta$ subsets with an $(s, N)$-sparsity pattern and plugging in the estimate of $I_p$ from Remark 5.1. (ii) follows from (iii) and (iv) of Proposition 5.7.

**Proposition 5.7.** Let $p \in (0, 1]$ and let $\Delta \subset \mathbb{N}$ with $|\Delta| = s$. Then, for all $g \in \mathcal{R}(D^*P_\Delta)$,

(i) if $\|Dg\|_{\ell^p} = 1$, then $\|Dg\|_{\ell^p}^p \leq \|(P_\Delta D)^{\dagger}\|^p I_p(\Delta, N)s^{1-p/2}$;

(ii) if $\|Dg\|_{\ell^p} = 1$, then $\|Dg\|_{\ell^p}^p \leq \|(P_\Delta D)^{\dagger}\|^p I_p(\Delta, N)s$.

Let $\{\Lambda_j\}_{j=1}^r$ be a partition for $\mathbb{N}$, and let $\Delta_j = \Lambda_j \cap \Delta$ and $s_j = |\Delta_j|$. Then, for all $g \in \mathcal{R}(D^*P_\Delta)$,

(iii) if $\|Dg\|_{\ell^p} = 1$, then

$$\|P_{\Lambda_n}Dg\|_{\ell^p}^p \leq \|(P_\Delta D)^{\dagger}\|^p \sum_{m=1}^r I_p(\Delta_m, \Lambda_n)s_m^{1-p/2}, \quad n = 1, \ldots, r;$$

(iv) if $\|Dg\|_{\ell^p} = 1$, then

$$\|P_{\Lambda_n}Dg\|_{\ell^p}^p \leq \|(P_\Delta D)^{\dagger}\|^p \sum_{m=1}^r I_p(\Delta_m, \Lambda_n)s_m, \quad n = 1, \ldots, r.$$

**Proof.** For (i), suppose that $\|D^*P_\Delta x\|_{\ell^p} = 1$.

$$\|D^*P_\Delta x\|_{\ell^p}^p = \sum_{k \in \mathbb{N}} \left| \sum_{j \in \Delta} x_j \langle \psi_k, \psi_j \rangle \right|^p \leq \sum_{k \in \mathbb{N}} \sum_{j \in \Delta} |x_j|^p |\langle \psi_k, \psi_j \rangle|^p \leq \left( \max_{j \in \Delta} \sum_{k \in \mathbb{N}} |\langle \psi_k, \psi_j \rangle|^p \right) \cdot \left( \sum_{j \in \Delta} |x_j|^p \right)^{1-p/2} \leq \left( \max_{j \in \Delta} \sum_{k \in \mathbb{N}} |\langle \psi_k, \psi_j \rangle|^p \right) \cdot \|(P_\Delta D)^{\dagger}\|^p \cdot \|D^*P_\Delta x\|_{\ell^p}^p \cdot s^{1-p/2},$$

where we have applied the Cauchy-Schwarz inequality in the penultimate line. Therefore,

$$\|D^*P_\Delta x\|_{\ell^p}^p \leq I(\Delta, N)(P_\Delta D)^{\dagger}\|^p \cdot s^{1-p/2}.$$

To show (ii), if we instead assume that $\|D^*P_\Delta x\|_{\ell^\infty} = 1$, then since

$$\|D^*P_\Delta x\|_{\ell^p}^p = \|D^*\|_{\ell^p}^p \|x\|_{\ell^p}^p \leq \|D^*\|_{\ell^p}^p s^{p/2},$$

by plugging this into the last line of (5.2), we obtain

$$\|D^*P_\Delta x\|_{\ell^p}^p \leq I(\Delta, \Lambda)(P_\Delta D)^{\dagger}\|^p s.$$

The proof of (iii) is similar to the above: if $\|D^*P_\Delta x\| = 1$, then for each $n = 1, \ldots, r$,

$$\|P_{\Lambda_n}D^*P_\Delta x\|_{\ell^p}^p \leq \sum_{m=1}^r \left( \max_{j \in \Delta_m} \sum_{k \in \Lambda} |\langle \psi_k, \psi_j \rangle|^p \right) \cdot \sum_{j \in \Delta_m} |x_j|^p \leq \sum_{m=1}^r \max_{j \in \Delta_m} \sum_{k \in \Lambda} |\langle \psi_k, \psi_j \rangle|^p \cdot \|P_{\Delta_m}x\|_{\ell^p}^p \cdot |\Delta_m|^{1-p/2} \leq \|(P_\Delta D)^{\dagger}\|^p \cdot \sum_{m=1}^r I(\Delta_m, \Lambda_n) s_m^{1-p/2},$$

(5.3)
Finally, to show (iv), if \( \|D^*P_\Delta x\|_{\ell^\infty} = 1 \), then for each \( n = 1, \ldots, r \),

\[
\|P_{\Delta_n} DD^* P_{\Delta} x\|_{\ell^p} \leq \sum_{m=1}^r \left( \max_{j \in \Delta_m} \sum_{k \in A} |\langle \psi_k, \psi_j \rangle|^{p} \right) \cdot \sum_{j \in \Delta_m} |x_j|^p
\leq \sum_{m=1}^r \sum_{j \in \Delta_m} \left( \max_{j \in \Delta_m} \sum_{k \in A} |\langle \psi_k, \psi_j \rangle|^{p} \right) \cdot \|P_{\Delta_m} x\|_{\ell^p} \cdot |\Delta_m|.
\]

(5.4)

\[
\leq \|(P_{\Delta} DD^* P_{\Delta})^\dagger\|_{\ell^{\infty} \to \ell^{\infty}} \cdot \sum_{m=1}^r I(\Delta_m, \Lambda_n) \cdot s_m.
\]

\[\square\]

6 Conditions for stable and robust recovery

Given \( \Delta \subset \mathbb{N} \) and some \( f \in \mathcal{H} \), the following proposition presents conditions under which one is guaranteed robust and stable recovery up to \( \|P_{\Delta}^{-1} Df\| \).

**Proposition 6.1** (Dual certificate). Let \( f \in \mathcal{H} \). Let \( \Delta \subset \mathbb{N} \) be such that \( |\Delta| = s \). Let \( \mathcal{W} = \mathcal{R}(D^* P_{\Delta}) \).

For \( r \in \mathbb{N} \), let \( q = \{q_j\}_{j=1}^r \in (0, 1]^r \) and let \( \{\Omega_k\}_{k=1}^r \) be disjoint subsets of \( \mathbb{N} \). Let \( \Omega = \Omega_1 \cup \cdots \cup \Omega_r \).

Suppose that

(i) \( \|QW(V^*(q_1^{-1} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1} P_{\Omega_r})VQW - QW\| < \frac{1}{4} \),

(ii) \( \sup_{j \in \mathbb{N}} \|P_{\Omega(j)} DQW^* (q_1^{-1} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1} P_{\Omega_r})VQW D^* P_{\Omega(j)}\| < \frac{1}{2} \),

and that there exists \( \rho = V^* P_{\Omega} w \) and \( L > 0 \) with the following properties.

(iii) \( \|D^* P_{\Delta} \text{sgn} (P_{\Delta} D x) - QW \rho\| \leq q^{1/2} / 8 \),

(iv) \( \|P_{\Delta} DQW^*\|_{\ell^\infty} \leq 1/2 \),

(v) \( \|\rho\|_{\ell^2} \leq L \sqrt{s} \).

Let \( y \in \ell^2(\mathbb{N}) \) be such that \( \|P_{\Omega_1} V f - y\| \leq \delta \). Then, any minimizer \( \hat{f} \in \mathcal{H} \) of (3.2) satisfies

\[
\|f - \hat{f}\|_{\mathcal{H}} \lesssim \delta \left( q^{-1/2} + L \sqrt{\delta} \right) + \|P_{\Delta}^{-1} Df\|_{\ell^1}.
\]

(6.1)

**Proof.** Since \( D \) is an isometry, \( D^* D \) is the identity on \( \mathcal{H} \). Given any \( g \in \mathcal{H} \),

\[
Q_{\mathcal{W}}^{-1} g = Q_{\mathcal{W}}^{-1} D^* g = Q_{\mathcal{W}}^{-1} D^* P_{\Delta} g
\]

(6.2)

since \( Q_{\mathcal{W}} \) is the orthogonal projection onto \( \mathcal{R}(D^* P_{\Delta}) \). So, using the assumption that \( \|D\|_{\mathcal{H} \to \ell^2} \leq 1 \), for any \( g \in \mathcal{H} \),

\[
\|g\| \leq \|QW g\| + \|Q_{\mathcal{W}}^{-1} g\| \leq \|QW g\|_{\mathcal{H}} + \|P_{\Delta}^{-1} Dg\|_{\ell^1}.
\]

Now, let \( h = f - \hat{f} \). To bound \( \|h\|_{\mathcal{H}} \), it suffices to derive bounds for \( \|QW h\| \) and \( \|P_{\Delta}^{-1} Dh\|_{\ell^1} \).

Let \( V_{\Omega, q} = QW V^*(q_1^{-1} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1} P_{\Omega_r})VQW \). We first observe that (i) implies that \( V_{\Omega, q} \) has a bounded inverse on \( QW(\mathcal{H}) \), with

\[
\|(QW V^*(q_1^{-1} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1} P_{\Omega_r})VQW)^{-1}\|_{\mathcal{H} \to \mathcal{H}} \leq \frac{4}{3},
\]

and

\[
\|(q_1^{-1/2} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1/2} P_{\Omega_r})VQW\| \leq \sqrt{\frac{5}{4}}.
\]

Observe also that

\[
\|P_{\Omega} V h\| \leq \|P_{\Omega} V f - y\| + \|P_{\Omega} V \hat{f} - y\| \leq 2\delta.
\]

18
By applying the above observations, we have that
\[
\|Q^\ast h\| = \left\| V^{-1}_{\Omega, \alpha} V_{\Omega, \alpha} q h \right\|
\leq \left\| V^{-1}_{\Omega, \alpha} \right\|_{\mathcal{H}} \|Q^\ast V^\ast (q_1^{-1} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1} P_{\Omega_r}) V(h - Q^\ast_W h)\|
\leq \frac{4}{3} \left( \sqrt[3]{q^{-1/2}} + \frac{5}{4} \left\| (q_1^{-1/2} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1/2} P_{\Omega_r}) V Q^\ast_W h \right\| \right) .
\] (6.3)

Also, by (ii),
\[
\left\| (q_1^{-1/2} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1/2} P_{\Omega_r}) V Q^\ast_W h \right\| = \left\| (q_1^{-1/2} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1/2} P_{\Omega_r}) V Q^\ast_W D^\ast P^\perp h \right\|_{\ell^2}
\leq \sup_{j \in \Delta^s} \left\| (q_1^{-1/2} P_{\Omega_1} \oplus \cdots \oplus q_r^{-1/2} P_{\Omega_r}) V Q^\ast_W D^\ast e_j \right\|_{\ell^2} \|P^\perp h\|_{\ell^2} \leq \sqrt{\frac{5}{4}} \|P^\perp h\|_{\ell^1} .
\]

Plugging this into (6.3) yields
\[
\|Q^\ast h\| \leq 2 \left( q^{-1/2} + \|P^\perp h\|_{\ell^1} \right) .
\] (6.4)

So, to bound \(\|h\|_{\mathcal{H}}\), it suffices to bound \(\|P^\perp h\|_{\ell^1}\).

Observe that
\[
\|Df\|_{\ell^1} = \|P^\perp D(f + h)\|_{\ell^1} + \|P^\perp D(f + h)\|_{\ell^1}
\geq \|P^\perp D h\|_{\ell^1} - 2 \|P^\perp D f\|_{\ell^1} + \|P^\perp D f\|_{\ell^1} + \|P^\perp D h\|_{\ell^1}
= \|P^\perp D h\|_{\ell^1} - 2 \|P^\perp D f\|_{\ell^1} + \|Df\|_{\ell^1} + \|P^\perp D h\|_{\ell^1} + \|P^\perp D h\|_{\ell^1} + \|P^\perp D h\|_{\ell^1} .
\]

Since \(\hat{f}\) is a minimizer, we can deduce that
\[
\|P^\perp h\|_{\ell^1} \leq \|P^\perp D h, \text{sgn}(P^\perp D f)\| + 2 \|P^\perp D f\|_{\ell^1} .
\]

Using the existence of \(\rho = V^\ast P_{\Omega} v\) and recalling that \(Q^\perp_W = Q^\perp_W D^\ast P^\perp h\) from (6.2), we have that
\[
\|P^\perp D h, \text{sgn}(P^\perp D f)\| = \| (h, D^\ast \text{sgn}(P^\perp D f) \| < (h, \text{sgn}(P^\perp D f)\|
\leq \|(h, D^\ast \text{sgn}(P^\perp D f) - Q^\perp_W \| + \| (h, \text{sgn}(P^\perp D f)\|
\leq \|Q^\perp_W h\|_{\ell^1} \|D^\ast \text{sgn}(P^\perp D f) - Q^\perp_W \|_{\ell^1} + \|P^\perp V h\|_{\ell^1} \|w\|_{\ell^1} + \| (D^\ast P^\perp h, Q^\perp_W \|_{\ell^1} + \|D^\ast P^\perp h, Q^\perp_W \|_{\ell^1}
\leq \frac{\sqrt{5}}{4} \|Q^\perp_W h\|_{\ell^1} + 2 \delta L \sqrt{\gamma} + \frac{1}{4} \|P^\perp D h\|_{\ell^1}
\leq \|P^\perp D h\|_{\ell^1} + \frac{3}{2} \|P^\perp h\|_{\ell^1} ,
\]

where we have used the bound on \(\|Q^\perp_W h\|_{\ell^1}\) from (6.4) to obtain the last inequality. Therefore,
\[
\|P^\perp h\|_{\ell^1} \leq 8 \|P^\perp D f\|_{\ell^1} + \delta (1 + 8L \sqrt{\gamma}) .
\]

Finally, combining this with (6.4) yields
\[
\|h\|_{\ell^1} \leq \delta \left( 2q^{-1/2} + 3 (1 + 8L \sqrt{\gamma}) \right) + 16 \|P^\perp D f\|_{\ell^1} .
\]

\[\square\]

**Proposition 6.2** (Dual certificate for the unconstrained problem). Consider the setting of Proposition 6.1 and assume that conditions (i)-(v) are satisfied. Let \(\alpha > 0\) and suppose that \(f \in \mathcal{H}\) is a minimizer of
\[
\inf_{g \in \mathcal{H}} \alpha \|Dg\|_{\ell^1} + \|P^\perp V g - y\|_{\ell^2}^2 ,
\]

19
where $y \in \ell^2(\mathbb{N})$ is such that $\|P_\Omega Vf - y\| \leq \delta$. Then,

$$\|f - \hat{f}\|_\mathcal{H} \leq \frac{\delta^2}{\alpha} + \alpha \left( \frac{1}{\sqrt{q}} + L\sqrt{\kappa} \right)^2 + \delta \left( \frac{1}{\sqrt{q}} + L\sqrt{\kappa} \right) + \|P_\Delta^h Df\|_{\ell^1}.$$  

Thus, if $\alpha = \sqrt{q}\delta$, then

$$\|f - \hat{f}\|_\mathcal{H} \leq \delta \left( \frac{1}{\sqrt{q}} + L\sqrt{\kappa} + \sqrt{q}L^2\kappa \right) + \|P_\Delta^h Df\|_{\ell^1}.$$  

**Proof.** Let $h = f - \hat{f}$, just as in Proposition 6.1

$$\|h\|_\mathcal{H} \leq \|QWh\|_\mathcal{H} + \|P_\Delta^h Dh\|_{\ell^1}$$

and it suffices to bound the two terms on the right side of the inequality. We first consider $\|QWh\|_\mathcal{H}$. By applying assumptions (i) and (ii), we can proceed as in the proof of Proposition 6.1 to derive

$$\|QWh\|_\mathcal{H} \leq \frac{4}{3} \left( \sqrt{\frac{3}{2}} q^{1/2} ||P_\Omega Vh||_{\ell^2} + \frac{5}{4} \|P_\Delta^h Dh\|_{\ell^1} \right).$$

(6.5)

Then, by letting $\lambda = \|y - P_\Omega V\hat{f}\|_{\ell^2}$ and observing that

$$\|P_\Omega Vh\|_{\ell^2} \leq \|P_\Omega Vf - y\|_{\ell^2} + \|P_\Omega V\hat{f} - y\|_{\ell^2} \leq \delta + \lambda,$$

we have that

$$\|QWh\|_\mathcal{H} \leq \frac{4}{3} \left( \sqrt{\frac{3}{2}} q^{1/2} (\delta + \lambda) + \frac{5}{4} \|P_\Delta^h Dh\|_{\ell^1} \right) \leq 2 \left( \frac{(\delta + \lambda)}{\sqrt{q}} + \|P_\Delta^h Dh\|_{\ell^1} \right).$$

(6.6)

To bound $\|P_\Delta^h Dh\|_{\ell^1}$, first observe that

$$\alpha\|D\hat{f}\|_{\ell^1} \geq \alpha \|P_\Delta^h Dh\|_{\ell^1} - 2\alpha\|P_\Delta^h Df\|_{\ell^1} + \alpha\|Df\|_{\ell^1} + \alpha \Re \langle P_\Delta Dh, \sgn(P_\Delta Df) \rangle$$

$$+ \lambda^2 - \lambda^2 + \|y - P_\Omega V\hat{f}\|_{\ell^2}^2 - \delta^2.$$  

Since $\hat{f}$ is a minimizer, it follows that $\alpha\|D\hat{f}\|_{\ell^1} + \lambda^2 \leq \alpha\|Df\|_{\ell^1} + \|y - P_\Omega V\hat{f}\|_{\ell^2}^2$ and therefore,

$$\alpha \|P_\Delta^h Dh\|_{\ell^1} + \lambda^2 \leq 2\alpha \|P_\Delta^h Df\|_{\ell^1} + \alpha \left( \langle P_\Delta Dh, \sgn(P_\Delta Df) \rangle + \delta^2 \right).$$

(6.7)

In the same way as in the proof of Proposition 6.1, we may apply the properties of the dual certificate $\rho = V^*P_\Omega w$ to bound $\langle P_\Delta Dh, \sgn(P_\Delta Df) \rangle$, so that the following holds.

$$\langle P_\Delta Dh, \sgn(P_\Delta Df) \rangle \leq \frac{\sqrt{q}}{8} \|QWh\|_\mathcal{H} + \|P_\Omega Vh\|_{\ell^2} \|w\|_{\ell^2} + \frac{1}{4} \|P_\Delta^h Dh\|_{\ell^1}.$$

By inserting the bound from (6.6), recalling that $\|P_\Omega Vh\|_{\ell^2} \leq \delta + \lambda$ and that $\|w\|_{\ell^2} \leq L\sqrt{\kappa}$, it follows that

$$\langle P_\Delta Dh, \sgn(P_\Delta Df) \rangle \leq (\delta + \lambda) \left( \frac{1}{4} + L\sqrt{\kappa} \right) + \frac{1}{2} \|P_\Delta^h Dh\|_{\ell^1}.$$  

Plugging this bound into (6.7) now yields

$$\lambda^2 + \frac{\alpha}{2} \|P_\Delta^h Dh\|_{\ell^1} \leq 2\alpha \|P_\Delta^h Df\|_{\ell^1} + \alpha (\delta + \lambda) \left( \frac{1}{4} + L\sqrt{\kappa} \right) + \delta^2.$$  

(6.8)
This implies that
\[
\lambda^2 - \alpha \left( \frac{1}{4} + L\sqrt{\kappa} \right) \lambda - \delta^2 - \alpha \delta \left( \frac{1}{4} + L\sqrt{\kappa} \right) \leq 0
\]
and by applying the quadratic formula and observing that \( \lambda \geq 0 \), it follows that
\[
\lambda \leq \frac{\alpha/4 + L\alpha\sqrt{\kappa} + \sqrt{(\alpha/4 + L\alpha\sqrt{\kappa})^2 + 4 \left( \delta^2 + \alpha\delta(1/4 + L\sqrt{\kappa}) + 2\alpha\|P^\perp \Delta f\|_{\ell^1} \right)}}{2}
\]
Note that \( \|\cdot\|_2 \leq \|\cdot\|_1 \), and so,
\[
\lambda \leq \frac{\alpha/4 + L\alpha\sqrt{\kappa} + \delta + \sqrt{\alpha\delta(1/4 + L\sqrt{\kappa}) + \sqrt{2\alpha\|P^\perp \Delta f\|_{\ell^1}}}}{2}
\]
where the second inequality comes from the fact that \( \sqrt{a}b \leq (a+b)/2 \) for any \( a, b \geq 0 \). We also know from (6.8) that
\[
\|P^\perp \Delta f\|_{\ell^1} \leq 4\|P^\perp \Delta f\|_{\ell^1} + (\delta + \lambda) \left( \frac{1}{2} + 2L\sqrt{\kappa} \right) + \frac{2\delta^2}{\alpha}.
\]
By combining this with the bound from (6.6),
\[
\|h\|_H \leq \|Q_{\omega}h\|_H + \|P^\perp \Delta h\|_{\ell^1}
\leq 3 \left( 4\|P^\perp \Delta f\|_{\ell^1} + \delta \left( \frac{1}{2} + 2L\sqrt{\kappa} \right) + \frac{2\delta^2}{\alpha} \right)
+ \frac{2\delta}{\sqrt{q}} + \left( \frac{2}{\sqrt{q}} + 3 \left( \frac{1}{2} + 2L\sqrt{\kappa} \right) \right) \lambda
\leq \|P^\perp \Delta f\|_{\ell^1} + \delta(1 + L\sqrt{\kappa}) + \frac{\delta^2}{\alpha} + \frac{\delta}{\sqrt{q}} + \left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right) \lambda.
\]
Recalling (6.9),
\[
\left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right) \lambda \leq \left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right) \left( \alpha(1 + L\sqrt{\kappa}) + \delta + \sqrt{\alpha\|P^\perp \Delta f\|_{\ell^1}} \right)
\leq \left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right) \left( \alpha(1 + L\sqrt{\kappa}) + \delta + \frac{\alpha}{q} + \|P^\perp \Delta f\|_{\ell^1} \right) + \alpha(1 + L\sqrt{\kappa})
\leq \alpha \left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right)^2 + \delta \left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right) + \|P^\perp \Delta f\|_{\ell^1}.
\]
Therefore,
\[
\|h\|_H \leq \frac{\delta^2}{\alpha} + \alpha \left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right)^2 + \delta \left( \frac{1}{\sqrt{q}} + 1 + L\sqrt{\kappa} \right) + \|P^\perp \Delta f\|_{\ell^1}.
\]

7 Overview of the proof of Theorem 4.4

The remainder of this paper is focussed on the proof of Theorem 4.4 and we begin by setting some notation which will be used throughout.

Let \( V, D \in B(\mathcal{H}, L^2(\mathbb{N})) \) be isometries. Let \( f \in \mathcal{H} \). For \( r \in \mathbb{N}, M \in \mathbb{N} \) and \( N \in \mathbb{N} \), let \( \mathcal{N} = \{N_k\}_{k=1}^r \in \mathbb{N}^r, M = \{M_k\}_{k=1}^r \in \mathbb{N}^r, s = \{s_k\}_{k=1}^r \in \mathbb{N}^r, m = \{m_k\}_{k=1}^r \in \mathbb{N}^r, \) with
- \( 0 = M_0 < M_1 < \cdots < M_r =: M \), and let \( \Gamma_k = (M_{k-1}, M_k] \cap \mathbb{N} \) and \( \Omega_k \sim \text{Ber}(q_k, \Gamma_k) \).
Definition 7.1. Let $\Omega = \Omega_{\text{Ber}}$, where $\Omega_{\text{Ber}} := \{\delta_j : j \in \Gamma_k\}$ with $\Gamma_k = \mathbb{N} \cap (M_{k-1}, \ldots, M_k]$, and $\delta_j$ are independent random variables such that $\mathbb{P}(\delta_j = 1) = m_k/(M_k - M_{k-1})$ and $\mathbb{P}(\delta_j = 0) = 1 - m_k/(M_k - M_{k-1})$. The Bernoulli sampling set $\Omega_{\text{Ber}}$ described will be denoted by $\Omega_{\text{Ber}}^* \sim \text{Ber}(m_k/(M_k - M_{k-1}), \Gamma_k)$ and we say that $\Omega_{\text{Ber},M,m} = \Omega_{\text{Ber}}^* \cup \cdots \cup \Omega_{\text{Ber}}^{r}$ with $\Omega_{\text{Ber}}^{r}$ is a Bernoulli $(M, m)$-sampling scheme.

As explained in [8, II.C] (see also [19]), the probability that one of the conditions of Proposition 6.1 fails for $\Omega = \Omega_{\text{Ber},M,m}$ chosen uniformly at random is up to a constant bounded from above by the probability that one of these conditions fails under the Bernoulli $(M, m)$-sampling scheme, $\Omega = \Omega_{\text{Ber}}^*$.

So, to prove Theorem 4.4 it suffices to show that conditions (i) to (v) of Proposition 6.1 hold with probability exceeding $1 - \epsilon$ with $\Omega = \Omega_{\text{Ber},M,m}$ satisfying the following assumption.

Assumption 7.2. Let $\mathcal{L} = (\log(\epsilon^{-1}) + 1) \log(\tilde{M}^{-1} k_{\max}^2 \|D^*\|_{l^\infty \to l^\infty})$ and

$$B = \max \left\{ \|DQ_{\tilde{V}}^* D^*\|_{l^\infty \to l^\infty}, \sqrt{\|DQ_{\tilde{W}} D^*\|_{l^\infty \to l^\infty}} \max_{i=1}^r \left\{ \sum_{t=1}^r \|P_{\Lambda_t} DQ_{\mathcal{R}(D^*, P_{\Lambda_t})} D^* P_{\Lambda_t}\|_{l^\infty \to l^\infty} \right\} \right\}.$$

Let

$$\tilde{M} = \min \left\{ i \in \mathbb{N} : \max_{j \geq 1} \left| P_{\tilde{M}} V D^* e_j \right|_{l^\infty} \leq \frac{q}{8 \sqrt{k_{\max}}}, \max_{j \geq 1} \left| Q_{\mathcal{R}(D^*, P_{\Lambda})} D^* e_j \right| \leq \frac{\sqrt{q} \epsilon}{4} \right\}.$$

Suppose that

\begin{enumerate}[(a)]
\item $\|DQ_{\tilde{V}}^* P_{\tilde{M}}^+ V Q_{\tilde{W}} D^*\|_{l^2 \to l^\infty} \leq \sqrt{\min_{k_{\max}} \log^{-1/2}(4q^{-1} \sqrt{k_{\max}} \tilde{M})}$.
\item $\|DQ_{\tilde{V}}^* P_{\tilde{M}}^+ V Q_{\tilde{W}}^* T^{-1} D^*\|_{l^2 \to l^\infty} \leq \frac{1}{\sqrt{k_{\max}}}$.
\item For $k = 1, \ldots, r$,

$$q_k \geq \sqrt{r} L B \sum_{l=1}^r \mu_{M,k,l}^2 \kappa_l.$$
\end{enumerate}
For \( k = 1, \ldots, r \), \( q_k \geq L \hat{q}_k \) with \( \{q_k\}_{k=1}^{r} \) satisfying
\[
1 \gtrsim r B^2 \sum_{k=1}^{r} (\hat{q}_k - 1) \mu_{N,M}^2(k,j) \hat{k}_k, \quad j = 1, \ldots, r.
\]

Note that this assumption is strictly weaker than the assumptions of Theorem 4.4, and by showing that conditions (i) to (v) of Proposition 6.1, we will prove that the error bound (6.1) holds for one support set \( \Delta \). So, by ensuring that the conditions of this assumption hold over all \( \Delta \) sets which are \((N,s)\) sparse patterns (as required by Theorem 4.4), we can conclude that (6.1) holds for any \((N,s)\) sparse support sets.

Under this assumption,

- §9 will show that conditions (iii) to (v) of Proposition 6.1 are satisfied with probability exceeding \( 1 - \frac{5}{6} \epsilon \);
- §10 will show that conditions (i) and (ii) of Proposition 6.1 are satisfied with probability exceeding \( 1 - \frac{\epsilon}{6} \);
- §8 will present some preliminary results for use in §9 and §10.

The proof of Corollary 4.5

Once we have shown that the conditions of Proposition 6.1 hold with probability exceeding \( 1 - \epsilon \), the conclusion of Proposition 6.2 automatically follows and we may conclude Corollary 4.5.

8 Preliminary results

In this section, we present four propositions which will be applied to show that the conditions of Proposition 6.1 are satisfied with high probability under the conditions of Theorem 4.4 with a Bernoulli sampling scheme. The results in this section are derived using Talagrand’s inequality and Bernstein inequalities (for random variables and random matrices) which we state below.

**Theorem 8.1.** (Talagrand [27, Cor. 7.8]) There exists a number \( K \) with the following property. Consider \( n \) independent random variables \( X_i \) valued in a measurable space \( \Omega \) and let \( \mathcal{F} \) be a (countable) class of measurable functions on \( \Omega \). Let \( Z \) be the random variable \( Z = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i) \) and define
\[
S = \sup_{f \in \mathcal{F}} \|f\|_{\infty}, \quad V = \sup_{f \in \mathcal{F}} \mathbb{E} \left( \sum_{i=1}^{n} f(X_i)^2 \right).
\]
If \( \mathbb{E}(f(X_i)) = 0 \) for all \( f \in \mathcal{F} \) and \( i \leq n \), then, for each \( t > 0 \), we have
\[
\mathbb{P}(\|Z - \mathbb{E}(Z)\| \geq t) \leq 3 \exp \left( -\frac{1}{2} \frac{t^2}{K^2} \log \left( 1 + \frac{1}{V + S^2 \mathbb{E}(Z)} \right) \right),
\]
where \( Z = \sup_{f \in \mathcal{F}} |\sum_{i \leq n} f(X_i)| \).

**Theorem 8.2** (Bernstein inequality for random variables [19]). Let \( Z_1, \ldots, Z_M \in \mathbb{C} \) be independent random variables with zero mean such that \( |Z_j| \leq K \) almost surely for all \( l = 1, \ldots, M \) and some constant \( K > 0 \). Assume also that \( \sum_{j=1}^{M} \mathbb{E} |Z_j|^2 \leq \sigma^2 \) for some constant \( \sigma^2 > 0 \). Then for \( t > 0 \),
\[
\mathbb{P} \left( \left| \sum_{j=1}^{M} Z_j \right| \geq t \right) \leq 4 \exp \left( -\frac{t^2/4}{\sigma^2 + Kt/(3\sqrt{2})} \right).
\]
If \( Z_1, \ldots, Z_M \in \mathbb{R} \) are real instead of complex random variables, then
\[
\mathbb{P} \left( \left| \sum_{j=1}^{M} Z_j \right| \geq t \right) \leq 2 \exp \left( -\frac{t^2/2}{\sigma^2 + Kt/3} \right).
\]
Theorem 8.3 (Bernstein inequality for rectangular matrices [26]). Let $Z_1, \ldots, Z_M \in \mathbb{C}^{d_1 \times d_2}$ be independent random matrices such that $\mathbb{E}Z_j = 0$ for each $j = 1, \ldots, M$ and $\|Z_j\|_{2 \to 2} \leq K$ almost surely for each $j = 1, \ldots, M$ and some constant $K > 0$. Let

$$
\sigma^2 := \max\left\{ \left\| \sum_{j=1}^M \mathbb{E}(Z_j Z_j^*) \right\|_{\ell^2 \to \ell^2}, \left\| \sum_{j=1}^M \mathbb{E}(Z_j^* Z_j) \right\|_{\ell^2 \to \ell^2} \right\}.
$$

Then, for $t > 0$,

$$
P\left( \left\| \sum_{j=1}^M Z_j \right\|_{\ell^2 \to \ell^2} \geq t \right) \leq 2(d_1 + d_2) \exp\left( \frac{-t^2/2}{\sigma^4 + K t/3} \right)
$$

Proposition 8.4. Let $g \in \mathcal{W}$ and let $\alpha > 0$ and $\gamma \in [0, 1]$. Suppose that

$$
\|TD(QWV^*P_{[M]}VQW - QW)D^*T^{-1}\|_{\ell^2 \to \ell^2} \leq \alpha/2.
$$

Then

$$
P\left( \|TD(QWV^*(q_1^{-1}P_{\Omega_1} \cdots \otimes q_r^{-1}P_{\Omega_r})VQW - QW)g\|_{\ell^2} \geq \gamma \|TDg\|_{\ell^2} \right) \leq \gamma
$$

provided that

$$
\sqrt{rB} \log \left( \frac{3}{\gamma} \right) \sum_{k=1}^r \mu_{N,M}(k, l) \kappa l \lesssim \alpha, \quad k = 1, \ldots, r,
$$

and

$$
rB^2 \log \left( \frac{3}{\gamma} \right) \sum_{k=1}^r (q_k^{-1} - 1) \mu_{N,M}(k, l) \kappa k \lesssim \alpha^2, \quad l = 1, \ldots, r,
$$

where $B^2 = \|DQW^*D^*\|_{\ell^\infty \to \ell^\infty} \max_{l=1}^r \|P_{\Omega_l} DQW^*D^*P_{\Omega_l}\|_{\ell^\infty \to \ell^\infty}$.

Proof. Without loss of generality, assume that $\|TDg\|_{\ell^2} = 1$. Let $\{\delta_j\}_{j=1}^M$ be random Bernoulli variables such that $P(\delta_j = 1) = \tilde{q}_j$ where $\tilde{q}_j = q_k$ for $j = M_{k-1} + 1, \ldots, M_k$. Then,

$$
TD(QWV^*(q_1^{-1}P_{\Omega_1} \otimes \cdots \otimes q_r^{-1}P_{\Omega_r})VQW - QW)g
$$

$$
= \sum_{j=1}^M (\tilde{q}_j^{-1} - 1)TDQW^*(e_j \otimes \overline{\tau}_j)VQWg + TD(QWV^*P_{[M]}VQW - QW)g,
$$

where $\otimes$ is the Kronecker product. Since $D^*D = I$ and since (8.1) holds by assumption,

$$
\|TD(QWV^*P_{[M]}VQW - QW)g\|_{\ell^2} \leq \|TD(QWV^*P_{[M]}VQW - QW)D^*T^{-1}TDg\|_{\ell^2}
$$

$$
\leq \|TD(QWV^*P_{[M]}VQW - QW)D^*T^{-1}\|_{\ell^2 \to \ell^2} \leq \alpha/2.
$$

So, it suffices to show that

$$
P\left( \left\| \sum_{j=1}^M Y_j \right\|_{\ell^2} \geq \alpha/2 \right) \leq \gamma,
$$

where for each $j = 1, \ldots, M$, $Y_j = (\tilde{q}_j^{-1} - 1)DQW^*(e_j \otimes \overline{\tau}_j)Vg$. We will aim to apply Talagrand’s inequality (Theorem 8.1) to obtain this probability bound.

Let $\mathcal{G}$ be a countable set of vectors in the unit ball of $\ell^2(\mathbb{N})$ and for each $\zeta \in \mathcal{G}$, define the linear functionals $\hat{\zeta}_1, \hat{\zeta}_2 : \ell^2(\mathbb{N}) \to \mathbb{R}$ by

$$
\hat{\zeta}_1(y) = \text{Re} \langle y, \zeta \rangle, \quad \hat{\zeta}_2(y) = -\text{Re} \langle y, \zeta \rangle, \quad \forall y \in \ell^2(\mathbb{N}).
$$

Let $\mathcal{F} = \{\hat{\zeta}_1, \hat{\zeta}_2 : \zeta \in \mathcal{G}\}$. Then, $Z := \sup_{f \in \mathcal{F}} \sum_{j=1}^M f(Y_j) = \left\| \sum_{j=1}^M Y_j \right\|_{\ell^2}$. 

24
To bound $S = \max_j \|Y_j\|_{\ell^2}$:

$$\|Y_j\|_{\ell^2} \leq \tilde{q}_j^{-1}\|TDQ_WV^*(e_j \otimes \tau_j)Vg\|_{\ell^2} = \tilde{q}_j^{-1} \sup_{\|x\|_{\ell^2} = 1} |\langle TDQ_WV^*(e_j \otimes \tau_j)Vg, x \rangle|$$

$$= \tilde{q}_j^{-1} \sup_{\|x\|_{\ell^2} = 1} |\langle Vg, e_j \rangle \langle TDQ_WV^*e_j, x \rangle|$$

For each $j \in \Gamma_k$,

$$|\langle Vg, e_j \rangle| = |\langle VD^*Dg, e_j \rangle| \leq \sum_{l=1}^r |\langle VD^*P_{\Lambda_l}Dg, e_j \rangle| \leq \sum_{l=1}^r \mu(P_{\Lambda_l} VD^*P_{\Lambda_l}) \|P_{\Lambda_l}Dg\|_{\ell^1}.$$ 

Observe that $\|TDg\|_{\ell^2} = 1$ implies $\|P_{\Lambda_l}Dg\|_{\ell^2} \leq \sqrt{\kappa_l}$ for each $l = 1, \ldots, r$. Furthermore, by (4) of Corollary 3.4 this implies that $\|P_{\Lambda_l}Dg\|_{\ell^1} \leq \kappa_l \sqrt{\tau}$. So, it follows that

$$|\langle Vg, e_j \rangle| \leq \sqrt{\tau} \sum_{l=1}^r \mu(P_{\Lambda_l} VD^*P_{\Lambda_l}) \kappa_l.$$ 

Also, if $\|x\|_{\ell^2} = 1$,

$$|\langle TDQ_WV^*e_j, x \rangle|^2 \leq \sum_{l=1}^r \|P_{\Lambda_l} TDQ_WV^*e_j\|^2_{\ell^2} = \sum_{l=1}^r \frac{1}{r \kappa_l} \|P_{\Lambda_l} DQ_WD^*DV^*e_j\|^2_{\ell^2}$$

$$\leq \sum_{l=1}^r \frac{1}{r \kappa_l} \|P_{\Lambda_l} DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2} \|DV^*e_j\|^2_{\ell^\infty} \leq \|DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2}^2 \mu^2(P_{\Lambda_l} VD),$$

where the last inequality follows because $\|P_{\Lambda_l}Df\|_{\ell^2} \leq \kappa_l$ for all $f \in W$ with $\|Df\|_{\ell^\infty} \leq 1$. Therefore,

$$\max_j \|Y_j\|_{\ell^2} \leq \max_{k=1}^M \|DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2} \sqrt{\tau} \sum_{l=1}^r \mu_{N,M}(k,l) \kappa_l.$$ 

To bound $V = \sup_{f \in F} \mathbb{E} \sum_{j=1}^M f(Y_j)^2 = \sup_{\xi \in \mathcal{G}} \mathbb{E} \sum_{j=1}^M |\langle \xi, Y_j \rangle|^2$:

$$V \leq \sup_{\xi \in \mathcal{G}} \sum_{j=1}^M (q_j^{-1} - 1) |\langle e_j, Vg \rangle|^2 |\langle V^*e_j, QW^*T \xi \rangle|^2$$

$$\leq \sup_{\xi \in \mathcal{G}} \sum_{k=1}^M (q_k^{-1} - 1) \|P_{\Lambda_k} Vg\|^2_{\ell^2} \max_{j \in \Gamma_k} |\langle V^*e_j, QW^*T \xi \rangle|^2$$

$$\leq \sum_{k=1}^M \sum_{r=1}^\infty \max_{j \in \Gamma_k} |\langle TDQ_WV^*e_j, x \rangle|^2.$$ 

By combining $\|P_{\Lambda_l}Dg\|_{\ell^2} \leq \sqrt{\kappa_l}$ (which follows from $\|TDg\|_{\ell^2} = 1$) with the definition of $\kappa_k$, we obtain $\|P_{\Lambda_l} Vg\|^2_{\ell^2} \leq \tau \kappa_k$. Also, for each $j \in \Gamma_k$,

$$\|TDQ_WV^*e_j\|^2_{\ell^2} = \sum_{l=1}^r \frac{1}{r \kappa_l} \|P_{\Lambda_l} DQ_WV^*e_j\|^2_{\ell^2}$$

$$= \sum_{l=1}^r \frac{1}{r \kappa_l} \|P_{\Lambda_l} DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2} \|DV^*e_j\|_{\ell^\infty} \|P_{\Lambda_l} DQ_WV^*e_j\|_{\ell^2}$$

$$\leq \sum_{l=1}^r \frac{1}{r \kappa_l} \|DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2} \|DV^*e_j\|_{\ell^\infty} \sum_{l=1}^r \|P_{\Lambda_l} DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2} \|P_{\Lambda_l} V^*e_j\|_{\ell^\infty}$$

$$\leq \sum_{l=1}^r \frac{1}{r \kappa_l} \|DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2} \sum_{l=1}^r \|P_{\Lambda_l} DQ_WD^*\|_{\ell^\infty \rightarrow \ell^2} \mu_{N,M}(k,l)$$
where we have used, from the definition of the $\kappa_j$’s, $\|P_{\alpha} Df\|_{\ell^2} \leq \sqrt{r_t}$ whenever $t = 1, \ldots, r$, $f \in W$ and $\|Df\|_{\ell^\infty} \leq 1$. Therefore,

\[
\mathcal{V} \leq r \|DQ_W D^*\|_{\ell^\infty \rightarrow \ell^\infty} \sum_{l=1}^{r} \sum_{k=1}^{r} (q_k^{-1} - 1) \frac{\mu_{N,M}^2(k,l)}{q_k} \leq r \hat{B}^2 \max_{l=1}^{r} \sum_{k=1}^{r} (q_k^{-1} - 1) \mu_{N,M}^2(k,l) \hat{\kappa}_k,
\]

where we have applied

\[
\frac{1}{r} \|DQ_W D^*\|_{\ell^\infty \rightarrow \ell^\infty} \sum_{l=1}^{r} \sum_{t=1}^{r} \|P_{\alpha} DQ_W D^* P_{\alpha}\|_{\ell^\infty \rightarrow \ell^\infty} \leq \|DQ_W D^*\|_{\ell^\infty \rightarrow \ell^\infty} \max_{l=1}^{r} \sum_{t=1}^{r} \|P_{\alpha} DQ_W D^* P_{\alpha}\|_{\ell^\infty \rightarrow \ell^\infty} =: \hat{B}^2.
\]

- To bound $E(Z) = E\left[ \left\| \sum_{j=1}^{M} Y_j \right\| \right]$:

\[
E \left( \left\| \sum_{j=1}^{M} Y_j \right\|_{\ell^2}^2 \right) = \sum_{j=1}^{M} E \left[ \left\| Y_j \right\|_{\ell^2}^2 \right] = \sum_{j=1}^{M} (q_j^{-1} - 1) \|TDQ_W V^* e_j\|_{\ell^2}^2 \|\langle V g, e_j \rangle\|_{\ell^2}^2 \leq \sum_{k=1}^{r} (q_k^{-1} - 1) \|P_{\alpha} V g\|_{\ell^\infty}^2 \max_{j \in \mathcal{K}} \|TDQ_W V^* e_j\|_{\ell^2}^2.
\]

This is the same upper bound as obtained in (8.5), so from the bound on $\mathcal{V}$, we have that

\[
E \left( \left\| \sum_{j=1}^{M} Y_j \right\|_{\ell^2}^2 \right) \leq \hat{B}^2 \max_{l=1}^{r} \sum_{k=1}^{r} (q_k^{-1} - 1) \mu_{N,M}^2(k,l) \hat{\kappa}_k.
\]

Finally, by Jensen’s inequality,

\[
E \left[ \sum_{j=1}^{M} Y_j \right] \leq \sqrt{E \left( \left\| \sum_{j=1}^{M} Y_j \right\|_{\ell^2}^2 \right)} \leq \sqrt{\hat{B}^2 \max_{l=1}^{r} \sum_{k=1}^{r} (q_k^{-1} - 1) \mu_{N,M}^2(k,l) \hat{\kappa}_k}.
\]

Let $C = \max \{C_1, 4C_2/\alpha\}$, where

\[
C_1 = \|DQ_W D^*\|_{\ell^\infty \rightarrow \ell^\infty} \sqrt{r} \max_{l=1}^{r} \sum_{k=1}^{r} \mu_{N,M}^2(k,l) \kappa_l,
\]

and

\[
C_2 = r \hat{B}^2 \max_{l=1}^{r} \sum_{k=1}^{r} (q_k^{-1} - 1) \mu_{N,M}^2(k,l) \hat{\kappa}_k.
\]

Note that $C \geq \mathcal{S}$, $\alpha C/4 \geq \mathcal{V}$. Suppose that $C_2 \leq \alpha^2/16$, then $E(Z) \leq \alpha/4$, since $(E(Z))^2 \leq C_2$. By applying Talagrand’s inequality with the upper bound of $C \geq \mathcal{S}$ and $E(Z) \leq \alpha/4$, $2$

\[
P \left( Z \geq \frac{\alpha}{2} \right) \leq P \left( Z \geq \frac{\alpha}{4} + E(Z) \right) \geq P \left( |Z - E(Z)| \leq \frac{\alpha}{4} \right) \leq 3 \exp \left( -\frac{\alpha}{4KC} \log \left( 1 + \frac{C\alpha/4}{\sqrt{\mathcal{V} + C\alpha/4}} \right) \right) \leq 3 \exp \left( -\frac{\alpha}{4KC} \log \left( 1 + \frac{C\alpha/4}{C\alpha/4 + C\alpha/4} \right) \right) \leq 3 \exp \left( -\frac{\alpha}{4KC} \log \left( \frac{3}{2} \right) \right),
\]

26
where $K$ is the constant from Talagrand’s inequality. So, $P\left(Z \geq \frac{\alpha}{2}\right) \leq \gamma$ provided that
\[
C \log \left(\frac{3}{\gamma}\right) \leq \frac{\alpha}{4K} \log \left(\frac{3}{2}\right),
\]
as well as $C_2 \leq \alpha^2/16$. Therefore, the require result would follow provided that
\[
\|DQ_WD^*\|_{\ell^\infty,\ell^\infty} \sqrt{r} \log \left(\frac{3}{2}\right) \sum_{l=1}^r \mu_{N,M}(k,l) \kappa_l \leq \frac{\alpha}{4K} \log \left(\frac{3}{2}\right), \quad k = 1, \ldots, r,
\]
and
\[
\tilde{B}^2 \log \left(\frac{3}{2}\right) \sum_{k=1}^r (q_k^{-1} - 1) \mu_{N,M}^2(k,l) \tilde{k}_l \leq \frac{\alpha^2}{16} \min \left\{1, \frac{1}{K} \log \left(\frac{3}{2}\right)\right\}, \quad l = 1, \ldots, r.
\]

**Proposition 8.5.** Fix $g \in \mathcal{H}$ and let $\alpha > 0$ and $\gamma \in [0,1]$. Suppose that
\[
\|DQ_WV^*P_MVQ_WD^*T^{-1}\|_{\ell^2,\ell^\infty} \leq \frac{\alpha}{2}.
\]
Let $B = \|DQ_WV^*\|_{\ell^\infty,\ell^\infty}$. Let
\[
\tilde{M} = \min \left\{ i \in \mathbb{N} : \max_{j \geq 1} 2\sqrt{k_{\max}} \cdot \max_{k=1}^r q_k^{-1} \cdot \|P_MV^*e_j\|_{\ell^2} \leq \alpha \right\}, \quad k_{\max} = r \max \{\kappa_j\}_{j=1}^r.
\]
Then $\tilde{M}$ is finite and
\[
P\left(\|P_M^\perp DQ_WV^*(q_1^{-1}P_{\Omega_1} \oplus \cdots \oplus q_r^{-1}P_{\Omega_r})VQ_Wg\|_{\ell^\infty} \geq \alpha \|TDg\|_{\ell^2}\right) \leq \gamma,
\]
provided that
\[
\sqrt{r} \tilde{B} \log \left(\frac{4\tilde{M}}{\gamma}\right) q_k^{-1} \sum_{l=1}^r \mu_{N,M}^2(k,l) \kappa_l \leq \alpha, \quad k = 1, \ldots, r
\]
and
\[
r \tilde{B}^2 \log \left(\frac{4\tilde{M}}{\gamma}\right) \sum_{k=1}^r (q_k^{-1} - 1) \mu_{N,M}^2(k,j) \tilde{k}_l \leq \alpha^2, \quad j = 1, \ldots, r.
\]

**Proof.** Without loss of generality, assume that $\|TDg\|_{\ell^2} = 1$. Let $\{\delta_j\}_{j=1}^M$ be random Bernoulli variables such that $P(\delta_j = 1) = \tilde{q}_j$ where $\tilde{q}_j = q_k$ for $j = M_{k-1} + 1, \ldots, M_k$. Observe that
\[
Q_{\tilde{V}W}^\perp = Q_W^\perp D^*D = Q_{\tilde{V}W}^\perp P_M^\perp D,
\]
so we have that
\[
P_M^\perp DQ_WV^*(q_1^{-1}P_{\Omega_1} \oplus \cdots \oplus q_r^{-1}P_{\Omega_r})Vg
\]
\[
= \sum_{j=1}^M (\tilde{q}_j^{-1} - 1)P_M^\perp DQ_W^*D^*D^*V^*(e_j \otimes \tilde{\tau}_j)Vg + P_M^\perp Q_W^*V^*P_MVQ_WD^*T^{-1}TDg.
\]
Since $\|P_M^\perp DQ_WV^*P_MVQ_WD^*T^{-1}\|_{\ell^2,\ell^\infty} \leq \alpha/2$, it suffices to show that
\[
P\left(\left\|\sum_{j=1}^M (\tilde{q}_j^{-1} - 1)P_M^\perp DQ_W^*V^*(e_j \otimes \tilde{\tau}_j)Vg\right\|_{\ell^\infty} > \alpha/2\right) \leq \gamma.
\]
For each $i \in \Delta$ and $j = 1, \ldots, M$, let
\[
Z_j^i = (P_M^\perp DQ_W^*V^*(e_j \otimes \tilde{\tau}_j)Vg, e_i).
\]
For each \( i \in \Delta^c \), we will first apply Bernstein’s inequality (Theorem 5.2) to consider upper bounds for 
\( \mathbb{P} \left( |\sum_{j=1}^{M} Z_j^i| \geq \alpha \right) \). Observe that

\[
\mathbb{E} \left( |Z_j^i|^2 \right) = (q_j^{-1} - 1) |(P_{\Lambda \Delta}^\perp DQ_{\alpha}^\perp V^*(e_j \otimes \mathbf{1}) V g, e_i)|^2 \\
= (q_j^{-1} - 1) |(V^* e_j, Q_{\alpha}^\perp P_{\Lambda \Delta}^\perp D e_i)|^2 |(e_j, V g)|^2.
\]

Let \( B = \|DQ_{\alpha}^\perp D^*\|_{\ell^\infty \rightarrow \ell^\infty} \). Then,

\[
\sum_{j=1}^{M} \mathbb{E}(|Z_j^i|^2) \leq \sup_{\|\alpha\|_1 \leq B} \sum_{k=1}^{r} (q_k^{-1} - 1) \|P_{\Lambda \Delta} V D^* \alpha\|_{\ell^\infty \rightarrow \ell^\infty}^2 \|P_{\Lambda \Delta} V g\|_{\ell^2}^2 \\
\leq B \sup_{\|\alpha\|_1 \leq B} \sum_{k=1}^{r} (q_k^{-1} - 1) \mu(P_{\Lambda \Delta} V D^*) \|P_{\Lambda \Delta} V D^* \alpha\|_{\ell^\infty \rightarrow \ell^\infty} \|P_{\Lambda \Delta} V g\|_{\ell^2}^2 \\
\leq B \sup_{\|\alpha\|_1 \leq B} \sum_{l \in \mathbb{N}} |\alpha_l| \sum_{k=1}^{r} (q_k^{-1} - 1) \mu(P_{\Lambda \Delta} V D^*) \|P_{\Lambda \Delta} V D^* e_l\|_{\ell^\infty \rightarrow \ell^\infty} \|P_{\Lambda \Delta} V g\|_{\ell^2}^2 \\
= B \sup_{\|\alpha\|_1 \leq B} \sum_{l \in \mathbb{N}} |\alpha_l| \sum_{k=1}^{r} (q_k^{-1} - 1) \mu(P_{\Lambda \Delta} V D^*) \|P_{\Lambda \Delta} V D^* e_l\|_{\ell^\infty \rightarrow \ell^\infty} \|P_{\Lambda \Delta} V g\|_{\ell^2}^2 \\
\leq B^2 \sup_{\|\alpha\|_1 \leq B} \sum_{r \in \mathbb{N}} (q_k^{-1} - 1) \mu_{N,M}(k, j) \|P_{\Lambda \Delta} V g\|_{\ell^2}^2 \\
\leq r B^2 \sum_{r \in \mathbb{N}} (q_k^{-1} - 1) \mu_{N,M}(k, j) \hat{k}_k =: \sigma^2,
\]

where the last line follows because \( \|TDg\|_{\ell^2} = 1 \) implies that \( \|P_{\Lambda \Delta} Dg\|_{\ell^2} \leq \sqrt{K \kappa} \), and by definition of \( \hat{k}_k, \|P_{\Lambda \Delta} V g\|_{\ell^2} \leq r \hat{k}_j \).

Also, we have that

\[
|Z_j^i| \leq q_j^{-1} |(V^* e_j, Q_{\alpha}^\perp V D^* D g)| \\
\leq \max_{k=1}^{r} q_k^{-1} \mu(P_{\Lambda \Delta} V Q_{\alpha}^\perp V D^* P_{\Lambda \Delta}^\perp) \|P_{\Lambda \Delta} D g\|_{\ell^2} \\
\leq \|DQ_{\alpha}^\perp D^*\|_{\ell^\infty \rightarrow \ell^\infty} \|\mu_{N,M}(k, l)\| \|P_{\Lambda \Delta} D g\|_{\ell^2} \\
\leq \sqrt{r} \|DQ_{\alpha}^\perp D^*\|_{\ell^\infty \rightarrow \ell^\infty} \|\mu_{N,M}(k, l)\| \kappa_l =: K,
\]

where the last line follows because \( \|P_{\Lambda \Delta} Dg\|_{\ell^2} \leq \sqrt{K \kappa} \) along with (4) of Corollary 5.4 implies that \( \|P_{\Lambda \Delta} Dg\|_{\ell^2} \leq \sqrt{r \kappa} \). Let \( T \subset \mathbb{N} \) be such that

\[
\mathbb{P} \left( \sup_{i \in T} \sum_{j=1}^{M} Z_j^i \geq \alpha \right) = 0
\]

and suppose that \( |T^c| \leq \bar{M} \). Then, by Theorem 8.2 and the union bound,

\[
\mathbb{P} \left( \sup_{i \in \Delta} \sum_{j=1}^{M} Z_j^i \geq \frac{\alpha}{2} \right) \leq \mathbb{P} \left( \sup_{i \in T^c} \sum_{j=1}^{M} Z_j^i \geq \frac{\alpha}{2} \right) \leq 4 |T^c| \exp \left( -\frac{\alpha^2 / 64}{\sigma^2 + K \alpha / (6 \sqrt{2})} \right).
\]
which is true provided that
\[
\log \left( \frac{4\tilde{M}}{\gamma} \right) \sigma^2 \leq \frac{\alpha^2}{128}, \quad \log \left( \frac{4\tilde{M}}{\gamma} \right) K \leq \frac{\alpha}{8},
\]
which are simply the assumptions of this theorem.

It remains to show that such as set \( \Upsilon^c \) exists: First note that \( \|Dg\| \leq \|T^{-1}\|\|TDg\| \leq \sqrt{\kappa_{\text{max}}} \) with \( \kappa_{\text{max}} = r \max \{\kappa_j\}_{j=1}^r \). So, using the fact that \( D \) and \( V \) are isometries and hence of norm 1, notice that
\[
\sum_{j=1}^M \left| q_k^{-1} - 1 \right| P_{\tilde{\kappa}}^j DQ_W V^*(q_1^{-1} P_{\Omega_1} \otimes \cdots \otimes q_r^{-1} P_{\Omega_r}) V g, e_i \right| \\
\leq \|Dg\| \|DQ_W V^*(q_1^{-1} \Omega_1 \otimes \cdots \otimes q_r^{-1} \Omega_r) V Q_W^* D^* P_{\tilde{\kappa}}^j e_i \| \leq \sqrt{\kappa_{\text{max}}} \max_{k=1}^r q_k^{-1} \|P_{\tilde{M}}|VQ_W^* D^* e_i \| \to 0
\]
as \( i \to \infty \) since \( P_{\tilde{M}}|VQ_W^* D^* \) is of finite rank. Thus, for \( \alpha > 0 \), it suffices to let
\[
\Upsilon^c := \left\{ i \in \mathbb{N} : \sqrt{\kappa_{\text{max}}} \max_{k=1}^r q_k^{-1} \|P_{\tilde{M}}|VQ_W^* D^* e_i \| \geq \alpha \right\}
\]
which is a finite set. To conclude this proof, observe that \( |\Upsilon^c| \leq \tilde{M} < \infty \). \( \square \)

**Proposition 8.6.** Let \( \alpha > 0 \) and \( \gamma \in [0, 1] \). Suppose that
\[
\left\| Q_W V^* P_{\tilde{M}}^j V Q_W \right\|_{H \mapsto H} \leq \alpha/2.
\]  
Then,
\[
\mathbb{P} \left( \|Q_W V^*(q_1^{-1} P_{\Omega_1} \otimes \cdots \otimes q_r^{-1} P_{\Omega_r}) V Q_W - Q_W V^* V Q_W \|_{H \mapsto H} \geq \alpha \right) \leq \gamma,
\]
provided that
\[
q_k \geq \alpha^{-2} \log \left( \frac{4\tilde{M}}{\gamma} \right) \sum_{i=1}^r \beta_{N,M}(k,l) \kappa_l, \quad k = 1, \ldots, r.
\]

**Proof.** Let \( \tilde{q}_j = q_j \) for \( j \in \Gamma_k \). Let \( \{\delta_j\}_{j=1}^r \) be Bernoulli random variables such that \( \mathbb{P}(\delta_j = 1) = \tilde{q}_j \).

\[
\left\| Q_W V^*(q_1^{-1} P_{\Omega_1} \otimes \cdots \otimes q_r^{-1} P_{\Omega_r}) V Q_W - Q_W V^* V Q_W \right\| \leq \left\| \sum_{j=1}^M \tilde{q}_j^{-1} \delta_j Q_W V^* (e_j \otimes \bar{e}_j) V Q_W \right\|
\]
\[
\leq \left\| \sum_{j=1}^M (\tilde{q}_j^{-1} \delta_j - 1) DQ_W V^* (e_j \otimes \bar{e}_j) V Q_W D^* \right\| + \left\| Q_W V^* P_{\tilde{M}}^j V Q_W \right\|
\]
\[
\leq \left\| \sum_{j=1}^M (\tilde{q}_j^{-1} \delta_j - 1) DQ_W V^* (e_j \otimes \bar{e}_j) V Q_W D^* \right\| + \alpha/2.
\]

Let
\[
U = \sum_{j=1}^M (\tilde{q}_j^{-1} \delta_j - 1)(e_j \otimes \bar{e}_j).
\]

Then, for each \( J \in \mathbb{N}, \left\| DQ_W V^* UVQ_W D \right\| \) is bounded above by
\[
\left\| P_{[j]} DQ_W V^* UVQ_W D^* P_{[j]} \right\| + \left\| DQ_W V^* UVQ_W D^* P_{[j]} \right\| + \left\| P_{[j]} DQ_W V^* UVQ_W D^* \right\|
\]
\[
\leq \left\| P_{[j]} DQ_W V^* UVQ_W D^* P_{[j]} \right\| + 2 \left\| P_{[j]} DQ_W \right\| \|U\|
\]
\[
\leq \left\| P_{[j]} DQ_W V^* UVQ_W D^* P_{[j]} \right\| + 2 q^{-1} \left\| P_{[j]} DQ_W \right\|
\]

29
where \( q = \min \{ q_j \} \) for each \( j \in \mathbb{N} \). Note that since \( DQ_W \) has finite rank, \( \left\| P_{\mathcal{H}} DQ_W \right\| \to 0 \) as \( J \to \infty \). Let
\[
\tilde{M} = \min \left\{ j \in \mathbb{N} : 8 \left\| P_{\mathcal{H}} DQ_W \right\| \leq q_0 \right\}.
\]
Then, it suffices to show that
\[
\mathbb{P} \left( \left\| P_{\mathcal{H}} DQ_W V^*U VQ_W D^* P_{\mathcal{H}} \right\| > \alpha / 4 \right) \leq \gamma.
\]
For each \( j = 1, \ldots, M \), define \( Z_j = (\tilde{q}_j^{-1} \delta_j - 1) P_{\mathcal{H}} DQ_W V^* (e_j \otimes \tilde{\pi}_j) V Q_W D^* P_{\mathcal{H}} \). We will aim to apply Theorem 8.3 to derive the following.
\[
P \left( \left\| \sum_{j=1}^M Z_j \right\| \geq \frac{\alpha}{4} \right) \leq \gamma.
\]
Notice that \( \{ Z_j \}_{j=1}^M \) are independent mean-zero matrices. Let \( \xi_j = P_{\mathcal{H}} DQ_W V^* e_j \). Then,
\[
\left\| Z_j \right\| \leq \tilde{q}_j^{-1} \left\| \xi_j \otimes \tilde{\xi}_j \right\| \leq \tilde{q}_j^{-1} \left\| \xi_j \right\|^2.
\]
To bound this, note that for each \( j = 1, \ldots, M \),
\[
\left\| \xi_j \right\|^2 = \sup_{\alpha \in \mathcal{C}^M} \left| \langle DQ_W V^* e_j, \alpha \rangle \right|^2 = \sup_{\alpha \in \mathcal{C}^M, \| \alpha \|_2 = 1} \left| \langle DQ_W D^* V Q_W D^* e_j, \alpha \rangle \right|^2
\]
\[
\leq \sup_{\alpha \in \mathcal{C}^M, \| \alpha \|_2 = 1} \sum_{l=1}^r \left| \langle P_{\mathcal{M}} DQ_W D^* e_j, \alpha \rangle \right|^2
\]
\[
\leq \sup_{g \in \mathcal{W}, \| g \|_1 = 1} \sum_{l=1}^r \left\| P_{\mathcal{M}} DQ_W D^* e_j \right\| \left\| P_{\mathcal{M}} Dg \right\|_1 = \sup_{g \in \mathcal{W}, \| g \|_1 = 1} \sum_{l=1}^r \mu_{\mathcal{N}, \mathcal{M}}^2 (j, l) \left\| P_{\mathcal{M}} Dg \right\|_1^2.
\]
So, since \( \| Dg \|_1 = 1 \) implies that \( \| P_{\mathcal{M}} Dg \|_1 \leq \kappa_l (\mathbf{N}, s) \) for \( l = 1, \ldots, r \) by (4) of Corollary 5.4 we have that
\[
\left\| Z_j \right\| \leq \max_{j=1}^r \tilde{q}_j^{-1} \sup_{g \in \mathcal{W}} \sum_{l=1}^r \mu_{\mathcal{N}, \mathcal{M}}^2 (j, l) \kappa_l (\mathbf{N}, s) =: K.
\]
Also,
\[
\left\| \sum_{j=1}^M \mathbb{E}(Z_j^* Z_j) \right\| \leq \sup_{\| x \|_2 = 1} \left\| \sum_{j=1}^M (\tilde{q}_j^{-1} - 1) \langle \xi_j \otimes \tilde{\xi}_j, x \rangle \right\|
\]
\[
\leq \sup_{\| x \|_2 = 1} \left( \max_{k=1}^r \left\{ \left( \tilde{q}_k^{-1} - 1 \right)^2 \right\} \sum_{j=1}^M \left\| e_j, V D^* P_{\mathcal{H}} x \right\| \right)
\]
\[
= \max_{k=1}^r \left\{ \left( \tilde{q}_k^{-1} - 1 \right)^2 \right\} \leq K.
\]
Thus, by Theorem 8.3
\[
\mathbb{P} \left( \left\| Q_W V^* (q_1^{-1} P_{\mathcal{H}} \oplus \cdots \oplus q_r^{-1} P_{\mathcal{H}}) V Q_W - Q_W V^* V Q_W \right\|_{\mathcal{H} \to \mathcal{H}} \geq \alpha \right) \leq 4 \tilde{M} \exp \left( -\frac{\alpha^2 / 8}{K + K \alpha / 6} \right) \leq \gamma
\]
provided that
\[
\log \left( \frac{4M}{\gamma} \right) K \leq \frac{\alpha^2}{16}, \quad \log \left( \frac{4M}{\gamma} \right) K \leq \alpha/2,
\] (8.9)
which are implied by the given assumptions.

\[\square\]

**Proposition 8.7.** Let \( \alpha > 0 \) and let \( \gamma \in [0, 1] \). Let
\[
\tilde{M} = \min \left\{ i \in \mathbb{N} : \sup_{j \geq i} \left| P[|M|] V D^* e_j \right|_{\ell^2} + \left| Q_{R(D^* P_{\gamma})} D^* e_j \right|_{\ell^2} \leq \sqrt{\frac{5q}{4}} \right\}.
\]

Then \( \tilde{M} \) is finite and
\[
P \left( \sup_{j \in \mathbb{N}} \left| P[j] D Q_{W} V^*(q_1^{-1} P_{\gamma_1} \oplus \ldots \oplus q_r^{-1} P_{\gamma_r}) V Q_{W} D^* P[j] \right| \leq \frac{5}{4} \right) \leq \gamma
\]
provided that for each \( k = 1, \ldots, r \) and each \( j \in \mathbb{N} \),
\[
1 \geq B^2 \max_{j=1} \left| q_k^{-1} \right| \mu_{N,N}^2 \log \left( \frac{2M}{\gamma} \right),
\]
where \( B = \|D Q_{W} D^* \|_{\ell^\infty \rightarrow \ell^\infty} \).

**Proof.** Let \( \{\delta_j\}_{j=1}^{M} \) be Bernoulli random variables such that \( P(\delta_j = 1) = q_j \) where \( q_j = q_k \) for \( j = \Gamma_k \). Observe that for each \( j \in \mathbb{N} \),
\[
\left| P[j] D Q_{W} V^*(q_1^{-1} \Omega_1 \oplus \ldots \oplus q_r^{-1} \Omega_r) V Q_{W} D^* P[j] \right| \leq \sum_{k=1}^{M} (q_k^{-1} \delta_k - 1) \left| (V Q_{W} D^* e_j, e_k) \right|^2 + 1,
\]
where we have applied \( \|V\| = \|D\| = 1 \) in the last line. For each \( j \in \mathbb{N} \) and \( k = 1, \ldots, M \), define
\[
Z_k^j = (q_k^{-1} \delta_k - 1) \left| (V Q_{W} D^* e_j, e_k) \right|^2. \]
To prove this proposition, we need to derive conditions under which
\[
P \left( \sup_{j \in \mathbb{N}} \left| \sum_{k=1}^{M} Z_k^j \right| \geq \frac{1}{4} \right). \quad (8.10)
\]

We first seek to apply Theorem 8.2 to analyse \( P(\sum_{k=1}^{M} Z_k^j > \frac{1}{4}) \) for each \( j \in \mathbb{N} \). Observe that
\[
\left| Z_k^j \right| \leq \sup_{j \in \mathbb{N}} \left| q_k^{-1} \delta_k - 1 \right| \mu^2 (P_j V Q_{W} D^* P[j]) =: K,
\]
and
\[
\sum_{k=1}^{M} E(\left| Z_k^j \right|^2) = \sum_{k=1}^{M} (q_k^{-1} - 1) \left| (V Q_{W} D^* e_j, e_k) \right|^4
\]
\[
\leq \max_{t=1}^{r} (q_t^{-1} - 1) \mu^2 (P_t V Q_{W} D^* P[j]) \left| V Q_{W} D^* e_j \right|^2
\]
\[
\leq \sup_{j \in \mathbb{N}} \max_{t=1}^{r} (q_t^{-1} - 1) \mu (P_t V Q_{W} D^* P[j])^2 =: \sigma^2.
\]

Thus, by applying Theorem 8.2,
\[
P \left( \left| \sum_{k=1}^{r} Z_k^j \right|^2 \geq \frac{1}{4} \right) \leq 2 \exp \left( \frac{-1/32}{\sigma^2 + K/12} \right).
\]
Thus, it suffices to let \( k \) for each \( k \in T \). This section will show that, with high probability, one can construct \( \rho \in \mathcal{R}(V^*P)(\Omega) \) which is a finite set. Observe also that \( \| P_{[M]}V^*D^*e_j \|_{\ell^2} \leq \frac{1}{q} \left( \| P_{[M]}V^*D^*e_j \|_{\ell^2} + \| Q_{P_{[M]}D^*}D^*e_j \|_{\ell^2} \right)^2 \to 0 \) as \( j \to \infty \). Therefore, it suffices to let

\[
\Upsilon^c = \left\{ j \in \mathbb{N} : \| P_{[M]}V^*D^*e_j \|_{\ell^2} + \| Q_{P_{[M]}D^*}D^*e_j \|_{\ell^2} > \sqrt{\frac{5q}{4}} \right\}
\]

which is a finite set. Observe also that \( |\Upsilon^c| \leq \tilde{M} < \infty \). Therefore, by applying the union bound,

\[
\mathbb{P} \left( \sup_{j \in \mathbb{N}} \| P_{[M]}Q_{V^*}(\Omega) \|_{\ell^2} \leq \frac{3}{2} \right) \leq 2\tilde{M} \exp \left( -\frac{1}{3\tilde{M}} \sigma^2 + \frac{K}{12} \right).
\]

Thus, it suffices to let, for each \( k = 1, \ldots, r \) and each \( j \in \mathbb{N} \),

\[
1 \gtrsim (q_k^{-1} - 1) \mu(\alpha_k V_{k}^*D^*e_j) \leq \log \left( \frac{2M}{\gamma} \right).
\]

Let \( B = \| DQ_{V^*}D^*e_j \|_{\ell^2} \). Finally, the following observation concludes the proof of this proposition.

\[
(q_k^{-1} - 1) \mu(\alpha_k V_{k}^*D^*e_j) \leq B \sup_{\| \alpha \|_{\ell^2} \leq B} (q_k^{-1} - 1) \mu(\alpha_k V_{k}^*D^*) \leq B \sup_{\| \alpha \|_{\ell^2} \leq B} \sum_{j \in \mathbb{N}} |\alpha_j| (q_k^{-1} - 1) \mu(\alpha_k V_{k}^*D^*) \leq B^2 \max_{j=1}^{\max(q_k^{-1} - 1) \mu_{\alpha_{k}(k,j)}(q_k)} \left( \frac{2M}{\gamma} \right).
\]

\[
\square
\]

### 9 Constructing the dual certificate

This section will show that, with high probability, one can construct \( \rho \in \mathcal{R}(V^*P_{[M]}) \) which satisfies conditions (iii) to (v) of Proposition 5.1 if \( \Omega = \Omega_{\alpha_{[M]}} \) is a Bernoulli multi-level sampling scheme satisfying Assumption 7.2.

As explained in [2], the sampling model of \( \Omega = \Omega_1 \cup \cdots \cup \Omega_r \) with \( \Omega_i \sim \text{Ber}(q_i, \Gamma_i) \) is equivalent to the following sampling model.

\[
\Omega_k = \Omega_k^1 \cup \cdots \cup \Omega_k^r, \quad \Omega_k^j \sim \text{Ber}(q_k^j, \Gamma_k), \quad k = 1, \ldots, r, \quad j = 1, \ldots, \mu,
\]

for \( \mu \in \mathbb{N} \) and \( \{ q_k^j \}_{j=1}^{\mu} \) such that

\[
(1 - q_k^1)(1 - q_k^2) \cdots (1 - q_k^\mu) = 1 - q_k.
\]

We will consider this alternative sampling model throughout this section so that we can apply the golfing scheme of [22] to construct the dual certificate described in Proposition 6.3. This section consists of the following steps:
1. Define the dual certificate.

2. Show that the constructed dual certificate satisfies conditions (iii) to (v) of Proposition 6.1 provided that certain events occur.

3. Show that the events described in step 2 occur with high probability.

**Definition of the dual certificate.** Let $\gamma = \epsilon / 6$. Let $\mathcal{L} = \log(4q^{-1} \sqrt{\mu} \|DD\|_{\ell^\infty})$. Define $\mu, \nu \in \mathbb{N}, \{\alpha_j\}_{j=1}^\mu$ and $\{\beta_j\}_{j=1}^\nu$ as follows.

$$
\mu = 8[3\nu + \log(\gamma^{-1/2})], \quad \nu = \log \left( 8q^{-1} \sqrt{\mu} \|DD\|_{\ell^\infty} \right),
$$

$$
q^1_k = q_k^1 = \frac{1}{4} q_k, \quad q^k_k = q^2_k = \cdots = q^\mu_k, \quad k = 1, \ldots, r,
$$

$$
\alpha_1 = \alpha_2 = \frac{1}{2}, \quad \alpha_i = \frac{1}{2}, \quad i = 3, \ldots, \mu,
$$

$$
\beta_1 = \beta_2 = \frac{1}{4}, \quad \beta_i = \frac{1}{4}, \quad i = 3, \ldots, \mu.
$$

For $j = 1, \ldots, \mu$, define $U_j : \mathcal{B}(\ell^2(\mathbb{N}), \ell^2(\mathbb{N}))$ by

$$
U_j = \frac{1}{q_j^1} P_{\Omega_1} \oplus \cdots \oplus \frac{1}{q_r^j} P_{\Omega_r^j}.
$$

Let $Z_0 = D^* \text{sgn}(P_\Delta Df)$ and for $i = 1, 2$, define

$$
Z_i = Z_0 - Q_W Y_i, \quad Y_i = \sum_{j=1}^i V^* U_j V Z_{j-1}.
$$

Let $\Theta_1 = \{1\}$, $\Theta_2 = \{1, 2\}$ and for $i \geq 3$, define

$$
\Theta_i = \begin{cases} 
\Theta_{i-1} \cup \{i\} & \text{||TD}(Z_{i-1} - V^* U_i V Z_{i-1})||_{\ell^2} \leq \alpha_i \|TDZ_{i-1}\|_{\ell^2} \\
\Theta_{i-1} & \text{otherwise.}
\end{cases}
$$

$$
Y_i = \begin{cases} 
\sum_{j \in \Theta_i} V^* U_j V Z_{j-1} & i \in \Theta_i \\
Y_{i-1} & \text{otherwise.}
\end{cases}
$$

$$
Z_i = \begin{cases} 
Z_0 - Q_W Y_i & i \in \Theta_i \\
Z_{i-1} & \text{otherwise.}
\end{cases}
$$

Note that $Z_i \in \mathcal{W}$ for each $i = 1, \ldots, \mu$. Define the following events.

$$
A_i : \quad ||TD(Z_{i-1} - V^* U_i V Z_{i-1})||_{\ell^2} \leq \alpha_i \|TDZ_{i-1}\|_{\ell^2}, \quad i = 1, 2,
$$

$$
B_i : \quad ||P_\Delta^i DQ_W^i V^* U_i V Z_{i-1}\|_{\ell^\infty} \leq \beta_i \|TDZ_{i-1}\|_{\ell^2}, \quad i = 1, 2,
$$

$$
B_3 : \quad |\Theta_\mu| \geq \nu,
$$

$$
B_4 : \quad \cap_{i=1}^2 A_i \cap \cap_{i=1}^3 B_i.
$$

Let $\tau(j)$ denote the $j^{th}$ element of $\Theta_\mu$ (in order of appearance).

**Properties of the dual certificate.** Suppose that $B_4$ occurs, and let $\rho = Y_{\tau(\nu)}$. By definition, $\rho = V^* P_{\Omega w}$ for some $w \in \ell^2(\mathbb{N})$. We now show that $\rho$ satisfies (iii) and (iv) of Proposition 6.1 and derive an upper bound on $||w||_{\ell^2}$. By definition,

$$
Z_{\tau(i)} = Z_0 - Q_W \sum_{j \in \Theta(i)} V^* U_j V Z_{j-1} = (Q_W - Q_W V^* U_{\tau(i)} V) Z_{\tau(i-1)}.
$$

(9.2)
1. Since $D^*D = I$, by construction of $\rho$,

$$
\|Z_0 - QW\rho\| = \|Z_0 - QWY_{\tau(\nu)}\| = \|Z_{\tau(\nu)}\|
= \|D^*D(QW - QWV^*U_{\tau(\nu)}VQW)Z_{\tau(\nu-1)}\|
= \|T^{-1}TD(QW - QWV^*U_{\tau(\nu)}VQW)Z_{\tau(\nu-1)}\|.
$$

Recalling the definition of $(\alpha_i)_{i=1}^\nu$ and recalling that $\kappa_{\text{max}} = r \max \{\kappa_j\}$ and $\|T^{-1}\| \leq \sqrt{\kappa_{\text{max}}}$, it follows that

$$
\|Z_0 - QW\rho\| \leq \sqrt{\kappa_{\text{max}}} \|TD(Z_{\tau(\nu-1)} - V^*U_{\tau(\nu)}VQWZ_{\tau(\nu-1)})\|_{L^2}
\leq \sqrt{\kappa_{\text{max}}} \|TDZ_{\tau(\nu-1)}\|_{L^2}
\leq \sqrt{\kappa_{\text{max}}} \|DD^*\|_{\ell^\infty \rightarrow \ell^\infty} \prod_{j=1}^\nu \alpha_{\tau(j)}
\leq \sqrt{\kappa_{\text{max}}} \|DD^*\|_{\ell^\infty \rightarrow \ell^\infty} 2^{-\nu} \leq \frac{\sqrt{\nu}}{8},
$$

by our choice of $\nu$. Note that to get from the third line to the forth line, we observe that $\|DD^*\|_{\ell^\infty \rightarrow \ell^\infty}$, then recall from (3) of Corollary 5.4 that $\|P_{\lambda_j}Dg\|_{L^2} \leq \kappa_j$ for all $g = D^*P_{\lambda}x$ with $\|Dg\|_{L^\infty} \leq 1$. Therefore,

$$
\|TDZ_0\|_{L^2}^2 = \sum_{j=1}^\nu \|P_{\lambda_j}DZ_0\|_{L^2} \leq \|DD^*\|_{\ell^\infty \rightarrow \ell^\infty}.
$$

2. Recalling our definition of $(\beta_i)_{i=1}^\nu$ and the estimate used in the previous step to bound $\|TDZ_{\tau(j)}\|_{L^2}$,

$$
\|P_{\lambda}^*DQ_W\rho\|_{L^\infty} \leq \sum_{j=1}^\nu \|P_{\lambda}^*DQ_WV^*U_{\tau(j)}VQWZ_{\tau(j-1)}\|_{L^\infty}
\leq \sum_{j=1}^\nu \beta_{\tau(j)} \|TDZ_{\tau(j-1)}\|_{L^2}
\leq \|DD^*\|_{\ell^\infty \rightarrow \ell^\infty} \sum_{j=1}^\nu \beta_{\tau(j)} \prod_{i=1}^{j-1} \alpha_{\tau(i)}
\leq \frac{1}{4} \left( 1 + \frac{1}{2\sqrt{\nu}} + \frac{L}{2\nu} + \cdots + \frac{L}{L^2\nu-1} \right) \leq \frac{1}{2}.
$$

3. By definition, $\rho = V^*P_{\Omega^k}w$ with $w = \sum_{j=1}^\nu w_j$ and $w_j = U_{\tau(j)}VZ_{\tau(j-1)}$. For each $j = 1, \ldots, \nu$,

$$
\|w_j\|^2 = \langle U_{\tau(j)}VZ_{\tau(j-1)}, U_{\tau(j)}VZ_{\tau(j-1)} \rangle = \sum_{k=1}^r \frac{1}{(q_k^{\tau(j)})^2} \|P_{\Omega_k^{\tau(j)}}VZ_{\tau(j-1)}\|^2
\leq K_{\tau(j)} \sum_{k=1}^r \frac{1}{q_k^{\tau(j)}} \left( V^*P_{\Omega_k^{\tau(j)}}VZ_{\tau(j-1)}, Z_{\tau(j-1)} \right) = K_{\tau(j)} \left( V^*U_{\tau(j)}VZ_{\tau(j-1)}, Z_{\tau(j-1)} \right),
$$

where $K_{\tau(j)} = \max \left\{ 1/q_k^{\tau(j)} : k = 1, \ldots, r \right\}$. Using (9.2), we have that

$$
\langle V^*U_{\tau(j)}VZ_{\tau(j-1)}, Z_{\tau(j-1)} \rangle
= \langle V^*U_{\tau(j)}VZ_{\tau(j-1)} - Z_{\tau(j-1)}, Z_{\tau(j-1)} \rangle + \langle Z_{\tau(j-1)}, Z_{\tau(j-1)} \rangle
= \langle Z_{\tau(j)}, Z_{\tau(j-1)} \rangle + \langle Z_{\tau(j-1)}, Z_{\tau(j-1)} \rangle.
$$
Since \( I = D^*D \) and \( \|T^{-1}\| \leq \kappa_{\max} \), we have that
\[
\| (V^*U_{(j)} V Z_{(j-1)}, Z_{(j-1)}) \| = \| (DZ_{(j)}, DZ_{(j-1)} + DZ_{(j-1)}) \|
\leq \| DZ_{(j-1)} \| (\| DZ_{(j)} \| + \| DZ_{(j-1)} \|) 
\leq \kappa_{\max} \| T DZ_{(j-1)} \| \| F \| + \| T DZ_{(j-1)} \| \| F \| .
\]
Using \( \| T DZ_{(j)} \| \leq \alpha_{\tau(j)} \| T DZ_{(j-1)} \| \) we obtain,
\[
\| (V^*U_{(j)} V Z_{(j-1)}, Z_{(j-1)}) \| \leq \kappa_{\max} (\alpha_{\tau(j)} + 1) \left( \prod_{i=1}^{j-1} \alpha_{\tau(i)} \right)^2 .
\]
Therefore,
\[
\| w_j \| \leq \sqrt{K_{\tau(j)} \kappa_{\max} (\alpha_{\tau(j)} + 1) \left( \prod_{i=1}^{j-1} \alpha_{\tau(i)} \right)} .
\]
Recall that for \( k = 1, \ldots, r, q_k^j = q_k / 4 \) for \( j = 1, 2 \) and \( q_k^j = \tilde{q}_k \) for all \( j \geq 3 \). Let \( K = \max_{k=1}^{r} \{ q_k^{-1} \} \) and for \( j \geq 3 \), first note that \( (1 - q_1^1) \cdots (1 - q_k^j) = 1 - q_k \) implies that \( q_k^1 + \cdots + q_k^j \geq q_k \). So, we have that \( 2(\mu - 2)\tilde{q}_k \geq q_k \). By our choice of \( \mu \), this implies that
\[
2 \left( 8 \left[ 3 \log \left( 8q^{-1} \sqrt{\kappa_{\max} M} \|DD^*\|_{\ell_\infty} \right) + \log(\gamma^{-1}) \right] - 2 \right) \tilde{q}_k \geq q_k ,
\]
and for \( j \geq 3 \),
\[
K_{\tau(j)} \leq 2 \left( 8 \left[ 3 \log \left( 8q^{-1} \sqrt{\kappa_{\max} M} \|DD^*\|_{\ell_\infty} \right) + \log(\gamma^{-1}) \right] - 2 \right) K .
\]
So, \( \| w_j \| \) are bounded as follows.
\[
\| w_1 \| \leq 2 \sqrt{K_{\tau(1)} \kappa_{\max}} \sqrt{1 + \frac{1}{2L^{1/2}}} , \quad \| w_2 \| \leq 2 \sqrt{K_{\tau(2)} \kappa_{\max}} \sqrt{1 + \frac{1}{2L^{1/2}}} \sqrt{\frac{1}{2L^{1/2}}} ,
\]
and for \( j \geq 3 \),
\[
\| w_j \| \leq 2 \sqrt{3} \sqrt{K_{\tau(j)} \kappa_{\max}} \sqrt{\frac{1}{2^{j-1} L^{1/2}}} \left( 8 \left[ 3 \log \left( 8q^{-1} \sqrt{\kappa_{\max} M} \|DD^*\|_{\ell_\infty} \right) + \log(\gamma^{-1}) \right] - 2 \right) .
\]
Summing these terms yields
\[
\| w \| \leq \sqrt{K_{\tau(j)} \kappa_{\max}} \left( \frac{\log \left( 8q^{-1} \sqrt{\kappa_{\max} M} \|DD^*\|_{\ell_\infty} \right) + \log(6/\gamma)}{\log(4q^{-1} \sqrt{\kappa_{\max} M} \|DD^*\|_{\ell_\infty})} \right) .
\]
To show that conditions (iii) to (v) of Proposition 6.1 are satisfied with probability exceeding \( 1 - 5\gamma = 1 - 5\epsilon/6 \) under Assumption 7.2 we will show that \( P(A_i^\tau) < \gamma \) for \( i = 1, 2 \) and \( P(B_j^\tau) < \gamma \) for \( j = 1, 2, 3 \).

**Proof of** \( P(B_j^\tau) < \gamma \). Define the random variables \( X_1, \ldots, X_{\mu - 2} \) by
\[
X_j = \begin{cases} 0 & \Theta_{j+2} \neq \Theta_{j+1}, \\ 1 & \text{otherwise}. \end{cases}
\]
Observe that
\[
P(B_j^\tau) = P(\| \Theta_{\mu} \| < \nu) = P(\| X_1 + \cdots + X_{\mu - 2} - \mu - \nu \|).
\]
Although \( \{ X_j \}_{j=1}^{\mu - 2} \) are not independent random variables, from [1] Eqn. (7.80) - (7.85) the above probability can be controlled by independent binary random variables and the standard Chernoff bound, so that \( P(B_j^\tau) \leq \gamma \) provided that
\[
\frac{1}{4} \geq P(X_j = 1| X_i = \cdots = X_k = 1) \quad (9.3)
\]
for all $j = 1, \ldots, \mu - 2$ and $l_1, \ldots, l_k \in \{1, \ldots, \mu - 2\}$ such that $j \notin \{l_1, \ldots, l_k\}$ and $\mu \geq 8[3^\nu + \log(\gamma^{-1/2})]$. It remains to verify that (9.3) holds with $p = 1/4$. Observe that $X_j = 0$ whenever

$$
\|TD(QW - QWV^*U_jVQW)Z_{i-1}\|_{\ell^2} \leq \frac{1}{2}\|TDZ_{i-1}\|_{\ell^2}
$$

and

$$
\|P_k^\perp DQ^\perp_W V^* U_j VQW Z_{i-1}\|_{\ell^\infty} \leq \frac{\ell_2}{4}\|TDZ_{i-1}\|_{\ell^2}
$$

for $i = j + 2$. Thus, (9.3) holds with $p = 1/4$ if

$$
P\left(\|TD(QW - QWV^*U_jVQW)Z_{i-1}\|_{\ell^2} > \frac{1}{2}\|TDZ_{i-1}\|_{\ell^2}\right) \leq \frac{1}{8},
$$

(9.4)

and

$$
P\left(\|P_k^\perp DQ^\perp_W V^* U_j VQW Z_{i-1}\|_{\ell^\infty} > \frac{\ell_2}{4}\|TDZ_{i-1}\|_{\ell^2}\right) \leq \frac{1}{8}.
$$

(9.5)

By Proposition 8.4, (9.4) is implied by (C1) below, and by Proposition 8.5, (9.5) is implied by (C2) below.

(C1) Let

$$
\|TD(QWV^*P_{[M]}VQW - QW)D^*T^{-1}\|_{\ell^2} \leq \frac{1}{4},
$$

(9.6)

$$
\tilde{q}_k \gtrsim \sqrt{B} \sum_{l=1}^{r} \mu_{N,M}(k,l) \kappa_l, \quad k = 1, \ldots, r,
$$

(9.7)

and

$$
1 \gtrsim rB^2 \sum_{k=1}^{r} (\tilde{q}_k^{-1} - 1) \mu_{N,M}(k,j) \tilde{\kappa}_k, \quad j = 1, \ldots, r.
$$

(9.8)

(C2) Let

$$
\|P_k^\perp DQ^\perp_W V^* P_{[M]} VQW D^*T^{-1}\|_{\ell^2} \leq \frac{\ell_2}{8},
$$

(9.9)

$$
\tilde{q}_k \gtrsim \sqrt{B} \log \left(\frac{32\tilde{M}}{\ell}\right) \sum_{l=1}^{r} \mu_{N,M}(k,l) \kappa_l, \quad k = 1, \ldots, r,
$$

(9.10)

and

$$
1 \gtrsim rB^2 \log \left(\frac{32\tilde{M}}{\ell}\right) \sum_{k=1}^{r} (\tilde{q}_k^{-1} - 1) \mu_{N,M}^2(k,j) \tilde{\kappa}_k, \quad j = 1, \ldots, r.
$$

(9.11)

It remains to show that (C1) and (C2) are implied by Assumption 7.2. First, (9.6) and (9.9) are implied by (a) and (b) of Assumption 7.2 respectively because $\|T^{-1}\| \leq \sqrt{\kappa_{\max}}$ and $\|T\| \leq 1/\sqrt{\kappa_{\min}}$. We now show that (d) of Assumption 7.2 implies conditions (9.8) and (9.11). Since $(1 - q_k^l) \cdots (1 - q_k^{\nu}) = 1 - q_k$ implies that $q_k^1 + \cdots + q_k^\nu \geq q_k$, by our choice of $q_k^l = q_k^l = q_k/4$ and $q_k^l = \tilde{q}_k$ for $j \geq 3$, it follows that $2(\mu - 2)\tilde{q}_k \geq q_k$. From (d) of Assumption 7.2 we have that for some appropriate constant $C$, $q_k \gtrsim rB(\log(\epsilon^{-1}) + 1) \log(M\sqrt{\kappa_{\max}}) \tilde{q}_k$ such that $\{q_k\}_{k=1}^r$ satisfies,

$$
1 \gtrsim \sum_{k=1}^{r} (\tilde{q}_k^{-1} - 1) \mu_{N,M}^2(k,j) \tilde{\kappa}_k, \quad j = 1, \ldots, r.
$$

So,

$$
2(8[3\log(8q^{-1}\tilde{M}\sqrt{\kappa_{\max}}\|DD^*\|_{\ell^{\infty} \to \ell^{\infty}}) + \log(\gamma^{-1})] - 2)\tilde{q}_k
$$

$$
\geq q_k \gtrsim (\log(\epsilon^{-1}) + 1) \log(8q^{-1}\tilde{M}\sqrt{\kappa_{\max}}\|DD^*\|_{\ell^{\infty} \to \ell^{\infty}}) \tilde{q}_k.
$$
Since $\gamma = \epsilon/6$, it follows that $\tilde{q}_k \gtrsim \hat{q}_k$. Thus, it follows that given any $j = 1, \ldots, r$,

$$1 \gtrsim \sum_{k=1}^{r} (\tilde{q}_k^{-1} - 1) \mu_{N,M}^2(k, j) \hat{\kappa}_k$$

$$\gtrsim \sum_{k=1}^{r} (\tilde{q}_k^{-1} - 1) \mu_{N,M}^2(k, j) \hat{\kappa}_k$$

as required.

Finally, we show that the remaining conditions (9.7) and (9.10) are implied by (c) of Assumption 7.2. Recall that (c) imposes that for some appropriate constant $C$ and each $k = 1, \ldots, r$,

$$q_k \geq C (\log(\epsilon^{-1}) + 1) \log(8\tilde{M}q^{-1}\sqrt{\kappa_{\max}} \|DD^*\|_{L^\infty \to L^\infty}) \sum_{l=1}^{r} \mu_{N,M}^2(k, l) \kappa_k.$$

Since

$$2(8\lceil 3 \log(8\tilde{M}q^{-1}\sqrt{\kappa_{\max}} \|DD^*\|_{L^\infty \to L^\infty}) + \log(\gamma^{-1}) \rceil - 2) \tilde{q}_k \geq q_k,$$

it follows that

$$\tilde{q}_k \gtrsim \sum_{l=1}^{r} \mu_{N,M}^2(k, l) \kappa_l,$$

as required.

**Proof of $P(A_c^i) \leq \gamma$ for $i = 1, 2$.** Recall from (a) of Assumption 7.2 that

$$\|TDQW^*P_{[M]}VQW^*D^*\|_{L^2 \to L^2} \leq \frac{\sqrt{\kappa_{\min}}}{4\sqrt{\kappa_{\max}}},$$

which implies (8.1) of Proposition 8.4 with $\alpha = \alpha_1$. So, by Proposition 8.4

$$P(A_c^i) \leq \gamma$$

provided that

$$q_k \gtrsim \mathcal{L}^{1/2} \log(3/\gamma) \sqrt{B} \sum_{l=1}^{r} \mu_{N,M}^2(k, l) \kappa_l, \quad k = 1, \ldots, r$$

and

$$1 \gtrsim \mathcal{L} \log(3/\gamma) B^2 \sum_{k=1}^{r} (q_k^{-1} - 1) \mu_{N,M}^2(k, j) \hat{\kappa}_k, \quad j = 1, \ldots, r.$$

These two conditions are implied by (c) and (d) of Assumption 7.2.

**Proof of $P(B_c^i) \leq \gamma$ for $i = 1, 2$.** Recall from (b) of Assumption 7.2 that

$$\|DQW^*P_{[M]}VQW^*D^*\|_{L^2 \to L^\infty} \leq \frac{1}{8\sqrt{\kappa_{\max}}},$$

which implies (8.6) with $\alpha = \beta_1$. So, by Proposition 8.5

$$P(B_c^i) \leq \gamma$$

provided that

$$1 \gtrsim \sqrt{B} \log \left( \frac{4\tilde{M}}{\gamma} \right) \sum_{k=1}^{r} (q_k^{-1} - 1) \mu_{N,M}^2(k, j) \hat{\kappa}_k, \quad j = 1, \ldots, r,$$

and

$$q_k \gtrsim B^2 \log \left( \frac{4\tilde{M}}{\gamma} \right) \sum_{l=1}^{r} \mu_{N,M}^2(k, l) \kappa_l, \quad k = 1, \ldots, r.$$

These two conditions are implied by (c) and (d) of Assumption 7.2.
10 Properties of the subsampled matrix

In this section we show that conditions (i) and (ii) of Proposition 6.1 are satisfied with probability exceeding $1 - \epsilon/6$ under Assumption 7.2.

Recall that conditions (i) and (ii) of Proposition 6.1 are

(i) $\|QWV^*(q_1^{-1}P_{\Omega_1} \oplus \cdots \oplus q_r^{-1}P_{\Omega_r})VQW - QW\|_{\mathcal{H} \to \mathcal{H}} < \frac{1}{4}$.

(ii) $\sup_{j \in \mathbb{N}} \|P_{\{j\}} DQWV^*(q_1^{-1}\Omega_1 \oplus \cdots \oplus q_r^{-1}\Omega_r)VQW D^*P_{\{j\}}\|_{\ell^2 \to \ell^2} < \frac{5}{4}$.

It is sufficient to show that

$$P \left( \|QWV^*(q_1^{-1}P_{\Omega_1} \oplus \cdots \oplus q_r^{-1}P_{\Omega_r})VQW - QW\|_{\mathcal{H} \to \mathcal{H}} \geq \frac{1}{4} \right) < \epsilon/12. \quad (10.1)$$

and

$$P \left( \sup_{j \in \mathbb{N}} \|P_{\{j\}} DQWV^*(q_1^{-1}\Omega_1 \oplus \cdots \oplus q_r^{-1}\Omega_r)VQW D^*P_{\{j\}}\|_{\ell^2 \to \ell^2} \geq \frac{5}{4} \right) \leq \epsilon/12. \quad (10.2)$$

The fact that (10.1) holds under Assumption 7.2 follows from Proposition 8.6.

To see that Proposition 8.7 implies that (10.2) under Assumption 7.2, first note that by our choice of $\tilde{M}$ and Proposition 8.7 (10.2) follows if for each $k, j = 1, \ldots, r$,

$$1 \gtrsim (q_k^{-1} - 1)B^2 \mu_{\mathcal{N}, \tilde{M}}(k, j)^2 \log \left( \frac{12\tilde{M}}{\epsilon} \right). \quad (10.3)$$

which is implied by (d) of Assumption 7.2.

11 Concluding remarks

Recent works [38, 1] have identified the need for further theoretical development on the use of variable density sampling in compressed sensing. Furthermore, variable density sampling schemes are dependent not only on sparsity but also the sparsity structure of the underlying signal. To address this, [1] showed that in the case where the sparsifying operator is associated with an orthonormal basis, by considering levels of the sampling and sparsifying operators, the amount of subsampling possible can be described in terms of the local coherences between the different sections and the sparsity of the underlying signal within each level. This paper presented an extension of this result to the case where the sparsifying operator is constructed from a tight frame. By defining the notions of localized sparsity and localized level sparsities, we derived a recovery guarantee for multi-level sampling patterns based on local coherences and localized level sparsities. One direction of future work would be to apply our abstract result to analyse the use of multi-level sampling schemes in the case of Fourier sampling with some multi-scale analysis operator such as wavelet frames and shearlets. By deriving estimates on the local coherences of such operators, one can expect to obtain a better understanding on how to exploit sparsity structure to subsample. Finally, although this paper considered only the case of a tight frame regularizer, this does not seem to be necessary in practice and it is likely that that similar estimates to Theorem 4.4 can be derived by considering the canonical dual operator of $D$.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/H023348/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis. The author would like to thank Anders Hansen for constructive comments.
Appendix

A The discrete Haar wavelet frame

The discrete Haar frame of redundancy two is defined as follows. Let $N = 2^p$ for some $p \in \mathbb{N}$ and $\{c_0\} \cup \{h_{k,j} : k = 0, \ldots, p - 1, j = 1, \ldots, 2^p\}$ be the discrete Haar basis for $\mathbb{C}^N$. Specifically, $c_0 = 2^{-p/2}(1, \ldots, 1)$ and for $l = 0, \ldots, p - 1$ and $k = 1, \ldots, 2^k$,

$$h_{l,k}[j] = \begin{cases} 2^{(l-p)/2} & j = k2^{p-l} + 1, \ldots, k2^{p-l} + 2^{p-l-1} \\ 2^{(l-p)/2} & j = k2^{p-l} + 2^{p-l-1} + 1, \ldots, k2^{p-l} + 2^{p-l} \\ 0 & \text{otherwise} \end{cases}, \quad 1 \leq j \leq 2^p.$$  

Let $\tilde{c}_0 = c_0$, and for each $k = 0, \ldots, p - 1$, $j = 1, \ldots, 2^k$, let

$$\tilde{h}_{k,j}[n] = \begin{cases} h_{k,j}[n-1] & n = 2, \ldots, N \\ h_{k,j}[N] & n = 1 \end{cases}.$$  

The two discrete Haar wavelet frame of redundancy two is defined by

$$\{2^{-1/2}c_0\} \cup \{2^{-1/2}h_{k,j} : k = 0, \ldots, p - 1, j = 1, \ldots, 2^k\} \cup \{2^{-1/2}\tilde{c}_0\} \cup \{2^{-1/2}\tilde{h}_{k,j} : k = 0, \ldots, p - 1, j = 1, \ldots, 2^k\}.$$  

For analysis purposes, we will order these frame elements in increasing order of scaling with

$$\{\varphi_j\}_{j=1}^{2N} = \left\{2^{-1/2}c_0, 2^{-1/2}c_0, 2^{-1/2}h_{0,1}, 2^{-1/2}\tilde{h}_{0,1}, \ldots, 2^{-1/2}h_{k,j}, \tilde{h}_{k,j}, 2^{-1/2}h_{k,j+1}, \tilde{h}_{k,j+1}, \ldots \right\}$$  

and let $Dx = (\langle x, \varphi_k \rangle)_{k=1}^{2N}$. Note that $D^*D = I$.

The following lemma shows that $\|DD^*\|_{\ell^\infty \to \ell^\infty}$ can be upper bounded independently of $N$.

**Lemma A.1.** $\|DD^*\|_{\ell^\infty \to \ell^\infty} \leq 2/(\sqrt{2} - 1)$.

**Proof.** Let $\varphi = 2^{-1/2}\tilde{h}_{m,n}$. Then, since $\{\tilde{h}_{i,j}\}_{i,j}$ is an orthonormal system,

$$\sum_j |\langle \varphi_j, \varphi_l \rangle| = \frac{1}{2} \sum_{l,k} |\tilde{h}_{m,n}, h_{l,k}| = \frac{1}{2} \sum_{l,k} |\tilde{h}_{m,n} - h_{m,n}, h_{l,k}| + \frac{1}{2} \sum_{l,k} |h_{m,n}, h_{l,k}|$$

$$\frac{1}{2} + \frac{1}{2} \sum_{l,k} |\tilde{h}_{m,n} - h_{m,n}, h_{l,k}|.$$  

Now,

$$\tilde{h}_{m,n} - h_{m,n}[j] = \begin{cases} 2^{(m-p)/2} & j \in \{n2^p - 1, n2^p - m + 2^p - m - 1\} \\ 2^{(m-p)/2 + 1} & j = n2^p - m + 2^p - m - 1 \\ 0 & \text{otherwise} \end{cases}.$$  

Note that for each $l$, $\langle \tilde{h}_{m,n} - h_{m,n}, h_{l,k} \rangle \neq 0$ for at most 3 values of $k$ in $\{0, \ldots, 2^l - 1\}$, and

$$\sum_k |\langle \tilde{h}_{m,n} - h_{m,n}, h_{l,k} \rangle| \leq 4 \cdot 2^{(m+l)/2 - p}.$$  

Therefore,

$$\sum_{l,k} |\tilde{h}_{m,n} - h_{m,n}, h_{l,k}| \leq \frac{4 \cdot 2^{m/2}}{2^p} \sum_{l=0}^{p-1} 2^{l/2} \leq \frac{4}{\sqrt{2} - 1},$$  

39
and
\[
\sum_j |\langle \varphi_j, \varphi_l \rangle| \leq \frac{2}{\sqrt{2} - 1}.
\]

The case where \( \varphi_l = 2^{-1/2} h_{n,m} \) for some \( n, m \) can be approached similarly with the same upper bound. Thus, \( \| DD^* \|_{l^\infty} = \sup_j \| DD^* e_j \|_{l^1} \leq \frac{2}{\sqrt{2} - 1} . \)

\[\square\]

**B  Existence of minimizers**

**Proposition B.1.** Let \( \Omega \subset \mathbb{N} \) be finite and let \( y \in \ell^2(\mathbb{N}) \). There exists \( g_* \in \mathcal{H} \) such that
\[
g_* = \text{argmin} \| Dg \|_{l^1} \text{ subject to } \| P_\Omega V g - y \|_{l^2} \leq \delta.
\]

**Proof.** Let \( (f_n)_{n \in \mathbb{N}} \subset \mathcal{H} \) be a minimizing sequence such that \( \| P_\Omega V f_n - y \|_{l^2} \leq \delta \) for each \( n \in \mathbb{N} \) and
\[
\| Df_n \|_{l^1} \to \inf_{g \in \mathcal{H}} \| Dg \|_{l^1} \text{ subject to } \| P_\Omega V g - y \| \leq \delta, \quad n \to \infty.
\]

This implies that \( (Df_n)_{n \in \mathbb{N}} \) is a bounded sequence in \( \ell^1(\mathbb{N}) \). Since the dual of \( c_0(\mathbb{N}) \) (the Banach space of sequences converging to zero) is \( \ell^1(\mathbb{N}) \), and the unit ball of \( \ell^1 \) is weak*- compact, there exists \( x \in \ell^1(\mathbb{N}) \) and a subsequence \( (Df_{n_k})_{k \in \mathbb{N}} \) such that \( Df_{n_k} \rightharpoonup x \) as \( k \to \infty \), and for each \( z \in c_0(\mathbb{N}) \), \( \langle Df_{n_k}, z \rangle \to \langle x, z \rangle \) as \( k \to \infty \).

Since \( DD^* \in B(\ell^\infty(\mathbb{N}), \ell^\infty(\mathbb{N})) \), it follows that given any \( z \in c_0(\mathbb{N}) \), \( DD^* z \in c_0(\mathbb{N}) \). To see this, note that given any \( \delta > 0 \), we can choose \( N_1 \in \mathbb{N} \) such that \( \| P_{[N_1]}^2 z \|_{l^\infty} \leq \delta/(2 \| DD^* \|_{l^1 \to l^1}) \) for all \( N \geq N_1 \). Furthermore, for this choice of \( N_1 \), we can choose \( N_2 \) such that \( \| P_{[N_1]} DD^* e_k \|_{l^1} \leq \delta/(2 \| e \|_{l^\infty}) \) for all \( k \geq N_2 \). Thus, for all \( k \geq \max \{N_1, N_2\} \),
\[
|\langle D^* Dz, e_k \rangle| \leq |\langle D^* P_{[N_1]}^2 Dz, e_k \rangle| + |\langle D^* P_{[N_1]}^2 Dz, e_k \rangle| \leq |\langle D^* Dz, e_k \rangle| + \| P_{[N_1]} z \|_{l^\infty} \| P_{[N_1]} DD^* e_k \|_{l^1} \leq \delta.
\]

Therefore, \( D^* Dz \in c_0(\mathbb{N}), \langle Df_{n_k}, DD^* z \rangle \to \langle x, DD^* z \rangle \) as \( k \to \infty \) and consequently, \( DD^* Df_{n_k} \rightharpoonup DD^* x \). This implies that,
\[
\lim_{k \to \infty} \inf \| Df_{n_k} \|_{l^1} \geq \| DD^* x \|_{l^1}.
\]

Furthermore, because \( Df_{n_k} \) converges weakly to \( x \) in \( \ell^2(\mathbb{N}) \) and \( P_\Omega V D^* \) is a compact operator (since it is of finite rank), \( P_\Omega V f_{n_k} = P_\Omega V D^* Df_{n_k} \to P_\Omega V D^* x \) as \( k \to \infty \). So, \( \| P_\Omega V D^* x - y \|_{l^2} \leq \delta \). Thus, \( g_* := D^* x \) is a minimizer.

\[\square\]

**References**

[1] B. Adcock, A. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. Preprint, 2014.

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

[3] M. Benning, L. Gladden, D. Holland, C.-B. Schönlieb, and T. Valkonen. Phase reconstruction from velocity-encoded mri measurements—a survey of sparsity-promoting variational approaches. Journal of Magnetic Resonance, 238:26–43, 2014.

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

[6] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randell. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011.

[7] E. J. Candès and Y. Plan. A probabilistic and ripless theory of compressed sensing. *Information Theory, IEEE Transactions on*, 57(11):7235–7254, 2011.

[8] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[9] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[10] E. Cordero and K. Gröchenig. Localization of frames ii. *Applied and Computational Harmonic Analysis*, 17(1):29–47, 2004.

[11] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Process.*, 46:886–902, 1998.

[12] 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.

[13] I. Daubechies, B. Han, A. Ron, and Z. Shen. Framelets: Mra-based constructions of wavelet frames. *Applied and computational harmonic analysis*, 14(1):1–46, 2003.

[14] M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *Image Processing, IEEE Transactions on*, 14(12):2091–2106, 2005.

[15] B. Dong, Z. Shen, et al. Mra based wavelet frames and applications. *IAS Lecture Notes Series, Summer Program on The Mathematics of Image Processing, Park City Mathematics Institute*, 2010.

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

[17] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007.

[18] M. Fornasier and K. Gröchenig. Intrinsic localization of frames. *Constructive Approximation*, 22(3):395–415, 2005.

[19] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. Springer, 2013.

[20] P. Grohs. Intrinsic localization of anisotropic frames. *Applied and Computational Harmonic Analysis*, 35(2):264–283, 2013.

[21] P. Grohs and S. Vigogna. Intrinsic localization of anisotropic frames ii: \( \alpha \)-molecules. *Journal of Fourier Analysis and Applications*, 21(1):182–205, 2014.

[22] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566, 2011.

[23] S. Jaffard. Propriétés des matrices bien localisées près de leur diagonale et quelques applications. In *Annales de l’institut Henri Poincaré (C) Analyse non linéaire*, volume 7, pages 461–476. Gauthier-Villars, 1990.

[24] F. Krahmer, D. Needell, and R. Ward. Compressive sensing with redundant dictionaries and structured measurements. *CoRR*, abs/1501.03208, 2015.
[25] 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.

[26] 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.

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

[28] 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.

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

[30] C. Poon. On the role of total variation in compressed sensing. *SIAM Journal on Imaging Sciences*, 8(1):682–720, 2015.

[31] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *Signal Processing Letters, IEEE*, 18(10):595–598, 2011.

[32] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *Information Theory, IEEE Transactions on*, 54(5):2210–2219, 2008.

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

[34] M. Rudelson and R. Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.

[35] V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candes, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences*, 109(26):E1679–E1687, 2012.

[36] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[37] Q. Wang, M. Zenge, H. E. Cetingul, E. Mueller, and M. S. Nadar. Novel sampling strategies for sparse mr image reconstruction. *Proc. Int. Soc. Mag. Res. in Med.*, (22), 2014.

[38] R. Ward and F. Krahmer. Stable and robust sampling strategies for compressive imaging. 2013.

[39] Y. Wiaux, L. Jacques, G. Puy, A. Scaife, and P. Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 395(3):1733–1742, 2009.

[40] P. Wojtaszczyk. Stability and instance optimality for gaussian measurements in compressed sensing. *Foundations of Computational Mathematics*, 10(1):1–13, 2010.