Appl. Comput. Harmon. Anal. 55 (2021) 1–40

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Uniform recovery in infinite-dimensional compressed sensing and applications to structured binary sampling

Ben Adcock [a], Vegard Antun [b,*], Anders C. Hansen [c,b]

[a] *Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada*
[b] *Department of Mathematics, University of Oslo, P.O box 1053, Blindern, 0316 Oslo, Norway*
[c] *DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom*

ARTICLE INFO

Article history:
Received 2 May 2019
Received in revised form 11 April 2021
Accepted 12 April 2021
Available online 15 April 2021
Communicated by Thomas Strohmer

MSC:
94A20
42C40
42C10
15B52

Keywords:
Infinite-dimensional compressed sensing
Uniform recovery
Walsh sampling
Wavelet recovery
Sparsity in levels
Local coherence

ABSTRACT

Infinite-dimensional compressed sensing deals with the recovery of analog signals (functions) from linear measurements, often in the form of integral transforms such as the Fourier transform. This framework is well-suited to many real-world inverse problems, which are typically modeled in infinite-dimensional spaces, and where the application of finite-dimensional approaches can lead to noticeable artefacts. Another typical feature of such problems is that the signals are not only sparse in some dictionary, but possess a so-called local sparsity in levels structure. Consequently, the sampling scheme should be designed so as to exploit this additional structure. In this paper, we introduce a series of uniform recovery guarantees for infinite-dimensional compressed sensing based on sparsity in levels and so-called multilevel random subsampling. By using a weighted ℓ^1-regularizer we derive measurement conditions that are sharp up to log factors, in the sense that they agree with the best known measurement conditions for oracle estimators in which the support is known a priori. These guarantees also apply in finite dimensions, and improve existing results for unweighted ℓ^1-regularization. To illustrate our results, we consider the problem of binary sampling with the Walsh transform using orthogonal wavelets. Binary sampling is an important mechanism for certain imaging modalities. Through carefully estimating the local coherence between the Walsh and wavelet bases, we derive the first known recovery guarantees for this problem.

1. Introduction

Compressive sensing (CS), introduced by Candès, Romberg & Tao in [1] and Donoho in [2], has been an area of substantial research during the last decade. The key observation, which lays the foundation for this

* Corresponding author.
E-mail addresses: ben_adcock@sfu.ca (B. Adcock), vegarant@math.uio.no (V. Antun), ach70@cam.ac.uk (A.C. Hansen).

https://doi.org/10.1016/j.acha.2021.04.001

line of research, is that a sparse vectors $x \in \mathbb{C}^M$, can be recovered from $m < M$ linear measurements, under suitable conditions on the measurements, using convex optimization algorithms [3,4].

Imaging has been one of the most successful areas of application of CS. However, in this area, the sparsity assumption is typically too general, to explain the performance of CS algorithms in practice; instead, the correct assumption is to impose a certain sparsity pattern. Examples include all applications using Fourier samples – such as Magnetic Resonance Imaging (MRI) [5–7], surface scattering [8], Computerized Tomography (CT) and electron microscopy – as well as applications using binary sampling, e.g. fluorescence microscopy [9], lensless imaging [10] and numerous other optical imaging modalities [11–13]. Natural images, when sparsified via a wavelet (or more generally, X-let) transform, are not only sparse, but have specific sparsity structure [14,15]. For wavelets, which will be our sparsifying transform in this paper, natural images have coefficients where most of the large entries are concentrated at the coarse scales, and progressively fewer at the fine scales (termed *asymptotic sparsity* in [14]).

In the presence of structured sparsity, it is natural to ask how best to promote this additional structure. In [14] it was proposed to do this via the sampling operator. Wavelets partition the frequency space into dyadic bands corresponding to distinct scales. Hence, by choosing Fourier samples in these bands corresponding to the local sparsities, one obtains a structured sampling scheme – a so-called *multilevel sampling* scheme – which promotes the asymptotic sparsity structure. The practical benefits of such schemes have been demonstrated in [15] for various different imaging modalities, including MRI, Nuclear Magnetic Resonance (NMR) spectroscopy, fluorescence microscopy and Helium Atom Scattering. Theoretical analysis has been presented in [14] (nonuniform recovery) and [16,17] (uniform recovery in the finite-dimensional setting).

1.1. Main results

This paper has two main objectives. First, we generalize existing uniform recovery guarantees [16,17] from the finite-dimensional to the infinite-dimensional setting. This extension is important for practical imaging. Although much of the compressive imaging literature considers the recovery of discrete images (i.e. finite-dimensional arrays) from discrete measurements (e.g. the discrete Fourier transform), modalities such as MRI, NMR and others are naturally analog, and hence better modeled over the continuum (i.e. functions, and the continuous Fourier transform). Indeed, as we will see in Section 2.3, discretizing such a problem leads to measurement mismatch [18], and in the case of wavelet recovery, the wavelet crime [19, Page 232], both of which can introduce artefacts in the reconstruction [20]. In this paper, we consider signals as functions $f \in L^2([0,1))$ and work with continuous integral transforms, thus avoiding these pitfalls.

In our theoretical analysis, we also improve the uniform recovery guarantee given in previous works [16,17]. Unlike previous results, our recovery guarantees are, up to log factors, optimal: specifically, they agree with those of the best known measurement conditions for oracle least-square estimator based on *a priori* knowledge of the support [21]. We do this by replacing the standard ℓ^1-minimization decoder by a certain weighted ℓ^1-minimization decoder; an idea originally proposed in [22].

Our second objective is to consider binary sampling. Previous works have addressed the case of (discrete or continuous) Fourier sampling. Yet many imaging modalities, e.g. fluorescence microscopy and lensless imaging, require binary sampling operators. To do so, we replace the Fourier transform

$$\mathcal{F}f(\omega) := \int_{[0,1)} f(x) e^{-2\pi \mathrm{i} \omega x} \,\mathrm{d}x, \qquad f \in L^2([0,1)),$$

by the binary *Walsh transform*

$$\mathcal{W}f(n) := \int_{[0,1)} f(x) w_n(x) \,\mathrm{d}x, \qquad f \in L^2([0,1))$$

where $w_n\colon [0,1) \to \{+1,-1\}$, $n \in \mathbb{Z}_+ := \{0,1,\dots\}$ denote the Walsh functions. This is a widely used sampling operator in binary imaging [9,10], and often goes under the name of Hadamard sampling in the discrete case. Working with this continuous transform, we provide analogous guarantees for binary sampling to those for Fourier sampling. As a side note, we remark that working in the continuous setting also simplifies the analysis (specifically, the derivation of so-called *local coherence* estimates) over working directly with the discrete setup.

We note that in this paper we only consider recovery guarantees for one dimensional functions. We expect that the setup for higher dimensional function will deviate slightly from what we present here, and we save this discussion for future work.

The outline of the remainder of this paper is as follows. We commence in Section 2 by reviewing previous work, and in particular, the existing finite-dimensional theory. We then introduce an abstract infinite-dimensional model for isometries U acting on $\ell^2(\mathbb{N})$ in Section 3. Here we will derive sufficient conditions for such operators to provide uniform recovery guarantees. In Section 4 we continue this work by finding conditions for which the cross-Gramian U between a wavelet and Walsh basis satisfies these conditions. Finally in Sections 5 and 6 we present proofs of our main results.

2. Sparsity in levels in finite dimensions

2.1. Notation

We call a vector $x \in \mathbb{C}^N$ s-sparse if $|\mathrm{supp}(x)| \le s$, where $\mathrm{supp}(x) = \{i : x_i \neq 0\}$. We write $A \lesssim B$ if there exists a constant $C > 0$, independent of all relevant parameters, so that $A \le CB$, and similarly for $A \gtrsim B$. Furthermore we define the following projection operator.

Definition 2.1 *(Finite dimensional projection operator).* Let $N \in \mathbb{N}$ and $\Omega \subseteq \{1,\dots,N\}$. We let P_Ω denote an $N \times N$ or $|\Omega| \times N$ projection operator, depending on the context. Whenever $P_\Omega \in \mathbb{C}^{N\times N}$, it acts as follows

$$(P_\Omega x)_i = \begin{cases} x_i & \text{if } i \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

on a vector $x \in \mathbb{C}^N$. In the same way, we define the projection operator $P_\Omega \in \mathbb{C}^{|\Omega|\times N}$ by discarding all the zero entries of $P_\Omega x$ in (2.1). If $\Omega = \{N_{k-1}+1,\dots,N_k\}$ we write $P_{N_k}^{N_{k-1}} = P_{\{N_{k-1}+1,\dots,N_k\}}$, and simply P_{N_k} if $N_{k-1} = 0$.

2.2. Finite model

Let $V \in \mathbb{C}^{N\times N}$ be a measurement matrix, e.g., a Fourier or Hadamard matrix, denoted V_{Four} and V_{Had}, respectively, and let $\Omega \subset \{1,\dots N\}$ with $|\Omega| = m < N$. In a typical finite-dimensional CS setup we consider the recovery of a signal $x \in \mathbb{C}^N$ from measurements $y = P_\Omega V x + e \in \mathbb{C}^m$, where $e \in \mathbb{C}^m$ is a vector of measurement error. If x is sparse in a discrete wavelet basis, one then recovers its coefficients by solving the optimization problem

$$\underset{z\in\mathbb{C}^N}{\text{minimize}}\, \|z\|_1 \quad \text{subject to} \quad \|P_\Omega V \Psi^{-1} z - y\|_2 \le \eta \tag{2.2}$$

where $\Psi \in \mathbb{C}^{N\times N}$ is a discrete wavelet transform and $\eta \ge \|e\|_2$ is a noise parameter. Usually one would scale $V \in \mathbb{C}^{N\times N}$ so that it becomes orthonormal and choose an orthonormal wavelet basis, so that the matrix $U = V\Psi^{-1} = V\Psi^*$ acts as an isometry on $\mathbb{C}^N$. Here Ψ^* denotes the adjoint of Ψ.

Suppose that U is indeed an isometry. To obtain a uniform recovery guarantee for the above system, one typically first shows that the matrix $A = \frac{1}{\sqrt{p}} P_\Omega U \in \mathbb{C}^{m\times N}$, with $p = \frac{m}{N}$, satisfies the *Restricted Isometry Property* (RIP) with high probability.

Definition 2.2 *(RIP).* Let $1 \leq s \leq N$ and $A \in \mathbb{C}^{m\times N}$. The *Restricted Isometry Constant (RIC)* of order s is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \quad \forall x \in \Sigma_s,$$

where Σ_s denotes the set of s-sparse vectors in $\mathbb{C}^N$. If $0 \leq \delta < 1$ we say that A has the *Restricted Isometry Property (RIP)* of order s.

For matrices satisfying the RIP, it can be shown that ℓ^1-minimization can recover sparse vectors, as illustrated by the next theorem.

Theorem 2.3 *([4, Thm. 6.12]). Suppose the RIC δ_{2s} of order $2s$ of a matrix $A \in \mathbb{C}^{m\times N}$ satisfies $\delta_{2s} < 4/\sqrt{41} \approx 0.62$. Then for any $x \in \mathbb{C}^N$ and $e \in \mathbb{C}^m$ with $\|e\|_2 \leq \eta$, any solution $\hat{x} \in \mathbb{C}^N$ of*

$$\underset{z\in\mathbb{C}^N}{\text{minimize}}\, \|z\|_1 \quad \textit{subject to} \quad \|z-(Ax+e)\|_2 \leq \eta$$

satisfies

$$\|x-\hat{x}\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta$$

where $C, D > 0$ are constants dependent on δ_{2s} only and $\sigma_s(x)_1 = \inf\{\|x-z\|_1 : z \in \Sigma_s\}$.

For an isometry $U \in \mathbb{C}^{N\times N}$ the question of whether or not $P_\Omega U$ satisfies the RIP is related to the so-called *coherence* of U:

Definition 2.4 *(Coherence).* Let $U \in \mathbb{C}^{N\times N}$ be an isometry. The *coherence* of U is

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

With this term defined, we look at measurement conditions ensuring that $P_\Omega U$ satisfies the RIP. Combined with Theorem 2.3, this ensures stable and accurate recovery of sparse vectors using ℓ^1-minimization.

Theorem 2.5 *([4, Thm. 12.32, see also page 371]). Let $U \in \mathbb{C}^{N\times N}$ be an isometry and let $0 < \delta, \epsilon < 1$. Suppose $\Omega = \{t_1, \dots t_m\} \subseteq \{1, \dots, N\}$ where each t_k is chosen uniformly and independently at random from the set $\{1, \dots, N\}$. If*

$$m \gtrsim \delta^{-2} \cdot s \cdot N \cdot \mu(U) \cdot \left(\log(2m)\log(2N)\log^2(2s) + \log(\epsilon^{-1})\right)$$

then with probability $1-\epsilon$ the matrix $A = \frac{1}{\sqrt{p}} P_\Omega U \in \mathbb{C}^{m\times N}$, with $p = \frac{m}{N}$, satisfies the RIP of order s with $\delta_s \leq \delta$. The constant implied by $\gtrsim$ is universal, and does not depend on any of the parameters.

(We slightly abuse notation here in that we allow for possible repeats of the values t_i that make up Ω.) Thus if the coherence $\mu(U) \approx N^{-1}$ we obtain the RIP of order s using approximately s measurements up to constants and log factors.

Historical note 2.6. The RIP, introduced above, can be traced back to the work by Candès & Tao in [23] and [24]. The former introduces the so-called *uniform uncertainty principle*, and the latter defines the *restricted isometry constant*, which is similar to how we define the RIP above. In [24] they also derived sufficient conditions for exact recovery of s-sparse vectors via ℓ_1-minimization. These sufficient conditions have been improved several times, [25–27], resulting in Theorem 6.12 in [4]. See notes section in [4] for an in-depth discussion. Theorem 2.5 follows from Theorem 12.32 in [4], and the discussion around Example 3, page 371. Theorem 12.32 is based on the results in [28]. See notes section in [4, Chap. 12] for further details.

There are, however, two problems with this approach. First, in our setup, where $U = V\Psi^*$ is the product of a Fourier or Hadamard matrix and a discrete wavelet transform, the coherence $\mu(U) \approx 1$. Hence satisfying the RIP requires at least $m \approx N$ measurements. Second, the RIP asserts recovery for *all* s-sparse vectors of wavelet coefficients, and thus does not exploit any additional structure these coefficients possess. However, as stated, wavelet coefficients are highly structured: large wavelet coefficients tend to cluster at coarse scales, with coefficients at fine scales being increasingly sparse.

Motivated by this, the following structured sparsity model was introduced in [14]:

Definition 2.7 *(Sparsity in levels).* Let $\mathbf{M} = [M_1, \dots, M_r] \in \mathbb{N}^r$, $M_0 = 0$, with $1 \le M_1 < \cdots < M_r = M$ and let $\boldsymbol{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$. We say that the vector $x \in \mathbb{C}^M$ is sparse in levels if

$$|\mathrm{supp}(x) \cap \{M_{l-1}+1, \dots, M_l\}| \le s_l \quad \text{for } l = 1, \dots, r.$$

In this case we say that x is $(\boldsymbol{s}, \mathbf{M})$-sparse, where $\boldsymbol{s}$ and $\mathbf{M}$ are called the local sparsities and sparsity levels, respectively. We denote the set of all $(\boldsymbol{s}, \mathbf{M})$-sparse vectors by $\Sigma_{\boldsymbol{s},\mathbf{M}}$.

As noted above, randomly subsampling an isometry U is a poor measurement protocol for coherent problems such as Fourier–Wavelets. Instead, in [14] it was proposed to sample in the following structured way:

Definition 2.8 *(Multilevel random subsampling).* Let $\mathbf{N} = [N_1, \dots, N_r] \in \mathbb{N}^r$, where $1 \le N_1 < \cdots < N_r = N$ and $\boldsymbol{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$ with $m_k \le N_k - N_{k-1}$ for $k = 1, \dots, r$, and $N_0 = 0$. For each $k = 1, \dots, r$, let $\Omega_k = \{N_{k-1}+1, \dots, N_k\}$ if $m_k = N_k - N_{k-1}$ and if not, let $t_{k,1}, \dots, t_{k,m_k}$ be chosen uniformly and independently from the set $\{N_{k-1}+1, \dots, N_k\}$, and set $\Omega_k = \{t_{k,1}, \dots, t_{k,m_k}\}$. If $\Omega = \Omega_{\mathbf{N},\boldsymbol{m}} = \Omega_1 \cup \cdots \cup \Omega_r$ we refer to Ω as an $(\mathbf{N}, \boldsymbol{m})$-*multilevel subsampling scheme.*

In the definition above, notice that if $m_k = N_k - N_{k-1}$, then there is no randomness involved, and we let $\Omega_k = \{N_{k-1}+1, \dots, N_k\}$. That is, we fully sample level k. For Fourier or Walsh sampling with wavelet reconstruction, it is sometimes necessary to fully sample the r_0 first levels, since images seldom are sparse in wavelets at coarse scales. In many of our theorems, we have therefore included a parameter r_0, which accounts for the deterministic sampling of the first r_0 sampling levels. If $r_0 = 0$, we do not fully sample any level.

For this structured model, the following extension of the RIP was first introduced in [16].

Definition 2.9 *(RIPL).* Let $\boldsymbol{s}, \mathbf{M} \in \mathbb{N}^r$ be given local sparsities and sparsity levels, respectively. For a matrix $A \in \mathbb{C}^{m \times N}$ the *Restricted Isometry Constant in Levels (RICL)* of order $(\boldsymbol{s}, \mathbf{M})$, denoted $\delta_{\boldsymbol{s},\mathbf{M}}$, is the smallest $\delta \ge 0$ such that

$$(1-\delta)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta)\|x\|_2^2 \quad \forall x \in \Sigma_{\boldsymbol{s},\mathbf{M}}.$$

We say that A has the *Restricted Isometry Property in Levels (RIPL)* if $0 \le \delta < 1$.

We shall see that this leads to uniform recovery of all $(\boldsymbol{s}, \mathbf{M})$-sparse vectors, but first we define the *best* $(\boldsymbol{s}, \mathbf{M})$*-term approximation error* of $x \in \mathbb{C}^N$. That is

$$\sigma_{\boldsymbol{s},\mathbf{M}}(x)_p \coloneqq \inf\{\|x - z\|_p : z \in \Sigma_{\boldsymbol{s},\mathbf{M}}\}.$$

Theorem 2.10 *([16, Thm. 4.4]). Let $\boldsymbol{s}, \mathbf{M} \in \mathbb{N}^r$ be local sparsities and sparsity levels, respectively. Let $\alpha_{\boldsymbol{s},\mathbf{M}} = \max_{k,l=1,\dots,r} s_l/s_k$ and $s = s_1 + \cdots + s_r$. Suppose that the RICL $\delta_{2\boldsymbol{s},\mathbf{M}} \geq 0$ for the matrix $A \in \mathbb{C}^{m\times M}$ satisfies*

$$\delta_{2\boldsymbol{s},\mathbf{M}} < \frac{1}{\sqrt{r(\sqrt{\alpha_{\boldsymbol{s},\mathbf{M}}} + \frac{1}{4})^2 + 1}}. \tag{2.3}$$

Then, for $x \in \mathbb{C}^M$ and $e \in \mathbb{C}^m$ with $\|e\|_2 \leq \eta$, any solution $\hat{x}$ of

$$\underset{z\in\mathbb{C}^M}{\text{minimize}} \|z\|_1 \quad \textit{subject to} \quad \|z - (Ax + e)\|_2 \leq \eta$$

satisfies

$$\|x - \hat{x}\|_2 \leq (C + C'(r\alpha_{\boldsymbol{s},\mathbf{M}})^{1/4})\frac{\sigma_{s,\mathbf{M}}(x)_1}{\sqrt{s}} + (D + D'(r\alpha_{\boldsymbol{s},\mathbf{M}})^{1/4})\eta$$

where $C, C', D, D' > 0$ are constants which only dependent on $\delta_{2\boldsymbol{s},\mathbf{M}}$.

In [17] Li & Adcock investigated conditions under which a subsampled isometry $U \in \mathbb{C}^{N\times N}$ satisfies the RIPL. It was shown that the number of samples required to satisfy the RIPL was related to the so-called *local coherence* properties of U:

Definition 2.11. Let $U \in \mathbb{C}^{N\times N}$ be an isometry and $\mathbf{N}, \mathbf{M} \in \mathbb{N}^r$ be given sampling and sparsity levels. The *local coherence* of U is

$$\mu_{k,l} = \mu_{k,l}(\mathbf{N}, \mathbf{M}) = \max\{|U_{ij}|^2 : i = N_{k-1} + 1, \dots, N_k, j = M_{l-1} + 1, \dots, M_l\}.$$

Theorem 2.12 *([17, Thm. 3.2]). Let $U \in \mathbb{C}^{N\times N}$ be an isometry. Let $r \in \mathbb{N}$, $0 < \delta, \epsilon < 1$, and $0 \leq r_0 \leq r$. Let $\Omega = \Omega_{\mathbf{N},\boldsymbol{m}}$ be an $(\mathbf{N}, \boldsymbol{m})$-multilevel random subsampling scheme, and let $(\boldsymbol{s}, \mathbf{M})$ be given local sparsities and sparsity levels, respectively. Let $\tilde{m} = m_{r_0+1} + \ldots + m_r$ and $s = s_1 + \ldots + s_r$. Suppose that the m_ks satisfy*

$$m_k = N_k - N_{k-1}, \quad \textit{for } k = 1, \dots, r_0, \tag{2.4}$$

and

$$m_k \gtrsim \delta^{-2} \cdot (N_k - N_{k-1}) \cdot \left(\sum_{l=1}^{r} s_l \mu_{k,l}\right) \cdot \left(r \log(2\tilde{m}) \log(2N) \log^2(2s) + \log(\epsilon^{-1})\right) \tag{2.5}$$

for $k = r_0 + 1, \dots, r$. Then, with probability at least $1 - \epsilon$, the matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{p_1}} P_{\Omega_1} U \\ \vdots \\ \frac{1}{\sqrt{p_r}} P_{\Omega_r} U \end{bmatrix} \quad \textit{where } p_k = \frac{m_k}{N_k - N_{k-1}} \quad \textit{for } k = 1, \dots, r \tag{2.6}$$

satisfies the RIPL of order $(\boldsymbol{s}, \mathbf{M})$ *with constant* $\delta_{\boldsymbol{s},\mathbf{M}} \leq \delta$. *The constant implied by* $\gtrsim$ *is universal, and does not depend on any of the parameters.*

This theorem provides sufficient conditions on the number of local measurements m_k needed to ensure uniform recovery explicitly in terms of local sparsities s_k and local coherences $\mu_{k,l}$. In particular, if the local coherences are suitably well-behaved, then recovery may still be possible from highly subsampled measurements, even though the global coherence may be high (see next). Note that the condition (2.4), whereby the first r_0 sampling levels are saturated, models practical imaging scenarios where the low Fourier frequencies are typically fully sampled. For an in-depth discussion on the choice of r_0 we refer to [17, Sec. 3.2].

To illustrate this theorem, in [29] the authors consider the one-dimensional discrete Fourier sampling problem with sparsity in Haar wavelets. For the Haar wavelet basis we choose an ordering where the first level $\{M_0 + 1, M_1\} = \{1, 2\}$ consists of the scaling function and mother wavelet and the subsequent levels are chosen so that $\{M_{l-1} + 1, \dots, M_l\} = \{2^{l-1} + 1, \dots, 2^l\}$ consists of the wavelets at scale $l - 1$. This gives the sparsity levels

$$\mathbf{M} = [2^1, 2^2, \dots, 2^r],$$

where $r = \log_2(N)$ (assumed to be an integer). Next we define the entries in the Fourier matrix $V_{\text{Four}} \in \mathbb{C}^{N\times N}$ as

$$V_{\text{Four}} = \left(\frac{1}{\sqrt{N}} \exp(2\pi \mathrm{i}(j-1)\omega/N) \right)_{\omega=-N/2+1,\ j=1}^{N/2,\ N},$$

where we have started the ordering of the rows with negative indices for convenience. We define the sampling levels for the frequencies ω in dyadic bands with $W_1 = \{0, 1\}$ and

$$W_{k+1} = \{-2^k + 1, \dots, -2^{k-1}\} \cup \{2^{k-1} + 1, \dots, 2^k\}, \quad k = 1, \dots, r-1.$$

Notice that for a suitable reordering of the rows of V_{Four} these bands correspond to the sampling levels $\mathbf{N} = [2^1, 2^2, \dots, 2^r]$.

Theorem 2.13 *([17, Cor. 3.3]). Let* $N = 2^r$ *for some* $r \geq 1$ *and let* $U = V_{Four}\Psi^{-1} \in \mathbb{C}^{N\times N}$, *where* Ψ *is the Haar wavelet matrix. Let* $0 < \delta, \epsilon < 1$ *and let* $\mathbf{N} = \mathbf{M} = [2^1, \dots, 2^r]$. *Let* $m = m_1 + \cdots m_r$ *and* $s = s_1 + \cdots s_r$. *For each* $k = 1, \dots, r$ *suppose we draw* m_k *Fourier samples from band* W_k *randomly and independently, where*

$$m_k \gtrsim \delta^{-2} \cdot \left(\sum_{l=1}^{r} 2^{-|k-l|} s_l \right) \left(r \log(2m) \log(2N) \log^2(2s) + \log(\epsilon^{-1}) \right).$$

Then with probability at least $1 - \epsilon$ *the matrix* (2.6) *satisfies the RIPL with constant* $\delta_{s,\mathbf{M}} \leq \delta$. *The constant implied by* $\gtrsim$ *is universal, and does not depend on any of the parameters.*

Here, for convenience, we have taken $r_0 = 0$; see [17] for further discussion on this point.

2.3. Shortcomings

These results have two primary shortcomings, which we now discuss in further detail. The key issue is that they are limited to finite dimensions. As noted in Section 1, applying finite-dimensional recovery procedures

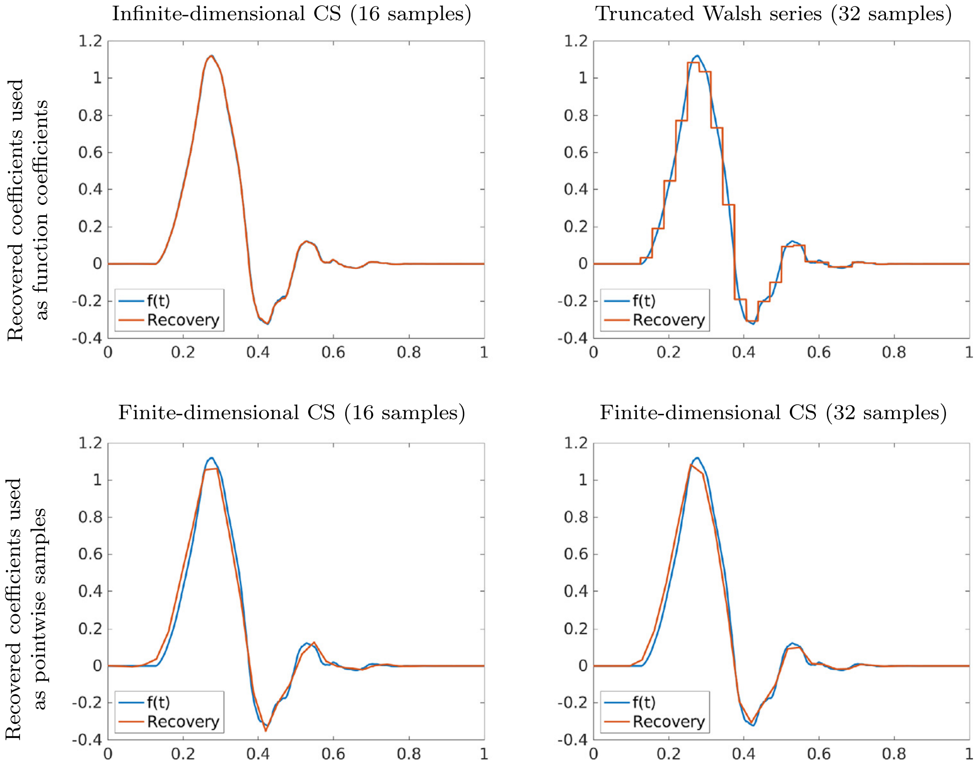

Fig. 1. Reconstructions (using Walsh samples) of $f(t) = \phi_{4,4}(t)$, where ϕ is the Daubechies scaling function, corresponding the wavelets with four vanishing moments. Upper left: Reconstruction from the first 16 Walsh samples using an infinite-dimensional CS model (described in Section 3). In this model, the reconstruction space is spanned by the $M = 32$ first wavelet and scaling functions. Upper right: Truncated Walsh series based on the first 32 Walsh samples. Lower left: Reconstruction from the first 16 Walsh samples using the finite-dimensional (32×32) CS model. Lower right: Reconstruction from the first 32 Walsh samples using the finite-dimensional (32×32) CS model. Note that the right images just correspond to different ways of visualizing the results. In particular, in the upper right image, we show a high-resolution Walsh function in an infinite-dimensional model. In the two lower images, we show vectors of length 32, and thereby commit the wavelet crime.

to analog problems can result in artefacts. For simplicity, let $N = 2^p$. We have argued that analog signals should be modeled as elements in $L^2([0,1))$, rather than $\mathbb{C}^N$. Yet, above we have tried to use discrete tools for recovering the signal $f \in L^2([0,1))$ by replacing $\mathcal{W}f$ and $\mathcal{F}f$ with V_{Had} and V_{Four}, respectively. Next we argue that this construction leads to both *measurement mismatch* and the *wavelet crime*.

Let $\chi_{[a,b)}$ denote step functions on the interval $[a,b)$ and set $\Delta_{k,p} = [k2^{-p}, (k+1)2^{-p})$. We see that replacing $\mathcal{W}f$ with $V_{\text{Had}} \in \mathbb{C}^{N\times N}$ is equivalent to replacing f by e.g. $\tilde{f} = \sum_{k=0}^{N-1} c_k \chi_{\Delta_{k,p}}$ for some $c \in \mathbb{C}^N$, since $\mathcal{W}\tilde{f} = V_{\text{Had}}c$. Clearly, $\mathcal{W}\tilde{f}$ will be a poor approximation to $\mathcal{W}f$. We refer to this as measurement mismatch.

Next let ϕ^0, ϕ^1 denote a scaling function and wavelet, respectively, and set $\phi^s_{j,k} = 2^{j/2}\phi^s(2^j \cdot -k)$ for $s \in \{0,1\}$. By construction the solution $\hat{x}$ of (2.1) will be the coefficients of a function $\hat{f}$ written in a basis consisting of both wavelets and scaling functions. Equivalently we can represent $\hat{f}$ in the basis $\{\phi^0_{j,k}\}_{k=0}^{N-1}$ using the coefficients $c = \Psi^{-1}\hat{x} \in \mathbb{C}^N$. The wavelet crime is whenever we let c represent pointwise samples of f i.e., $c_k = f(k/N)$.

What does this mean for reconstruction? To illustrate the issue we provide a similar example to the first numerical simulation in [30], showing how finite-dimensional compressed sensing fails to recover even a function that is 1-sparse (meaning it has only one non-zero coefficient) in its wavelet decomposition. Indeed, in Fig. 1 we consider the problem of recovering a function f from samples of the continuous Walsh transform. In particular, we choose $f(t) = \phi_{4,4}(t)$, where ϕ is the Daubechies scaling function, corresponding to the wavelet with four vanishing moments. To ensure that the wavelet decomposition of f is 1-sparse, we use a wavelet reconstruction basis starting at scale $J_0 = 4$ (see Definition 4.3 for complete setup). Fig. 1 shows the poor performance of CS using the discrete finite-dimensional setup when applied to a continuous problem. Conversely, the infinite-dimensional CS approach, which we develop in the next sections, gives a much higher fidelity reconstruction from exactly the same samples as used in the finite-dimensional case. In fact, the infinite-dimensional CS reconstruction recovers f perfectly up to numerical errors occurring from solving the optimization problem. We also observe the slightly paradoxical phenomenon in the finite-dimensional case: more samples do not improve performance. This is due to the fact that the finite-dimensional CS solution with full sampling coincides with the truncated Walsh series (direct inversion) approximation. This approximation is clearly highly suboptimal, as demonstrated in Fig. 1.

We note in passing that the above crimes stem from too early a discretization of the inverse problem. Our infinite-dimensional CS approach replaces $V_{\text{Had}}\Psi^{-1}$ by a finite section of the isometry $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ representing the change of basis between the continuous Fourier or Walsh transform and wavelet basis.

On a related note, even if one were to ignore the above issues, estimating the local coherences $\mu_{k,l}$ in the discrete setting for anything but the Haar wavelet becomes extremely complicated. Conversely, by moving to the continuous setting, these estimates become much easier to derive. We do this later in the paper for arbitrary Daubechies' wavelets with the Walsh transform.

The second shortcoming relates to Theorem 2.10. It says that we can guarantee recovery of all sparse signals provided the matrix $A \in \mathbb{C}^{m\times M}$ satisfies the RIPL with constant

$$\delta_{2\boldsymbol{s},\mathbf{M}} < \frac{1}{\sqrt{r(\sqrt{\alpha_{\boldsymbol{s},\mathbf{M}}}+\frac{1}{4})^2+1}}.$$

Here r is the number of levels and $\alpha_{\boldsymbol{s},\mathbf{M}} = \max_{k,l=1,\dots,r} s_l/s_k$ is the sparsity ratio. Inserting the above inequality into Theorem 2.12 gives a sampling condition of the form

$$m_k \gtrsim r \cdot \alpha_{\boldsymbol{s},\mathbf{M}} \cdot (N_k - N_{k-1}) \cdot \left(\sum_{l=1}^{r} \mu_{k,l} s_l\right) \cdot L$$

where L is the log factors. This means that the sparsity ratio $\alpha_{\boldsymbol{s},\mathbf{M}}$ will affect the sampling condition in all sampling levels. Thus for signals where we expect the local sparsities to vary greatly from level to level (e.g. wavelets) this will lead to an unreasonably high number of samples.

To overcome this problem, using an idea from [22], we replace the ℓ^1-regularizer in the optimization problem (2.2) with a weighted ℓ^1-regularizer. For a suitable choice of weights, this removes the factor of $\alpha_{\boldsymbol{s},\mathbf{M}}$ in the various measurement conditions. As we show, these guarantees are optimal up to constants and log factors.

3. Extensions to infinite dimensions

3.1. Notation

We will continue with the notation we introduced above, extended to infinite dimensions. That is, we still let P_Ω denote the projection onto the span of the canonical basis index by Ω (see Definition 2.1), but we now let it be an element in either $\mathcal{B}(\ell^2(\mathbb{N}))$ or $\mathcal{B}(\ell^2(\mathbb{N}), \mathbb{C}^{|\Omega|})$. Note that if $P_\Omega \in \mathcal{B}(\ell^2(\mathbb{N}))$, then $P_\Omega = P_\Omega^2 = P_\Omega^*$, and we simply write P_Ω. If, however, $P_\Omega \in \mathcal{B}(\ell^2(\mathbb{N}), \mathbb{C}^{|\Omega|})$, then $P_\Omega \neq P_\Omega^*$, since $P_\Omega^*\colon \mathbb{C}^{|\Omega|} \to \ell^2(\mathbb{N})$, however, with slight abuse of notation we still write P_Ω rather than P_Ω^*. Also, recall that if $\Omega = \{1,\dots,M\}$, we write P_M, rather than $P_{\{1,\dots,M\}}$, and if $\Omega = \{M+1,\dots,K\}$, we write P_K^M as before. Furthermore, recall that $P_M^\perp := I - P_M$, where I is the identity operator on $\mathbb{C}^N$ or $\ell^2(\mathbb{N})$, depending on the context.

Remark 3.1 *(Remark about projection adjoints and dimensions).* Throughout the document, we do not specify the dimensions of the projection operators, as the same operator can have different dimensions depending on the context. Furthermore, (as a consequence) we do not use the adjoint of a projection operator. For example, for an operator $U \in \mathcal{B}(\ell^2(\mathbb{N}))$, we write $P_\Omega U P_M$, rather than $P_\Omega U P_M^*$, and we will treat $P_\Omega U P_M$ either as an element in $\mathbb{C}^{|\Omega|\times M}$ or $\mathcal{B}(\ell^2(\mathbb{N}))$, depending on the context. This has the advantage (see many of the theorems) that we can apply $P_\Omega U P_M$ both to a sequence $x \in \ell^2(\mathbb{N})$ and a vector $z \in \mathbb{C}^M$. Finally, we use the same notation for finite dimensional matrices. That is, if $M < K$ and $A \in \mathbb{C}^{m\times K}$ is a finite dimensional matrix, we write AP_M, rather than AP_M^*, as in the infinite-dimensional case, and AP_M can be either a $m \times M$ or a $m \times K$ matrix, depending on the context.

We still assume that the signal f is an element of $L^2([0,1))$. As in the finite dimensional case, we call a vector $x \in \ell^2(\mathbb{N})$ $(\boldsymbol{s}, \mathbf{M})$-sparse if $P_M x$ is $(\boldsymbol{s}, \mathbf{M})$-sparse and $P_M^\perp x = 0$. Here $M = M_r$ and we refer to it as the *sparsity bandwidth* of x. For an isometry $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ we define the coherence of U as $\mu(U) = \sup\{|U_{ij}|^2 : i, j \in \mathbb{N}\}$.

3.2. Setup

Next we describe the setup for a general sampling basis $B_{\text{sa}} = \{b_1^{\text{sa}}, b_2^{\text{sa}}, b_3^{\text{sa}}, \ldots, \}$ and a sparsifying basis $B_{\text{sp}} = \{b_1^{\text{sp}}, b_2^{\text{sp}}, b_3^{\text{sp}}, \ldots, \}$, both assumed to be orthonormal bases of $L^2([0,1))$. In Section 4, we specialize this so that B_{sa} is the Walsh sampling basis and B_{sp} is a wavelet sparsifying basis. This will enable us to derive concrete recovery guarantees for f. The setup below is, however, completely general.

For the two bases B_{sa} and B_{sp} we can represent f using the coefficients $y = \{\langle f, b_n^{\text{sa}}\rangle\}_{n\in\mathbb{N}}$ and $x = \{\langle f, b_n^{\text{sp}}\rangle\}_{n\in\mathbb{N}}$, respectively. To change the representation from B_{sa} to B_{sp} we define the following matrix.

Definition 3.2. Let $B_{\text{sa}} = \{b_1^{\text{sa}}, b_2^{\text{sa}}, b_3^{\text{sa}}, \ldots, \}$ and $B_{\text{sp}} = \{b_1^{\text{sp}}, b_2^{\text{sp}}, b_3^{\text{sp}}, \ldots, \}$ be orthonormal bases for $L^2([0,1))$. The *change of basis matrix* $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ between B_{sa} and B_{sp} is the infinite matrix with entries

$$U_{ij} = \left\langle b_j^{\text{sp}}, b_i^{\text{sa}} \right\rangle$$

We will denote this matrix by $U = [B_{\text{sa}}, B_{\text{sp}}]$.

Notice in particular that since B_{sa} and B_{sp} are orthonormal, $U = [B_{\text{sa}}, B_{\text{sp}}]$ is an isometry on $\ell^2(\mathbb{N})$ and we can write $y = Ux$.

Next let $\Omega = \Omega_{\boldsymbol{m},\mathbf{N}} = \Omega_1 \cup \ldots \cup \Omega_r$ be a given multilevel random sampling scheme with $|\Omega| = m$. We refer to $N = N_r$ as the *sampling bandwidth* of Ω (as discussed in Section 3.3, this will be chosen in terms of the sparsity bandwidth to ensure stable truncation of U). Now define the matrix

$$H := \begin{bmatrix} 1/\sqrt{p_1} P_{\Omega_1} U \\ 1/\sqrt{p_2} P_{\Omega_2} U \\ \cdots \\ 1/\sqrt{p_r} P_{\Omega_r} U \end{bmatrix} \in \mathbb{C}^{m\times\infty}, \quad \text{where} \quad p_k = m_k/(N_k - N_{k-1}) \tag{3.1}$$

and we use the slightly unusual notation $\mathbb{C}^{m\times\infty}$ for the operators $\mathcal{B}(\ell^2(\mathbb{N}), \mathbb{C}^m)$. Due to the scaling factors $1/\sqrt{p_k}$ we consider scaled noisy measurements

$$\tilde{y} = DP_\Omega y + e = DP_\Omega U x + e = Hx + e \in \mathbb{C}^m, \tag{3.2}$$

where $D \in \mathbb{C}^{m\times m}$ is a diagonal matrix with the corresponding scaling factors $1/\sqrt{p_k}$ found in H along the diagonal and e is the measurement noise.

Suppose that x is approximately $(\boldsymbol{s}, \mathbf{M})$-sparse with sparsity bandwidth M. It is tempting to form the finite matrix $A = HP_M \in \mathbb{C}^{m\times M}$ and solve the minimization problem

$$\text{minimize} \, \|z\|_1 \quad \text{subject to} \quad \|Az - \tilde{y}\|_2 \leq \eta.$$

However, note that the truncation of H to A introduces an additional truncation error $HP_M^{\perp}x$. Indeed,

$$Ax - \tilde{y} = -(HP_M^{\perp}x + e),$$

and this poses a problem since for the above decoder we require $\eta \geq \|HP_M^{\perp}x + e\|_2$ in order for $P_M x$ to be a feasible point. For some applications we might have a rough estimate of $\|e\|_2$, but any estimate of $\|HP_M^{\perp}x\|_2$ would require a priori knowledge of x, the signal we are trying to recover. This is generally impossible. (We note in passing that there is some recent work [31] which derives CS recovery guarantees in the absence of feasibility of the target vector $P_M x$, but the application of this work to the sparse in levels model is not clear).

To overcome this issue, we will introduce a *data fidelity parameter* $K \geq M$ and assume we know $\|e\|_2$ so that we can let $\eta > \|e\|_2$. Then there will always exist a $K' \geq M$ such that $P_K x$ lies in the feasible set $\{z \in \mathbb{C}^K : \|Az - \tilde{y}\|_2 \leq \eta\}$ corresponding to the augmented matrix

$$A = HP_K \tag{3.3}$$

for all $K \geq K'$. In practice (for the general case) it will also be impossible to determine a sufficient value for K, but for fixed $\eta > \|e\|_2$ there will always exist a K, such that $P_K x$ is a feasible point. It should, however, be noted that there are special cases, such as Walsh sampling and wavelet recovery, where sufficient values for K are known; see Remark 4.9.

This aside, as previously mentioned, we also now modify the optimization problem to include weights. Specifically, let $\mathbf{M}, \boldsymbol{s} \in \mathbb{N}^r$ be given sparsity levels and local sparsities respectively. For positive weights $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{r+1})$ we define

$$\|x\|_{1,\boldsymbol{\omega}} := \sum_{l=1}^{r+1} \omega_l \|P_{M_l}^{M_{l-1}} x\|_1,$$

with $M_{r+1} = K$ for $x \in \mathbb{C}^K$. Notice that this weighted regularizer assigns constant weights on each sparsity level. With this in hand, our recovery procedure is

$$\text{minimize} \, \|z\|_{1,\boldsymbol{\omega}} \quad \text{subject to} \quad \|Az - \tilde{y}\|_2 \leq \eta,$$

with A as in (3.3) and $\eta \geq \|Ax - \tilde{y}\|_2$.

3.3. The balancing property

We now discuss the relation between the sampling and sparsity bandwidths N and M. From generalized sampling theory [30] we know that we must choose $N \geq M$ to obtain a stable mapping between the first N sampling basis functions and the first M sparsity basis functions. The degree of stability for this solution will depend of the so-called *balancing property*:

Definition 3.3. Let $U \colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be an isometry. Let $0 < \theta < 1$ and $N \geq M \geq 1$. Then U has the *balancing property* with constant θ if

$$\|P_M U^* P_N U P_M - P_M\|_2 \leq 1 - \theta.$$

Note that for general orthonormal bases, the balancing property many not hold for small values of N, even if $N \geq M$. However, it always holds for sufficiently large N (for fixed M) [30, Sec. 5]. Indeed, $P_M U^* P_N U P_M \to P_M U^* U P_M \equiv P_M$ in the operator norm, hence the balancing property holds with θ arbitrarily close to 1 for large enough N.

Below we shall see that this property will also affect our recovery guarantees, but it will be camouflaged as the quantity $\|G^{-1}\|_2$, where $G = \sqrt{P_M U^* P_N U P_M}$. This gives the following relation.

Lemma 3.4. *Let $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ be an isometry. Suppose for some fixed pair $M, N \in \mathbb{N}$ that U satisfies the balancing property with constant $0 < \theta < 1$. Then the matrix $G = \sqrt{P_M U^* P_N U P_M}$ is Hermitian and positive definite. Furthermore, G is invertible and*

$$\|G^{-1}\|_2 \leq 1/\sqrt{\theta}. \tag{3.4}$$

3.4. G-adjusted Restricted Isometry Property in Levels (G-RIPL)

Our theoretical analysis requires a RIP-type property for the matrix HP_M. However, as implied in the previous discussion, the finite matrix $P_N U P_M \in \mathbb{C}^{N \times M}$ (from which HP_M is constructed) is not an isometry for any $N \geq M$. In particular, unlike in finite dimensions $\mathbb{E}(P_M H^* H P_M) = P_M U^* P_N U P_M = G^2$ is not the identity. In order to handle this situation, we introduce the following generalization of the RIP:

Definition 3.5 *(G-RIPL).* Let $A \in \mathbb{C}^{m \times M}$, $G \in \mathbb{C}^{M \times M}$ be invertible, $\mathbf{M} = (M_1, \dots, M_r)$ be sparsity levels and $\mathbf{s} = (s_1, \dots, s_r)$ be local sparsities. The $\mathbf{s}^{\text{th}}$ *G-adjusted Restricted Isometry Constant in Levels (G-RICL)* $\delta_{\mathbf{s},\mathbf{M}}$ is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|Gx\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|Gx\|_2^2, \quad \forall x \in \Sigma_{\mathbf{s},\mathbf{M}}.$$

If $0 < \delta_{\mathbf{s},\mathbf{M}} < 1$ we say that the matrix A satisfies the *G-adjusted Restricted Isometry Property in Levels (G-RIPL)* of order $(\mathbf{s}, \mathbf{M})$.

The G-RIPL is of course completely general and can be stated for any G. However, in the following we let $G = \sqrt{P_M U^* P_N U P_M}$ and show that the matrix $A = HP_K$ (or equivalently, HP_M – note that $\Sigma_{\boldsymbol{s},\mathbf{M}}$ consists of vectors z with $P_M^{\perp} z = 0$) satisfies the G-RIPL for this particular G.

First, however, we show in Theorem 3.6 below that the G-RIPL implies uniform recovery. For this, we introduce the following notation:

$$S_{\boldsymbol{\omega},\boldsymbol{s}} := \sum_{l=1}^{r} \omega_l^2 s_l \quad \text{and} \quad \zeta_{\boldsymbol{s},\boldsymbol{\omega}} = \min_{l \in \{1,\dots,r\}} \omega_l^2 s_l.$$

Notice in particular that for the choice $\boldsymbol{\omega} = (1, \dots, 1, \omega_{r+1})$ we have $S_{\boldsymbol{\omega},\boldsymbol{s}} = s_1 + \ldots + s_r$ and for the choice $\boldsymbol{\omega} = (s_1^{-1/2}, \dots, s_r^{-1/2}, \omega_{r+1})$ we have $S_{\boldsymbol{\omega},\boldsymbol{s}} = r$. Finally, we let $\kappa(G) = \|G\|_2 \|G^{-1}\|_2$ denote the condition number of G.

Theorem 3.6 *(The G-RIPL implies uniform recovery). Let $A \in \mathbb{C}^{m \times K}$, $G \in \mathbb{C}^{M \times M}$ with $K \geq M$ and let $\mathbf{M}, \boldsymbol{s} \in \mathbb{N}^r$ be given sparsity levels and local sparsities, respectively. Let $\boldsymbol{\omega} \in \mathbb{R}^{r+1}$ be positive weights. Suppose $AP_M \in \mathbb{C}^{m \times M}$ satisfies the G-RIPL of order $(\boldsymbol{t}, \mathbf{M})$ with constant $\delta_{\boldsymbol{t},\mathbf{M}} \leq 1/2$ and*

$$t_l = \min\left\{ M_l - M_{l-1},\ 2\left\lceil \frac{4\kappa(G)^2 S_{\boldsymbol{\omega},\boldsymbol{s}}}{\omega_l^2} \right\rceil \right\} \quad \textit{for } l = 1, \dots, r. \tag{3.5}$$

Let

$$\omega_{r+1} \geq \sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}} \left(\frac{1}{3} \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)^{-1} + 2\sqrt{2}\|AP_K^M\|_{1\to 2}\|G^{-1}\|_2 \right). \tag{3.6}$$

Let $\eta \geq 0$, $x \in \mathbb{C}^K$, $e \in \mathbb{C}^m$ with $\|e\|_2 \leq \eta$ and set $y = Ax + e$. Then any solution $\hat{x}$ of the optimization problem

$$\underset{z\in\mathbb{C}^K}{\text{minimize}} \, \|z\|_{1,\boldsymbol{\omega}} \quad \textit{subject to} \quad \|Az - y\|_2 \leq \eta \tag{3.7}$$

satisfies

$$\|x - \hat{x}\|_{1,\boldsymbol{\omega}} \leq C\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + D\|G^{-1}\|_2\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\eta \tag{3.8}$$

$$\|x - \hat{x}\|_2 \leq (1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}) \left(C\frac{\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + D\|G^{-1}\|_2\eta \right) \tag{3.9}$$

where $C = 2(2+\sqrt{3})/(2-\sqrt{3})$, $D = 8\sqrt{2}/(2-\sqrt{3})$ and $\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} = \inf\{\|x - z\|_{1,\boldsymbol{\omega}} : z \in \Sigma_{\boldsymbol{s},\mathbf{M}}\}$.

Notice that the condition on δ in the above theorem is fundamentally different from the condition found in Theorem 2.10. In the latter one requires $\delta_{2\boldsymbol{s},\mathbf{M}} < (r(\sqrt{\alpha_{\boldsymbol{s},\mathbf{M}}} + \frac{1}{4})^2 + 1)^{-1/2}$ where $\alpha_{\boldsymbol{s},\mathbf{M}} = \max_{k,l=1,\dots,r} s_k/s_l$ is the sparsity ratio. Thus for sparsity levels where the local sparsities vary greatly, this bound will be unreasonably small.

In the above theorem we have removed this sparsity ratio term, by setting $\delta = 1/2$, and require $\delta_{\boldsymbol{t},\mathbf{M}} \leq \delta$ where $t_l \geq 2\left\lceil 4\kappa(G)^2 S_{\boldsymbol{\omega},\boldsymbol{s}} w_l^{-2}\right\rceil$. For the unweighted case this leads to a condition of the form

$$t_l \geq 2\left\lceil 4\kappa(G)^2(s_1 + \ldots + s_r)\right\rceil,$$

which does not take the local sparsity into account, since each t_l would have to be greater than the total sparsity of the signal. However, by considering the weights $\boldsymbol{\omega} = (s_1^{-1/2}, \dots, s_r^{-1/2}, \omega_{r+1})$ we obtain a condition of the form

$$t_l \geq 2\left\lceil 4\kappa(G)^2 r s_l\right\rceil,$$

where t_l is independent of s_k for $k \neq l$. Thus, for this choice of weights each t_l only depend on s_l, something which gives greater flexibility for signals where the sparsity ratio $\alpha_{\boldsymbol{s},\mathbf{M}}$ is large. Moreover, this also means that we can write the requirement in Theorem 3.6 as $\delta_{2\lceil 4\kappa(G)^2 r\boldsymbol{s}\rceil,\mathbf{M}} \leq 1/2$, and avoid the sparsity ratio term $\alpha_{\boldsymbol{s},\mathbf{M}}$ as was the problem in Theorem 2.10.

Finally, recall that ω_{r+1} is the last weight in the weighted ℓ^1-norm $\|\cdot\|_{1,\boldsymbol{\omega}}$, and notice that in the above theorem we have introduced a lower bound for ω_{r+1} in (3.6) which depends on AP_K^M. This is necessary, since the G-RIPL only put conditions on the first M columns of $A \in \mathbb{C}^{m\times K}$. Thus, to ensure that the minimizer (of length K) of (3.7) does not have a large component in the last $K - M$ entires it is needed to ensure that the weight ω_{r+1} satisfies the lower bound in (3.6).

3.5. Sufficient condition for the G-RIPL

In Definition 2.11 we defined the local coherence $\mu_{k,l}$ of an isometry $U \in \mathbb{C}^{N\times N}$. We extend this to isometries $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ in the exact same way

$$\mu_{k,l} = \mu_{k,l}(\mathbf{N},\mathbf{M}) = \max\{|U_{ij}|^2 : i = N_{k-1}+1,\dots,N_k, j = M_{l-1}+1,\dots,M_l\}.$$

This yields the following theorem providing sufficient conditions on the number of samples, required for $A = HP_M$ to have the G-RIPL.

Theorem 3.7 *(Subsampled isometries and the G-RIPL). Let $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ be an isometry, and let $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ be an $(\mathbf{N},\mathbf{m})$-multilevel random sampling scheme with r levels. Let $\mathbf{M}, \boldsymbol{s} \in \mathbb{N}^r$ be sparsity levels and local sparsities, respectively. Let $\epsilon, \delta \in (0,1)$ and let $0 \le r_0 \le r$, with $\tilde{m} = m_{r_0+1} + \cdots + m_r$. Let $s = s_1 + \cdots + s_r$ and*

$$L = r \cdot \log(2\tilde{m}) \cdot \log(2N) \cdot \log^2(2s) + \log(\epsilon^{-1}).$$

Suppose $G = \sqrt{P_M U^* P_N U P_M}$ is non-singular. If

$$m_k = N_k - N_{k-1}, \quad k = 1, \dots, r_0, \tag{3.10}$$

and

$$m_k \gtrsim \delta^{-2} \cdot \|G^{-1}\|_2^2 \cdot (N_k - N_{k-1}) \cdot \left(\sum_{l=1}^{r} \mu_{k,l} \cdot s_l \right) \cdot L, \tag{3.11}$$

for $k = r_0 + 1, \dots, r$ then with probability at least $1 - \epsilon$, the matrix

$$A = \begin{bmatrix} 1/\sqrt{p_1} P_{\Omega_1} U P_M \\ \vdots \\ 1/\sqrt{p_r} P_{\Omega_r} U P_M \end{bmatrix} \quad \text{where } p_k = \frac{m_k}{N_k - N_{k-1}} \quad \text{for } k = 1, \dots, r \tag{3.12}$$

satisfies the G-RIPL of order $(\mathbf{s}, \mathbf{M})$ with constant $\delta_{\mathbf{s},\mathbf{M}} \le \delta$. The constant implied by $\gtrsim$ is universal, and does not depend on any of the parameters.

We notice that these measurement conditions are identical to the once we find in Theorem 2.12, except the extra term $\|G^{-1}\|_2$. However, Theorem 2.12, concerns the RILP, which is a special case of the G-RIPL with $G = I$. It is, therefore, natural that this term disappears.

3.6. Overall recovery guarantee

Combining Theorem 3.6 and Theorem 3.7, we now find measurement conditions which ensure recovery of $(\boldsymbol{s}, \mathbf{M})$-sparse signals.

Corollary 3.8 *(Overall recovery guarantee for subsampled isometries). Let $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ be an isometry, and let $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ be an $(\mathbf{N},\mathbf{m})$-multilevel random sampling scheme with r levels. Let $\mathbf{M}, \boldsymbol{s} \in \mathbb{N}^r$ be sparsity levels and local sparsities, respectively, and let $\boldsymbol{\omega} = [s_1^{-1/2}, \dots, s_r^{-1/2}, \omega_{r+1}]$ be weights. Let $\epsilon \in (0,1)$ and $0 \le r_0 \le r$. Let $m = m_1 + \ldots + m_r$, $\tilde{m} = m_{r_0+1} + \cdots + m_r$, $s = s_1 + \cdots + s_r$, and*

$$L = r \cdot \log(2\tilde{m}) \cdot \log(2N) \cdot \log^2(2s) + \log(\epsilon^{-1}).$$

Let $H \in \mathbb{C}^{m \times \infty}$ be as in (3.1) and set $A = HP_K$. Let $x \in \ell^2(\mathbb{N})$, $e_1 \in \mathbb{C}^m$ and $\eta > 0$. Set $e = HP_K^{\perp} x + e_1$ and $\tilde{y} = Ax + e$. Suppose

(i) we choose M and N so that U satisfies the balancing property with constant $0 < \theta < 1$,
(ii) we choose $\eta \ge \|e_1\|$ and K so that $\|HP_K^{\perp} x\|_2 \le \eta'$,
(iii) the weight ω_{r+1} satisfies

$$\omega_{r+1} \ge \sqrt{r} \left(\frac{1}{3(1 + r^{1/4})} + 2\sqrt{\frac{2}{\theta}} \|AP_K^M\|_{1 \to 2} \right),$$

(iv) the m_k's satisfy $m_k = N_k - N_{k-1}$ for $k = 1, \dots, r_0$ and

$$m_k \gtrsim \theta^{-2} \cdot r \cdot (N_k - N_{k-1}) \cdot \left(\sum_{l=1}^{r} \mu_{k,l} s_l \right) \cdot L \quad \text{for } k = r_0 + 1, \dots, r. \tag{3.13}$$

Then with probability $1 - \epsilon$ any solution $\hat{x}$ of the optimization problem

$$\underset{z \in \mathbb{C}^K}{\text{minimize}} \, \|z\|_{1,\boldsymbol{\omega}} \quad \text{subject to} \quad \|Az - \tilde{y}\|_2 \le \eta + \eta'$$

satisfies

$$\|P_K x - \hat{x}\|_{1,\boldsymbol{\omega}} \le C \sigma_{\boldsymbol{s},\mathbf{M}}(P_K x)_{1,\boldsymbol{\omega}} + D \frac{\sqrt{r}}{\sqrt{\theta}} (\eta + \eta') \tag{3.14}$$

$$\|P_K x - \hat{x}\|_2 \le (1 + r^{1/4}) \left(C \frac{\sigma_{\boldsymbol{s},\mathbf{M}}(P_K x)_{1,\boldsymbol{\omega}}}{\sqrt{r}} + D \frac{1}{\sqrt{\theta}} (\eta + \eta') \right) \tag{3.15}$$

where $C = 2(2 + \sqrt{3})/(2 - \sqrt{3})$ and $D = 8\sqrt{2}/(2 - \sqrt{3})$.

Suppose that x is exactly $(\boldsymbol{s}, \mathbf{M})$-sparse. Then the above theorem guarantees exact recovery of x via weighted ℓ^1 minimization subject to the corresponding measurement condition. We note in passing that this measurement condition is the same, up to the log terms, as sufficient conditions derived recently in [21] for oracle estimators that assume *a priori* knowledge of $\text{supp}(x)$. While it is unknown whether or not these are strictly sharp, this does at lead indicate that the proposed weighted ℓ^1-minimization problem, in particular, with the specific weights used above, is an appropriate approach.

4. Recovery guarantees for Walsh sampling with wavelet reconstruction

Having presented the abstract infinite-dimensional CS framework in full generality, the remainder of the paper is devoted to its application to the case of binary sampling with the Walsh transform with sparsity in orthogonal wavelet bases. In particular, this means that the isometry U is now known, and all quantities depending on U will be estimated to derive concrete measurement conditions for $(\boldsymbol{s}, \mathbf{M})$-sparse wavelet recovery using Walsh sampling. Our goal is to derive theorems similar to Theorem 2.12 for different wavelets and combine these theorems with Corollary 3.8 to get overall recovery guarantees.

We start this section by describing the general setup of Walsh functions and wavelets in Sections 4.1 and 4.2, before presenting the main recovery guarantees in Sections 4.3 and 4.4.

4.1. Walsh functions

For any number $n \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ there exists a unique dyadic expansion

$$n = n_1 2^0 + n_2 2^1 + \ldots + n_j 2^{j-1} + \cdots$$

where $n_j \in \{0, 1\}$ for $j \in \mathbb{N}$. Similarly any $x \in [0, 1)$ can be written in its dyadic form as

$$x = x_1 2^{-1} + x_2 2^{-2} + \cdots + x_j 2^{-j} +$$

with $x_j \in \{0, 1\}$ for all $j \in \mathbb{N}$. For a dyadic rational number x this expansion is not unique, as one may use either a finite expansion, or an infinite expansion where $x_i = 1$ for all $i \ge k$ for some $k \in \mathbb{N}$. In such cases we always consider the finite expansion. In practice this means that we have removed countably many singletons from $[0, 1)$.

Definition 4.1. Let $n \in \mathbb{Z}_+$ and $x \in [0,1)$. The *Walsh function* $w_n \colon [0,1) \to \{+1,-1\}$ is given by

$$w_n(x) := (-1)^{\sum_{j=1}^{\infty}(n_j+n_{j+1})x_j} \tag{4.1}$$

We note that in the literature there are different definitions of Walsh functions, and the above definition is sometimes called the *sequency-ordered* Walsh function. It has the advantage that on the interval $[0,1)$, w_n has n sign changes. Another popular choice is the *Paley-ordered* Walsh function defined by $w_n^P := (-1)^{\sum_{j=1}^{\infty} n_j x_j}$. The Paley-ordered Walsh function of sometimes more practical to work with for establishing theoretical properties of Walsh functions. However, it has the disadvantage that its frequency, which we take to be the number of sign changes, is not equal to n. In the following we therefore work with the sequency-ordered Walsh function. For an in-depth discussion of the different orderings of Walsh functions we refer to [32, Chap. 1].

We note that the 2^r first Walsh functions gives rise to the entries in the Hadamard matrix

$$(V_{\text{Had}})_{i,j} = w_{i-1}((j-1)/2^r) \quad \text{where} \quad i,j = 1,\dots,2^r.$$

Definition 4.2 *(Walsh basis).* Define the *Walsh basis* as

$$B_{\text{wh}} := \{w_n : n \in \mathbb{Z}_+\}$$

where “wh” is an abbreviation for Walsh-Hadamard.

We note that this is an orthonormal basis of $L^2([0,1))$, see, e.g., [33, Chap. 2.6].

4.2. Wavelet transform

Let $\phi\colon \mathbb{R} \to \mathbb{R}$ and $\psi\colon \mathbb{R} \to \mathbb{R}$ be an orthonormal scaling function and wavelet [34], respectively, with minimal support, corresponding to a multiresolution analysis (MRA). Note that this could both be the classical “Daubechies wavelet” with a minimum-phase or “symlets” which are close to being symmetric, but with a larger phase [35, Page 294]. Let

$$\phi_{j,k}(x) := 2^{j/2}\phi(2^j x - k) \quad \text{and} \quad \psi_{j,k}(x) := 2^{j/2}\psi(2^j x - k) \tag{4.2}$$

denote the scaled and translated versions.

A wavelet ψ is said to have ν vanishing moments if

$$\int_{-\infty}^{\infty} x^k \psi(x)\,\mathrm{d}x = 0 \qquad \text{for } 0 \le k < \nu.$$

For orthogonal wavelets with minimum support, the support depends on the number of vanishing moments. That is

$$\operatorname{supp}(\phi) = \operatorname{supp}(\psi) = [-\nu+1, \nu]. \tag{4.3}$$

While this system constitutes an orthonormal basis of $L^2(\mathbb{R})$, in our case we require an orthonormal basis of $L^2([0,1))$. There exist several constructions of wavelets on the interval, but we only consider periodic extensions and the orthogonal *boundary wavelets* introduced by Cohen, Daubechies and Vial in [36], which preserves the number of vanishing moments.

For wavelets on the interval we need to replace the 2ν wavelets/scaling functions intersecting the boundaries at each scale, with their corresponding boundary-corrected counterparts. We postpone the formal definition of periodic and boundary wavelets until we need it, in the proof sections. But to simplify the notation let

$$\phi^0_{j,k} := \begin{cases} \phi^{\text{boundary}}_{j,k} & \text{for } k \in \{0,\dots,\nu-1\} \\ \phi_{j,k} & \text{for } k \in \{\nu,\dots,2^j-\nu-1\} \\ \phi^{\text{boundary}}_{j,k} & \text{for } k \in \{2^j-\nu,\dots,2^j-1\} \end{cases},$$

$$\phi^1_{j,k} := \begin{cases} \psi^{\text{boundary}}_{j,k} & \text{for } k \in \{0,\dots,\nu-1\} \\ \psi_{j,k} & \text{for } k \in \{\nu,\dots,2^j-\nu-1\} \\ \psi^{\text{boundary}}_{j,k} & \text{for } k \in \{2^j-\nu,\dots,2^j-1\} \end{cases},$$

where $\phi^{\text{boundary}}_{j,k}$ and $\psi^{\text{boundary}}_{j,k}$ are either a periodic wavelet/scaling function or the boundary wavelet/scaling functions introduced in [36]. For the former extension we say that $\phi^s_{j,k}$, $s \in \{0,1\}$ "originate from a *periodic wavelet*" while for the latter we say that it "originate from a *boundary wavelet*".

We will throughout assume that $J_0 \in \mathbb{Z}_+$ satisfies $2^{J_0} \geq 2\nu$ in case $\nu \geq 2$, while $J_0 \geq 0$ for $\nu = 1$. This will ensure that for each $j \geq J_0$ there exists at least one $k \in \{0,\dots,2^j-1\}$ such that $\text{supp}(\phi_{j,k}) = \text{supp}(\psi_{j,k}) \subseteq [0,1)$ for all $j \geq J_0$.

Definition 4.3. For a fixed number of vanishing moments ν, minimum wavelet decomposition J_0 and a boundary extension which is either periodic or boundary wavelets, let $\phi^s_{j,k}$ be the corresponding wavelets and scaling functions. We define

$$B^{J_0,\nu}_{\text{wave}} = \left\{\phi^0_{J_0,0},\dots,\phi^0_{J_0,2^{J_0}-1},\phi^1_{J_0,0},\dots,\phi^1_{J_0,2^{J_0}-1},\phi^1_{J_0+1,0},\dots,\phi^1_{J_0+1,2^{J_0+1}-1},\dots\right\}$$

Both B_{wh} and $B^{J_0,\nu}_{\text{wave}}$ are orthonormal bases for $L^2([0,1))$. See, e.g., [35, Thm. 7.6] for the periodic extension and [36] for the vanishing moments preserving boundary extension.

4.3. Recovery guarantees

From Section 3 there are four unknown factors depending on U which need to be estimated. These are the local coherences $\mu_{k,l}$, the norm $\|HP^M_K\|_{1\to 2}$ where H is given by (3.1), the condition number $\kappa(G) = \|G\|_2\|G^{-1}\|_2$ and the factor $\|G^{-1}\|_2$ found in condition (3.11).

For the two latter factors we have $G = \sqrt{P_M U^* P_N U P_M}$. Furthermore we know that $\|G\|_2 \leq 1$ since U is an isometry. In practice we therefore only need to determine an upper bound $\|G^{-1}\|_2$ and from Lemma 3.4 we know that $\|G^{-1}\|_2 \leq 1/\sqrt{\theta}$, where $0 < \theta < 1$ is the balancing property constant. In other words, it suffices to determine when the balancing property holds with a given θ.

The following three propositions estimate these quantities for the case $U = [B_{\text{wh}}, B^{J_0,\nu}_{\text{wave}}]$.

Proposition 4.4. *Let $U = [B_{\text{wh}}, B^{J_0,\nu}_{\text{wave}}]$. For each $\theta \in (0,1)$, there exists a constant $q = q_\theta \geq 0$, such that U satisfies*

$$\|P_{2^k} U^* P_{2^{k+q_\theta}} U P_{2^k} - P_{2^k}\|_2 \leq 1-\theta \quad \textit{for all } k \in \mathbb{N}.$$

That is, U satisfies the balancing property with constant θ, for each pair $\{N = 2^{k+q_\theta},\ M = 2^k\}_{k\in\mathbb{N}}$.

Note that Proposition 4.4 is a consequence of Theorem 1.1 in [37], see Section 6.6 for details.

Proposition 4.5. *Let* $U = [B_{\mathrm{wh}}, B_{\mathrm{wave}}^{J_0,\nu}]$ *with* $\nu \geq 3$ *and let*

$$\mathbf{M} = [2^{J_0+1}, \dots, 2^{J_0+r}] \text{ and } \mathbf{N} = [2^{J_0+1}, \dots, 2^{J_0+r-1}, 2^{J_0+r+q}] \text{ with } q \geq 0,$$

be sparsity and sampling levels, respectively. Then the local coherences of U *scale like*

$$\mu_{k,l} \lesssim 2^{-J_0-k} 2^{-|l-k|}.$$

Proposition 4.6. *Let* $U = [B_{\mathrm{wh}}, B_{\mathrm{wave}}^{J_0,\nu}]$ *and let* $\mathbf{M}, \mathbf{N} \in \mathbb{N}^r$ *be sparsity and sampling levels. Let* $\Omega = \Omega_{\boldsymbol{m},\mathbf{N}}$ *be a multilevel random sampling scheme, and let* H *be as in* (3.1)*. Then*

$$\|HP_K^{\perp}\|_{1\to 2} \lesssim \sqrt{\frac{N}{K}}.$$

We can now present the two main theorems in this section. We point out that these are only valid for $\nu \geq 3$ vanishing moments. For $\nu = 1$, the corresponding wavelet is the Haar wavelet, and will be considered in the next subsection. For $\nu = 2$, the coherence of $U = [B_{\mathrm{wh}}, B_{\mathrm{wave}}^{J_0,2}]$ does not decay as fast as for the other wavelets. Whether this is because our coherence bounds are not sharp enough for this wavelet or if it is because the coherence of $U = [B_{\mathrm{wh}}, B_{\mathrm{wave}}^{J_0,2}]$ decays more slowly is not known. We do, however, present some numerics in Section 6.5 which indicate that it is potentially the latter.

Theorem 4.7. *Let* $U = [B_{\mathrm{wh}}, B_{\mathrm{wave}}^{J_0,\nu}]$ *with* $\nu \geq 3$ *and let*

$$\mathbf{M} = [2^{J_0+1}, \dots, 2^{J_0+r}] \text{ and } \mathbf{N} = [2^{J_0+1}, \dots, 2^{J_0+r-1}, 2^{J_0+r+q}] \text{ with } q \geq 0,$$

be sparsity and sampling levels, respectively. Let $\boldsymbol{s} \in \mathbb{N}^r$ *be local sparsities. Suppose* q *is chosen so that* U *satisfies the balancing property with constant* $0 < \theta < 1$ *and set* $G = \sqrt{P_M U^* P_N U P_M}$*. Let* $\epsilon, \delta \in (0,1)$ *and let* $0 \leq r_0 \leq r$*, with* $\tilde{m} = m_{r_0+1} + \cdots + m_r$*. Let* $s = s_1 + \cdots + s_r$ *and*

$$L = r \cdot \log(2\tilde{m}) \cdot \log(2N) \cdot \log^2(2s) + \log(\epsilon^{-1}).$$

If

$$m_k = N_k - N_{k-1}, \quad k = 1, \dots, r_0, \tag{4.4}$$

and

$$m_k \gtrsim \delta^{-2} \cdot \theta^{-1} \cdot 2^{q \max\{k+1-r,0\}} \cdot \left(\sum_{l=1}^{r} 2^{-|k-l|} s_l \right) \cdot L$$

for $k = r_0 + 1, \dots, r$*, then with probability at least* $1 - \epsilon$*, the matrix in* (3.12) *satisfies the G-RIPL of order* $(\mathbf{s}, \mathbf{M})$ *with constant* $\delta_{\mathbf{s},\mathbf{M}} \leq \delta$*.*

We notice the similarity between this theorem and Theorem 2.12, which considers the finite-dimensional Fourier-Haar wavelet problem. In particular, the same type of local measurement conditions is required to ensure a RIPL-like condition for the matrix A. The main difference is that since we now consider a finite section of an isometry $U \in \mathcal{B}(\ell^2(\mathbb{N}))$, we get the extra term θ^{-1} from the balancing property between N and M, and the extra factor $2^q = N/M$ in the r'th sampling level. Since $N = M$ and $\theta = 1$ in the finite-dimensional setup, this is natural.

We note that while the constant implied by $\gtrsim$ in Theorem 3.7 is just a numerical constant independent of all parameters, the constant implied by $\gtrsim$ in the above theorem depends on the wavelet. In particular, Theorem 4.7 follows from Theorem 3.7 by using an upper bound $\gtrsim \mu_{k,l}$ on the local coherences (see Theorem 6.8). The constant implied by $\gtrsim$ can, therefore, vary if we change the wavelet basis. The same comment also holds for Theorem 4.8 below.

Next, we present an overall recovery guarantee for Walsh sampling and orthonormal wavelet reconstruction with $\nu \geq 3$ vanishing moments in infinite dimensions. Notice that the measurement conditions are identical to the once above, except for the extra factor r, introduced by our particular choice of weights. As discussed at the end of Section 3.4, this factor is inevitable if we want to get measurement conditions independent of the total sparsity $s = s_1 + \ldots s_r$.

Theorem 4.8. *Let $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,\nu}]$ with $\nu \geq 3$ and let*

$$\mathbf{M} = [2^{J_0+1}, \ldots, 2^{J_0+r}] \text{ and } \mathbf{N} = [2^{J_0+1}, \ldots, 2^{J_0+r-1}, 2^{J_0+r+q}], \text{ with } q \geq 0$$

be sparsity and sampling levels, respectively. Let $\boldsymbol{s} \in \mathbb{N}^r$ be local sparsities, $\boldsymbol{\omega} = (s_1^{-1/2}, \ldots, s_r^{-1/2}, \omega_{r+1})$ be weights and let $\boldsymbol{m} \in \mathbb{N}^r$ be sampling densities. Let $\epsilon \in (0,1)$ and let $0 \leq r_0 \leq r$. Let $m = m_1 + \ldots + m_r$, $\tilde{m} = m_{r_0+1} + \cdots + m_r$, $s = s_1 + \ldots + s_r$, and

$$L = r \cdot \log(2\tilde{m}) \cdot \log(2N) \cdot \log^2(2s) + \log(\epsilon^{-1}).$$

Let $H \in \mathbb{C}^{m \times \infty}$ be as in (3.1) and set $A = HP_K$. Let $x \in \ell^2(\mathbb{N})$, $e_1 \in \mathbb{C}^m$ and $\eta > 0$. Set $e = HP_K^{\perp}x + e_1$ and $\tilde{y} = Ax + e$. Suppose

(i) we choose $q = q_\theta$ as in Proposition 4.4 so that U satisfies the balancing property with constant $0 < \theta < 1$,
(ii) we choose $\eta \geq \|e_1\|$ and K so that $\|HP_K^{\perp}x\|_2 \leq \eta'$,
(iii) the weight ω_{r+1} satisfies

$$\omega_{r+1} \geq \sqrt{r}\left(\frac{1}{3(1+r^{1/4})} + 2\sqrt{\frac{2}{\theta}}\|AP_K^M\|_{1\to 2}\right),$$

(iv) the m_k's satisfy $m_k = N_k - N_{k-1}$ for $k = 1, \ldots, r_0$ and

$$m_k \gtrsim \theta^{-2} \cdot r \cdot 2^{q\max\{k+1-r,0\}}\left(\sum_{l=1}^{r} 2^{-|k-l|}s_l\right) \cdot L \quad \text{for } k = r_0+1, \ldots, r. \tag{4.5}$$

Then with probability $1-\epsilon$ any solution $\hat{x}$ of the optimization problem

$$\underset{z \in \mathbb{C}^K}{\text{minimize}} \|z\|_{1,\boldsymbol{\omega}} \quad \text{subject to} \quad \|Az - \tilde{y}\|_2 \leq \eta + \eta'$$

satisfies

$$\|P_K x - \hat{x}\|_{1,\boldsymbol{\omega}} \leq C\sigma_{\boldsymbol{s},\mathbf{M}}(P_K x)_{1,\boldsymbol{\omega}} + D\frac{\sqrt{r}}{\sqrt{\theta}}(\eta + \eta') \tag{4.6}$$

$$\|P_K x - \hat{x}\|_2 \leq (1 + r^{1/4})\left(C\frac{\sigma_{\boldsymbol{s},\mathbf{M}}(P_K x)_{1,\boldsymbol{\omega}}}{\sqrt{r}} + D\frac{1}{\sqrt{\theta}}(\eta + \eta')\right) \tag{4.7}$$

where $C = 2(2+\sqrt{3})/(2-\sqrt{3})$ and $D = 8\sqrt{2}/(2-\sqrt{3})$.

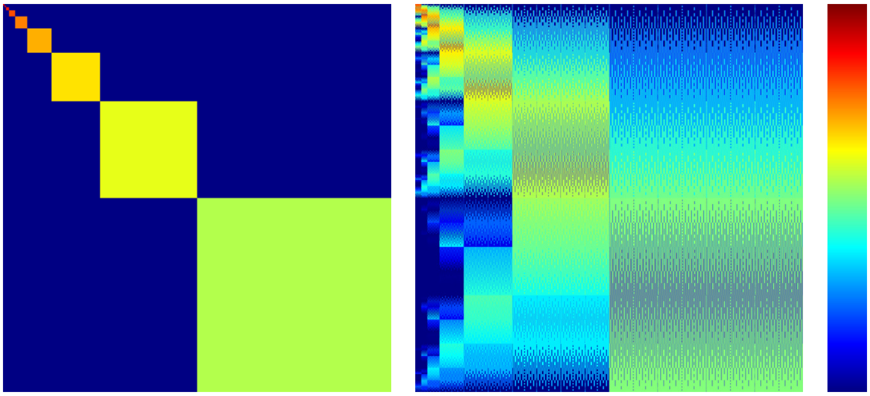

Fig. 2. The absolute values in log scale of the matrix $P_M U P_M$ for $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,\nu}]$, with $\nu = 1$ (left) and $\nu = 4$ (middle). The rightmost image is the colorbar. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Remark 4.9. Note that the second condition (ii) can be guaranteed using Proposition 4.6. Indeed, it suffices for K to satisfy

$$\frac{\|P_K^{\perp} x\|_1}{\sqrt{K}} \lesssim \frac{\eta'}{\sqrt{N}}.$$

Hence, given any *a priori* estimates on the decay of the coefficients x (such as in the case of wavelets), one can use this to determine a suitable K.

4.4. Uniform recovery for Haar wavelets

Below we shall see that for the Haar wavelet, $P_N U P_N$ will be an isometry for $N = 2^r$ where $r \in \mathbb{N}$. This can also be seen from Fig. 2, where $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,\nu}]$ is perfectly block diagonal for $\nu = 1$. This means that the G-RIPL, reduces to the I-adjusted RIPL, or simply the RIPL, which we know from the finite dimensional case. Notice in particular that we also avoid any considerations where $K > M = N$ as above, since $HP_M^{\perp} = 0$.

Proposition 4.10. *Let $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,1}]$ and let $N = 2^k$, for some $k \in \mathbb{N}$ with $k \geq J_0 + 1$. Then $P_N U P_N$ is an isometry on $\mathbb{C}^N$.*

Proposition 4.11. *Let $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,1}]$ and let $\mathbf{M} = \mathbf{N} = [2^{J_0+1}, \dots, 2^{J_0+r}]$ be sparsity and sampling levels, respectively. Then the local coherences of U are*

$$\mu_{kl} = \begin{cases} 2^{-J_0-k+1} & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases}$$

These propositions have several consequences. First, since $P_N U P_N$ is an isometry, we have that the balancing property constant $\theta = 1$, and since $M = N$, we have that $N/M = 1 = 2^0$, which implies that the q in Theorems 4.7 and 4.8 are zero. Furthermore, due to the block diagonal structure of $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,1}]$, we shall see below that the measurements in level k now only depend on the local sparsity s_k, rather than an exponentially decaying sum of local sparsities of the form $\sum_{l=1}^{r} 2^{-|k-l|} s_l$, which we are used to from the above theorems. This follows immediately from Proposition 4.11 where the local coherences $\mu_{kl} = 0$ for $k \neq l$.

Theorem 4.12. *Let $U = [B_{\mathrm{wh}}, B^{J_0,1}_{\mathrm{wave}}]$ and let $\mathbf{M} = \mathbf{N} = [2^{J_0+1}, \dots, 2^{J_0+r}]$ be sparsity and sampling levels. Let $s \in \mathbb{N}^r$ be local sparsities and $\boldsymbol{m} \in \mathbb{N}^r$ be local sampling densities. Let $\epsilon, \delta \in (0, 1)$ and $0 \le r_0 \le r$. Let $\tilde{m} = m_{r_0+1} + \ldots + m_r$ and $s = s_1 + \ldots + s_r$. Suppose that the m_k's satisfy $m_k = N_k - N_{k-1}$ for $k = 1, \dots, r_0$ and*

$$m_k \gtrsim \delta^{-2} s_k \left(r \log(2\tilde{m}) \log(2N) \log^2(2s) + \log(\epsilon^{-1}) \right), \qquad \text{for } k = r_0 + 1, \dots, r. \tag{4.8}$$

If $\Omega = \Omega_{\mathbf{N},\boldsymbol{m}}$ is an $(\mathbf{N}, \boldsymbol{m})$-multilevel random sampling scheme, then with probability at least $1-\epsilon$ the matrix (3.12) satisfies the RIPL of order $(\boldsymbol{s}, \mathbf{M})$ with constant $\delta_{\boldsymbol{s},\mathbf{M}} \le \delta$. The constant implied by $\gtrsim$ is universal, and does not depend on any of the parameters.

Proof. Using Proposition 4.10 we know that $P_N U P_N$ is an isometry. Thus inserting the local coherences from Proposition 4.11 into (2.5) in Theorem 2.12 gives the result. As the wavelet is fixed, it follows from Theorem 2.12 that the numerical constant implied by $\gtrsim$ is independent of all parameter choices. □

Finally, we present the overall recovery guarantee for Walsh sampling and Haar wavelet recovery. Notice that the block diagonal structure of $U = [B_{\mathrm{wh}}, B^{J_0,1}_{\mathrm{wave}}]$ allows us to remove many of the technical conditions of Theorem 4.8, yet the measurement conditions are identical to the one from Theorem 4.12 above, except for the extra r factor, introduced by our particular choice of weights.

Theorem 4.13. *Let $U = [B_{\mathrm{wh}}, B^{J_0,1}_{\mathrm{wave}}]$ and let $\mathbf{M} = \mathbf{N} = [2^{J_0+1}, \dots, 2^{J_0+r}]$ be sparsity and sampling levels. Let $\boldsymbol{s} \in \mathbb{N}^r$ be local sparsities, $\boldsymbol{\omega} = (s_1^{1/2}, \dots, s_r^{1/2})$ be weights and $\boldsymbol{m} \in \mathbb{N}^r$ be local sampling densities. Let $\epsilon \in (0, 1)$ and let $0 \le r_0 \le r$. Let $m = m_1 + \ldots + m_r$, $\tilde{m} = m_{r_0+1} + \dots + m_r$ and $s = s_1 + \ldots + s_r$. Suppose we sample $m_k = N_k - N_{k-1}$ for $k = 1, \dots, r_0$ and*

$$m_k \gtrsim r \cdot s_k \cdot \left(r \log(2\tilde{m}) \log(2N) \log^2(2s) + \log(\epsilon^{-1}) \right),$$

for $k = r_0 + 1, \dots, r$. Let $\Omega = \Omega_{\mathbf{N},\boldsymbol{m}}$ be an $(\mathbf{N}, \boldsymbol{m})$-multilevel random sampling scheme and let $H \in \mathbb{C}^{m\times\infty}$ be as in (3.1) with $A = HP_M$. Let $x \in \ell^2(\mathbb{N})$ and $e \in \mathbb{C}^m$ with $\|e\|_2 \le \eta$ for some $\eta \ge 0$. Set $\tilde{y} = Ax + e$. Then any solution $\hat{x}$ of the optimization problem

$$\underset{z\in\mathbb{C}^M}{\text{minimize}}\, \|z\|_{1,\boldsymbol{\omega}} \quad \text{subject to} \quad \|Az - \tilde{y}\|_2 \le \eta$$

satisfies

$$\|P_M x - \hat{x}\|_{1,\boldsymbol{\omega}} \le C\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}} + D\sqrt{r}\eta$$

$$\|P_M x - \hat{x}\|_2 \le (1 + r^{1/4}) \left(C \frac{\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}}}{\sqrt{r}} + D\eta \right)$$

with probability $1-\epsilon$, where $C = 2(2 + \sqrt{3})/(2 - \sqrt{3})$ and $D = 8\sqrt{2}/(2 - \sqrt{3})$. The constant implied by $\gtrsim$ is universal, and does not depend on any of the parameters.

Proof. Proposition 4.10 gives $G = \sqrt{P_M U^* P_N U P_M} = \sqrt{I} = I$. Next notice that $S_{\boldsymbol{\omega},\boldsymbol{s}} = r$ and that $P_M x \in \{z \in \mathbb{C}^M : \|Az - \tilde{y}\|_2 \le \eta\}$ since $\|HP_M^{\perp}\| = 0$. Using Theorem 3.6 we see that we can guarantee recovery of $(\boldsymbol{s}, \mathbf{M})$-sparse vectors, if A satisfies the RIPL with constant $\delta_{\boldsymbol{t},\mathbf{M}} \le 1/2$, where $t_l = \min\{M_l - M_{l-1}, 8rs_l\}$. Using Theorem 4.12 gives the result. □

5. Proofs of results in Section 3

When deriving uniform recovery guarantees via the RIP, it is typical to proceed as follows. First, one shows that the RIP implies the so-called *robust Null space Property (rNSP)* of order s (see Def. 4.17 in [4]). Second, one the shows that the rNSP implies stable and robust recovery. Thus the line of implications reads

$$\text{(RIP)} \implies \text{(rNSP)} \implies \text{(uniform recovery)}.$$

A similar line of implications holds for the RIPL and the corresponding *robust Null Space Property in levels (rNSPL)*; see Def. 3.6 in [16].

Both of the recovery guarantees for matrices satisfying the rNSP and rNSPL consider minimizers of the unweighed quadratically-constrained basis pursuit (QCBP) optimization problem. In our setup we consider minimizers of the weighted QCBP. We have therefore generalized the rNSPL to what we call the weighted robust null space property in levels.

For the sufficient condition for the G-RIPL in Theorem 3.7, the proof follows along similar lines as in [17]. We only sketch the main differences here.

5.1. The weighted rNSPL and norm bounds

For a set $\Theta \subseteq \{1, \ldots, M\}$ and a vector $x \in \mathbb{C}^M$ we let the vector x_Θ be given by

$$(x_\Theta)_i = \begin{cases} x_i & i \in \Theta \\ 0 & i \notin \Theta \end{cases}.$$

We also define

$$E_{\boldsymbol{s},\mathbf{M}} = \{\Theta \subseteq \{1, \ldots, M\} : |\Theta \cap \{M_{l-1}+1, \ldots, M_l\}| \leq s_l, \text{ for } l = 1, \ldots, r\}.$$

Definition 5.1 *(Weighted rNSP in levels).* Let $\mathbf{M}, \boldsymbol{s} \in \mathbb{N}^r$ be sparsity levels and local sparsities, respectively. For positive weights $\boldsymbol{\omega} \in \mathbb{R}^{r+1}$, we say that $A \in \mathbb{C}^{m \times M}$ satisfies the *weighted robust Null Space Property in Levels* (weighted rNSPL) of order $(\boldsymbol{s}, \mathbf{M})$ with constants $0 < \rho < 1$ and $\gamma > 0$ if

$$\|x_\Theta\|_2 \leq \frac{\rho \|x_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + \gamma \|Ax\|_2 \tag{5.1}$$

for all $x \in \mathbb{C}^M$ and all $\Theta \in E_{\boldsymbol{s},\mathbf{M}}$.

Lemma 5.2 *(Weighted rNSPL implies $\ell^{(1,\boldsymbol{\omega})}$-distance bound). Suppose that $A \in \mathbb{C}^{m \times M}$ satisfies the weighted rNSPL of order $(\boldsymbol{s}, \mathbf{M})$ with constants $0 < \rho < 1$ and $\gamma > 0$. Let $x, z \in \mathbb{C}^M$. Then*

$$\|z - x\|_{1,\boldsymbol{\omega}} \leq \frac{1+\rho}{1-\rho} \left(2\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}} - \|x\|_{1,\boldsymbol{\omega}}\right) + \frac{2\gamma}{1-\rho} \sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}} \|A(z-x)\|_2. \tag{5.2}$$

Proof. Let $v = z - x$ and $\Theta \in E_{\boldsymbol{s},\mathbf{M}}$ be such that $\|x_{\Theta^c}\|_{1,\boldsymbol{\omega}} = \sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}}$. Then

$$\begin{aligned} \|x\|_{1,\boldsymbol{\omega}} + \|v_{\Theta^c}\|_{1,\boldsymbol{\omega}} &\leq 2\|x_{\Theta^c}\|_{1,\boldsymbol{\omega}} + \|x_\Theta\|_{1,\boldsymbol{\omega}} + \|z_{\Theta^c}\|_{1,\boldsymbol{\omega}} \\ &= 2\|x_{\Theta^c}\|_{1,\boldsymbol{\omega}} + \|x_\Theta\|_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}} - \|z_\Theta\|_{1,\boldsymbol{\omega}} \\ &\leq 2\sigma_{\mathbf{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|v_\Theta\|_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}}, \end{aligned}$$

which implies that

$$\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}} \leq 2\sigma_{\mathbf{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}} - \|x\|_{1,\boldsymbol{\omega}} + \|v_\Theta\|_{1,\boldsymbol{\omega}}. \tag{5.3}$$

Now consider $\|v_\Theta\|_{1,\boldsymbol{\omega}}$. By the weighted rNSPL, we have

$$\|v_\Theta\|_{1,\boldsymbol{\omega}} \leq \sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|v_\Theta\|_2 \leq \rho\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}} + \sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\gamma\|Av\|_2.$$

Hence (5.3) gives

$$\|v_\Theta\|_{1,\boldsymbol{\omega}} \leq \rho\left(2\sigma_{\mathbf{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}} - \|x\|_{1,\boldsymbol{\omega}} + \|v_\Theta\|_{1,\boldsymbol{\omega}}\right) + \sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\gamma\|Av\|_2,$$

and after rearranging we get

$$\|v_\Theta\|_{1,\boldsymbol{\omega}} \leq \frac{\rho}{1-\rho}\left(2\sigma_{\mathbf{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}} - \|x\|_{1,\boldsymbol{\omega}}\right) + \frac{\gamma}{1-\rho}\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|Av\|_2.$$

Therefore, using this and (5.3) once more, we deduce that

$$\begin{aligned}
\|z-x\|_{1,\boldsymbol{\omega}} &= \|v_\Theta\|_{1,\boldsymbol{\omega}} + \|v_{\Theta^c}\|_{1,\boldsymbol{\omega}} \\
&\leq 2\|v_\Theta\|_{1,\boldsymbol{\omega}} + \left(2\sigma_{\mathbf{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}} - \|x\|_{1,\boldsymbol{\omega}}\right) \\
&\leq \frac{1+\rho}{1-\rho}\left(2\sigma_{\mathbf{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|z\|_{1,\boldsymbol{\omega}} - \|x\|_{1,\boldsymbol{\omega}}\right) + \frac{2\gamma}{1-\rho}\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|A(z-x)\|_2,
\end{aligned}$$

which gives the result. □

Lemma 5.3 *(Weighted rNSPL implies ℓ^2 distance bound). Suppose that $A \in \mathbb{C}^{m\times M}$ satisfies the weighted rNSPL of order $(\mathbf{s},\mathbf{M})$ with constants $0<\rho<1$ and $\gamma>0$. Let $x, z \in \mathbb{C}^M$. Then*

$$\|z-x\|_2 \leq \left(\rho + (1+\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\frac{\|z-x\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\gamma\|A(z-x)\|_2. \tag{5.4}$$

Proof. Let $v = z - x$ and $\Theta = \Theta_1 \cup \cdots \cup \Theta_r$, where $\Theta_l \subseteq \{M_{l-1}+1, \ldots, M_l\}$, $|\Theta_l| = s_l$ is the index set of the largest s_l coefficients of $P_{M_l}^{M_{l-1}} v$ in absolute value. Then

$$\|v_{\Theta_l}\|_2 = \sqrt{\sum_{i\in\Theta_l}|v_i|^2} \geq \sqrt{s_l}\min_{i\in\Theta_l}|v_i| \geq \sqrt{s_l}\max_{\substack{M_{l-1}<i\leq M_l \\ i\notin\Theta_l}}|v_i|, \quad l = 1,\ldots,r,$$

which gives

$$\begin{aligned}
\|v_{\Theta^c}\|_2^2 &= \sum_{l=1}^r \sum_{\substack{M_{l-1}<i\leq M_l \\ i\notin\Theta_l}} |v_i|^2 \leq \sum_{l=1}^r \max_{\substack{M_{l-1}<i\leq M_l \\ i\notin\Theta_l}}|v_i| \sum_{\substack{M_{l-1}<i\leq M_l \\ i\notin\Theta_l}}|v_i| \\
&\leq \sum_{l=1}^r \frac{\|v_{\Theta_l}\|_2}{\sqrt{s_l}} \sum_{\substack{M_{l-1}<i\leq M_l \\ i\notin\Theta_l}}|v_i| \leq \max_{l=1,\ldots,r}\left\{\frac{\|v_{\Theta_l}\|_2}{\omega_l\sqrt{s_l}}\right\}\sum_{l=1}^r \omega_l \sum_{\substack{M_{l-1}<i\leq M_l \\ i\notin\Theta_l}}|v_i| \\
&\leq \max_{l=1,\ldots,r}\left\{\frac{\|v_{\Theta_l}\|_2}{\omega_l\sqrt{s_l}}\right\}\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}}.
\end{aligned}$$

Since $\|v_{\Theta_l}\|_2 \leq \|v_\Theta\|_2$ we deduce that

$$\|v_{\Theta^c}\|_2 \leq \sqrt{\frac{\|v_\Theta\|_2\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\min_{l=1,\dots,r}\{\omega_l\sqrt{s_l}\}}} = \sqrt{\frac{\|v_\Theta\|_2\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\sqrt{\zeta_{\boldsymbol{s},\boldsymbol{\omega}}}}}.$$

Applying Young's inequality $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$, we obtain

$$\|v_{\Theta^c}\|_2 \leq \frac{(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}}{2}\frac{\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + \frac{(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}}{2}\|v_\Theta\|_2.$$

Hence

$$\|v\|_2 \leq \|v_\Theta\|_2 + \|v_{\Theta^c}\|_2 \leq \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\|v_\Theta\|_2 + \frac{(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}}{2}\frac{\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}}.$$

We now use the weighted rNSPL to get

$$\|v\|_2 \leq \left(\rho + (1+\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\frac{\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\gamma\|Av\|_2.$$

To complete the proof, we use the inequality $\|v_{\Theta^c}\|_{1,\boldsymbol{\omega}} \leq \|v\|_{1,\boldsymbol{\omega}}$. □

5.2. Weighted rNSPL implies uniform recovery

Theorem 5.4. *Let $\mathbf{M}, \boldsymbol{s} \in \mathbb{N}^r$ be sparsity levels and local sparsities, respectively, and let $\boldsymbol{\omega} \in \mathbb{R}^{r+1}$ be positive weights. Let $A \in \mathbb{C}^{m\times K}$ and suppose that AP_M satisfies the weighted rNSP in levels of order $(\boldsymbol{s}, \mathbf{M})$ with constants $\rho = \sqrt{3}/2$ and $\gamma > 0$. Let $x \in \mathbb{C}^K$, with $K > M$ and $e \in \mathbb{C}^m$ with $\|e\|_2 \leq \eta$. Set $y = Ax + e$. If*

$$\omega_{r+1} \geq \sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\left(\frac{1}{3(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4})} + 2\gamma\|AP_K^M\|_{1\to 2}\right) \tag{5.5}$$

then any solution $\hat{x}$ of the optimization problem

$$\operatorname*{minimize}_{z\in\mathbb{C}^K} \|z\|_{1,\boldsymbol{\omega}} \quad \textit{subject to} \quad \|Az - y\|_2 \leq \eta \tag{5.6}$$

satisfies

$$\begin{aligned}\|x - \hat{x}\|_{1,\boldsymbol{\omega}} \leq& C\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + D\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\eta \\ \|x - \hat{x}\|_2 \leq& \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)\left(C\frac{\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + D\gamma\eta\right),\end{aligned}$$

where $C = 2(2+\sqrt{3})/(2-\sqrt{3})$ and $D = 8/(2-\sqrt{3})$.

Proof. Recall that $\rho = \sqrt{3}/2$, and notice that this gives $C/2 = (1+\rho)/(1-\rho)$ and $D/2 = 2/(1-\rho)$. Next we consider the bound (5.5), and note that this bound implies

$$\omega_{r+1} \geq \gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|AP_K^M\|_{1\to 2}/\rho \tag{5.7}$$

$$1 + 2\rho \geq 1 + 2\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|AP_K^M\|_{1\to 2}/\omega_{r+1} \tag{5.8}$$

$$1+\rho \geq 1-\rho+2\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|AP_K^M\|_{1\to 2}/\omega_{r+1} \tag{5.9}$$

$$\frac{C}{2} \geq 1+\frac{D}{2}\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|AP_K^M\|_{1\to 2}/\omega_{r+1}. \tag{5.10}$$

We also note that (5.5) implies

$$\begin{aligned}\omega_{r+1} &\geq \left(\frac{1}{3(1+(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4})}+2\gamma\|AP_K^M\|_{1\to 2}\right)\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}} \\ &\geq \left(\frac{2}{C(1+(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4})}+\frac{D}{C}\gamma\|AP_K^M\|_{1\to 2}\right)\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\end{aligned}$$

which can be written as

$$(1+(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4})(C/2)\frac{1}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} \geq \Big((D/2)(1+(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4})\gamma\|AP_K^M\|_{1\to 2}+1\Big)/\omega_{r+1}. \tag{5.11}$$

Next set $v = x-\hat{x}$ and consider the $\ell^{(1,\boldsymbol{\omega})}$-bound. First notice that since AP_M satisfies the weighted rNSPL, Lemma 5.2 gives

$$\|P_M v\|_{1,\boldsymbol{\omega}} \leq (C/2)\left(2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}}+\|P_M\hat{x}\|_{1,\boldsymbol{\omega}}-\|P_M x\|_{1,\boldsymbol{\omega}}\right)+(D/2)\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|AP_M v\|_2. \tag{5.12}$$

Here the last term can be bounded by

$$\|AP_M v\|_2 \leq \|Av+y-y\|_2+\|AP_K^M v\|_2 \leq 2\eta+\frac{\|AP_K^M\|_{1\to 2}}{\omega_{r+1}}\|P_K^M v\|_{1,\boldsymbol{\omega}} \tag{5.13}$$

$$\leq 2\eta+\frac{\|AP_K^M\|_{1\to 2}}{\omega_{r+1}}\left(\|P_K^M x\|_{1,\boldsymbol{\omega}}+\|P_K^M\hat{x}\|_{1,\boldsymbol{\omega}}\right), \tag{5.14}$$

since both x and $\hat{x}$ are feasible. We also observe that

$$\begin{aligned}2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}}-\|P_M x\|_{1,\boldsymbol{\omega}}+\|P_K^M x\|_{1,\boldsymbol{\omega}} &= 2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}}+2\|P_K^M x\|_{1,\boldsymbol{\omega}}-\|x\|_{1,\boldsymbol{\omega}} \\ &= 2\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}}-\|x\|_{1,\boldsymbol{\omega}}\end{aligned} \tag{5.15}$$

Combining (5.12), (5.14), (5.10) and (5.15) gives

$$\begin{aligned}\|v\|_{1,\boldsymbol{\omega}} \leq& \|P_M v\|_{1,\boldsymbol{\omega}}+\|P_K^M x\|_{1,\boldsymbol{\omega}}+\|P_K^M\hat{x}\|_{1,\boldsymbol{\omega}} \\ \leq& (C/2)\big(2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}}+\|P_M\hat{x}\|_{1,\boldsymbol{\omega}}-\|P_M x\|_{1,\boldsymbol{\omega}}\big)+\|P_K^M x\|_{1,\boldsymbol{\omega}}+\|P_K^M\hat{x}\|_{1,\boldsymbol{\omega}} \\ &+(D/2)\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\|AP_M v\|_2 \\ \leq& (C/2)\big(2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}}+\|P_M\hat{x}\|_{1,\boldsymbol{\omega}}-\|P_M x\|_{1,\boldsymbol{\omega}}\big)+D\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\eta \\ &+\left(1+(D/2)\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\frac{\|AP_K^M\|_{1\to 2}}{\omega_{r+1}}\right)\left(\|P_K^M x\|_{1,\boldsymbol{\omega}}+\|P_K^M\hat{x}\|_{1,\boldsymbol{\omega}}\right) \\ \leq& (C/2)\big(2\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}}+\|\hat{x}\|_{1,\boldsymbol{\omega}}-\|x\|_{1,\boldsymbol{\omega}}\big)+D\gamma\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}\eta.\end{aligned}$$

Using that $\hat{x}$ is a minimizer of (5.6) gives the desired bound.

We now consider the ℓ^2-bound. First note that

$$\|v\|_2 \leq \|P_M v\|_2+\|P_K^M v\|_2 \leq \|P_M v\|_2+\frac{1}{\omega_{r+1}}\|P_K^M v\|_{1,\boldsymbol{\omega}}. \tag{5.16}$$

We shall also need

$$\begin{aligned}&(\rho + (1+\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2)\frac{2}{1-\rho} + (1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2)\\ =&(D/4)\big(2\rho + (1+\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4} + (1-\rho) + (1-\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\big)\\ =&(D/4)\big((1+\rho) + \tfrac{1}{2}(3+\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\big)\\ \leq&(D/2)\big(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\big).\end{aligned} \tag{5.17}$$

Again, since AP_M satisfies the weighted rNSPL we can apply Lemma 5.3, Lemma 5.2 and inequality (5.17) to obtain the bound

$$\begin{aligned}\|P_M v\|_2 \leq& \left(\rho + (1+\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\frac{\|P_M v\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\gamma\|AP_M v\|_2\\ \leq& \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)(C/2)\frac{2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}} + \|P_M\hat{x}\|_{1,\boldsymbol{\omega}} - \|P_M x\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}}\\ &+ \left(\rho + (1+\rho)(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\frac{2\gamma}{1-\rho}\|AP_M v\|_2\\ &+ \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}/2\right)\gamma\|AP_M v\|_2\\ \leq& \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)(C/2)\frac{2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}} + \|P_M\hat{x}\|_{1,\boldsymbol{\omega}} - \|P_M x\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}}\\ &+(D/2)\left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)\gamma\|AP_M v\|_2.\end{aligned} \tag{5.18}$$

Combining (5.16), (5.18), (5.14), (5.11) and (5.15) now gives

$$\begin{aligned}\|v\|_2 \leq& \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)(C/2)\frac{2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}} + \|P_M\hat{x}\|_{1,\boldsymbol{\omega}} - \|P_M x\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}}\\ &+ (D/2)\left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)\gamma\|AP_M v\|_2 + \frac{1}{\omega_{r+1}}\|P_K^M v\|_{1,\boldsymbol{\omega}}\\ \leq& \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)(C/2)\frac{2\sigma_{\boldsymbol{s},\mathbf{M}}(P_M x)_{1,\boldsymbol{\omega}} + \|P_M\hat{x}\|_{1,\boldsymbol{\omega}} - \|P_M x\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}}\\ &+ \left((D/2)\left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)\gamma\left\|AP_K^M\right\|_{1\to 2} + 1\right)\frac{\|P_K^M x\|_{1,\boldsymbol{\omega}} + \|P_K^M\hat{x}\|_{1,\boldsymbol{\omega}}}{\omega_{r+1}}\\ &+ \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)D\gamma\eta\\ \leq& \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)(C/2)\frac{2\sigma_{\boldsymbol{s},\mathbf{M}}(x)_{1,\boldsymbol{\omega}} + \|\hat{x}\|_{1,\boldsymbol{\omega}} - \|x\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}}\\ &+ \left(1 + (S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4}\right)D\gamma\eta\end{aligned}$$

Using that $\hat{x}$ is a minimizer of (5.6) completes the proof. □

5.3. G-RIPL implies weighted rNSPL

Theorem 5.5. *Let $A \in \mathbb{C}^{m\times M}$ and let $G \in \mathbb{C}^{M\times M}$ be invertible. Let $\mathbf{M} \in \mathbb{N}^r$ be sparsity levels, $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{N}^r$ be local sparsities and let $\boldsymbol{\omega} \in \mathbb{R}^r$ be positive weights. Let $0 < \rho <$ and suppose that A satisfies the G-RIPL of order $(\boldsymbol{t}, \mathbf{M})$ with constant $0 < \delta_{\boldsymbol{t},\mathbf{M}} < 1$, where*

$$t_l = \min\left\{M_l - M_{l-1}, 2\left\lceil\left(\frac{1+\delta_{\boldsymbol{t},\mathbf{M}}}{1-\delta_{\boldsymbol{t},\mathbf{M}}}\right)\frac{\kappa(G)^2}{\rho^2\omega_l^2}S_{\boldsymbol{\omega},\boldsymbol{s}}\right\rceil\right\}, \quad \textit{for } l = 1,\dots,r. \tag{5.19}$$

Then A satisfies the weighted rNSP in levels of order $(\boldsymbol{s},\mathbf{M})$ with constants $0 < \rho < 1$ and $\gamma = \|G^{-1}\|_2/\sqrt{1-\delta}$.

Proof. Let $x \in \mathbb{C}^M$ and let $\Theta = \Theta_1 \cup \cdots \cup \Theta_r$, where Θ_l is the set of the largest s_l indices of $P_{M_l}^{M_{l-1}}x$ in absolute value. If $t_l = M_l - M_{l-1}$, let $T_{l,0} = \{M_{l-1}+1,\dots,M_l\}$ and let $T_{l,k} = \emptyset$ for $k \geq 1$. For $t_l < M_l - M_{l-1}$ let $T_{l,0}$ be the index set of the largest $t_l/2$ values of $|P_{M_l}^{M_{l-1}}x|$, and let $T_{l,1}$ be the index set of the next $t_l/2$ largest values and so forth. In the case where there are less than $t_l/2$ values left at iteration k, we let $T_{l,k}$ be the remaining indices. Let $T_k = T_{1,k} \cup \cdots \cup T_{r,k}$ and let $T_{\{0,1\}} = T_0 \cup T_1$. Since $\Theta \subseteq T_{\{0,1\}}$ we have

$$\|x_\Theta\|_2^2 \leq \|x_{T_{\{0,1\}}}\|_2^2 \leq \|G^{-1}\|_2^2\|Gx_{T_{\{0,1\}}}\|_2^2 \leq \frac{\|G^{-1}\|_2^2}{1-\delta}\|Ax_{T_{\{0,1\}}}\|_2^2 \tag{5.20}$$

where $\delta = \delta_{\boldsymbol{t},\mathbf{M}}$. Note that

$$Ax_{T_{\{0,1\}}} = Ax - \sum_{k\geq 2} Ax_{T_k}.$$

Then using the triangle inequality, and that $x_{T_k} \in \Sigma_{\boldsymbol{t},\mathbf{M}}$, we get

$$\begin{aligned}\|Ax_{T_{\{0,1\}}}\|_2^2 &= \|Ax\|_2 + \sum_{k\geq 2}\|Ax_{T_k}\|_2 \\ &\leq \|Ax\|_2 + \sqrt{1+\delta}\sum_{k\geq 2}\|Gx_{T_k}\|_2 \\ &\leq \|Ax\|_2 + \sqrt{1+\delta}\|G\|_2\sum_{k\geq 2}\|x_{T_k}\|_2.\end{aligned}$$

Set $\Delta = \{l \in \{1,\dots,r\} : t_l < M_l - M_{l-1}\}$ and notice that $T_{l,k} = \emptyset$ for $l \in \{1,\dots,r\}\setminus\Delta$ and $k \geq 1$. Thus for $k \geq 2$ we get

$$\begin{aligned}\|x_{T_k}\|_2^2 &= \sum_{l\in\Delta}\|x_{T_{l,k}}\|_2^2 \leq \sum_{l\in\Delta}\frac{2\|x_{T_{l,k-1}}\|_1^2}{t_l} = \sum_{l\in\Delta}\frac{2\|x_{T_{l,k-1}}\|_1^2\omega_l^2}{t_l\omega_l^2} \\ &\leq \frac{\sum_{l\in\Delta}2\|x_{T_{l,k-1}}\|_{1,\boldsymbol{\omega}}^2}{\min_{l\in\Delta}\{\omega_l^2 t_l\}} \leq \frac{2\|x_{T_{k-1}}\|_{1,\boldsymbol{\omega}}^2}{\min_{l\in\Delta}\{\omega_l^2 t_l\}}.\end{aligned}$$

Therefore

$$\begin{aligned}\|Ax_{T_{\{0,1\}}}\|_2 &\leq \|Ax\|_2 + \frac{\sqrt{2(1+\delta)}\|G\|_2}{\sqrt{\min_{l\in\Delta}\{\omega_l^2 t_l\}}}\sum_{k\geq 2}\|x_{T_{k-1}}\|_{1,\boldsymbol{\omega}} \\ &\leq \|Ax\|_2 + \frac{\sqrt{2(1+\delta)}\|G\|_2}{\sqrt{\min_{l\in\Delta}\{\omega_l^2 t_l\}}}\|x_{T_0^c}\|_{1,\boldsymbol{\omega}} \\ &\leq \|Ax\|_2 + \frac{\sqrt{1+\delta}\|G\|_2}{\min_{l\in\Delta}\{\omega_l\sqrt{t_l/2}\}}\|x_{\Theta^c}\|_{1,\boldsymbol{\omega}}.\end{aligned}$$

Combining this with (5.20) results in

$$\begin{aligned}\|x_\Theta\|_2 \leq &\sqrt{\frac{1+\delta}{1-\delta}}\|G\|_2\|G^{-1}\|_2 \frac{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}}{\min_{l\in\Delta}\{\omega_l\sqrt{t_l/2}\}}\frac{\|x_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + \frac{\|G^{-1}\|_2}{\sqrt{1-\delta}}\|Ax\|_2 \\ \leq &\rho\frac{\|x_{\Theta^c}\|_{1,\boldsymbol{\omega}}}{\sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}} + \frac{\|G^{-1}\|_2}{\sqrt{1-\delta}}\|Ax\|_2.\end{aligned} \tag{5.21}$$

By construction of Θ, we have that $\|x_{\Theta'}\|_2 \leq \|x_\Theta\|_2$ and $\|x_{\Theta^c}\|_2 \leq \|x_{(\Theta')^c}\|_2$ for any $\Theta' \in E_{\boldsymbol{s},\mathbf{M}}$. Combining this with Equation (5.21) establishes the weighted rNSPL of order $(\boldsymbol{s},\mathbf{M})$ with $0 < \rho < 1$ and $\gamma = \|G^{-1}\|_2/\sqrt{1-\delta}$. □

5.4. Proof of Theorem 3.6

Proof of Theorem 3.6. First notice that for $0 < \delta \leq 1/2$ we have

$$\frac{1+\delta}{1-\delta} \leq 3.$$

Hence using Theorem 5.5 with $0 < \delta_{\boldsymbol{t},\mathbf{M}} \leq \delta \leq 1/2$ and $\rho = \sqrt{3}/2$ we see that Equation (5.19) simplifies to Equation (3.5). This implies that AP_M satisfies the weighted rNSPL of order $(\boldsymbol{s},\mathbf{M})$, with constants $\rho = \sqrt{3}/2$ and $\gamma = \sqrt{2}\|G^{-1}\|_2$. Now since

$$\omega_{r+1} \geq \sqrt{S_{\boldsymbol{\omega},\boldsymbol{s}}}(\tfrac{1}{3}(1+(S_{\boldsymbol{\omega},\boldsymbol{s}}/\zeta_{\boldsymbol{s},\boldsymbol{\omega}})^{1/4})^{-1} + 2\sqrt{2}\|AP_K^M\|_{1\to 2}\|G^{-1}\|_2)$$

we know from Theorem 5.4 that any solution $\hat{x}$ of (3.7) satisfies (3.8) and (3.9). □

5.5. Proof of Theorem 3.7

Proof of Theorem 3.7. We recall that $U \in \mathcal{B}(\ell^2)$ is an isometry and that

$$A = \begin{bmatrix} 1/\sqrt{p_1}P_{\Omega_1}UP_M \\ 1/\sqrt{p_2}P_{\Omega_2}UP_M \\ \vdots \\ 1/\sqrt{p_r}P_{\Omega_r}UP_M \end{bmatrix} \in \mathbb{C}^{m\times M}, \quad \text{where} \quad p_k = m_k/(N_k - N_{k-1}),$$

and $m = m_1 + \ldots + m_r$. Note that

$$\|Ax\|^2 - \|Gx\|^2 = \langle (A^*A - G^*G)x, x\rangle.$$

Now let

$$D_{\mathbf{s},\mathbf{M},G} = \left\{\eta \in \mathbb{C}^M : \|G\eta\|_2 \leq 1,\ |\operatorname{supp}(\eta) \cap \{M_{k-1}+1,\ldots,M_k\}| \leq s_k,\ k=1,\ldots,r\right\},$$

and define the following seminorm on $\mathbb{C}^{M\times M}$:

$$|||B|||_{\mathbf{s},\mathbf{M},G} := \sup_{z\in D_{\mathbf{s},\mathbf{M},G}} |\langle Bz, z\rangle|.$$

Then we see that

$$\delta_{\mathbf{s},\mathbf{M}} = |||A^*A - G^*G|||_{\mathbf{s},\mathbf{M}}. \tag{5.22}$$

Notice also that $p_k = 1$ and $\Omega_k = \{N_{k-1}+1, \dots, N_k\}$ for $k = 1, \dots, r_0$. Next notice that the matrix P_{Ω_k} can be written as

$$P_{\Omega_k} = \sum_{i=1}^{m_k} e_{t_{k,i}} e^*_{t_{k,i}},$$

where $\{e_i\}_{i=1}^{\infty}$ is the standard basis on $\ell^2(\mathbb{N})$. It now follows that

$$A^*A = \sum_{k=1}^{r} \frac{1}{p_k} P_M U^* P_{\Omega_k} U P_M = \sum_{k=1}^{r} \frac{1}{p_k} \sum_{i=1}^{m_k} P_M U^* e_{t_{k,i}} e^*_{t_{k,i}} U P_M \tag{5.23}$$

$$= P_M U^* P_{N_{r_0}} U P_M + \sum_{k=r_0+1}^{r} \sum_{i=1}^{m_k} X_{k,i} X^*_{k,i}, \tag{5.24}$$

where $X_{k,i}$ are random vectors given by $X_{k,i} = \frac{1}{\sqrt{p_k}} P_M U^* e_{t_{k,i}}$. Note that the $X_{k,i}$ are independent, and also that

$$\begin{aligned}
\mathbb{E}(A^*A) &= P_M U^* P_{N_{r_0}} U P_M + \sum_{k=r_0+1}^{r} \sum_{i=1}^{m_k} \mathbb{E}\left(X_{k,i} X^*_{k,i}\right) \\
&= P_M U^* P_{N_{r_0}} U P_M + \sum_{k=r_0+1}^{r} \frac{m_k}{p_k(N_k - N_{k-1})} \sum_{j=N_{k-1}+1}^{N_k} P_M U^* e_j e_j^* U P_M \\
&= P_M U^* P_{N_{r_0}} U P_M + P_M U^* P^{N_{r_0}}_{N_r} U P_M \\
&= P_M U^* P_N U P_M \qquad (5.25) \\
&= G^* G, \qquad (5.26)
\end{aligned}$$

where $G \in \mathbb{C}^{M \times M}$ is non-singular by assumption. Hence, combining this with (5.22) and (5.23), we see that

$$\delta_{\mathbf{s},\mathbf{M}} = \left|\left|\left| \sum_{k=r_0+1}^{r} \sum_{i=1}^{m_k} \left(X_{k,i} X^*_{k,i} - \mathbb{E}(X_{k,i} X^*_{k,i}) \right) \right|\right|\right|_{\mathbf{s},\mathbf{M}}. \tag{5.27}$$

Having detailed the setup, the remainder of the proof now follows along very similar lines to that of [17, Thm. 3.2]. Hence we only sketch the details.

The first step is to estimate $\mathbb{E}\left(\delta_{\mathbf{s},\mathbf{M}}\right)$. Using the standard techniques of symmetrization, Dudley's inequality, properties of covering numbers, and arguing as in [17, Sec. 4.2], we deduce that

$$\mathbb{E}\left(\delta_{\mathbf{s},\mathbf{M}}\right) \le D + D^2, \qquad D = C_1 \sqrt{\frac{rQ\|G^{-1}\|_2^2 \log(2\tilde{m}) \log(2M) \log^2(2s)}{m}}, \tag{5.28}$$

where $C_1 > 0$ is a universal constant, $\tilde{m} = \sum_{k=r_0+1}^{r} m_k$, and

$$Q = \max_{k=r_0+1,\dots,r} \sum_{l=1}^{r} \frac{\mu_{k,l} s_l}{p_k}. \tag{5.29}$$

In particular,

$$\mathbb{E}\left(\delta_{\mathbf{s},\mathbf{M}}\right) \le \delta/2,$$

provided

$$C_2 Q \left\|G^{-1}\right\|_2^2 \delta^{-2} r \log(2\tilde{m}) \log(2M) \log^2(2s) \leq 1, \tag{5.30}$$

where $C_2 > 0$ is a universal constant. Using this, Talagrand's theorem and using the fact that $\|P_N U P_M\|_2 \leq \|U\|_2 = 1$ (see [17, Sec. 4.3]) we deduce that

$$\mathbb{P}(\delta_{\mathbf{s},\mathbf{M}} \geq \delta) \leq \exp\left(-3\delta^2/(8(3+7\delta)Q\left\|G^{-1}\right\|_2^2)\right).$$

In particular,

$$\mathbb{P}(\delta_{\mathbf{s},\mathbf{M}} \geq \delta) \leq \epsilon,$$

provided

$$\frac{80}{3} Q \left\|G^{-1}\right\|_2^2 \delta^{-2} \log(\epsilon^{-1}) \leq 1.$$

Combining this with (5.29) and (5.30) now completes the proof. □

5.6. Proof of Corollary 3.8 and Lemma 3.4

Proof of Corollary 3.8. We must ensure that all the conditions are met to be able to apply Theorem 3.6 with $P_K x$.

First notice that for weights $\boldsymbol{\omega} = (s_1^{-1/2}, \ldots, s_r^{-1/2}, \omega_{r+1})$ we have $S_{\boldsymbol{\omega},\boldsymbol{s}} = r$ and $\zeta_{\boldsymbol{s},\boldsymbol{\omega}} = 1$. Next we note that condition (ii) implies that $P_K x$ is a feasible point since $\|HP_K x - \tilde{y}\|_2 \leq \|HP_K^{\perp} x\|_2 + \|e_1\|_2 \leq \eta + \eta'$.

Let $G = \sqrt{P_M U^* P_N U P_M}$. Combining condition (i) and Lemma 3.4 gives $\|G^{-1}\|_2 \leq 1/\sqrt{\theta}$ and since $\|G\|_2 \leq 1$ we also have $\kappa(G) = \|G\|_2 \|G^{-1}\|_2 \leq 1/\sqrt{\theta}$. Inserting the above equalities and inequalities into the weight condition for ω_{r+1} in Theorem 3.6 gives condition (iii).

Next we must ensure that AP_M satisfies the G-RIPL of order $(\boldsymbol{t}, \mathbf{M})$ with $\delta_{\boldsymbol{t},\mathbf{M}} \leq 1/2$ where

$$t_l = \min\left\{M_l - M_{l-1}, 2\left\lceil 4\theta^{-1} r s_l \right\rceil\right\}. \tag{5.31}$$

According to Theorem 3.7 this occurs with probability $1 - \epsilon$ if the m_k's satisfy condition (iv). The error bounds (3.14) and (3.15) now follows directly from Theorem 3.6. □

Proof of Lemma 3.4. First notice that the balancing property is equivalent to requiring

$$\sigma_M(P_N U P_M) \geq \sqrt{\theta} \tag{5.32}$$

where $\sigma_M(P_N U P_M)$ is the Mth largest singular value of $P_N U P_M$. Indeed, since U is an isometry, the matrix $P_M - P_M U^* P_N U P_M$ is nonnegative definite, and therefore

$$\|P_M U^* P_N U P_M - P_M\|_2 = \sup_{x \in \mathbb{C}^M, \|x\|_2 \leq 1} \langle (P_M - P_M U^* P_N U P_M) x, x \rangle \tag{5.33}$$

$$= \sup_{x \in \mathbb{C}^M, \|x\|_2 \leq 1} \left(\|P_M x\|_2 - \|P_N U P_M x\|_2\right) \tag{5.34}$$

$$= 1 - \inf_{x \in \mathbb{C}^M, \|x\|_2 = 1} \|P_N U P_M x\|_2 \tag{5.35}$$

This gives (5.32). Next let $G = \sqrt{P_M U^* P_N U P_M}$ and notice that $\sigma_M(G) = \sigma_M(P_N U P_M)$. This gives $\|G^{-1}\|_2 = 1/\sigma_M(G) \leq 1/\sqrt{\theta}$. □

6. Proofs of results in Section 4

In Section 4 we found concrete recovery guarantees for the Walsh sampling and wavelet reconstruction, using the theorems in Section 3. The key to deriving Walsh-wavelet recovery guarantees boils down to estimating the quantities $\mu_{k,l}$, $||HP_K^M||_{1\to 2}$ and $||G^{-1}||_2 \leq \frac{1}{\sqrt{\theta}}$. All of these quantities depend directly on $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,\nu}]$, and to control them we have to estimate how the entries of U changes for varying n, j, k and s. We therefore start this section by setting up notation for wavelets on the interval and stating some useful properties of Walsh functions. Then in Section 6.3 and 6.4 we will estimate $\mu_{k,l}$, followed by a discussion of the sharpness of this estimate for $\nu = 2$ in Section 6.5. We will then finish in Section 6.6 by estimating $||HP_K^M||_{1\to 2}$, show how θ scales for varying M and N, and prove Theorem 4.7 and 4.8.

6.1. Wavelets on the interval and regularity

In section 4.2 we introduced orthogonal wavelets on the real line, but we did not make any formal definitions of the wavelets we used at the boundaries of the interval $[0,1)$. Next we consider the two boundary extensions, *periodically* and *boundary wavelets*. To simplify the exposition we define the following sets

$$\Lambda_{\nu,j,\text{left}} \coloneqq \{0,\dots,\nu-1\}, \qquad \Lambda_{\nu,j,\text{mid}} \coloneqq \{\nu,\dots,2^j-\nu-1\},$$
$$\Lambda_{\nu,j,\text{right}} \coloneqq \{2^j-\nu,\dots,2^j-1\} \qquad \Lambda_j = \Lambda_{\nu,j,\text{left}} \cup \Lambda_{\nu,j,\text{mid}} \cup \Lambda_{\nu,j,\text{right}}$$

At each scale $j \geq J_0$, the periodic wavelet basis consists of the usual wavelets and scaling functions $\psi_{j,k}$, $\phi_{j,k}$ for $k \in \Lambda_{\nu,j,\text{mid}}$ and the periodic extended functions $\phi_{j,k}^{\text{per}}$ and $\psi_{j,k}^{\text{per}}$ for $k \in \Lambda_{\nu,j,\text{left}} \cup \Lambda_{\nu,j,\text{right}}$. These are defined as

$$\phi_{j,k}^{\text{per}} \coloneqq \phi_{j,k}|_{[0,1)} + \phi_{j,2^j+k}|_{[0,1)} \qquad \text{for } k \in \Lambda_{\nu,j,\text{left}} \tag{6.1}$$
$$\phi_{j,k}^{\text{per}} \coloneqq \phi_{j,k-2^j}|_{[0,1)} + \phi_{j,k}|_{[0,1)} \qquad \text{for } k \in \Lambda_{\nu,j,\text{right}} \tag{6.2}$$

and similarly for $\psi_{j,k}^{\text{per}}$. Strictly speaking we could have defined these periodic extensions only for $k = 0,\dots,\nu-2$ and $k = 2^j-\nu+1,\dots,2^j-1$, but to unify the notation for both boundary extensions we have chosen the former.

Next we have the boundary wavelet basis with ν vanishing moments. This wavelet basis consists of the same interior wavelets as the periodic basis, but with 2ν boundary scaling and wavelet functions.

$$\phi_k^{\text{left}}, \phi_k^{\text{right}}, \psi_k^{\text{left}}, \psi_k^{\text{right}}, \qquad \text{for } k = 0,\dots,\nu-1.$$

As for the interior functions we also define the scaled versions as

$$\begin{aligned} \phi_{j,k}^{\text{left}}(x) \coloneqq 2^{j/2}\phi_k^{\text{left}}(2^j x),\ \phi_{j,k}^{\text{right}}(x) \coloneqq 2^{j/2}\phi_k^{\text{right}}(2^j x), \\ \psi_{j,k}^{\text{left}}(x) \coloneqq 2^{j/2}\psi_k^{\text{left}}(2^j x),\ \psi_{j,k}^{\text{right}}(x) \coloneqq 2^{j/2}\psi_k^{\text{right}}(2^j x). \end{aligned} \tag{6.3}$$

The names ‘left’ and ‘right’ correspond to the support of these functions. That is

$$\operatorname{supp} \phi_{j,k}^{\text{left}} = \operatorname{supp} \psi_{j,k}^{\text{left}} = [0, 2^{-j}(\nu+k)]$$
$$\operatorname{supp} \phi_{j,k}^{\text{right}} = \operatorname{supp} \psi_{j,k}^{\text{right}} = [2^{-j}(2^j-\nu-k), 1]$$

for $k = 0,\dots,\nu-1$.

In the following we shall see that all of our results holds for both periodic and boundary wavelets, but their treatment in some of the proofs differs slightly. To make the treatment as unified as possible we make the following definition.

Table 1
The Lipschitz regularity of Daubechies wavelets with ν vanishing moments. See [34, Page 239].

ν	2	3	4
α	0.55	1.08	1.61

Definition 6.1. We say that $\phi^s_{j,k}$, $s \in \{0,1\}$ "originates from a periodic wavelet" if

$$\phi^0_{j,k} := \begin{cases} \phi^{\text{per}}_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{left}} \\ \phi_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{mid}} \\ \phi^{\text{per}}_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{right}} \end{cases}, \qquad \phi^1_{j,k} := \begin{cases} \psi^{\text{per}}_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{left}} \\ \psi_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{mid}} \\ \psi^{\text{per}}_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{right}} \end{cases}.$$

We say that $\phi^s_{j,k}$ "originates from a boundary wavelet" if

$$\phi^0_{j,k} := \begin{cases} \phi^{\text{left}}_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{left}} \\ \phi_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{mid}} \\ \phi^{\text{right}}_{j,2^j-1-k} & \text{for } k \in \Lambda_{\nu,j,\text{right}} \end{cases}, \qquad \phi^1_{j,k} := \begin{cases} \psi^{\text{left}}_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{left}} \\ \psi_{j,k} & \text{for } k \in \Lambda_{\nu,j,\text{mid}} \\ \psi^{\text{right}}_{j,2^j-1-k} & \text{for } k \in \Lambda_{\nu,j,\text{right}} \end{cases}.$$

With these functions defined now for both boundary extensions, the definition of $B^{J_0,\nu}_{\text{wave}}$ is also clear. Next we make a note on the regularity of these orthogonal wavelets.

Definition 6.2. Let $\alpha = k + \beta$, where $k \in \mathbb{Z}_+$ and $0 < \beta \leq 1$. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be uniformly Lipschitz α if f is k-times continuously differentiable and for which the k^{th} derivative $f^{(k)}$ is Hölder continuous with exponent β, i.e.,

$$|f^{(k)}(x) - f^{(k)}(y)| < C|x-y|^{\beta}, \quad \forall x, y \in \mathbb{R}$$

for some constant $C > 0$.

Generally we shall be interested in wavelets which are uniformly Lipschitz $\alpha \geq 1$. For $\nu = 1$ we have the discontinuous Haar wavelet which is not uniformly Lipschitz. For $\nu = 2, 3, 4$ we have the values of α found in Table 1. As can be seen from the table, we have $\alpha \geq 1$ for $\nu = 3$ and $\nu = 4$.

To show that $\alpha \geq 1$ also for $\nu = 5, 6, 7, \ldots$, we can use the following argument. Let $q > 0$ and recall that $\mathcal{F}f$ denotes the Fourier transform of $f \in L^2(\mathbb{R})$. A well-known fact is that if $|\mathcal{F}f(t)| \lesssim (1+|t|)^{1-q}$ for all $t \in \mathbb{R}$, then f is uniformly Lipschitz with constant α for any $\alpha < q$ [34, Page 216]. For orthonormal wavelets with minimal compact support, as considered in this paper, it can be shown (see [34, Eq. (7.1.23)]) that q is exactly equal to

$$q = \nu - 1 - \frac{\log|P_\nu(\frac{3}{4})|}{2\log 2} \quad \text{where} \quad P_\nu(t) = \sum_{n=0}^{\nu-1} \binom{\nu-1+n}{n} t^n.$$

Moreover, we have that $\nu^{-1/2}3^{\nu-1} \leq P_\nu(3/4) \leq 3^{\nu-1}$ (see [34, Page 226]), so that a lower bound for q is $q \geq \nu - 1 - (\nu-1)\log 3/(2\log 2) = (\nu-1)(1 - \log 3/(2\log 2)) \geq 0.205 \cdot (\nu - 1)$. This implies that $q > 1$ for $\nu \geq 6$. For $\nu = 5$ it can be shown by direct calculation of $P_5(3/4)$, that $q \approx 1.17$ (see also [34, Table at p. 226]).

There are several ways of estimating the regularity α of Daubechies wavelets and the values for α in Table 1 are estimated using a different argument. See [34, Chap. 7] for an in-depth treatment of this topic.

We end this paragraph by noticing that each of the boundary functions $\phi_k^{\text{left}}, \phi_k^{\text{right}}$ and $\psi_k^{\text{left}}, \phi_k^{\text{right}}$ are constructed as finite linear combinations of the interior scaling function ϕ and wavelet ψ. Thus all of these boundary functions have the same regularity as ϕ and ψ.

6.2. Properties of Walsh functions

Definition 6.3. For $x, y \in [0, 1)$, write $x = \sum_{i=1}^{\infty} x_i 2^{-i}$ and $y = \sum_{i=1}^{\infty} y_i 2^{-i}$ for $x_i, y_i \in \{0, 1\}$. We define

$$x \oplus y \coloneqq \sum_{i=1}^{\infty} (x_i \oplus y_i) 2^{-i}$$

where $x_i \oplus y_i \coloneqq |x_i - y_i|$.

Proposition 6.4. *For $j, m, n \in \mathbb{Z}_+$ and $x, y \in [0, 1)$, the Walsh function satisfies the following properties*

$$\int_0^1 w_n(x) w_m(x) \,\mathrm{d}x = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \tag{6.4}$$

$$w_n(x \oplus y) = w_n(x) w_n(y) \tag{6.5}$$

$$w_n(2^{-j} x) = w_{\lfloor n/2^j \rfloor}(x) \tag{6.6}$$

Proof. Equation (6.5) and (6.4) can be found in any standard text on Walsh functions e.g., [33], whereas the last follows by inserting j zeros in front of x's dyadic expansion. □

Next we prove a short lemma which will be useful later.

Lemma 6.5. *For $n \in \mathbb{N}$, let $p \geq 0$ be an integer such that $2^p \leq n < 2^{p+1}$. Denote the interval $\Delta_{k,r} = [2^{-r}k, 2^{-r}(k + 1))$ for $k \in \{0, \dots, 2^r - 1\}$, $r \in \mathbb{Z}_+$. Then w_n is constant on each of the intervals $\Delta_{k,p+1}$, $k \in \{0, \dots, 2^{p+1} - 1\}$, and for each pair of intervals $(\Delta_{2k,p+1}, \Delta_{2k+1,p+1})$ $k \in \{0, \dots, 2^p - 1\}$, w_n attains the value 1 on exactly one of them and -1 on the other. w_0 is constant equal to 1 on all of $[0, 1)$.*

Proof. Let $x, z \in [0, 1)$ and $n \in \mathbb{N}$ have binary representations $(x_i)_{i\in\mathbb{N}}$, $(z_i)_{i\in\mathbb{N}}$ and $(n_i)_{i\in\mathbb{N}}$, respectively, and recall that

$$w_n(x) = (-1)^{\sum_{j=1}^{\infty}(n_j + n_{j+1})x_j}. \tag{6.7}$$

Since $2^p \leq n < 2^{p+1}$, we know that $n_{p+1} = 1$ and $n_i = 0$ for $i > p + 1$. Let $k \in \{0, \dots, 2^p - 1\}$ and notice $\Delta_{k,p} = \Delta_{2k,p+1} \cup \Delta_{2k+1,p+1}$. We also notice that all $x \in \Delta_{k,p}$ have the same p first binary digits and that if $x \in \Delta_{2k,p+1}$, then $x_{p+1} = 0$, whereas for $x \in \Delta_{2k+1,p+1}$ we have $x_{p+1} = 1$. For $x, z \in \Delta_{k,p+1}$, $k \in \{0, \dots, 2^{p+1} - 1\}$, we then have $x_i = z_i$ for $i = 1, \dots, p + 1$, which implies that w_n is constant on $\Delta_{k,p+1}$ since $n_i = 0$ for $i > p + 1$. Furthermore, we know that the sum in (6.7), contains the term $(n_{p+1} + n_{p+2})x_{p+1} = (1 + 0)x_{p+1}$. Since $x_{p+1} = 0$ for $x \in \Delta_{2k,p+1}$ and $x_{p+1} = 1$ for $x \in \Delta_{2k+1,p+1}$, $k \in \{0, \dots, 2^p - 1\}$, it follows that w_n have different sign on the two intervals. That $w_0 \equiv 1$ follows immediately. □

6.3. Bounding the inner product $|\langle \phi_{j,k}^s, w_n \rangle|$

The entries in $U = [B_{\text{wh}}, B_{\text{wave}}^{J_0,\nu}]$, consist of $\langle \phi_{j,k}^s, w_n \rangle$ for different values of j, ks and n. Thus in order to determine the local coherences we need to find a upper bound of this inner product. Next we derive such a bound for $\nu \geq 2$ vanishing moments and discuss its sharpness. For $\nu = 1$ we determine the magnitude of each matrix entry explicitly.

Lemma 6.6. *Let $w_n \in B_{\mathrm{wh}}$ and let $\phi^s_{j,k} \in B^{J_0,\nu}_{\mathrm{wave}}$ for $\nu \geq 2$. For $j \geq J_0$, $s \in \{0,1\}$ and $k \in \Lambda_j$ we have*

$$\left|\langle \phi^s_{j,k}, w_n \rangle\right| \leq 2^{-j/2} 2\nu \max_{l \in \Gamma_k} \left\{ \left| \mathcal{W}\left[\phi^s(\cdot + l)\big|_{[0,1)}\right] \left(\left\lfloor \frac{n}{2^j} \right\rfloor \right) \right| \right\} \tag{6.8}$$

where

$$\Gamma_k = \begin{cases} \{0, \dots, \nu + k - 1\} & \textit{for } k \in \Lambda_{\nu,j,\mathrm{left}}; \\ \{-\nu + 1, \dots, \nu - 1\} & \textit{for } k \in \Lambda_{\nu,j,\mathrm{mid}}; \\ \{k - \nu + 1, \dots, 2^j - 1\} & \textit{for } k \in \Lambda_{\nu,j,\mathrm{right}}. \end{cases}$$

and

$$\phi^s = \begin{cases} \phi^{left}_k & \textit{if } k \in \Lambda_{\nu,j,\mathrm{left}} \\ \phi & \textit{if } k \in \Lambda_{\nu,j,\mathrm{mid}} \\ \phi^{right}_{2^j-1-k} & \textit{if } k \in \Lambda_{\nu,j,\mathrm{right}} \end{cases}, \ \textit{for } s = 0 \quad \textit{and} \quad \phi^s = \begin{cases} \psi^{left}_k & \textit{if } k \in \Lambda_{\nu,j,\mathrm{left}} \\ \psi & \textit{if } k \in \Lambda_{\nu,j,\mathrm{mid}} \\ \psi^{right}_{2^j-1-k} & \textit{if } k \in \Lambda_{\nu,j,\mathrm{right}} \end{cases}, \ \textit{for } s = 1$$

if $\phi^s_{j,k}$ originates from a boundary wavelet and

$$\Gamma_k = \{-\nu + 1, \dots, \nu - 1\}, \quad \phi^s = \phi \textit{ for } s = 0 \textit{ and } \phi^s = \psi \textit{ for } s = 1$$

if $\phi^s_{j,k}$ originates from a periodic wavelet.

Proof. First notice that for any $x \in [0,1)$ and $k \in \{0, 1, \dots, 2^j - 1\}$ we have

$$\begin{aligned} \frac{x}{2^j} + \frac{k}{2^j} &= \sum_{i=j}^{\infty} x_{i-j+1} 2^{-i-1} + \sum_{i=1}^{j} k_i 2^{-j-1+i} \\ &= \sum_{i=j}^{\infty} x_{i-j+1} 2^{-i-1} \oplus \sum_{i=1}^{j} k_i 2^{-j-1+i} = \frac{x}{2^j} \oplus \frac{k}{2^j}. \end{aligned} \tag{6.9}$$

Next, we consider the interior wavelets $\phi^s_{j,k}$, i.e., $k \in \Lambda_{\nu,j,\mathrm{mid}}$. We start by noticing that for $k \in \Lambda_{\nu,j,\mathrm{mid}}$, $\mathrm{supp}(\phi^s_{j,k}) = [2^{-j}(-\nu+1+k), 2^{-j}(\nu+k)]$ and $k+l \in \{0, \dots, 2^j-1\}$ for $k \in \Lambda_{\nu,j,\mathrm{mid}}$ and $l \in \{-\nu+1, \dots, \nu-1\}$. Combining this and Equations (6.9) and (6.5) we get

$$\begin{aligned} \langle \phi^s_{j,k}, w_n \rangle &= \int_0^1 \phi^s_{j,k}(x) w_n(x)\, \mathrm{d}x \\ &= \int_{2^{-j}(-\nu+1+k)}^{2^{-j}(\nu+k)} 2^{j/2} \phi^s(2^j x - k) w_n(x)\, \mathrm{d}x \\ &= 2^{-j/2} \int_{-\nu+1}^{\nu} \phi^s(x)\, w_n\left(\frac{x+k}{2^j}\right) \mathrm{d}x \end{aligned}$$

$$= 2^{-j/2} \sum_{l=-\nu+1}^{\nu-1} \int_0^1 \phi^s(x+l)\, w_n\left(\frac{x+l+k}{2^j}\right) \mathrm{d}x \tag{6.10}$$

$$= 2^{-j/2} \sum_{l=-\nu+1}^{\nu-1} \int_0^1 \phi^s(x+l)\, w_n\left(\frac{x}{2^j} \oplus \frac{l+k}{2^j}\right) \mathrm{d}x$$

$$= 2^{-j/2} \sum_{l=-\nu+1}^{\nu-1} w_n\left(\frac{l+k}{2^j}\right) \int_0^1 \phi^s(x+l)\, w_n\left(\frac{x}{2^j}\right) \mathrm{d}x$$

$$= 2^{-j/2} \sum_{l=-\nu+1}^{\nu-1} w_n\left(\frac{l+k}{2^j}\right) \mathcal{W}\left[\phi^s_{0,-l}\big|_{[0,1)}\right]\left(\left\lfloor \frac{n}{2^j} \right\rfloor\right)$$

Taking the absolute value and using the triangle inequality now gives

$$|\langle \phi^s_{j,k}, w_n \rangle| \leq 2^{-j/2} 2\nu \max_{l\in\Gamma_k} \left\{ \left| \mathcal{W}\left[\phi^s(\cdot+l)\big|_{[0,1)}\right]\left(\left\lfloor \frac{n}{2^j} \right\rfloor\right) \right| \right\}$$

The key in the above argument is to restrict the integral to the support of $\phi^s_{j,k}$, perform a change of variable and then apply Equation (6.9) to introduce the $\oplus$ operator. That $\operatorname{supp} \phi^s_{j,k} \subset [0,1]$ ensures that we can apply Equation (6.9) also for $k \in \Lambda_{\nu,j,\text{left}} \cup \Lambda_{\nu,j,\text{right}}$. The wavelets on the boundaries are therefore handled in the same manner. We note that the factor 2ν remains the same also for these wavelets since their support will be an interval or the union of two intervals, with total length at most 2ν in both cases. □

Lemma 6.7 *([38]). Let $f : [0,1) \to \mathbb{R}$ be uniformly Lipschitz $0 < \alpha \leq 1$. Then*

$$|\mathcal{W}f(n)| = \left| \int_0^1 f(x) w_n(x)\, \mathrm{d}x \right| \lesssim (n+1)^{-\alpha}$$

for $n \in \mathbb{Z}_+$.

Proof. This lemma follows from Lemma 6.2 in [38]. For completeness of this paper we redo the main steps. Suppose $n \neq 0$ and let $p \geq 0$ be an integer be such that $2^p \leq n < 2^{p+1}$. For convenience we denote the interval $\Delta_{k,p} := [2^{-p}k, 2^{-p}(k+1))$. Due to the Lipschitz regularity we know that there exits a constant $C > 0$ such that $f(x) \leq f(t) + C|t-x|^\alpha$ for all $x, t \in [0,1)$. Hence, for each $k \in \{0, \dots, 2^p - 1\}$ we have

$$\sup_{x\in\Delta_{k,p}} f(x) \leq f(2^{-p}k + 2^{-(p+1)}) + C2^{-(p+1)\alpha}$$

$$\sup_{x\in\Delta_{k,p}} -f(x) \leq -f(2^{-p}k + 2^{-(p+1)}) + C2^{-(p+1)\alpha}$$

Next notice that $\Delta_{k,p} = \Delta_{2k,p+1} \cup \Delta_{2k+1,p+1}$. From Lemma 6.5 we know that on each interval $\Delta_{k,p}$, w_n is constant equal to 1 on one of the subintervals $\Delta_{2k,p+1}, \Delta_{2k+1,p+1}$, and equal to -1 on the other. Hence

$$\left| \int_{\Delta_{k,p}} f(x) w_n \mathrm{d}x \right| \leq 2^{-p} \left| \left(f(2^{-p}k + 2^{-(p+1)}) C2^{-(p+1)\alpha} \right) + \left(-f(2^{-p}k + 2^{-(p+1)}) + C2^{-(p+1)\alpha} \right) \right|$$

$$\leq C2^{-p-p\alpha}.$$

Thus we get

$$|\mathcal{W}f(n)| \leq \sum_{k=0}^{2^p-1} \left| \int_{\Delta_{k,p}} f(x) w_n \mathrm{d}x \right| \leq C 2^{-p\alpha} \leq \frac{2C}{(n+1)^\alpha}$$

as desired. □

Theorem 6.8. *Let $\phi^s_{l,t} \in B^{J_0,\nu}_{\text{wave}}$ with $\nu \geq 3$ and let $w_n \in B_{\text{wh}}$. For $l \geq J_0$ and $2^k \leq n < 2^{k+1}$ with $k \in \mathbb{Z}_+$, we have*

$$|\langle \phi^s_{l,t}, w_n \rangle|^2 \lesssim 2^{-k} 2^{-|l-k|}$$

for all $t \in \Lambda_l$ and $s \in \{0,1\}$. For $n = 0$ the bound hold with $k = 0$.

Proof. Recall from the discussion at the end of Section 6.1, that for $\nu \geq 3$ the Lipschitz regularity α of ψ and ϕ is lower bounded by $\alpha \geq 1$. Moreover, the same hold for all boundary functions, since these are finite linear combinations of ψ and ϕ. Thus, it is clear that $\phi^s_{l,t}$ has regularity α if ψ and ϕ have it, regardless of boundary extension.

To obtain the bound above, we combine Lemma 6.6 and Lemma 6.7 with $\alpha = 1$. This gives

$$\left|\langle \phi^s_{l,t}, w_n \rangle\right|^2 \leq 2^{-l} 4\nu^2 \max_{a \in \Gamma_t} \left\{ \left| \mathcal{W}\left[\phi^s(\cdot + a)|_{[0,1)} \right] \left(\left\lfloor \frac{n}{2^l} \right\rfloor \right) \right|^2 \right\} \tag{6.11}$$

$$\lesssim 2^{-l} \frac{1}{(\lfloor \frac{n}{2^l} \rfloor + 1)^2} \leq 2^{-l} \frac{1}{(\lfloor 2^{k-l} \rfloor + 1)^2} \leq 2^{-k} 2^{-|l-k|} \tag{6.12}$$

where Γ_t is as in Lemma 6.6, and depends on the boundary extension. □

Theorem 6.9. *Let $w_n \in B_{\text{wh}}$ and let $\phi^s_{l,t} \in B^{J_0,1}_{\text{wave}}$ for $l \geq J_0$ and $t \in \Lambda_l$. Then*

$$|\langle \phi^0_{l,t}, w_n \rangle|^2 = \begin{cases} 2^{-l} & \text{if } n < 2^l \\ 0 & \text{otherwise} \end{cases}$$

$$|\langle \phi^1_{l,t}, w_n \rangle|^2 = \begin{cases} 2^{-l} & \text{if } 2^l \leq n < 2^{l+1} \\ 0 & \text{otherwise} \end{cases}.$$

Proof. These equalities can be found in either [38] or [39, Lemmas 1 and 2]. □

6.4. Proofs of Proposition 4.5, 4.10 and 4.11

Using the above results we are now able to determine the local coherences of $U = [B_{\text{wh}}, B^{J_0,\nu}_{\text{wave}}]$.

Proof of Proposition 4.5. We use the bound found in Theorem 6.8. Recall that $\mathbf{M} = [2^{J_0+1}, \ldots, 2^{J_0+r}]$ and $\mathbf{N} = [2^{J_0+1}, \ldots, 2^{J_0-1+r}, 2^{J_0+r+q}]$. For fixed $l \in \{1, \ldots, r\}$ and $k \in \{2, \ldots, r\}$ we have

$$\mu_{k,l} = \max \left\{ |\langle \phi^s_{J_0-1+l,t}, w_n \rangle|^2 : {}^{N_{k-1} \leq n < N_k}_{t \in \Lambda_{J_0-1+l}, s \in \{0,1\}, \text{ if } l=1}_{s=1 \text{ if } l>1} \right\},$$

$$\lesssim 2^{-(J_0-1+k)} 2^{-|(J_0-1+l)-(J_0-1+k)|} \lesssim 2^{-J_0-k} 2^{-|l-k|}.$$

For $l \in \{1, \ldots, r\}$ and $k = 1$ we have $N_0 = 0$. This gives

$$\begin{aligned}\mu_{k,l} &= \max\left\{ |\langle \phi^s_{J_0-1+l,t}, w_n\rangle|^2 : \substack{0\le n<N_1 \\ t\in\Lambda_{J_0-1+l}, s\in\{0,1\},\text{ if } l=1 \\ s=1 \text{ if } l>1} \right\}, \\ &\lesssim 2^{-(J_0-1+l)} \lesssim 2^{-J_0-k}2^{-|l-k|}. \quad \square\end{aligned}$$

Proof of Proposition 4.10. Since both $B^{J_0,1}_{\text{wave}}$ and B_{wh} are orthonormal, $U = [B_{\text{wh}}, B^{J_0,1}_{\text{wave}}]$ is an isometry on $\ell^2(\mathbb{N})$, i.e., $U^*U = I \in \mathcal{B}(\ell^2(\mathbb{N}))$. Let $N = 2^k$ for some $k \in \mathbb{N}$ with $k \ge J_0 + 1$. Using Theorem 6.9 we see that

$$P_N^{\perp} U P_N^* = \left\{ \langle \phi^s_{j,t}, w_n\rangle : \substack{n\ge 2^k \\ s=1, J_0\le j<k, t\in\Lambda_j, \text{ or} \\ s=0, j=J_0, t\in\Lambda_{J_0}} \right\} = 0.$$

Next, for clarity, let $P_N = P_N^{\mathbb{N}} \in \mathcal{B}(\ell^2(N))$ and interpret $P_N : \ell^2(\mathbb{N}) \to \mathbb{C}^N$. This gives

$$\begin{aligned}(P_N U P_N^*)^*(P_N U P_N^*) &= P_N U^* P_N^* P_N U P_N^* = P_N P_N^{\mathbb{N}} U^* P_N^{\mathbb{N}} U P_N^{\mathbb{N}} P_N^* \\ &= P_N P_N^{\mathbb{N}} P_N^* = I \in \mathbb{C}^{N\times N}. \quad \square\end{aligned}$$

Proof of Proposition 4.11. We use the bound found in Theorem 6.9. Recall that $\mathbf{M} = \mathbf{N} = [2^{J_0+1}, \dots, 2^{J_0+r}]$. For fixed $k, l \in \{1, \dots, r\}$ we have that

$$\begin{aligned}\mu_{k,l} &= \max\left\{ |\langle \phi^s_{J_0-1+l,t}, w_n\rangle|^2 : \substack{N_{k-1}\le n<N_k \\ t\in\Lambda_{J_0-1+l}, s\in\{0,1\},\text{ if } l=1, \\ s=1 \text{ if } l>1} \right\} \\ &= \begin{cases} 2^{-J_0-l+1} & \text{if } l = k \\ 0 & \text{otherwise} \end{cases}. \quad \square\end{aligned}$$

6.5. About the sharpness of the local coherence bounds

As can be seen from Proposition 4.11, the coherence bounds for $\nu = 1$ are sharp. However, for $\nu \ge 2$, we have not discussed their sharpness. In fact, none of the results in this paper consider the case for $\nu = 2$ vanishing moments. The reason for this is that these wavelets have a Lipschitz regularity $\alpha \approx 0.55$, which means that the bound in Theorem 6.8 would have less rapid decay if we had included these wavelets in the theorem. To simplify the presentation we have chosen to exclude them.

We will argue that Theorem 6.8 does not seem to extend to wavelets with $\nu = 2$ vanishing moments. Let $\mathbf{M} = \mathbf{N} = [2^{J_0+1}, \dots, 2^{J_0+r}]$ and $U = [B_{\text{wh}}, B^{J_0,\nu}_{\text{wave}}]$ for $\nu \ge 2$. Notice that setting $\nu = 2$ does only affect the local coherence estimates $\mu_{k,l}$ for $k \ge l$. For $k < l$, the local coherences are unaffected by the regularity of the wavelet. This follows from Lemma 6.6, by using that $|\mathcal{W}[\phi^s(\cdot + l)](0)| \approx 1$. Next consider the case where $k \ge l$, then Theorem 6.8 suggests that $\mu_{k,l}/\mu_{k+1} \approx 4$ for $\nu \ge 3$.

We now consider Table 2 and notice that for $\nu = 2$, all of the 18 entries in Table 2 have values less than 4. This suggest that the bound in Theorem 6.8 does not extend to the case of $\nu = 2$ vanishing moments. From the same table we also observe that for $\nu = 4$, the bound in Theorem 6.8 seem to be quite sharp. While there are a few entries that are less than 4, most are very close, if not larger than this value.

6.6. Proofs of remaining results in Section 4

Proof of Proposition 4.4. This proposition is a consequence of Theorem 1.1 in [37]. The mentioned theorem is only stated for the boundary wavelets preserving vanishing moments. We start this proof by showing how Theorem 1.1 in [37] implies the desired result. We then comment on how to extend Theorem 1.1 to periodic wavelets as well.

Table 2
Left: Fraction between the local coherences for $U = [B_{\text{wh}}, B^{3,2}_{\text{wave}}]$ and $\mathbf{M} = \mathbf{N} = [2^4, \dots, 2^{11}]$. Right: Fraction between the local coherences for $U = [B_{\text{wh}}, B^{4,4}_{\text{wave}}]$ and $\mathbf{M} = \mathbf{N} = [2^5, \dots, 2^{12}]$.

$\mu_{k,l}/\mu_{k+1,l}$	$l=1$	$l=2$	$l=3$
$k=2$	3.017		
$k=3$	2.532	1.854	
$k=4$	3.292	2.532	1.846
$k=5$	3.653	3.293	2.534
$k=6$	3.828	3.653	3.293
$k=7$	3.914	3.828	3.654
$k=8$	3.957	3.914	3.828

$\mu_{k,l}/\mu_{k+1,l}$	$l=1$	$l=2$	$l=3$
$k=2$	4.342		
$k=3$	6.160	3.439	
$k=4$	3.643	6.202	3.503
$k=5$	4.060	3.639	6.286
$k=6$	3.961	4.064	3.632
$k=7$	4.004	3.960	4.070
$k=8$	3.996	4.004	3.960

Let $\mathcal{S}_N = \operatorname{span}\{w_n : n = 0, \dots, N-1\}$ and $\mathcal{R}_M$ be the span of the M first functions in $B^{J_0,\nu}_{\text{wave}}$. The subspace cosine angle between $\mathcal{S}_N$ and $\mathcal{R}_M$ is defined as

$$\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) = \inf_{f \in \mathcal{R}_M, \|f\|=1} \|P_{\mathcal{S}_N} f\| \quad \text{where} \quad \omega(\mathcal{R}_M, \mathcal{S}_N) \in [0, \pi/2],$$

and $P_{\mathcal{S}_N}$ is the projection operator onto $\mathcal{S}_N$. As both B_{wh} and $B^{J_0,\nu}_{\text{wave}}$ are orthonormal bases, the synthesis and analysis operators are unitary. We therefore have

$$\inf_{f \in \mathcal{R}_M, \|f\|=1} \|P_{\mathcal{S}_N} f\| = \inf_{x \in \mathbb{C}^M, \|x\|_2=1} \|P_N U P_M x\|_2. \tag{6.13}$$

Furthermore notice that by Equation (5.35) and the definition of the balancing property, we have

$$\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) = \inf_{x \in \mathbb{C}^M, \|x\|_2=1} \|P_N U P_M x\|_2 \geq \theta, \tag{6.14}$$

if and only if U satisfies the balancing property with constant $\theta \in (0,1)$ for N and M. Note that (6.14) is equivalent to $1/\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) \leq 1/\theta$, where $1/\theta > 1$. Next for $M \in \mathbb{N}$ and $\gamma > 1$ we define the *stable sampling rate* as

$$\Theta(M, \gamma) = \min(N \in \mathbb{N} : 1/\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) < \gamma).$$

We have thus shown that U satisfies the balancing property with constant $\theta \in (0,1)$ for N and M if and only if $N \geq \Theta(M, \theta^{-1})$.

Rearranging the terms we see that if N, M satisfy the stable sampling rate of order $\gamma = 1/\theta > 1$ then U satisfies the balancing property with constant θ for N and M.

Theorem 1.1 in [37] states that for $M = 2^r$, $r \in \mathbb{N}$ and for all $\gamma > 1$ there exists a constant $S_\gamma > 1$ (dependent on γ, but not on M), such that whenever $N \geq S_\gamma M$, then $1/\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) < \gamma$. Therefore, we have the relation $\Theta(M, \gamma) \leq S_\gamma M = \mathcal{O}(M)$. Hence if $q = \lceil \log_2 S_{1/\theta} \rceil$ we see that the proposition holds with $N = 2^{k+q} \geq S_{1/\theta} 2^k > 2^k = M$.

Next, we comment on how Theorem 1.1 in [37] can be extended to periodic wavelets. We start by noticing that in Equation (4.7) in [37], the inner product $\langle \phi^s_{j,k}, w_n \rangle$ is split in the same way as we do in Equation (6.10). Thus, replacing the boundary wavelets with periodic wavelets results in a slightly different outer sum indices. Next, we follow the computations in the proof, with slightly different indices, until we reach Equation (4.15). Since we only consider periodic extensions, the proof simplifies, since all $\widehat{\phi_i}^{\text{W}}$ in the proof will now be pieces of the scaling function alone and not scaling function and scaling boundary functions. Hence inequality in Equation (4.15) in [37] still holds for periodic wavelets. The remaining parts of the proof of Theorem 1.1. in [37] is identical. □

Proof of Proposition 4.6. Using Theorem 6.8, we see that $\mu(P_N U P_K^\perp) \lesssim K^{-1}$. This gives

$$\begin{aligned}
\|HP_K^\perp x\|_2^2 &= \sum_{k=1}^r \frac{N_k - N_{k-1}}{m_k} \sum_{i\in\Omega_k} \left| \sum_{j>K} U_{ij} x_j \right|^2 \\
&\leq \sum_{k=1}^r \frac{N_k - N_{k-1}}{m_k} \sum_{i\in\Omega_k} \left(\sum_{j>K} \sqrt{\mu(P_N U P_K^\perp)} |x_j| \right)^2 \\
&\leq \sum_{k=1}^r (N_k - N_{k-1}) \mu(P_N U P_K^\perp) \left(\sum_{j>K} |x_j| \right)^2 \\
&\leq N\mu(P_N U P_K^\perp) \left(\sum_{j>K} |x_j| \right)^2 \lesssim \frac{N}{K} \|x\|_1^2. \quad \square
\end{aligned}$$

Proof of Theorem 4.7. First recall that $\mathbf{M} = [2^{J_0+1}, \ldots, 2^{J_0+r}]$ and $\mathbf{N} = [2^{J_0+1}, \ldots, 2^{J_0+r-1}, 2^{J_0+r+q}]$ where $q \geq 0$ is chosen so that U satisfies the balancing property with constant $0 < \theta < 1$. From Lemma 3.4 we therefore have $\|G^{-1}\|_2 \leq 1/\sqrt{\theta}$.

From Theorem 3.7 we know that the matrix A in equation (3.12) satisfies the G-RIPL with $\delta_{\boldsymbol{s},\mathbf{M}} \leq \delta$, with probability at least $1-\epsilon$, provided the sample densities $\boldsymbol{m} \in \mathbb{N}^r$ satisfy $m_k = N_k - N_{k-1}$ for $k = 1, \ldots, r_0$, and

$$m_k \gtrsim \delta^{-2} \cdot \|G^{-1}\|_2^2 \cdot (N_k - N_{k-1}) \cdot \left(\sum_{l=1}^r \mu_{k,l} \cdot s_l \right) \cdot L, \tag{6.15}$$

for $k = r_0 + 1, \ldots, r$. Next notice that $N_k - N_{k-1} = 2^{J_0+k-1}$ for $k = 2, \ldots, r-1$, while $N_r - N_{r-1} = 2^{J_0+r}(2^q - 2^{-1})$ and $N_1 - N_0 = 2^{J_0+1}$. Using the local coherences $\mu_{k,l}$ from Proposition 4.5 we obtain

$$\begin{aligned}
(N_k - N_{k-1}) \left(\sum_{l=1}^r \mu_{k,l} s_l \right) &\lesssim 2^{J_0+k} 2^{q\max\{k+1-r,0\}} \left(\sum_{l=1}^r 2^{-J_0-k} 2^{-|l-k|} s_l \right) \\
&= 2^{q\max\{k+1-r,0\}} \left(\sum_{l=1}^r 2^{-|k-l|} s_l \right).
\end{aligned}$$

Inserting this and $\|G^{-1}\|_2^2 \leq \theta^{-1}$ into (6.15) leads to the sampling condition in Theorem 4.7. $\square$

Proof of Theorem 4.8. The theorem is identical to Corollary 3.8, except that we have fixed $\mathbf{M}$ and $\mathbf{N}$. The concrete values for these have been inserted in condition (*iv*) together with the local coherences $\mu_{k,l}$. The computation of this can be found in the proof above. $\square$

Acknowledgments

The authors would like to thank Simone Brugiapaglia, Simon Foucart, Rémi Gribonval, Øyvind Ryan and Laura Thesing for useful discussions and comments. BA acknowledges support from the Natural Sciences and Engineering Research Council of Canada through grant 611675. ACH acknowledges support from the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L003457/1, a Royal Society University Research Fellowship, and the Philip Leverhulme Prize (Leverhulme Trust, 2017).

References

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

[2] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[3] Y.C. Eldar, G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.
[4] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, 1st edition, Springer - Birkhäuser, 2013.
[5] P.E. Larson, S. Hu, M. Lustig, A.B. Kerr, S.J. Nelson, J. Kurhanewicz, J.M. Pauly, D.B. Vigneron, Fast dynamic 3D MR spectroscopic imaging with compressed sensing and multiband excitation pulses for hyperpolarized 13C studies, Magn. Reson. Med. 65 (3) (2011) 610–619.
[6] M. Lustig, D. Donoho, J.M. Pauly, Sparse MRI: the application of compressed sensing for rapid MR imaging, Magn. Reson. Med. 58 (6) (2007) 1182–1195.
[7] M. Lustig, D.L. Donoho, J.M. Santos, J.M. Pauly, Compressed sensing MRI, IEEE Signal Process. Mag. 25 (2) (2008) 72–82.
[8] A. Jones, A. Tamtögl, I. Calvo-Almazán, A. Hansen, Continuous compressed sensing for surface dynamical processes with helium atom scattering, Sci. Rep. 6 (2016) 27776.
[9] V. Studer, J. Bobin, M. Chahid, H.S. Mousavi, E. Candés, M. Dahan, Compressive fluorescence microscopy for biological and hyperspectral imaging, Proc. Natl. Acad. Sci. 109 (26) (2012) E1679–E1687.
[10] A. Zomet, S.K. Nayar, Lensless imaging with a controllable aperture, in: Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, 2006, pp. 339–346.
[11] G.R. Arce, D.J. Brady, L. Carin, H. Arguello, D.S. Kittle, Compressive coded aperture spectral imaging: an introduction, IEEE Signal Process. Mag. 31 (1) (2014) 105–115.
[12] M.E. Gehm, D.J. Brady, Compressive sensing in the EO/IR, Appl. Opt. 54 (8) (2015) C14–C22.
[13] R.M. Willett, R.F. Marcia, J.M. Nichols, Compressed sensing for practical optical imaging systems: a tutorial, Opt. Eng. 50 (7) (2011).
[14] B. Adcock, A.C. Hansen, C. Poon, B. Roman, Breaking the coherence barrier: a new theory for compressed sensing, Forum Math. Sigma 5 (2017).
[15] B. Roman, A.C. Hansen, B. Adcock, On asymptotic structure in compressed sensing, arXiv:1406.4178, 2014.
[16] A. Bastounis, A.C. Hansen, On the absence of uniform recovery in many real-world applications of compressed sensing and the restricted isometry property and nullspace property in levels, SIAM J. Imaging Sci. 10 (1) (2017) 335–371.
[17] C. Li, B. Adcock, Compressed sensing with local structure: uniform recovery guarantees for the sparsity in levels class, Appl. Comput. Harmon. Anal. 46 (3) (2019) 453–477.
[18] Y. Chi, L.L. Scharf, A. Pezeshki, A.R. Calderbank, Sensitivity to basis mismatch in compressed sensing, IEEE Trans. Signal Process. 59 (5) (2011) 2182–2195.
[19] G. Strang, T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, 1996.
[20] M. Guerquin-Kern, L. Lejeune, K.P. Pruessmann, M. Unser, Realistic analytical phantoms for parallel magnetic resonance imaging, IEEE Trans. Med. Imaging 31 (3) (2012) 626–636.
[21] B. Adcock, C. Boyer, S. Brugiapaglia, On oracle-type local recovery guarantees in compressed sensing, Inf. Inference (2020) 1–49.
[22] Y. Traonmilin, R. Gribonval, Stable recovery of low-dimensional cones in Hilbert spaces: one RIP to rule them all, Appl. Comput. Harmon. Anal. 45 (1) (2018) 170–205.
[23] E.J. Candès, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, IEEE Trans. Inf. Theory 52 (12) (2006) 5406–5425.
[24] E.J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215.
[25] T.T. Cai, L. Wang, G. Xu, Shifting inequality and recovery of sparse signals, IEEE Trans. Signal Process. 58 (3) (2010) 1300–1308.
[26] T.T. Cai, L. Wang, G. Xu, New bounds for restricted isometry constants, IEEE Trans. Inf. Theory 56 (9) (2010) 4388–4394.
[27] J. Andersson, J. Strömberg, On the theorem of uniform recovery of random sampling matrices, IEEE Trans. Inf. Theory 60 (3) (2014) 1700–1710.
[28] M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Commun. Pure Appl. Math. 61 (8) (2008) 1025–1045.
[29] B. Adcock, A.C. Hansen, B. Roman, A note on compressed sensing of structured sparse wavelet coefficients from subsampled Fourier measurements, IEEE Signal Process. Lett. 23 (5) (2016) 732–736.
[30] B. Adcock, A.C. Hansen, Generalized sampling and infinite-dimensional compressed sensing, Found. Comput. Math. 16 (5) (2016) 1263–1323.
[31] S. Brugiapaglia, B. Adcock, Robustness to unknown error in sparse regularization, IEEE Trans. Inf. Theory 64 (10) (2018) 6638–6661.
[32] K.G. Beauchamp, Walsh Functions and Their Applications, vol. 3, Academic Press, 1975.
[33] B. Golubov, A. Efimov, V. Skvortsov, Walsh Series and Transforms: Theory and Applications, vol. 64, Springer Science & Business Media, 1991.
[34] I. Daubechies, Ten Lectures on Wavelets, Vol. 61, SIAM, 1992.
[35] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, 3rd edition, Academic Press, 2008.
[36] A. Cohen, I. Daubechies, P. Vial, Wavelets on the interval and fast wavelet transforms, Appl. Comput. Harmon. Anal. 1 (1) (1993) 54–81.
[37] A. Hansen, L. Thesing, On the stable sampling rate for binary measurements and wavelet reconstruction, Appl. Comput. Harmon. Anal. 48 (2) (2020) 630–654.
[38] V. Antun, Coherence estimates between Hadamard matrices and Daubechies wavelets, Master's thesis, University of Oslo, 2016.
[39] L. Thesing, A.C. Hansen, Linear reconstructions and the analysis of the stable sampling rate, Sampl. Theory Signal Image Process. 17 (1) (2018) 103–126.