Generalized notions of sparsity and restricted isometry property. Part II: Applications

Marius Junge and Kiryung Lee

December 10, 2020

Abstract

Restricted isometry property (RIP) provides a near isometric map for sparse signals. RIP of structured random matrices has played a key role for dimensionality reduction and recovery from compressive measurements. In a companion paper, we have developed a unified theory for RIP of group structured measurement operators on generalized sparsity models. The implication of the extended result will be further discussed in this paper in terms of its pros and cons over the conventional theory. We first show that the extended RIP theory enables the optimization of sample complexity over various relaxations of the canonical sparsity model. Meanwhile, the generalized sparsity model is no longer described as a union of subspaces. Thus the sparsity level is not sub-additive. This incurs that RIP of double the sparsity level does not imply RIP on the Minkowski difference of the sparsity model with itself, which is crucial for dimensionality reduction. We show that a group structured measurement operator provides an RIP-like property with additive distortion for non-sub-additive models. This weaker result can be useful for applications like locality-sensitive hashing. Moreover, we also present that the group structured measurements with random sign enables near isometric sketching on any set similar to the Gaussian measurements. Lastly, an extension of theory to infinite dimension is derived and illustrated over selected examples given by Lebesgue measure of support and Sobolev seminorms.

Keywords – Restricted isometry property, sparsity, dimensionality reduction, sketching, Banach space. **Mathematics Subject Classification** – 42B35, 46N30, 94A20.

1 Introduction

A linear operator $A : \mathbb{R}^N \to \mathbb{R}^m$ has *restricted isometry property* (RIP) on $\Gamma \subset \mathbb{R}^N$ if

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2, \quad \forall x \in \Gamma \, . \tag{1}$$

In other words, A preserves the norm of all vectors in Γ up to multiplicative distortion δ. A class of random matrices have been shown to provide RIP for sparse vectors with the number of measurements m scaling near optimally, i.e. proportional to the number of nonzero entries up to a logarithmic factor. In a companion paper [20], we provided a generalization of the RIP theory for structured random matrices that unified previous results [37, 36, 12, 27] and applies to a class of sparsity and measurement models. Specifically we introduced the notion of generalized sparsity model determined by a pair of Hilbert and Banach spaces so that Γ generalizes beyond existing models such as sparse vectors and low-rank matrices. Moreover the RIP theory in [20] applies to a group structured measurement operator that generalizes partial Fourier transform to a broader class including the Gabor and Radon transforms. In this sequel paper, we further discuss the implications of the results in [20] over a selected set of scenarios. Furthermore we will illustrate how the results extend to a sparsity model in infinite dimension.

For a self-contained introduction, we recall one specific case of the main results in [20] below. First, a generalized sparsity model is described as follows: Let K be a convex subset of the unit ball of ℓ_2^N and let X be the Banach induced by the Minkowski functional of K, i.e.

$$\|x\|_X := \inf\{\lambda > 0 \ : \ x \in \lambda K\} \, .$$

In other words, K is the unit ball in X. Then the sparsity model is determined by

$$\Gamma_s := \left\{ x \in X \cap \ell_2^N \ : \ \|x\|_X \leq \sqrt{s}\|x\|_2 \right\} \ , \tag{2}$$

where $s > 0$ is the parameter corresponding to the sparsity level. We say that a vector x is (K, s)-*sparse* if it belongs to Γ_s. For example, when $X = \ell_1^N$, the resulting Γ_s is a well-studied relaxation of the canonical sparsity model that consists of sparse vectors with at most s nonzero entries. RIP theory has been extended to the model in (2) for a certain class of Banach spaces. Particularly, in this paper, we consider Banach spaces of dual type 2, which is defined as follows.

Definition 1.1 (Banach space of type p [28])**.** *A Banach space Y is of type p if*

$$\left(\mathbb{E} \left\| \sum_j \varepsilon_j y_j \right\|_Y^p \right)^{1/p} \leq c \left(\sum_j \|y_j\|_Y^p \right)^{1/p} \tag{3}$$

holds for some numerical constant c and every finite sequence $(y_j) \subset Y$, where (ε_j) denotes a Rademacher sequence. The type p constant of Y, denoted by $T_p(Y)$, is the smallest constant c that satisfies (3)*.*

A generalization of partial Fourier is given by general group actions. Specifically the measurements are obtained through a fixed measurement functional from randomly selected samples of the orbit of the input by given group actions. To simplify the expression for the number of measurements providing RIP, we impose an additional structure to Banach space X so that the corresponding model is invariant under the group actions. This is formally stated as follows.

Definition 1.2 (G-invariance)**.** *A set K is G-invariant if $\sigma(g)K := \{\sigma(g)x : g \in G,\ x \in K\}$ coincides with K for all $g \in G$.*

Given the generalized sparsity model in (2) for Banach space X of dual type 2 together with the G-invariance of K, the following theorem presents the main result of [20] in this particular case.

Theorem 1.3 (A paraphrased version of [20, Theorem 5.3])**.** *Let $K \subset B_2^N$ be a convex body and X be the Banach space induced by the Minkowski functional of K. Let $\sigma : G \to O_N$ be a continuous group homomorphism of a finite group G to the orthogonal group O_N. Let $g_1, \ldots, g_m \in G$ be independent copies of a uniform random variables on G. Let $\eta \in X^*$ be fixed. Suppose that the following conditions hold: i) X^* has type 2; ii) K is G-invariant; iii) $\sigma(g)^*\eta$ is isotropic. Then there exists a numerical constant c such that for $\delta \in (0,1)$*

$$\sup_{\substack{\|x\|_X \leq \sqrt{s} \\ \|x\|_2 = 1}} \left| \frac{1}{m} \sum_{j=1}^{m} |\langle \eta, (\sigma(g_j)x \rangle|^2 - \|x\|_2^2 \right| \leq \delta$$

holds with probability $1 - \zeta$ provided

$$m \geq c\delta^{-2} s \max\left\{T_2(X^*)^2(1+\ln m)^5, \ln(\zeta^{-1})\right\} \|\eta\|_{X^*}^2 . \tag{4}$$

The result of Theorem 1.3 generalizes existing RIP theory [37, 36, 12, 27] in terms of both the restriction model and measurement operator, which applies to a broader context of applications and also helps improve the RIP analysis of certain type of measurement operators. However, the generalized models lose the useful sub-additive structures in the conventional models. This ends up with a penalty in applying the result to dimensionality reduction. This paper will discuss the pros and cons of Theorem 1.3 in the above perspectives. It will also present an extension of Theorem 1.3 to infinite dimension. The contents are summarized in a simplified form below and further elaborated in the following sections.

1.1 RIP of partial discrete short-time Fourier transform:

The first vignette shows how Theorem 1.3 can be utilized to improve the number of measurements for RIP. Recall that the group structured measurement model in Theorem 1.3 generalizes the Fourier transform. We will focus on the action of a partial discrete short-time Fourier transform (DSTFT) on the canonical sparsity model. Let $\eta \in \mathbb{R}^N$ denote the window of length L. We assume that the nonzero entries of η decays exponentially with parameter α, i.e. the jth largest magnitude is proportional to $j^{-\alpha}$. This decay model describes the behavior of the Poisson window. We consider partial DSTFT with η as above. Moreover, we consider relaxations of the canonical sparsity model consisting of all r-sparse vectors with at most r nonzero entries. Let $1 < q \leq 2$ and q' satisfy $1/q + 1/q' = 1$. Since an r-sparse $x \in \ell_2^N$ satisfies $\|x\|_{q'} \leq r^{1/2-1/q'} \|x\|_2$, it follows that Γ_s in (2) with $s = r^{1-2/q'}$ includes the set of r-sparse vectors. In other words, Γ_s relaxes the canonical sparsity model. The following result is a direct consequence of Theorem 1.3 on this particular scenario while the number of measurements is optimized over the choice of $X = \ell_q^N$ for $1 < q \leq 2$.

Corollary 1.4. *Suppose that $\eta \in \mathbb{R}^N$ is supported on the first L coordinates and the magnitudes of η decay exponentially with $\alpha \leq 1/\ln(N/r)$. Then the random partial DSTFT with η satisfies RIP on r-sparse vectors with high probability if the number of measurements m scales proportional to r up to a logarithmic factor.*

Indeed the optimization over X is crucial to derive the result with a less demanding requirement on α. With fixed $X = \ell_q^N$ for $q' = \ln N$, which approximates the case of $X = \ell_1^N$, RIP by the same number of measurements is obtained when the decay parameter satisfies $\alpha \leq 1/\ln N$. Therefore it allows a significantly slower decay of the magnitudes of η than Corollary 1.4. The gain is obtained from the flexibility of Theorem 1.3 that allows various relaxations of the canonical sparsity model given by choice of Banach space X.

1.2 RIP on non-sub-additive sparsity models

Conventional sparsity models such as sparse vectors, low-rank matrices, and their generalization to atomic models [9] are described as a union of subspaces. These models are sub-additive in the sense that the difference of two s-sparse vectors is $2s$-sparse, i.e. the sparsity level is sub-additive. Unfortunately, this property does not hold for the generalized sparsity model in (2), which is given by a nonconvex cone. Particularly, compared to the conventional sparsity models, a central drawback in the generalization is that the difference $x-y$ of two (K,s)-sparse vectors x and y is no longer $(K,2s)$-sparse. In fact, the adversarial instance of $x-y$ can attain the maximum (trivial) sparsity level. Consequently, RIP on Γ_{2s} does not necessarily imply RIP on $\Gamma_s - \Gamma_s$, which denotes the Minkowski sum of Γ_s and $-\Gamma_s$. In fact, the latter property plays a crucial role for dimensionality reduction that preserves the distance of any pair from Γ_s and recovery of signal in Γ_s by convex programming [42].

The second vignette shows that despite the lack of sub-additivity one can still obtain an RIP-like result for the generalized sparsity model if the measurement operator A satisfies "multiresolution" RIP, that is A satisfies RIP on $(K, 2^l s)$-sparse vectors with distortion $\max(2^{l/2}\delta, 2^l\delta^2)$ for all $\lfloor -\log_2 s \rfloor \le l \le \lceil \log_2(\left\| \mathrm{Id} : \ell_2^N \to X \right\|^2 /s) \rceil$.

Theorem 1.5. *Suppose that A satisfies multiresolution RIP. Let x and y be (K,s)-sparse vectors. If $\|Ax - Ay\|_2 \ge 4\sqrt{2}\delta$, then*

$$\left(1 - \frac{1}{\sqrt{2}}\right) \|Ax - Ay\|_2^2 \le \|x - y\|_2^2 \le \left(1 + \frac{1}{\sqrt{2}}\right) \|Ax - Ay\|_2^2 \ .$$

Otherwise $\|x - y\|_2 \le 8\delta$.

The consequence of Theorem 1.5 is obviously weaker than the analogous result by RIP on $\Gamma_s - \Gamma_s$. The former implies that A preserves the distance only when the images of $x - y$ via A is larger than certain threshold. Otherwise it can only say that two (K,s)-sparse vectors are close by being within certain distance. Particularly the map A restricted on $\Gamma_s - \Gamma_s$ may not be injective. Alternatively the result of Theorem 1.5 implies that A provides a near isometric map with additive distortion instead of multiplicative distortion. A similar phenomenon occurs in embedding with 1-bit quantization [18].

On a positive side, the group structured measurement operator in Theorem 1.3 provides the multiresolution RIP while the number of measurements scales in the same order. The weaker RIP result on $\Gamma_s - \Gamma_s$ with additive distortion can still be useful in some applications such as locality-sensitive hashing.

1.3 RIP of group structured measurement operator with random sign

In a companion paper [20], it has been shown that Theorem 1.3 provides near optimal RIP results on several examples of the generalized sparsity models, for which the number of measurements scales proportional to the sparsity level up to a logarithmic factor. However, for certain other instances, RIP holds with significantly more measurements. This is not the case for the Gaussian measurement operator. It provides RIP on any set Γ with the number of measurements scaling optimally with respect to the complexity of Γ [14]. This weakness of Theorem 1.3 compared to the Gaussian case can be mitigated by incorporating random sign into the measurement operator. Oymak et al. [34] showed that a measurement operator with multiresolution RIP on the canonical sparsity model following random sign provides RIP on an arbitrary set. By applying their result, we obtain RIP of the group structured measurements with random sign similar to the Gaussian measurement case.

Corollary 1.6. *Let A denote the composition of the measurement operator in Theorem 1.3 and a diagonal operator with random ± 1 entries. Then with high probability A satisfies RIP on all (K,s)-sparse vectors if m is proportional to the squared Gaussian width of $\Gamma_s \cap \mathbb{S}^{N-1}$ up to a logarithmic factor, where $\mathbb{S}^{N-1}$ denotes the unit sphere in ℓ_2^N.*

1.4 Extension to infinite-dimensional models

Lastly we illustrate how Theorem 1.3 extends to infinite dimension. The Hilbert space ℓ_2^N is substituted by a seminormed space given as a subspace of $L_2[0,1]$ equipped with a seminorm $\|f\|_{2,w} := (\sum_{k\in\mathbb{Z}} w_k \hat{f}[k]^2)^{1/2}$, where $\hat{f}[k]$ denotes the kth Fourier series coefficient and $(w_k)_{k\in\mathbb{Z}}$ is a nonnegative weight sequence. Then we derive an extension of Theorem 1.3 to the sparsity model given by $\|f\|_X \le \sqrt{s}\,\|f\|_{2,w}$.

Theorem 1.7. *Let $X \subset L_2[0,1]$ be a Banach space with unit ball K. Let $\sigma : G \to U$ be a continuous group homomorphism of a group G to the unitary group U. Let $g_1,\ldots,g_m$ be independent copies of a Haar-distributed random variable on G. Suppose that the following* *conditions hold: i) X^* has type 2; ii) X is a Banach lattice*[1]*; iii) K is G-invariant; iv) $\mathbb{E}\sum_{j=1}^m |\langle \eta, \sigma(g_j f)\rangle|_2^2 = m\|f\|_{2,w}^2$ for all $f \in L_2[0,1]$; and v) $\sigma(g)$ commutes with any point-wise operation on $f \in L_2[0,1]$. Then there exists a numerical constant c such that for $\delta \in (0,1)$*

$$\sup_{\substack{\|f\|_X \le \sqrt{s} \\ \|f\|_{2,w}=1}} \left| \frac{1}{m} \sum_{j=1}^m |\langle \eta, \sigma(g_j) f\rangle|_2^2 - \|f\|_{2,w}^2 \right| \le \delta$$

holds with probability $1-\zeta$ provided

$$m \;\ge\; c\delta^{-2} s \left\{ T_2(X^*)^2 (1+\ln d)^5 (1+\ln m)^5 + \ln(\zeta^{-1}) \right\} \|\eta\|_{X^*}^2 \;.$$

The result of Theorem 1.7 will become substantive through concrete examples of sparsity models given by a specific choice of Banach space X. Particularly we consider a model given by the Lebesgue measure of the support of f, which naturally extends the canonical sparsity in finite dimension. This model is further restricted by a Sobolev seminorm. We also illustrate the result over another class of continuous-time signals constrained by the Sobolev $(1,q)$-seminorm, which has been shown an effective regularizer for denoising similar to total variation.

The above infinite-dimensional sparsity models are compared to those appearing in the literature as follows. The spike model refers to a superposition of Dirac's impulses. Recovery of the spike model from Fourier coefficients has been extensively studied [40, 7, 19, 5, 6, 4]. There also exist extensions to signals of *finite rate of innovations* (FRI) [45, 13, 16, 35, 30–32] and to a sparsity model in a countable representation system [2]. Note that these models are either restricted to point measures supported on a set of measure zero or obtained via an approximation by a certain form of discretization. In contrast, the infinite-dimensional model in Theorem 1.7 does not suffer from these limitations and provides more flexibility in describing signals in infinite dimension. However, due to the

[1] A Banach space X is a lattice if $|f| \le |h|$ implies $\|f\|_X \le \|h\|_X$. As a consequence, the norm is invariant under point-wise sign change.

lack of sub-additivity of sparsity level, Theorem 1.7 does not imply RIP on the Minkowski difference of the sparsity model, which could have provided recovery guarantee by convex programming via the result by Traonmilin and Gribonval [42]. From this perspective, Theorem 1.7 is weaker than previous work on the other infinite-dimensional models.

1.5 Notation

The symbols $c, c_1, c_2, \ldots$ and $C, C_1, C_2, \ldots$ will be reserved to denote positive numerical constants, which may vary from line to line. We will use notation for various Banach spaces and norms. The norm of a Banach space X is denoted by $\|\cdot\|_X$. For brevity, we use the shorthand notation $\|\cdot\|_q$ for the norm of ℓ_q^N and $L_q[0,1]$. The distinction will be clear from the context. Moreover B_q^N and B_q will denote the unit ball of ℓ_q^N and $L_q[0,1]$, respectively. The identity operator will be denoted by Id. The operator norm will be denoted by $\|\cdot\|$ without any subscript. We also use the following two big-O notations: First $y = O(x)$ implies that there is a numerical constant $c > 0$ independent of two positive numbers x and y such that $y \leq cx$, which is also equivalently written as $y \lesssim x$ or $x \gtrsim y$. If $y \leq cx$ with c depends on a logarithmic function of x and y, then it will be denoted by $y = \tilde{O}(x)$.

2 Optimized analysis over relaxations of canonical sparsity model

The first vignette will demonstrate how Theorem 1.3 can be leveraged to optimize the number of group structured measurements that provides a near isometric map on sparse signals. Here we consider the canonical sparsity model that constrains the number of nonzero entries and its relaxations given by a set of Banach spaces ℓ_q^N for $q \in [1, 2)$. The approach will be illustrated over random partial discrete short-time Fourier transform (DSTFT) with a decaying window. The optimization is carried out over the choice of the

Banach space ℓ_q^N.

We start with showing how DSTFT is described by a set of group actions. Let $\eta \in \mathbb{R}^N$ represent a window of length L, i.e. $\eta[l] = 0$ for $L \le l < N$, where the time index system is zero-based and modulo N. The windowed discrete Fourier transform (DFT) of $x \in \mathbb{R}^N$ with time-shift $t \in \mathbb{Z}_N$ and at frequency $k \in \mathbb{Z}_N$ is given by

$$c(t,k) = \sum_{l=0}^{N-1} x[l]\eta[l-t]e^{-\mathrm{i}2\pi kl/N} \ . \tag{5}$$

The windowed DFT in (5) is indeed described by a set of group actions as follows. Let σ denote the group homomorphism of $\mathbb{Z}_N \times \mathbb{Z}_N$ to the orthogonal group given O_N given by

$$\sigma : (t,k) \mapsto \Lambda^t \mathrm{Sh}^k \ , \tag{6}$$

where $\Lambda : \mathbb{C}^N \to \mathbb{C}^N$ denotes the modulation operator defined by

$$\Lambda(e_l) = e^{\mathrm{i}2\pi l/N} e_l \ , \quad \forall l = 1, \dots, N \ ,$$

and $\mathrm{Sh} : \mathbb{C}^N \to \mathbb{C}^N$ acts as the circular shift modulo N, i.e.

$$\mathrm{Sh}(e_l) = \begin{cases} e_{l+1} & 1 \le l \le N-1 \ , \\ e_1 & l = N \ . \end{cases}$$

Here $e_1, \dots, e_N$ denote the standard basis vectors in $\mathbb{C}^N$. Then the windowed DFT coefficient in (5) is written as a group structured measurement given by

$$c(t,k) = \langle \eta, \sigma(t,k)x \rangle \ .$$

Alternatively, the entire set of measurement functionals is described as the orbit of η over G, that is $\{\sigma(t,k)^* \eta \ : (t,k) \in G\}$.

The windowed DFT generalizes DFT as the latter is a special case with $\eta = (1, \dots, 1) \in \mathbb{R}^N$, where $c(t,k)$ is uniquely determined by k regardless the choice of t. However, we are more interested in the scenario where $L < N$, which corresponds to DSTFT. Moreover, windows those providing a certain decay property such as Hann window or Gaussian window are commonly used in practice. In what follows, we consider a Poisson-like window η with exponentially decaying magnitudes.

From Theorem 1.3 we obtain a sufficient number of random DSTFT measurements that provides a near isometry on sparse signals. The following corollary is a consequence of Theorem 1.3 for a Banach space X that approximates ℓ_1^N.

Corollary 2.1. *Let $(t_1, k_1), \ldots, (t_m, k_m)$ be independent copies of a uniform random variable on $\mathbb{Z}_N \times \mathbb{Z}_N$. For $\eta = (\eta_1, \ldots, \eta_N) \in \mathbb{R}^N$, with $(\eta_j^{\downarrow})_{1 \le j \le N}$ denoting the rearrangement of $(|\eta_j|)_{1 \le j \le N}$ in the non-increasing order, suppose that there exist $\alpha \in (0, 1/2)$, $\beta > 0$, $\gamma \ge 1$, and $L < N$ such that i) $\beta j^{-\alpha} \le \eta_j^{\downarrow} \le \beta\gamma j^{-\alpha}$ for $1 \le j \le L$; ii) $\eta_j^{\downarrow} = 0$ for $L < j \le L$; iii) $\|\eta\|_2 = \sqrt{N}$. Then there exists a numerical constant $c > 0$ such that*

$$\mathbb{P}\left\{ \sup_{\|x\|_0 \le r,\ \|x\|_2 = 1} \left| \frac{1}{m} \sum_{j=1}^{m} \left| \langle \sigma(t_j, k_j)^* \eta, x \rangle \right|^2 - \|x\|_2^2 \right| \ge \delta \right\} \le \zeta \tag{7}$$

provided

$$m \ge c\delta^{-2} r \max\{(1 + \ln L)^2 (1 + \ln m)^5,\ \ln(\zeta^{-1})\} \cdot \frac{\gamma^2 N^{2\alpha}(1 - 2\alpha)}{1 - L^{-1+2\alpha}} \cdot \left(\frac{N}{L}\right)^{1-2\alpha}.$$

Proof. We apply Theorem 1.3 for $X = \ell_q^N$ with $q = 3\ln L/(3\ln L - 1)$. First we verify that the assumptions of Theorem 1.3 are satisfied. Note that $X^* = \ell_{q'}^N$, where $q' = 3\ln L > 2$ for any $L \ge 2$. Therefore X^* has type 2 and the type 2 constant $T_2(X^*)$ is upper-bounded by $\sqrt{q'}$ [8, Lemma 3]. Moreover, the unit ball $K = B_q^N$ is G-invariant, where the group homomorphism σ is from $\mathbb{Z}_N \times \mathbb{Z}_N$ to O_N as in (6). This is deduced from the following two observations: Since Λ is represented as a diagonal matrix whose diagonal entries are complex numbers of unit modulus, the ℓ_q-norm is invariant under Λ. It is also straightforward to see that the ℓ_q-norm is invariant under Sh. Lastly, as shown in [20, Section 4.2.1], since σ is irreducible, i.e. the only subspaces invariant under group actions are $\{0\}$ and X, it follows that $\sigma(t_j, k_j)^* \eta$ is an isotropic random vector.

It remains to derive an upper bound on $\|\eta\|_{q'}$. By the assumptions on η, we have

$$N = \|\eta\|_2^2 \geq \beta^2 \sum_{j=1}^{L} j^{-2\alpha} \geq \frac{\beta^2(L^{1-2\alpha}-1)}{1-2\alpha} \,. \tag{8}$$

For the particular choice of $q' = 3\ln L$, since η has only L nonzero entries, it follows that

$$\|\eta\|_\infty \leq \|\eta\|_{q'} \leq L^{1/q'} \|\eta\|_\infty \,.$$

Since $L^{1/3\ln L} = e^{1/3}$, the two norms are equivalent up to a numerical constant. Therefore we deduce

$$\|\eta\|_{q'}^2 \leq \beta^2\gamma^2 \lesssim \frac{\gamma^2 N(1-2\alpha)}{L^{1-2\alpha}-1} \,.$$

□

Recall that the set $\{x \in \mathbb{R}^N : \|x\|_1 \leq \sqrt{r}\,\|x\|_2\}$ is a well-studied relaxation of the canonical sparsity model consisting of r-sparse vectors. Therefore, in Corollary 2.1, the Banach space X was chosen so that it accurately approximates ℓ_1^N. However, as shown in the following corollary, the number of measurements providing a near isometric map can be significantly reduced over that in Corollary 2.1 by a naïve approach. We optimize it over relaxations of the canonical sparsity model given by Banach space X.

Corollary 2.2. *Suppose the hypothesis of Corollary 2.1 holds. Then there exists a numerical constant $c > 0$ such that (7) holds if*

$$m \geq c\delta^{-2} r \max\{\alpha^{-2}(1+\ln m)^5, \ln(\zeta^{-1})\} \cdot \frac{\gamma^2(1-2\alpha)(1+\ln L)^{2\alpha}}{1-L^{-1+2\alpha}} \cdot \left(\frac{N}{r}\right)^{2\alpha} \cdot \left(\frac{N}{L}\right)^{1-2\alpha} \,. \tag{9}$$

Proof. Let $1 < q < 2$ and $X = \ell_q^N$. We first verify that the assumptions of Theorem 1.3 hold. It follows that $X^* = \ell_{q'}^N$ with $2 < q' < \infty$ is of type 2 and the type 2 constant $T_2(X^*)$ is upper bounded by $\sqrt{q'}$ [8, Lemma 3]. Furthermore, similarly to the proof of Corollary 2.1, we also have the G-invariance of B_q^N and the isotropy of $\sigma(t_j, k_j)^*\eta$. Indeed, the arguments there remain valid for any $q \geq 1$.

Since any r-sparse x satisfies $\|x\|_{q'} \leq r^{1/2-1/q'} \|x\|_2$, it follows that the set of (B_q^N, s)-sparse vectors where $s = r^{1-2/q'}$ provides a relaxation of the canonical sparsity model consisting of r-sparse vectors. Therefore, by Theorem 1.3, the assertion in (7) holds if

$$m \geq c\,\delta^{-2} r^{1-2/q'} \max\left\{(q')^2(1+\ln m)^5, \ln(\zeta^{-1})\right\} \|\eta\|_{X^*}^2 . \tag{10}$$

It remains to derive an upper bound on $\|\eta\|_{X^*}$. We recall that the normalization of η implies (8). Next we choose $q' = 1/\alpha$ and compute $r^{1-2/q'} \|\eta\|_{q'}^2$ as follows:

$$\begin{aligned} r^{1-2/q'} \|\eta\|_{q'}^2 &\leq r^{1-2/q'} \gamma^2 \beta^2 \left(\sum_{j=1}^{L} j^{-q'\alpha} \right)^{2/q'} \\ &\leq \frac{\gamma^2(1-2\alpha) r^{1-2\alpha} N (1+\ln L)^{2\alpha}}{L^{1-2\alpha}-1} = r \left(\frac{L}{r}\right)^{2\alpha} \frac{\gamma^2(1-2\alpha) N (1+\ln L)^{2\alpha}}{L(1-L^{-1+2\alpha})} . \end{aligned} \tag{11}$$

Then (9) is obtained as a sufficient condition for (10) by plugging in the upper bound in (11). □

Corollary 2.2 provides a tighter lower bound on the number of measurements than Corollary 2.1 by a factor of $((1+\ln L)/r)^{2\alpha}$. Alternatively, the two corollaries can be compared in terms of the requirement on the decay parameter α that achieves a near isometry with $\tilde{O}(r)$ measurements. Suppose that the window length L is proportional to N. According to the result by Corollary 2.1, α needs to be set to at least $1/\ln N$, whereas Corollary 2.2 relaxes the requirement to $\alpha = 1/\ln(N/r)$. The improvement is significant when the sparsity level is proportional to N. Therefore the optimized analysis over relaxations of the canonical sparsity model allows that the near restricted isometry property of random partial DSTFT applies to a broader class of window functions.

3 Sketching non-sub-additive sparsity models

Theorem 1.3 extends existing RIP theory to generalized sparsity models. However, this generalization comes at a cost that the model loses the sub-additivity of sparsity level. In other words, a near isometric map does not necessarily preserve the distance between two sparse vectors with the same distortion in preserving the length. The vignette in this section illustrates how this penalty can be overcome by leveraging the notion of *multiresolution restricted isometry property* (MRIP) [34].

In the conventional sparsity models, which are given by a union of subspaces, the function that measures the sparsity level of a vector is *sub-additive*. For example, in the canonical sparsity model, ℓ_0-pseudo-norm that counts the number of nonzero entries measures the sparsity level of a vector and satisfies $\|x + y\|_0 \leq \|x\|_0 + \|y\|_0$ for any x and y. The sparsity level function has the same sub-additivity for the low-rank matrix model and more generally for an atomic model [9].

The generalized sparsity model in Theorem 1.3 does not provide the sub-additivity of sparsity level. Therefore a near isometry on the set of $(K, 2s)$-sparse vectors does not necessarily imply the same property on the Minkowski difference of the set of (K, s)-sparse vectors with itself. For inverse problems with a generalized sparsity model, the latter property is crucial for deriving a performance guarantee of recovery methods [42]. Theorem 1.3 only provides the former result and the extension to the Minkowski difference is not straightforward because of significant difference in geometry between conventional union-of-subspace models and the generalized model given by a nonconvex cone. We show that it is possible to obtain a weaker result providing a restricted near isometry with additive distortion instead of multiplicative distortion as in Theorem 1.3. The derivation will rely on a modified version of MRIP described below.

MRIP was originally proposed for the canonical sparsity model to analyze sketching of an arbitrary set with random sign [34]. As the name implies, MRIP consists of RIPs at various sparsity and distortion levels, where the two parameters vary simultaneously over a certain range. We consider its extension with respect to a general sparsity model given by a Banach space. Let X be a Banach space with unit ball $K \subset B_2^N$. Note that if K has a nonempty interior then there exists a number

$$s_{\max}(K) = \|\mathrm{Id} : \ell_2^N \to X\|^2 \tag{12}$$

such that $\|x\|_X \leq \sqrt{s_{\max}(K)}\|x\|_2$ holds for all x. For example, if $H = \ell_2^N$ and $X = \ell_1^N$, then $s_{\max}(K) = N$. Given a sparsity generalized sparsity model by K with the maximum sparsity level $s_{\max}(K)$, MRIP on this model is defined as follows.

Definition 3.1 (Multi-resolution RIP for (K, s)-sparse vectors)**.** *Let X be a Banach space with unit ball $K \subset B_2^N$ and $s_{\max}(K)$ be defined in (12). For $\delta > 0$ and $s \geq 1$, we say that $A \in \mathbb{C}^{m\times N}$ satisfies (K, s)-MRIP with distortion δ if*

$$\sup_{\|x\|_X \leq \sqrt{2^l s},\ \|x\|_2=1} \left|\|Ax\|_2^2 - \|x\|_2^2\right| \leq \max(2^{l/2}\delta, 2^l\delta^2)$$

holds for all $l \in \mathbb{Z}$ satisfying $\lfloor -\log_2 s\rfloor \leq l \leq \lceil \log_2(s_{\max}(K)/s)\rceil$.

Note that MRIP in Definition 3.1 is a generalization of the original definition by Oymak et al. [34], which is obtained by substituting the canonical sparsity model by a generalized model given by Banach space X. The generalized version in Definition 3.1 plays a key role in analyzing sketching of non-sub-additive sparsity models.

Moreover, by definition, RIP is a special instance within MRIP at the resolution by $l = 0$. Therefore one expects that a larger number of measurements are necessary to provide MRIP compared to that for RIP. For certain measurement maps, MRIP can be guaranteed by a number of measurements that scales in the same order compared to that for RIP. This is the case with the group structured measurements in Theorem 1.3, which will be shown in the end of this section.

The following theorem demonstrates that (K, s)-MRIP provides a restricted near isometry on the Minkowski difference of the set of (K, s)-sparse vectors in a weaker sense.

Theorem 3.2. *Let $K \subset B_2^N$ and X be the Banach space whose unit ball is K. For $\delta > 0$, $s \geq 1$, and $s_{\max}(K)$ defined in (12), suppose that $A \in \mathbb{C}^{m \times N}$ satisfies (K, s)-MRIP with distortion δ. Then we have the following results for any (K, s)-sparse unit vectors x and y in ℓ_2^N: If Ax and Ay are separated enough by satisfying*

$$\|Ax - Ay\|_2 \geq 4\sqrt{2}\delta \; , \tag{13}$$

then we have

$$\left(1 - \frac{1}{\sqrt{2}}\right) \|Ax - Ay\|_2^2 \leq \|x - y\|_2^2 \leq \left(1 + \frac{1}{\sqrt{2}}\right) \|Ax - Ay\|_2^2 \; . \tag{14}$$

Otherwise, if (13) is violated, then

$$\|x - y\|_2 \leq 8\delta \; . \tag{15}$$

Proof of Theorem 3.2. Let $h = x - y$ denote the difference between x and y. We first show that

$$\left| \|Ah\|_2^2 - \|h\|_2^2 \right| \leq \max \left\{ \frac{\sqrt{2}\delta \|x - y\|_X \|x - y\|_2}{\sqrt{s}}, \; \frac{2\delta^2 \|x - y\|_X^2}{s} \right\} \; . \tag{16}$$

It follows from the definition of $s_{\max}$ that there exists $l \in \mathbb{Z}$ such that

$$\lfloor -\log_2 s \rfloor \leq l \leq \lceil \log_2 (s_{\max}(K)/s) \rceil \tag{17}$$

and

$$2^l s < \frac{\|h\|_X^2}{\|h\|_2^2} \;\leq\; 2^{l+1} s \; . \tag{18}$$

This implies that h is $(K, 2^{l+1}s)$-sparse. Thus the MRIP assumption provides

$$\left| \|Ah\|_2^2 - \|h\|_2^2 \right| \;\leq\; \max\{\delta_{l+1}, \delta_{l+1}^2\} \|h\|_2^2 \; , \tag{19}$$

where $\delta_{l+1} := 2^{(l+1)/2}\delta$. If $\delta_{l+1} < 1$, then by (18) the right-hand side of (19) is upper-bounded by

$$\delta_{l+1} \|h\|_2^2 = 2^{(l+1)/2} \delta \|h\|_2^2 \leq \sqrt{2}\delta s^{-1/2} \|h\|_X \|h\|_2 \; . \tag{20}$$

Otherwise, if $\delta_{l+1} \geq 1$, then the right-hand side of (19) is bounded from above by

$$\delta_{l+1}^2 \|h\|_2^2 = 2^{l+1}\delta^2 \|h\|_2^2 \leq 2\delta^2 s^{-1} \|h\|_X^2 \; . \tag{21}$$

Plugging in (20) and (21) into (19) provides (16).

By the triangle inequality, we obtain

$$\|x - y\|_X \leq \|x\|_X + \|y\|_X \leq \sqrt{s}\,\|x\|_2 + \sqrt{s}\,\|y\|_2 \leq 2\sqrt{s} \; .$$

Therefore from (16) we can continue as

$$\big|\|Ah\|_2^2 - \|h\|_2^2\big| \leq \max\{2\sqrt{2}\delta\|h\|_2, 8\delta^2\} \; . \tag{22}$$

Then we proceed with the proof by considering the following two complementary cases.

Case 1: $\|h\|_2 \geq \|Ah\|_2$. It follows that $2\sqrt{2}\delta\|h\|_2 \geq 16\delta^2$ and the maximum in (22) is attained by the first term. Thus by (13) and (22), we have

$$\|h\|_2^2 - \|Ah\|_2^2 \leq 2\sqrt{2}\delta\|h\|_2 \leq \frac{\|h\|_2^2}{2} \; ,$$

which implies $\|h\|_2 \leq \sqrt{2}\|Ah\|_2$. This together with (13) and (22) provides

$$\big|\|Ah\|_2^2 - \|h\|_2^2\big| \leq 4\delta\|Ah\|_2 \leq \frac{\|Ah\|_2^2}{\sqrt{2}} \; . \tag{23}$$

Note that (23) is equivalent to (14).

Case 2: $\|h\|_2 \leq \|Ah\|_2$. It follows from (22) and (13) that

$$\big|\|Ah\|_2^2 - \|h\|_2^2\big| \leq 2\sqrt{2}\delta\|Ah\|_2 \leq \frac{\|Ah\|_2^2}{2} \; ,$$

which implies (14). Thus the first assertion is proved.

The second assertion is proved by contradiction. Suppose that $\|Ah\|_2 < 4\sqrt{2}\delta$ and $\|h\|_2 > 8\delta$ hold simultaneously. In this case the first term achieves the maximum in (22). Therefore (22) provides

$$\|h\|_2^2 - \|Ah\|_2^2 \leq 2\sqrt{2}\delta\|h\|_2 \leq \frac{\|h\|_2^2}{2\sqrt{2}} \; .$$

Then it follows that

$$\|h\|_2^2 \leq \frac{2\sqrt{2}}{2\sqrt{2}-1}\|Ah\|_2^2 < 64\delta^2 < \|h\|_2^2 \,,$$

which is a contradiction. Therefore, $\|Ah\|_2 < 4\sqrt{2}\delta$ implies $\|h\|_2 \leq 8\delta$. This completes the proof. $\square$

Theorem 3.2 implies that if the distance between Ax and Ay for two unit-norm sparse vectors x and y is larger than $4\sqrt{2}\delta$, then the distance between x and y in ℓ_2^N is equivalent to $\|Ax - Ay\|_2$ up to a constant factor. In other words, one can distinguish x and y from their linear measurements. However, if Ax and Ay are close by satisfying $\|Ax - Ay\|_2 < 4\sqrt{2}\delta$, then Theorem 3.2 only confirms that $\|x-y\|_2$ is less than 8δ, i.e., one cannot distinguish two similar sparse vectors x and y from their measurements. Obviously, this result is weaker than sketching any two sparse vectors (regardless of the amount of distance) given by the restricted isometry on a sub-additive sparsity model. However, this weak result applies to a broader class of signals and can be useful for certain applications. For example, in locality-sensitive hashing, if the centroids of clusters are well separated via the dimensionality reduction via A, then one can compute clustering in the compressed domain.

Remark 3.3. We did not attempt to optimize the constants in Theorem 3.2. The result can be stated with positive constants α and β that satisfy $\alpha > 2\sqrt{2}$ and $\alpha^2 < \beta(\beta - 2\sqrt{2})$ as follows: If $\|Ah\|_2 \geq \alpha\delta$, then

$$\left(1 - \frac{2\sqrt{2}}{\sqrt{\alpha(\alpha - 2\sqrt{2})}}\right)\|Ah\|_2^2 \leq \|h\|_2^2 \leq \left(1 + \frac{2\sqrt{2}}{\sqrt{\alpha(\alpha - 2\sqrt{2})}}\right)\|Ah\|_2^2 \,.$$

Otherwise, if $\|Ah\|_2 < \alpha\delta$, then $\|h\|_2 \leq \beta\delta$. One may optimize the constants α and β in order to further tighten the estimates.

Next, in the following corollary, we demonstrate that MRIP preserves the ℓ_2-norm of two (K, s)-sparse unit vectors x and y under a mild condition on the sparsity level of $x - y$.

Corollary 3.4. *Suppose that the hypothesis of Theorem 3.2 holds. For any $\varepsilon > 0$, we have*

$$\left|\|Ax - Ay\|_2^2 - \|x - y\|_2^2\right| \leq \min\left\{\frac{\|x-y\|_2^2}{1+\epsilon},\ 2\sqrt{2}\delta\|x-y\|_2\right\} \tag{24}$$

provided that $x, y \in \mathbb{S}^{N-1}$ are (K, s)-sparse and satisfy

$$\|x - y\|_X \leq \frac{\sqrt{s}\|x-y\|_2}{\sqrt{2}(1+\epsilon)\delta}\ . \tag{25}$$

Remark 3.5. Corollary 3.4 preserves the distance of two (K, s)-sparse vectors x, y by (24) if the sparsity level of the difference $x - y$ is below the threshold $s/2(1+\epsilon)^2\delta^2$, which is higher than the sparsity level s of each of x and y for small δ. The estimate in (24) implies that the distortion is strictly less than $\|x - y\|_2^2$, which implies a local injectivity.

Since the ℓ_2-norm is preserved by RIP up to a small multiplicative distortion δ, we can always compare two spare vectors after normalization. Suppose that $\|x\|_2 = \|y\|_2 = 1$. Then (24) also implies that the distortion is no larger than $2\sqrt{2}\delta\|x - y\|_2$. Although this distortion bound is more conservative than $\delta\|x - y\|_2^2$, which is available if the sparsity level is sub-additive, it can still be useful for certain applications. For example, similar deviation bounds have been used in the analysis of iterative optimization algorithms for matrix completion (see [10, Lemma 5] and [47, Lemma 8]).

Proof of Corollary 3.4 . The arguments in that for Theorem 3.2 are used to prove Corollary 3.4 . For the sake of claity, we repeat these arguments. Let $h = x - y$. By the definition of $s_{\max}$ that there exists $l \in \mathbb{Z}$ such that $\lfloor -\log_2 s\rfloor \leq l \leq \lceil \log_2(s_{\max}(K)/s)\rceil$ and

$$2^l s < \frac{\|h\|_X^2}{\|h\|_2^2}\ \leq\ 2^{l+1} s\ . \tag{26}$$

The upper bound in (26) implies that h is $(K, 2^{l+1}s)$-sparse. Thus by MRIP we have

$$\left|\|Ah\|_2^2 - \|h\|_2^2\right|\ \leq\ \max\{\delta_{l+1}, \delta_{l+1}^2\}\|h\|_2^2\ , \tag{27}$$

where $\delta_{l+1} := 2^{(l+1)/2}\delta$. From the lower bound in (26) together with the assumption in

(25) for $h = x - y$ provides

$$2^l s < \frac{\|h\|_X^2}{\|h\|_2^2} \leq \frac{s}{2(1+\epsilon)^2\delta^2} ,$$

which implies

$$\delta_{l+1} = 2^{(l+1)/2}\delta < \frac{1}{1+\epsilon} < 1 .$$

Therefore the upper bound in (27) reduces to

$$\delta_{l+1}\|h\|_2^2 = 2^{(l+1)/2}\delta\|h\|_2^2 \leq \frac{\sqrt{2}\delta\|h\|_X\|h\|_2}{\sqrt{s}} , \tag{28}$$

where the second step follows from the lower bound in (26). Recall that $\|h\|_X$ is upper-bounded by (25). Moreover since each of x and y is (K, s)-sparse, by the triangle inequality, we obtain

$$\|h\|_X \;\leq\; \|x\|_X + \|y\|_X \;\leq\; \sqrt{s}(\|x\|_2 + \|y\|_2) = 2\sqrt{s} .$$

The assertion is obtained by plugging in the minimum of these two upper bounds on $\|h\|_X$ into (28). $\square$

Next we show that Theorem 3.2 provides a recovery guarantee for the following optimization problem:

$$\begin{aligned} &\underset{\tilde{x}}{\text{minimize}} && \|\tilde{x}\|_X \\ &\text{subject to} && A\tilde{x} = Ax, \quad \|\tilde{x}\|_2 = 1 . \end{aligned} \tag{29}$$

Suppose that x satisfies $\|x\|_2 = 1$ and $\|x\|_X \leq \sqrt{s}$. Let $\hat{x}$ be the solution to (29). Since x is feasible in (29), we have $\|\hat{x}\|_X \leq \|x\|_X \leq \sqrt{s}$. Moreover, $\hat{x}$ also satisfies $\|\hat{x}\|_2 = 1$. Therefore, by Theorem 3.2, it follows that $\|\hat{x} - x\|_2 \leq 8\delta$.

However, the optimization in (29) is a nonconvex program with a spherical constraint. Furthermore, computing the X-norm may be expensive (e.g. certain tensor norms are NP-hard to compute. [17]). It will be a fruitful direction to pursue a guaranteed method and its practical implementation to solve (29).

Finally we conclude this section by showing that MRIP holds with high probability for the group structured measurements in Theorem 1.3.

Proposition 3.6. *Suppose that the hypothesis of Theorem 1.3 holds and $s_{\max}(K)$ is defined in (12). Then there exists a numerical constant c such that*

$$A = \frac{1}{\sqrt{m}} \begin{bmatrix} \sigma(g_1)\eta & \dots & \sigma(g_m)\eta \end{bmatrix}^* \in \mathbb{C}^{m \times N}$$

satisfies (K, s)-MRIP with distortion δ with probability $1 - \zeta$ provided

$$m \ \geq \ c\delta^{-2} s \max \left\{ T_2(X^*)^2 (1 + \ln m)^5, \ln\ln(s_{\max}(K)) + \ln(\zeta^{-1}) \right\} \|\eta\|_{X^*}^2 . \tag{30}$$

Proof. Fix $l \in \mathbb{N}$. Since it trivially holds that $2^l s/(2^{l/2}\delta)^2 = s/\delta^2$, it follows from Theorem 1.3 that there exists a numerical constant c such that (30) implies

$$\mathbb{P}\left\{ \sup_{\|x\|_0 \leq 2^l s, \ \|x\|_2 = 1} \left| \frac{1}{m} \sum_{j=1}^m |\langle \eta, \sigma(g_j) x \rangle|^2 - \|x\|_2^2 \right| \geq \max(2^{l/2}\delta, 2^l \delta^2) \right\} \leq \frac{\zeta}{\lceil \log_2(s_{\max}(K)) \rceil} .$$

Since l was arbitrary, the assertion is obtained by applying the union bound over $\lfloor -\log_2 s \rfloor \leq l \leq \lceil \log_2(s_{\max}(K)/s) \rceil$. $\square$

4 Sketching by group structured measurements with random sign

The third vignette demonstrates how Theorem 1.3 can be strengthened by introducing extra randomness to the group structured measurement operator. Specifically we show that the composition of the group structured measurement operator and the diagonal operator with random sign achieves a restricted near isometry on any set by the "Gordon-optimal" number of measurements. For the baseline of comparison, we first recall the result by Gordon [14] on the number of Gaussian measurements for sketching an arbitrary set.

Theorem 4.1 (Gordon's escape through the mesh [34, Theorem 1.2])**.** *Let $0 < \delta < 1$, T be a subset of the unit sphere $\mathbb{S}^{N-1}$, and $\xi_1, \dots, \xi_m$ be i.i.d.* $\mathrm{Normal}(0, b_m^{-2} I_N)$ *for $b_m =$*

$\sqrt{2}\Gamma(m/2)/\Gamma((m+1)/2)$, *where* Γ *denotes the Gamma function. Then*

$$\mathbb{P}\left\{\sup_{x\in T}\left|\sum_{j=1}^{m}|\langle\xi_j,x\rangle|^2-\|x\|_2^2\right|\geq\delta\right\}\leq\zeta$$

holds provided

$$m\geq\delta^{-2}\left(w(T)+\sqrt{2\ln(2/\zeta)}\right)^2 ,$$

where $w(T)$ *denotes the Gaussian width of* T *defined by*

$$w(T):=\mathbb{E}\sup_{x\in T}\langle\xi,x\rangle \ . \tag{31}$$

Remark 4.2. The Gaussian width in (31) satisfies $w(T)=w(\text{absconv}(T))$, where $\text{absconv}(T)$ denotes the absolute convex hull of D. It also coincides with the *Gaussian-summing norm*, also known as the ℓ-norm, of the identity operator from ℓ_2^N to Y, where Y is the Banach space with unit ball as $\text{absconv}(T)$ [11].

Further extensions of Theorem 4.1 showed that random matrices with either i.i.d. subgaussian entries or i.i.d. subgaussian rows also achieve a similar near optimal sample complexity result in Theorem 4.1 [21, 22, 29, 33]. However these random matrices do not provide a useful structure for fast computation. In contrast, the group structured measurement operator in Theorem 1.3 can describe Fourier transform and its generalization to a broader class including Gabor transform, short-time Fourier transform, and Radon transform. In a companion paper [20], we have shown that the number of group structured measurements m for RIP in Theorem 1.3 scales near optimally for certain sparsity models (e.g. sparsity models with respect to Banach spaces $X=\ell_1^N$ for the canonical sparsity or $X=S_1^n$ for low-rankness). However, in general, one might need a larger number of measurements for the group structured case than the Gaussian case by more than a logarithmic factor. To strength Theorem 1.3 so that it is comparable to Theorem 4.1, we adopt the idea of applying random sign before the structured measurement operator.

Oymak et al. [34] showed that the composition of a matrix providing the multiresolution RIP, which is formally defined below, and a random sign operator provides near isometric sketching of an arbitrary set where the number of measurements is slightly larger than that for Theorem 4.1 only by a logarithmic factor. This result generalizes an earlier work by Krahmer and Ward [23] that applies to any finite set. Below we provide a summary of the result by Oymak et al. [34] for the convenience of readers. Multiresolution RIP is formally defined as follows:

Definition 4.3 (Multiresolution RIP [34]). *We say that $A : \ell_2^N \to \ell_2^m$ satisfies multiresolution restricted isometry property (MRIP) with distortion level $\delta > 0$ and sparsity level $s \geq 1$ if*

$$\sup_{\|x\|_0 \leq 2^l s} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \leq \max(2^{l/2}\delta,\ 2^l \delta^2)$$

for all $l = 0, 1, \ldots, \lceil \log_2(N/s) \rceil$.

The following theorem by Oymak et al. [34] shows that a matrix with multiresolution RIP followed by random sign provides near isometric sketching.

Theorem 4.4 (Oymak et al. [34, Theorem 3.1]). *Let $T \subset \mathbb{S}^{N-1}$ and $D_\varepsilon \in \mathbb{R}^{N \times N}$ be a diagonal matrix with a Rademacher sequence on the diagonal. Then there exists a numerical constant c such that if $H \in \mathbb{R}^{m \times N}$ obeys MRIP with sparsity and distortion levels $s = 200(1 + \ln(\zeta^{-1}))$ and $\tilde{\delta} = c\delta / w(T)$, then for $\delta \in (0,1)$*

$$\mathbb{P}\left\{ \sup_{x \in T} \left| \|HD_\varepsilon x\|_2^2 - \|x\|_2^2 \right| \geq \delta \right\} \leq \zeta\ .$$

Combining the multiresolution RIP result for group structured measurements in Proposition 3.6 and Theorem 4.4 provides the following corollary.

Corollary 4.5. *Let $T \subset \mathbb{S}^{N-1}$ and D_ε be as in Theorem 4.4. Suppose that the hypothesis of Theorem 1.3 holds with $X = \ell_q^N$ for $q = \ln(N)/\ln(N/e)$. Then there exists a numerical constant c such that for $\delta \in (0,1)$*

$$\mathbb{P}\left\{ \sup_{x \in T} \left| \frac{1}{m} \sum_{j=1}^{m} |\langle D_\varepsilon \sigma(g_j)^* \eta, x \rangle|^2 - \|x\|_2^2 \right| \geq \delta \right\} \leq \zeta$$

provided

$$m \geq c\delta^{-2} w(T)^2 (1+\ln(\zeta^{-1}))^2 \max\left\{(1+\ln N)(1+\ln m)^5, 1+\ln(\zeta^{-1})\right\} \|\eta\|_\infty^2 .$$

Proof. The assertion follows by combining Proposition 3.6 for $K = B_q^N$ and Theorem 4.4. $\square$

To achieve near isometric sketching at the sample complexity comparable to Gordon's result, it is important for matrix A to satisfy RIP at a near optimal rate. Oymak et al. [34] considered a set of such random matrices including partial Fourier [37] and its generalization to a subsampled bounded orthogonal system [36]. Corollary 4.5 demonstrates that a group structure measurement operator, which is considered as a generalization of partial Fourier along a different perspective, also provides a Gordon-optimal isometric sketching by leveraging Theorem 1.3. For example, DSTFT in Section 2 does not belong to a set of bounded orthogonal systems.

For a special case of Corollary 4.5, we consider $T = K_s = \sqrt{s}K \cap \mathbb{S}^{N-1}$ for a fixed convex body $K \subset B_2^N$. Then we obtain the following corollary.

Corollary 4.6. *Suppose that the hypothesis of Corollary 4.5 holds. Then there exists a numerical constant c such that for $\delta \in (0,1)$*

$$\mathbb{P}\left\{ \sup_{\|x\|_X \leq \sqrt{s},\ \|x\|_2 = 1} \left| \frac{1}{m}\sum_{j=1}^m |\langle D_\varepsilon \sigma(g_j)^* \eta, x\rangle|^2 - \|x\|_2^2 \right| \geq \delta \right\} \leq \zeta$$

holds provided

$$m \geq c\delta^{-2} s w(K)^2 (1+\ln(\zeta^{-1}))^2 \max\left\{(1+\ln N)^2(1+\ln m)^5, 1+\ln(\zeta^{-1})\right\} \|\eta\|_\infty^2 . \quad (32)$$

Proof. Let $K_s = \sqrt{s}K \cap \mathbb{S}^{N-1}$. Then $w(K_s) \leq w(\sqrt{s}K) \leq \sqrt{s}w(K)$. The assertion follows from Corollary 4.5 as a special case where $T = K_s$. $\square$

Note that two key parameters $w(K)$ and $\|\eta\|_\infty$ determine the sufficient number of measurements for RIP in (32) for the generalized sparsity model given by Banach space X. The Gaussian width of the unit ball K describes the complexity of the model whereas the second parameter $\|\eta\|_\infty$ is corresponding to the incoherence of measurement functionals to the canonical sparsity model, which is totally independent of the model by X. Due to decoupling of the measurement operator and the sparsity model by random sign operator D_ε, the result holds without the G-invariance of the unit ball K in X. In other words, the group structured measurement operator with random sign provides a near isometry universally for any sparsity model.

One may deduce that it is always desirable to incorporate random sign to measurements. However, in certain scenarios, a group structured measurements without random sign can performs better due to its coupling to the sparsity model through the G-invariance. For example, in a high-dimensional case with very large N, sampling-based approximation of $\langle \eta, \sigma(g)x \rangle$ for selected group indices can accelerate sketching. Moreover, in an extreme case, where the model is built on an infinite-dimensional vector space, discretization similar to the Marcinkiewicz-type problem [41] is inevitable unless an equipment that directly takes measurements (like Fourier transform by a lens in optical imaging) is available. With the G-invariance of the model, random sampling can be utilized to apply empirical method universally to all selected group indices without increasing the approximation error. In contrast, incorporating random sign may lose the restrictive power of the model and result in significantly increased approximation error. If the Banach space X is a lattice, then it is also invariant under random sign and there will be no such penalty. However, there are models which are not invariant with entry-wise sign changes such as low-rank matrices/tensors and functions constrained by a Sobolev seminorm. Therefore we deduce that the group structured measurements tightly coupled with the sparsity model through G-invariance are still preferred over that with random sign in certain scenarios.

5 Extension to infinite-dimensional sparsity models

In the last section, we present an extension of Theorem 1.3 that applies to an infinite-dimensional "sparsity" model. We are interested in preserving (semi-) norms by finitely many measurements up to certain accuracy. Similar to its counterpart in finite dimension, this is possible only when the input is restricted appropriately. Particularly we consider sparsity models those given by coupling various (semi-) norms on the vector space of functions defined on the unit interval and measurements induced by a translation group. A focus will be given to explain how the theory for infinite-dimensional models deviates from the finite-dimensional counterpart. We start with an extension of Theorem 1.3 to an abstract infinite-dimensional model via the Fourier series, which will be made substantive with concrete examples that follow.

5.1 RIP on infinite-dimensional sparsity models

In Theorem 1.3, we considered a generalized sparsity model given as the set of vectors those satisfying $\|x\|_X \leq \sqrt{s}\|x\|_2$, where X is a Banach space in ℓ_2^N. In fact, the derivation of Theorem 1.3 does not rely on the fact that ℓ_2-norm satisfies the definiteness, i.e. $\|x\|_2 = 0$ implies $x = 0$. Moreover, it does not depend on ℓ_2^N being finite dimensional. Below we show that the result in Theorem 1.3 extends to an *infinite-dimensional* sparsity model in $L_2[0,1]$, which is obtained by substituting ℓ_2^N to a (semi-) normed space H_w and by choosing a Banach space X in $L_2[0,1]$. Specifically we consider a class of (semi-) norms on $L_2[0,1]$ defined by using the Fourier series representation. Let $(\hat{f}[k])_{k\in\mathbb{Z}}$ denote the Fourier series of the periodization of $f \in L_2[0,1]$, i.e.

$$\hat{f}[k] = \langle \psi_k, f \rangle = \int_0^1 \overline{\psi_k(t)} f(t) dt \,,$$

where $\psi_k : \mathbb{R} \to \mathbb{C}$ denotes the complex sinusoidal function defined by

$$\psi_k(t) = e^{\mathrm{i}2\pi kt} \,. \tag{33}$$

For a nonnegative sequence $(w_k)_{k\in\mathbb{Z}}$, we define a weighted (semi-) norm by

$$\|f\|_{2,w} \ = \ \left(\sum_{k\in\mathbb{Z}} w_k \left|\hat{f}[k]\right|^2\right)^{1/2} . \tag{34}$$

For example, if $w_k = k$ for all $k \in \mathbb{Z}$, then it follows that $\|f\|_{2,w} = \|f'\|_2$ for all $f \in C^1[0,1]$, where f' denotes a weak derivative of f. Let H_w denote the (semi-) normed space equipped with $\|\cdot\|_{2,w}$. Particularly if $w_k > 0$ for all $k \in \mathbb{Z}$, then $\|\cdot\|_{2,w}$ is a valid norm and H_w is a Hilbert space. Otherwise if $(w_k)_{k\in\mathbb{Z}}$ is finitely supported, then $\|\cdot\|_{2,w}$ is only a seminorm.

With the (semi-) norm defined in (34), we state the following theorem that extends Theorem 1.3 to an infinite-dimensional case.

Theorem 5.1. *Let X be the Banach space defined by the Minkowski functional of a convex body $K \subset B_2$, where B_2 denotes the unit ball in $L_2[0,1]$. Let $\sigma : G \to U$ be a continuous group homomorphism of a group G to the unitary group U acting on $L_2[0,1]$. Let $g_1, \ldots, g_m$ be independent copies of a Haar-distributed random variable on G. Let $u : X \to \ell_2^d$ be fixed. Suppose the following conditions hold: i) X^* has type 2; ii) X is a Banach lattice; iii) K is G-invariant; iv) $\mathbb{E}\sum_{j=1}^m \|u(\sigma(g_j)f)\|_2^2 = m\|f\|_{2,w}^2$ for all $f \in L_2[0,1]$, where $\|\cdot\|_{2,w}$ is a (semi-) norm as in (34); and v) $\sigma(g)$ commutes with any point-wise operation on $f \in L_2[0,1]$. Then there exists a numerical constant c such that for $\delta \in (0,1)$*

$$\sup_{\substack{\|f\|_X \le \sqrt{s} \\ \|f\|_{2,w}=1}} \left| \frac{1}{m}\sum_{j=1}^m \|u(\sigma(g_j)f)\|_2^2 - \|f\|_{2,w}^2 \right| \le \delta$$

holds with probability $1-\zeta$ provided

$$m \ \ge \ c\delta^{-2}s\left\{T_2(X^*)^2(1+\ln d)^5(1+\ln m)^5 + \ln(\zeta^{-1})\right\} \left\| \left(\sum_{l=1}^d |u^*(e_l)|^2\right)^{1/2} \right\|_{X^*}^2 , \tag{35}$$

where $e_1, \ldots, e_d$ denote the standard basis vectors in $\mathbb{R}^d$.

Theorem 5.1 can be thought of as an extension of Theorem 1.3 from the following perspectives: Most importantly, the sparsity model in Theorem 5.1 is built upon an infinite-dimensional vector space together with ℓ_2-norm substituted by a (semi-) norm $\|\cdot\|_{2,w}$, which may violate the definiteness. If the set $V = \{x : \|x\|_{2,w} = 0\}$ is a subspace, then the result of Theorem 5.1 implies a near isometry on the quotient space with respect to V. (One such case will be illustrated in the next section.) Furthermore the measurements in Theorem 5.1 are obtained via $u : L_2[0,1] \to \ell_2^d$ from an orbit of the input. Therefore each measurement is vector-valued whereas that in Theorem 1.3 is scalar-valued.

Proof of Theorem 5.1. It is straightforward to verify that [20, Theorem 2.1] still applies when an Hilbert space H in the original statement is substituted by a (semi-) normed space H_w with slight modifications described below. Specifically, the result of [20, Theorem 2.1] remains valid when the assumption

$$\sup_{\|f\|_H \le 1} \mathbb{E} \frac{1}{m} \sum_{j=1}^{m} \left\| u(\tau_{t_j} f) \right\|_2^2 \le 1 \tag{36}$$

is substituted by

$$\mathbb{E} \frac{1}{m} \sum_{j=1}^{m} \|u(\sigma(g_j) f)\|_2^2 = \|f\|_{2,w}^2 \,, \quad \forall f \,. \tag{37}$$

Indeed, the identity in (37) follows from the assumption that $\mathbb{E}\|u(\sigma(g_j)f)\|^2 = \|f\|_{2,w}^2$.

Set the parameter p in [20, Theorem 2.1] to 2 and choose the 1-homogeneous function α_d on linear maps from X to ℓ_2^d so that

$$\alpha_d : w \mapsto \left\| \left(\sum_{l=1}^{d} |w^*(e_l)|^2 \right)^{1/2} \right\|_{X^*} .$$

Since X is a Banach lattice, i.e. $\||f|\|_X = \|f\|_X$ for all $f \in X$, and X^* has type 2, it follows from [20, Theorem 3.11] that

$$M_{2,\alpha_d}(K) \lesssim T_2(X^*)(1 + \ln d)^{5/2} \,,$$

where $M_{2,\alpha_d}(K)$ is defined in [20, eq. (7)]. Therefore the assumptions of [20, Theorem 2.1] except (36) are satisfied. Recall that we substituted (36) by (37).

In the above setting, the modification of [20, Theorem 2.1] provides the tail bound in (42) if m satisfies

$$m \geq c\delta^{-2} s T_2(X^*)^2 (1+\ln d)^5 (1+\ln m)^5 \sup_{k\in\mathbb{N}} \left(\mathbb{E} \sup_{1\leq j\leq m} \alpha_d(u\tau_{t_j})^{2k} \right)^{1/k} \tag{38}$$

and

$$m \geq c\delta^{-2} s \sup_{k\in\mathbb{N}} \left(\mathbb{E} \sup_{1\leq j\leq m} \left\| u\tau_{t_j} : X \to \ell_2^d \right\|^{2k} \right)^{1/k} \tag{39}$$

for some constant c. It remains to show that (35) implies both (38) and (39).

First we show that $\alpha_d(u\tau_{t_j}) = \alpha_d(u)$ for all $j = 1, \ldots, m$. Then the last factor in the right-hand side of (38) simplifies to $\alpha_d(u)^2$. Indeed $\alpha_d(u\tau_{t_j})$ is written as

$$\alpha_d(u\tau_{t_j}) = \sup_{\|f\|_X \leq 1} \left\langle |f|, \left(\sum_{l=1}^d |\sigma(g_j)^* u^*(e_l)|^2 \right)^{1/2} \right\rangle .$$

Since $\sigma(g_j)$ commutes with both the point-wise square-root and point-wise magnitude operators, we have

$$\begin{aligned}
\left\langle |f|, \left(\sum_{l=1}^d |\sigma(g_j)^* u^*(e_l)|^2 \right)^{1/2} \right\rangle &= \left\langle |f|, \sigma(g_j)^* \left(\sum_{l=1}^d |u^*(e_l)|^2 \right)^{1/2} \right\rangle \\
&= \left\langle \sigma(g_j)|f|, \left(\sum_{l=1}^d |u^*(e_l)|^2 \right)^{1/2} \right\rangle \\
&= \left\langle \, |\sigma(g_j) f| \, , \left(\sum_{l=1}^d |u^*(e_l)|^2 \right)^{1/2} \right\rangle .
\end{aligned}$$

Therefore, by the assumption that K is G-invariant and X is a Banach lattice, we can continue as

$$\alpha_d(u\tau_{t_j}) = \sup_{\|f\|_X \leq 1} \left\langle |f|, \left(\sum_{l=1}^d |u^*(e_l)|^2 \right)^{1/2} \right\rangle = \alpha_d(u) .$$

Similarly we simplify the right-hand side of (39) by dropping the supremum over $k \in \mathbb{N}$. Since K is G-invariant, we have

$$\left\| u\sigma(g_j) : X \to \ell_2^d \right\| = \sup_{\|f\|_X \leq 1} \|u(\sigma(g_j) f)\|_2 = \sup_{\|f\|_X \leq 1} \|u(f)\|_2 = \left\| u : X \to \ell_2^d \right\|$$

for all $j = 1, \dots, m$. Furthermore, by Jensen's inequality, we have

$$\|u(f)\|_2 = \left(\sum_{l=1}^{d} |\langle e_l, u(f) \rangle|^2 \right)^{1/2} = \left(\sum_{l=1}^{d} |\langle u^*(e_l), f \rangle|^2 \right)^{1/2} \leq \left\langle \left(\sum_{l=1}^{d} |u^*(e_l)|^2 \right)^{1/2}, |f| \right\rangle ,$$

which, together with the fact that X is a Banach lattice, implies

$$\left\| u : X \to \ell_2^d \right\| = \sup_{\|f\|_X \leq 1} \|u(f)\|_2 \leq \sup_{\|f\|_X \leq 1} \left\langle \left(\sum_{l=1}^{d} |u^*(e_l)|^2 \right)^{1/2}, |f| \right\rangle = \alpha_d(u) .$$

Finally the assertion follows by plugging in the above estimates to (38) and (39). □

The next corollary follows from Theorem 5.1 as a special case where Banach space X is set to $L_q[0,1]$ for $1 < q \leq 2$ and group actions represent all circular time shifts as follows: Let τ_t denote the linear operator on $L_2[0,1]$ that maps f to its translation to the right by $t \in [0,1)$ modulo 1. Then $t \mapsto \tau_t$ can be considered as a group homomorphism from $[0,1)$ modulo 1 to a unitary group $\{\tau_t : t \in [0,1)\}$ consisting of all time shifts modulo 1.[2]

Corollary 5.2. *Let $q \in (1,2]$ and $0 < \delta < 1$. Let $\|\cdot\|_{2,w}$ be a (semi-) norm as in (34), where $(w_j)_{j\in\mathbb{Z}}$ is determined explicitly by $u : L_q[0,1] \to \ell_2^d$ as*

$$w_k \;=\; \|u(\psi_k)\|_2^2 \; . \tag{40}$$

Suppose that $t_1, \dots, t_m$ are independent copies of a uniform random variable on $[0,1)$. Then

$$\mathbb{E}\, \frac{1}{m} \sum_{j=1}^{m} \|u(\tau_{t_j} f)\|_2^2 = \|f\|_{2,w} \, , \quad \forall f \in L_2[0,1] \, . \tag{41}$$

Moreover there exists a numerical constant c such that

$$\mathbb{P} \left(\sup_{\substack{\|f\|_q \leq \sqrt{s} \\ \|f\|_{2,w}=1}} \left| \frac{1}{m} \sum_{j=1}^{m} \|u(\tau_{t_j} f)\|_2^2 - \|f\|_{2,w}^2 \right| \geq \delta \right) \leq \zeta \tag{42}$$

[2]A unitary group in infinite dimension is defined as an inductive limit.

holds provided

$$m \ \geq \ c\delta^{-2} s \left\{ q'(1+\ln d)^5 (1+\ln m)^5 + \ln(\zeta^{-1}) \right\} \left\| \left(\sum_{l=1}^{d} |u^*(e_l)|^2 \right)^{1/2} \right\|_{q'}^2 , \qquad (43)$$

where $e_1, \ldots, e_d$ *denote the standard basis vectors in* $\mathbb{R}^d$*.*

Proof. Let $j \in \{1, \ldots, m\}$ be arbitrary. Since ψ_j and ψ_k are complex sinusoids as in (33), we have

$$\mathbb{E}\langle \psi_k, (\tau_t^* u^* u \tau_t)\psi_j \rangle = \int_0^1 e^{\mathrm{i}2\pi(k-j)t} \langle \psi_k, (u^* u)(\psi_j) \rangle dt \ = \ \delta_{kj} \langle \psi_k, (u^* u)(\psi_j) \rangle ,$$

where δ_{kj} denotes the Kronecker delta. Then (40) implies that $\mathbb{E}\tau_t^* u^* u \tau_t$ is a Fourier multiplier with respect to $(w_k)_{k\in\mathbb{N}}$, i.e.

$$\mathbb{E}\tau_t^* u^* u \tau_t : f \mapsto \sum_{k\in\mathbb{Z}} w_k \psi_k \langle \psi_k, f \rangle .$$

Therefore we obtain

$$\mathbb{E}\|u(\tau_{t_j} f)\|_2^2 = \|f\|_{2,w}, \quad \forall j = 1, \ldots, m , \qquad (44)$$

which implies (41).

The second assertion is obtained as a consequence of Theorem 5.1. It only remains to verify that the assumptions of Theorem 5.1 are satisfied. Let q' satisfy $1/q + 1/q' = 1$. Then $L_{q'}[0,1]$ has type 2 with $T_2(L_{q'}[0,1]) \leq \sqrt{q'}$ [8, Lemma 3]. Furthermore, $L_q[0,1]$ is a Banach lattice and its unit ball is shift-invariant. The first assertion has already satisfied $\mathbb{E}\sum_{j=1}^m \|u(\tau_{t_j} f)\|_2^2 = m\|f\|_{2,w}^2$ for all f. Finally, it is straightforward to see that τ_{t_j} commutes with any point-wise operation. $\square$

Unlike the canonical sparsity model, where the sparsity level s counts the number of nonzero elements, the corresponding parameter in the infinite-dimensional sparsity model with Banach space $L_q[0,1]$ does not provide such a physical meaning in general. In the remainder of Section 5, we will illustrate the implication of Corollary 5.2 over two concrete examples of weight sequence $(w_k)_{k\in\mathbb{N}}$ with their matching sparsity models in finite dimensions. The assertion will be made substantive in the context of these examples providing physical interpretation.

5.2 Approximating continuous-time sparse signals

In the first example, we consider the acquisition of a continuous-time signal f that is sparsely supported within $[0,1)$ but not necessarily continuous. This model arises, for example, in imaging applications like seismology or functional magnetic resonance imaging, where singularities convey information. It is often considered satisfactory to acquire the signal up to certain frequency resolution, i.e. only the Fourier series coefficients $\hat{f}[k]$ at $k \in [-N,N)$ for some finite N. Let V denote a subspace of $L_2[0,1]$ given by

$$V = \Big\{ f \in L_2[0,1] \ : \ \hat{f}[k] = 0, \ \forall k \in [-N,N) \Big\}$$

and W be the quotient space $L_2[0,1]/V$. Then $\|\cdot\|_W$ is written as a seminorm $\|\cdot\|_{2,w}$ as in (34) for $(w_k)_{k\in\mathbb{Z}}$ given by

$$w_k = \begin{cases} 1 & -N \le k < N \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{45}$$

Our objective here is to construct a near isometric map from a subset of W into finite dimension, which will be made specific below. Particularly we show that it is possible to construct such a near isometric map with m group structured measurements for m much smaller than $2N$. The map will be restricted to a set of continuous-time sparse signals defined as

$$K_{\rho,\gamma} := \big\{ f \in L_2[0,1] \ : \ \|f'\|_2 \ \le \ \rho \|f\|_2, \ \lambda(\mathrm{supp}(f)) \ \le \ \gamma \big\} \ , \tag{46}$$

where f' denotes a weak derivative of f [1] and λ denotes the normalized Lebesgue measure on $[0,1)$. Note that $K_{\rho,\gamma}$ is restricted to sparse signals, those with small Sobolev $(1,2)$-seminorm relative to L_2-norm.

The next lemma shows that $K_{\rho,\gamma}$ is indeed a subset of the restricted domain in Corollary 5.2. Therefore, we can utilize Corollary 5.2 to derive a near isometric map with group structured measurements.

Lemma 5.3. *Let $(w_k)_{k\in\mathbb{Z}}$ and $K_{\rho,\gamma}$ be as in (45) and (46), respectively. Suppose that $\rho \leq N/2$. Then $f \in K_{\rho,\gamma}$ implies $\|f\|_q \leq \sqrt{s}\|f\|_{2,w}$ for $q \in (1,2]$ and $s = (1+4\rho^2/N^2)\gamma^{2/q-1}$.*

Proof. By Parseval's theorem, we have

$$\sum_{|k|\geq N} \left|\hat{f}[k]\right|^2 \;\leq\; N^{-2}\sum_{k}\left|k\hat{f}[k]\right|^2 \;=\; N^{-2}\|f'\|_2^2 \,.$$

This together with $\|f'\| \leq \rho\,\|f\|_2$ implies

$$\|f\|_2^2 \leq \|f\|_{2,w}^2 + \sum_{|k|\geq N}\left|\hat{f}[k]\right|^2 \leq \|f\|_{2,w}^2 + \frac{\rho^2\,\|f\|_2^2}{N^2}\,.$$

Therefore, by using the fact that $(1-t)^{-2} \leq (1+4t)$ for $0 < t \leq 1/4$ and by the assumption $N \geq 2\rho$, we deduce

$$\|f\|_2 \;\leq\; (1-\rho^2/N^2)^{-1}\|f\|_{2,w} \;\leq\; (1+4\rho^2/N^2)^{1/2}\|f\|_{2,w}\,.$$

Let E be the support of f and $\mathbb{1}_E(t)$ denote the indicator function of E. Since $|E| \leq \gamma$, by Hölder's inequality, we obtain

$$\|f\|_q^q \;\leq\; \int \mathbb{1}_E(t)|f(t)|^q dt \;\leq\; \gamma^{1-q/2}\left(\int |f(t)|^{q(2/q)}dt\right)^{q/2}.$$

This implies

$$\|f\|_q \;\leq\; \gamma^{1/q-1/2}\|f\|_2 \leq (1+4\rho^2/N^2)^{1/2}\gamma^{1/q-1/2}\|f\|_{2,w}\,.$$

Then the assertion follows by letting $s = (1+4\rho^2/N^2)\gamma^{2/q-1}$. $\square$

The following lemma presents a concrete example that belongs to $K_{\rho,\gamma}$, which consists of a superposition of narrow pulses. The parameters ρ and γ are explicitly determined by the number of pulses and by the common pulse shape.

Lemma 5.4. *Let ϕ be a positive function with $\phi(0) = 1$ and support contained in $[-1/2, 1/2)$. Let $\phi_T(x) = T\phi(Tx)$ and $t_1, \ldots, t_l \in (1/2T, 1-1/2T)$ be $1/T$-separated, i.e., $|t_j - t_{j'}| > 1/T$ for all $j \neq j'$. Then*

$$f(t) \;=\; \sum_{j=1}^{l} \alpha_j \phi_T(t - t_j)$$

satisfies

$$\lambda(\mathrm{supp}(f)) \;\leq\; \frac{l}{T}$$

and

$$\frac{\|f'\|_2}{\|f\|_2} \;=\; \frac{T\|\phi'\|_2}{\|\phi\|_2}\,.$$

Proof. The first assertion follows since f is the sum of disjointly supported functions. Moreover

$$\|f\|_p^p = \sum_{j=1}^{n} |\alpha_j|^p \|\phi_T\|_p^p = \sum_{j=1}^{n} |\alpha_j|^p T^{p-1} \|\phi\|_p^p\,.$$

Indeed, we deduce from a change of variable that

$$\|\phi_T\|_p^p \;=\; T^{p-1} \int |\phi(Tx)|^p \, T dx \;=\; T^{p-1} \|\phi\|_p^p\,.$$

This yields

$$\|f\|_p \;=\; \|\phi\|_p T^{1/p'} \left(\sum_{j=1}^{l} |\alpha_j|^p \right)^{1/p}\,.$$

Note that ϕ_T' is also disjointly supported, and hence

$$\left\|f'\right\|_p^p \;=\; \sum_{j=1}^{l} |\alpha_j|^p \left\|\phi_T'\right\|_p^p\,.$$

Then we note that

$$\int \left|\phi_T'(x)\right|^p dx \;=\; T^{2p-1} \int \left|\phi'(x)\right|^p dx \;=\; T^{2p-1} \left\|\phi'\right\|_p^p .$$

This implies

$$\left\|f'\right\|_p \;=\; \left(\sum_{j=1}^n |\alpha_j|^p \right)^{1/p} T^{2-1/p} \left\|\phi'\right\|_p \;=\; \frac{\|\phi'\|_p T \|f\|_p}{\|\phi\|_p} .$$

The second assertion follows as a special case. □

The signal model in Lemma 5.4 consists of a superposition of translations and dilations of ϕ with a compact support. This model can be considered as a generalization of cardinal B-spline [43]. Particularly the translation parameters $t_1, \ldots, t_l$ in Lemma 5.4 are not necessarily on a uniform grid whereas knots in cardinal B-spline are integer-valued.

Next we proceed with deriving a near isometric map on $K_{\rho.\gamma}$ by using Corollary 5.2. Note that there exists more than one linear operator u so that $(w_k)_{k\in\mathbb{Z}}$ given by (40) satisfies (45). We will consider two constructions of u. The following proposition employs u that takes partial sums of the input Fourier series over a partition of $[-N, N) \cap \mathbb{Z}$.

Proposition 5.5. *Let $0 < \delta < 1$, $(w_k)_{k\in\mathbb{Z}}$ be as in (45), and $u : L_2[0,1] \to \ell_2^d$ be defined by*

$$u : f \mapsto \left(\sum_{k\in\mathcal{J}_l} \hat{f}[k] \right)_{l=1}^d ,$$

where $\mathcal{J}_1, \ldots, \mathcal{J}_d$ partition $[-N, N) \cap \mathbb{Z}$ so that $L := \max_{1\le l\le d} |\mathcal{J}_l| \le \lceil 2N/d \rceil$. Suppose that $t_1, \ldots, t_m$ are independent copies of a uniform random variable on $[0,1)$ and $\rho \le N/2$. Then there exists a constant c such that

$$\mathbb{P}\left(\sup_{\substack{f\in K_{\rho,\gamma} \\ \|f\|_{2,w}=1}} \left| \frac{1}{m} \sum_{j=1}^m \|u(\tau_{t_j} f)\|_2^2 - \|f\|_{2,w}^2 \right| \ge \delta \right) \le \zeta \tag{47}$$

holds provided

$$m \ge \frac{c\delta^{-2}\gamma L N}{\max\{1, \gamma L\}} \ln\left(\frac{e}{\min\{1, \gamma L\}} \right) \left\{ \ln\left(\frac{e}{\min\{1, \gamma L\}} \right) (1+\ln d)^5 (1+\ln m)^5 + \ln(\zeta^{-1}) \right\} .$$

Proof of Proposition 5.5. Let $q \in (1,2]$. Then Lemma 5.3 implies that every $f \in K_{\rho,\gamma}$ satisfies

$$\|f\|_q \leq \sqrt{s}\|f\|_{2,w}$$

for $s = (1+4\rho^2/N^2)\gamma^{2/q-1}$. Note that $u(f)$ is written as

$$u(f) = (\langle h_l, f\rangle)_{l=1}^d \ ,$$

where

$$h_l = \sum_{k\in\mathcal{J}_l} \psi_k$$

for $l = 1,\dots,d$. For q' satisfying $1/q + 1/q' = 1$ and α_d as in Corollary 5.2, we have

$$\alpha_d(u) = \left\| \left(\sum_{l=1}^d |h_l|^2 \right)^{1/2} \right\|_{q'} \leq \sqrt{2}\,\mathbb{E} \left\| \sum_{l=1}^d \varepsilon_l h_l \right\|_{q'} \leq \sqrt{2q'd} \max_{1\leq l\leq d} \|h_i\|_{q'} \lesssim \sqrt{q'd} L^{1-1/q'}$$

for a Rademacher sequence $(\varepsilon_l)_{1\leq l\leq d}$, where the first and second inequalities respectively follow from Khintchine's inequality [15] and $T_2(L_{q'}[0,1]) \leq \sqrt{q'}$, and the last step follows from the fact that the inverse Fourier series operator from $\ell_q(\mathbb{N})$ to $L_{q'}[0,1]$ is a contraction (see e.g. [46, Theorem IV.1]). Furthermore, for any $t \in [0,1)$, we have

$$\|u(\tau_t f)\|_2^2 = \sum_{l=1}^d \sum_{k\in\mathcal{J}_l} \left| e^{-\mathrm{i}2\pi kt} \hat{f}[k] \right|^2 = \sum_{k=-N}^{N-1} \left| \hat{f}[k] \right|^2 = \|f\|_{2,w}^2 \ .$$

Therefore, by Corollary 5.2, it suffices to satisfy

$$m \ \geq \ c\delta^{-2} \left(1 + \frac{4\rho^2}{N^2} \right) \gamma L^2 q' d (\gamma L)^{-2/q'} \left\{ q'(1+\ln d)^5 (1+\ln m)^5 + \ln(\zeta^{-1}) \right\} \ .$$

We choose $q' = 2$ if $\gamma L > 1$ and $q' = -2\ln(\gamma L)$ otherwise. Then the assertion follows since $Ld \leq 4N$. ☐

Since $f \in V$ implies both $\|f\|_{2,w} = 0$ and $u(\tau_t f) = 0$ for all $t \in [0,1)$, the tail bound in (47) indeed implies that the measurement operator provides a near isometric map on $K_{\rho,\gamma}$ in the quotient space $W = L_2[0,1]/V$ with high probability. One may choose d large enough so that $\gamma L = O(1)$. Then the number of translations m for this result can be as small as γNL up to a logarithmic factor. Therefore the total number of measurements md satisfies $\tilde{O}(\gamma N^2)$. Furthermore, if $\gamma N = O(1)$, then the number of measurements scales sublinearly in the dimension $2N$ of the reconstruction space ℓ_2^{2N}. Such small $\gamma = O(1/N)$ can be interesting in certain infinite-dimensional scenarios like the one presented below.

In general, one cannot recover a sparse signal in an infinite-dimensional space from finitely many measurements. For example, when unknown sparse signal $f \in L_2[0,1]$ is supported on a set of nonzero measure, at least a subsequence of the Fourier series at the Landau rate is necessary [24]. Known exceptions include the case where the unknown f corresponds to a point measure, that is a superposition of finitely many Dirac's delta functions (e.g. see [45, 7, 40]). We compare Proposition 5.5 to these results as follows: In one hand, we have a slightly different goal to sketch continuous-time sparse signals up to finite frequency resolution instead of exact recovery. This approximation is well accepted in imaging applications. Meanwhile our model still avoid unnecessary discretization in the time domain. On the other hand, Proposition 5.5 applies to a much wider class of signals, which can be useful particularly in the context of imaging. Note that the previous results apply only to point measures supported on a null set. In contrast, Proposition 5.5 applies to continuous-time sparse signals whose support sets have nonzero measure. Modeling For example, curves in 2D signals or surfaces in 3D signals correspond to a set of measure zero (hence one can choose N arbitrarily large in Proposition 5.5). However they are not described as a finite superposition of Dirac's delta functions.

Recall that the construction of u in Proposition 5.5 is purely deterministic. The following result shows that the total number of measurements md for a near isometric map can be significantly reduced to $\tilde{O}(\gamma Nd)$ when the deterministic u in Proposition 5.5 is substituted by a random linear operator.

Proposition 5.6. *Suppose that the hypothesis of Proposition 5.5 holds except that* u :

$L_2[0,1] \to \ell_2^d$ *is defined by*

$$u : f \mapsto \left(\sum_{k \in \mathcal{J}_l} \varepsilon_k \hat{f}[k] \right)_{l=1}^{d} ,$$

where $(\varepsilon_k)_{-N \le k < N}$ *is a Rademacher sequence independent of everything else. Then there exists a numerical constant* c *such that* (47) *holds provided*

$$m \;\ge\; c\delta^{-2}\gamma N \left\{ \ln(\zeta^{-1}) + \ln(\gamma^{-1}) \right\}^2 (1 + \ln d)^5 (1 + \ln m)^5 \, .$$

Proof of Proposition 5.6. Note that $u(f)$ is written as

$$u(f) = (\langle h_l, f \rangle)_{l=1}^{d} \, ,$$

where

$$h_l = \sum_{k \in \mathcal{J}_l} \varepsilon_k \psi_k$$

for $l = 1, \ldots, d$. By construction, the corresponding sequence $(w_k)_{k \in \mathbb{Z}}$ by (40) satisfies (45).

Let α_d be as in Corollary 5.2, that is

$$\alpha_d(u) = \left\| \left(\sum_{l=1}^{d} |h_l|^2 \right)^{1/2} \right\|_{q'} \, .$$

Then $\alpha_d(u)$ is a random variable due to the randomness in u. We compute a tail bound on α_d as follows. Let $(\varepsilon'_l)_{1 \le l \le d}$ be a Rademacher sequence independent of everything else. Then

$$\mathbb{E}_{\varepsilon,\varepsilon'} \left\| \sum_{l=1}^{d} \varepsilon'_l h_l \right\|_{q'}^{q'} = \mathbb{E}_{\varepsilon,\varepsilon'} \left\| \sum_{l=1}^{d} \varepsilon'_l \sum_{k \in \mathcal{J}_l} \varepsilon_k \psi_k \right\|_{q'}^{q'} \lesssim \mathbb{E}_{\varepsilon} \left\| \sum_{k=-N}^{N-1} \varepsilon_k \psi_k \right\|_{q'}^{q'} , \tag{48}$$

where $(\varepsilon''_k)_{-N \le k < N}$ is a Rademacher sequences independent of everything else. By applying Khintchine's inequality to the right-hand side of (48), we obtain

$$\left(\mathbb{E}_{\varepsilon,\varepsilon'} \left\| \sum_{l=1}^{d} \varepsilon'_l h_l \right\|_{q'}^{q'} \right)^{1/q'} \lesssim \sqrt{q' N} \, . \tag{49}$$

Since q' was arbitrary, this upper bound indeed holds for any $q' \geq 2$. From (49), we derive an upper bound on the moments of $\alpha_d(u)$. For $r > q' \geq 2$, by Khintchine's inequality, we obtain

$$
\begin{aligned}
(\mathbb{E}\alpha_d(u)^r)^{1/r} &= \left(\mathbb{E} \left\| \left(\sum_{l=1}^{d} |h_l|^2 \right)^{1/2} \right\|_{q'}^{r} \right)^{1/r} \leq \sqrt{2} \left(\mathbb{E} \left\| \sum_{l=1}^{d} \varepsilon_l' h_l \right\|_{q'}^{r} \right)^{1/r} \\
&\leq \sqrt{2} \left(\mathbb{E} \left\| \sum_{l=1}^{d} \varepsilon_l' h_l \right\|_{r}^{r} \right)^{1/r} \lesssim \sqrt{rN} \,.
\end{aligned}
$$

For $r \leq q'$, by Kahane's inequality [25], we obtain

$$
\begin{aligned}
(\mathbb{E}\alpha_d(u)^r)^{1/r} &= \left(\mathbb{E} \left\| \left(\sum_{l=1}^{d} |h_l|^2 \right)^{1/2} \right\|_{q'}^{r} \right)^{1/r} \leq \sqrt{2} \left(\mathbb{E} \left\| \sum_{l=1}^{d} \varepsilon_l' h_l \right\|_{q'}^{r} \right)^{1/r} \\
&\leq \sqrt{2} \left(\mathbb{E} \left\| \sum_{l=1}^{d} \varepsilon_l' h_l \right\|_{q'}^{q'} \right)^{1/q'} \lesssim \sqrt{q'N} \,.
\end{aligned}
$$

Therefore $(\mathbb{E}\alpha_d(u)^r)^{1/r} \lesssim \max\left\{\sqrt{r}, \sqrt{q'}\right\} N^{1/2}$ for all $r \geq 1$, from which together with a consequence of Markov's inequality [12, Lemma A.1], we obtain that

$$
\alpha_d(u) \lesssim \sqrt{N\left(\ln(\zeta^{-1}) + q'\right)} \tag{50}
$$

holds with probability $1 - \zeta/2$. Conditioned on the event in (50), Corollary 5.2 provides that (47) holds with probability $1 - \zeta/2$ if

$$
m \;\geq\; c\delta^{-2}\left(1 + \frac{4\rho^2}{N^2}\right)\gamma N \gamma^{-2/q'}\left(\ln(\zeta^{-1}) + q'\right)\max\left\{q'(1+\ln d)^5(1+\ln m)^5, \ln(\zeta^{-1})\right\} \,.
$$

Finally the assertion follows by letting $q' = -2\ln\gamma$. □

5.3 Approximating Sobolev seminorm constrained signals

A total variation seminorm has been employed as an effective regularizer for denoising [38]. Particularly it is shown superior to other regularizers in terms of preserving edges without incurring severe blurring. Another well-known regularizer for denoising that provides similar strength is the qth power of the Sobolev $(1, q)$-seminorm for $q > 1$ [3], which is defined as the L_q norm of weak derivative [1]. Motivated by these regularized denoising, we propose two near isometric maps for signals $g \in L_2[0,1] \cap W^{1,q}[0,1]$ that satisfy

$$\left\|g'\right\|_q \leq \sqrt{s}\,\|g\|_2 \ , \tag{51}$$

where g' denotes a weak derivative of g, and

$$\hat{g}[0] = 0 \ , \tag{52}$$

where $W^{1,q}[0,1]$ denotes the Sobolev space equipped with the Sobolev $(1, q)$-seminorm. The first condition in (51) is crucial for providing another infinite-dimensional sparsity model, whereas the latter condition in (52) is for the sake of simple presentation and can be dropped with slight modifications of linear maps. In fact, the Sobolev $(1, q)$-seminorm becomes a valid norm in the subspace given by (52).

We first present a near isometric map given by time-domain measurements. The following theorem shows that one can approximate the L_2-norm of $g \in L_2[0,1] \cap W^{1,q}[0,1]$ satisfying (51) and (52) by finitely many random time samples.

Theorem 5.7. *Let $1 < q \leq 2$ and $0 < \delta < 1$. Suppose that $t_1, \ldots, t_m$ are independent copies of a uniform random variable on $[0,1)$. Then there exists a constant C_q depending only on q such that*

$$\mathbb{P}\left(\sup_{\substack{\|g'\|_q \leq \sqrt{s},\ \|g\|_2 = 1 \\ \hat{g}[0]=0}} \left| \frac{1}{m} \sum_{j=1}^{m} |g(t_j)|^2 - \|g\|_2^2 \right| \geq \delta \right) \leq \zeta$$

holds provided

$$m \geq C_q \delta^{-2} s \left\{ q'(1+\ln m)^5 + \ln(\zeta^{-1}) \right\} \ . \tag{53}$$

Moreover, (53) *also implies*

$$\mathbb{P}\left(\sup_{\substack{\|g'\|_q \leq \sqrt{s}, \\ \|g\|_2^2 - |\hat{g}[0]|^2 = 1}} \left| \frac{1}{m} \sum_{j=1}^{m} |g(t_j) - \hat{g}[0]|^2 - \left(\|g\|_2^2 - |\hat{g}[0]|^2 \right) \right| \geq \delta \right) \leq \zeta \ .$$

The result of Theorem 5.7 is similar to the Monte-Carlo method for Lipschitz continuous functions (e.g. see [44, Section 8.2]). Both results show that an empirical process approximates its expectation. Note that the signal model in Theorem 5.7 is based on the Sobolev $(1, q)$-seminorm, which is defined through weak derivatives, and include signals with singularities. This flexible signal model can accurately describe signals in imaging applications like neural activation in functional imaging or anomalies in anatomical imaging.

Proof of Theorem 5.7. For g satisfying $\|g'\|_q \leq \sqrt{s}$, $\|g\|_2 = 1$, and $\hat{g}[0] = 0$, let $f = g'$ denote a weak derivative of g. Furthermore, let $\|\cdot\|_{2,w}$ be a seminorm in (34) determined by $(w_k)_{k\in\mathbb{Z}}$ that satisfies

$$w_k = \frac{1}{4\pi^2 \max(1, k^2)} \ .$$

Then since $\hat{f}[k] = \mathfrak{i}2\pi k \hat{g}[k]$ for all $k \in \mathbb{Z}$ we have

$$\|f\|_{2,w}^2 = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{|\hat{f}[k]|^2}{4\pi^2 k^2} = \sum_{k\in\mathbb{Z}\setminus\{0\}} |\hat{g}[k]|^2 = \|g\|_2^2 = 1 \ .$$

We also have $\|f\|_q \leq \sqrt{s}\|f\|_{2,w}$. Moreover, the map $u : f \mapsto \langle h, f\rangle$ with

$$h = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\psi_k}{\mathfrak{i}2\pi k}$$

satisfies

$$u(f) = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\langle\psi_k, f\rangle}{\mathfrak{i}2\pi k} = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\hat{f}[k]}{\mathfrak{i}2\pi k} = \sum_{k\in\mathbb{Z}\setminus\{0\}} \hat{g}[k] = g(0) \ ,$$

where the last step used the fact $\hat{g}[0] = 0$. Since the weak derivative and shift operators commute, we also have

$$u(\tau_{-t_j} f) = g(t_j) \ .$$

Let α_d be as in Corollary 5.2. Note that $d = 1$, i.e. u is a linear functional. Therefore, $\alpha_d(u)$ simplifies and is upper-bounded by the Hausdorff-Young inequality as

$$\alpha_d(u) = \|h\|_{q'} \leq \frac{1}{2\pi}\left\|\left(\frac{1}{j}\right)_{j\in\mathbb{Z}\setminus\{0\}}\right\|_{\ell_q(\mathbb{Z}\setminus\{0\})} \leq \ C_q$$

for some constant C_q determined by q. Then Corollary 5.2 with the above estimates provides that

$$\mathbb{P}\left(\sup_{\|f\|_q\le\sqrt{s},\ \|f\|_{2,w}\ =\ 1}\left|\frac{1}{m}\sum_{j=1}^{m}|u(\tau_{-t_j}f)|^2-\|f\|_{2,w}^2\right|\ge\delta\right)\le\zeta$$

holds if (53) is satisfied. This indeed implies the first assertion. In general, without $\hat{g}[0]=0$, we have

$$\|g'\|_{2,w}^2+|\hat{g}[0]|^2=\|g\|_2^2$$

and

$$u(\tau_{t_j}g')+\hat{g}[0]=g(t_j)\ .$$

Therefore the second assertion is obtained by plugging in these identities to the first assertion. $\square$

Next we present another near isometric map, which is given by Fourier-domain measurements. In certain modalities like magnetic resonance imaging (MRI), measurements are acquired sequentially in the Fourier domain. Therefore one can evaluate time samples only after acquiring the full Fourier series. In such scenarios, it is preferable to design the measurement operator in the Fourier domain and the following theorem presents the analogous result.

Theorem 5.8. *Let $1<q\le 2$, $0<\delta<1$, and $\mathcal{I}_l=\{k\ :\ 2^{l-2}<|k|\le 2^{l-1}\}$ for $l\in\mathbb{N}$. Let $t_1,\ldots,t_m$ be independent copies of a uniform random variable on $[0,1)$. Then there exist a numerical constants c and a constant C_q that only depends on q such that*

$$\mathbb{P}\left\{\sup_{\substack{\|g'\|_q\le\sqrt{s},\ \|g\|_2=1\\ \hat{g}[0]=0}}\left|\frac{1}{m}\sum_{j=1}^{m}\sum_{1\le l<l_0}\left|\sum_{k\in\mathcal{I}_l}\widehat{\tau_{t_j}g}[k]\right|^2-\|g\|_2^2\right|\ge\delta\right\}\le\zeta$$

provided

$$l_0\ \ge\ \max\left\{1,c\ln\left(\delta^{-1}s\right)\right\}\tag{54}$$

and

$$m \geq C_q \delta^{-2} s \left(q' (1 + \ln m)^5 + \ln(\zeta^{-1}) \right\} \ . \tag{55}$$

The constant C_q in Theorem 5.8 is proportional to $f(q) = (q-1)^{-1/q}$. Recall that as q decreases toward 1, the sparsity model converges to the total-variation-sparsity model. However, since $f(q)$ increases in q, one cannot set q arbitrarily close to 1. For example, for $q \geq 1.1$, we have $f(q) \leq 8.12$.

Note that each scalar measurement in Theorem 5.8 is computed as the sum of Fourier series coefficients of a translation of g over a given dyadic interval. For example in MRI, this superposition can be obtained without access to the individual summands with an appropriate design of the pulse sequence. The total number of measurements $l_0 m$ scales at most $\tilde{O}(s \ln s)$. Specifically $\tilde{O}(s)$ random translations and $O(\ln s)$ measurements per translation suffice to invoke Theorem 5.8.

Proof of Theorem 5.8. By the triangle inequality, it suffices to show

$$\sup_{\substack{\|g'\|_q \leq \sqrt{s},\ \|g\|_2 = 1 \\ \hat{g}[0]=0,\ t \in [0,1)}} \sum_{l \geq l_0} \left| \sum_{k \in \mathcal{I}_l} \widehat{\tau_t g}[k] \right|^2 \leq \frac{\delta}{2} \tag{56}$$

and

$$\mathbb{P} \left\{ \sup_{\substack{\|g'\|_q \leq \sqrt{s},\ \|g\|_2 = 1 \\ \hat{g}[0]=0}} \left| \frac{1}{m} \sum_{j=1}^{m} \sum_{l=1}^{\infty} \left| \sum_{k \in \mathcal{I}_l} \widehat{\tau_{t_j} g}[k] \right|^2 - \|g\|_2^2 \right| \geq \frac{\delta}{2} \right\} \leq \zeta \tag{57}$$

hold.

First we show that (54) implies (56). Let $f = g'$ and $t \in [0,1)$. Since the weak derivative operator commutes with the shift operator, we have $\tau_t f = (\tau_t g)'$. Therefore it follows that

$\widehat{\tau_t f}[k] = \mathfrak{i} 2\pi k \widehat{\tau_t g}[k]$ for all $k \in \mathbb{Z}$. Then we have

$$\begin{aligned}
\sum_{l \geq l_0} \left| \sum_{k \in \mathcal{I}_l} \widehat{\tau_t g}[k] \right|^2 &= \sum_{l \,\geq\, l_0} \left| \left\langle \sum_{k \in \mathcal{I}_l} \frac{\psi_k}{2\pi k}, \tau_t f \right\rangle \right|^2 \\
&\leq \sum_{l \,\geq\, l_0} \|\tau_t f\|_q^2 \left\| \sum_{k \in \mathcal{I}_l} \frac{\psi_k}{2\pi k} \right\|_{q'}^2 \leq \frac{s}{4\pi^2} \sum_{l \,\geq\, l_0} \left(\sum_{k \in \mathcal{I}_l} |k|^{-q} \right)^{2/q} \qquad (58)
\end{aligned}$$

where the first two inequalities follow from Hölder's inequality and the Hausdorff-Young inequality. Furthermore, we have

$$\begin{aligned}
\left(\sum_{k \in \mathcal{I}_l} |k|^{-q} \right)^{2/q} &\leq \left(2 \int_{2^{l-1}}^{2^l} t^{-q} dt \right)^{2/q} \\
&= \left(\frac{2^{l(1-q)+1}(2^{q-1} - 1)}{q-1} \right)^{2/q} = 2^{-2l/q'} \underbrace{\left(\frac{2(2^{q-1} - 1)}{q-1} \right)^{2/q}}_{C_q'} .
\end{aligned}$$

Note that C_q' is monotone increasing in $q \in (1, \infty)$ and $C_q' \leq 4$ for all $q > 1$. Therefore (58) implies that there exists a numerical constant C such that

$$\sum_{l \geq l_0} \left| \sum_{k \in \mathcal{I}_l} \widehat{\tau_t g}[k] \right|^2 \leq \frac{s}{\pi^2} \sum_{l \,\geq\, l_0} 2^{-2l/q'} \;\leq\; C s 2^{-2l_0/q'} .$$

Then (56) follows by choosing l_0 so that

$$2^{-2l_0/q'} \;\leq\; \frac{\delta}{2Cs} ,$$

which is implied by (54).

Next we show that (55) implies (57) by using the following lemma, the proof of which is deferred to the end of this section.

Lemma 5.9. *Let $1 < q \leq 2$, $0 < \delta < 1$, and $\|\cdot\|_{2,w}$ defined by (34) and (40) from $u : f \mapsto (\langle h_l, f \rangle)_{l \in \mathbb{N}}$. Suppose that $t_1, \ldots, t_m$ are independent copies of a uniform random variable on $[0, 1)$. Then there exists a numerical constant c such that*

$$\mathbb{P} \left\{ \sup_{\|f\|_q \leq \sqrt{s},\ \|f\|_{2,w} = 1} \left| \frac{1}{m} \sum_{j=1}^{m} \sum_{l=1}^{\infty} \left\| u(\tau_{t_j} f) \right\|_2^2 - \|f\|_{2,w}^2 \right| \geq \delta \right\} \leq \zeta$$

holds provided

$$m \geq c\delta^{-2}s \left\{ q'(1+\ln m)^5 \left(\sum_{l=1}^{\infty} (1+\ln d_l)^{5/2} \|h_l\|_{q'} \right)^2 + \ln(\zeta^{-1}) \left(\sum_{l=1}^{\infty} \|h_l\|_{q'} \right)^2 \right\} . \tag{59}$$

We apply Lemma 5.9 with $(h_l)_{l\in\mathbb{N}}$ given by

$$h_l = \sum_{k\in\mathcal{I}_l} \frac{\psi_k}{2\pi k} .$$

Then $(w_k)_{k\in\mathbb{Z}}$ determined by u as in (40) satisfies

$$w_k = \|u(\phi_k)\|_2^2 = \sum_{l=1}^{\infty} |\langle h_l, \psi_k \rangle|^2 = \frac{1}{4\pi^2 k^2}$$

for all $k \neq 0$ and $w_0 = 0$. Let $f = g'$ be a weak derivative of $g \in L_2[0,1] \cap W^{1,q}[0,1]$. Since $\hat{f}[k] = \mathfrak{i}2\pi k \hat{g}[k]$ for all $k \in \mathbb{Z}$, we have $\hat{f}[0] = 0$. Thus we have

$$\|f\|_{2,w}^2 \;=\; \sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{|\hat{f}[k]|^2}{4\pi^2 k^2} \;=\; \sum_{j\in\mathbb{Z}\setminus\{0\}} |\hat{g}[k]|^2 \;=\; \|g\|_2^2 ,$$

where the last step follows since $\hat{g}[0] = 0$. Furthermore, by the Hausdorff-Young inequality, we have

$$\|h_l\|_{q'} = \left\| \sum_{2^{l-2}<|k|\leq 2^{l-1}} \frac{\psi_k}{2\pi k} \right\|_{q'} \leq \frac{1}{\pi} \left(\sum_{k\,\geq\,2^l} k^{-q} \right)^{1/q} = \frac{(2^l-1)^{-1/q'}}{\pi(q-1)^{1/q}} \leq \frac{2^{-(l-1)/q'}}{\pi(q-1)^{1/q}} .$$

Therefore (55) implies (59) and Lemma 5.9 provides (57). $\square$

Finally we provide the proof of Lemma 5.9.

Proof of Lemma 5.9. We adapt the proof of [20, Proposition 2.6] to prove Lemma 5.9. As shown earlier in the proof of Corollary 5.2, the weighted (semi-) norm satisfies

$$\mathbb{E} \left\| u(\tau_{t_j}(f)) \right\|_2^2 = \|f\|_{2,w}^2$$

for all $j = 1, \ldots, m$. For notational simplicity, let

$$D := \{ f : \|f\|_q \leq \sqrt{s},\ \|f\|_{2,w} = 1 \} \tag{60}$$

and

$$Z := \sup_{f\in D}\left|\frac{1}{m}\sum_{j=1}^{m}\|u(\tau_{t_j}f)\|_2^2 - \|f\|_{2,w}^2\right|.$$

For $r\in\mathbb{N}$, by the standard symmetrization technique [26, Lemma 6.3] we obtain

$$(\mathbb{E}Z^r)^{1/r} \lesssim \left(\mathbb{E}\sup_{f\in D}\left|\frac{1}{m}\sum_{j=1}^{m}\xi_j\|u(\tau_{t_j}f)\|_2^2\right|^r\right)^{1/r},$$

where $\xi_1,\ldots,\xi_m$ are i.i.d. Normal$(0,1)$. Conditioned on $t_1,\ldots,t_m$, by the triangle inequality, we obtain

$$\left(\mathbb{E}\sup_{f\in D}\left|\frac{1}{m}\sum_{j=1}^{m}\xi_j\|u(\tau_{t_j}f)\|_2^2\right|^r\right)^{1/r} \le \sum_{l=1}^{\infty}\left(\mathbb{E}\sup_{f\in D}\left|\frac{1}{m}\sum_{j=1}^{m}\xi_j|\langle h_l,\tau_{t_j}f\rangle|^2\right|^r\right)^{1/r}. \qquad (61)$$

For $l\in\mathbb{N}$, let $v_j : L_q[0,1]\to\ell_\infty^m$ denote a linear operator defined by $v_l : f\mapsto(\langle h_l,\tau_{t_j}f\rangle)_{1\le j\le m}$. Then for all $l\in\mathbb{N}$ by applying [20, Lemma 2.4] we obtain[3]

$$\left(\mathbb{E}_{\xi_1,\ldots,\xi_m}\sup_{f\in D}\left|\sum_{j=1}^{m}\xi_j|\langle h_l,\tau_{t_j}f\rangle|^2\right|^r\right)^{1/r} \lesssim \sqrt{s}\left(\sup_{f\in D}\sum_{j=1}^{m}|\langle h_l,\tau_{t_j}f\rangle|^2\right)^{1/2}\Big(\mathcal{E}_{2,1}(v_l)+\sqrt{r}\,\|v_l\|\Big),$$

where $\mathcal{E}_{2,1}$ denotes a weighted sum of dyadic entropy numbers [see 20, Lemma 2.3] that provides an upper bound of Talagrand's γ_2-functional [39]. Then by Hölder's inequality we can continue as follows:

$$\begin{aligned}(\mathbb{E}Z^r)^{1/r} &\lesssim \frac{\sqrt{s}}{m}\sum_{l=1}^{\infty}\left\{\mathbb{E}\left(\sup_{f\in D}\sum_{j=1}^{m}|\langle h_l,\tau_{t_j}f\rangle|^2\right)^r\right\}^{1/2r}\left\{\mathbb{E}\Big(\mathcal{E}_{2,1}(v_l)+\sqrt{r}\,\|v_l\|\Big)^{2r}\right\}^{1/2r}\\ &\lesssim \frac{\sqrt{s}}{m}\sup_{l\in\mathbb{N}}\left\{\mathbb{E}\left(\sup_{f\in D}\sum_{j=1}^{m}|\langle h_l,\tau_{t_j}f\rangle|^2\right)^r\right\}^{1/2r}\\ &\quad\cdot\left(\sum_{l=1}^{\infty}\left(\mathbb{E}\mathcal{E}_{2,1}(v_l)^{2r}\right)^{1/2r}+\sqrt{r}\left(\mathbb{E}\,\|v_l\|^{2r}\right)^{1/2r}\right).\end{aligned}$$

[3]The original version of [20, Lemma 2.4] is stated for a set D of unit vectors in a finite dimensional Hilbert space. However, since the proof does not rely on this assumption, the result of [20, Lemma 2.4] also applies to the set D defined in (60) in a semi-normed space.

Note that the factor after $\sqrt{s}/m$ of the right-hand side is bounded from above as

$$\begin{aligned}
\sup_{l\in\mathbb{N}} \left\{ \mathbb{E}\left(\sup_{f\in D} \sum_{j=1}^{m} |\langle h_l, \tau_{t_j} f\rangle|^2 \right)^r \right\}^{1/2r} &\le \left\{ \mathbb{E}\left(\sup_{f\in D, l\in\mathbb{N}} \sum_{j=1}^{m} |\langle h_l, \tau_{t_j} f\rangle|^2 \right)^r \right\}^{1/2r} \\
\le \left\{ \mathbb{E}\left(\sup_{f\in D} \sum_{j=1}^{m} \sum_{l=1}^{\infty} |\langle h_l, \tau_{t_j} f\rangle|^2 \right)^r \right\}^{1/2r} &= \left\{ \mathbb{E}\left(\sup_{f\in D} \sum_{j=1}^{m} \|u(\tau_{t_j} f)\|_2^2 \right)^r \right\}^{1/2r} \\
\le \left\{ \left(\mathbb{E}\left| \sup_{f\in D} \sum_{j=1}^{m} \|u(\tau_{t_j} f)\|_2^2 - \|f\|_{2,w}^2 \right|^r \right)^{1/r} + m \right\}^{1/2} &= \sqrt{m}\left\{ (\mathbb{E}Z^r)^{1/r} + 1 \right\}^{1/2} .
\end{aligned}$$

The last factor is also upper-bounded by the following facts: By the shift-invariance of the unit ball in $L_{q'}[0,1]$ we have

$$\|v_l\| = \max_{1\le j\le m} \|\tau_{t_j}^* h_l\|_{q'} = \|h_l\|_{q'} .$$

Moreover, by [20, Lemma 3.7], we also have

$$\mathcal{E}_{2,1}(v_l) \lesssim \sqrt{q'}(1+\ln m)^{5/2} \|h_l\|_{q'} .$$

Then by collecting the above upper bounds we obtain

$$(\mathbb{E}Z^r)^{1/r} \lesssim \frac{\sqrt{s}\,\varrho_1}{\sqrt{m}} + \frac{s\varrho_1^2}{m} + \sqrt{r}\cdot \frac{\sqrt{s}\,\varrho_2}{\sqrt{m}} + r\cdot \frac{s\varrho_2^2}{m} , \tag{62}$$

where

$$\varrho_1 = \sum_{l=1}^{\infty} \|h_q\|_{q'} , \quad \varrho_2 = \sum_{l=1}^{\infty} \sqrt{q'}(1+\ln m)^{5/2} \|h_l\|_{q'} .$$

Since r was arbitrary, by a consequence of Markov's inequality [12, Lemma A.1], (62) for all $r \ge 1$ implies

$$Z \lesssim \frac{\sqrt{s}\,\varrho_1}{\sqrt{m}} + \frac{s\varrho_1^2}{m} + \frac{\sqrt{s\ln(\zeta^{-1})}\,\varrho_2}{\sqrt{m}} + \frac{s\ln(\zeta^{-1})\varrho_2^2}{m} \tag{63}$$

with probability $1-\zeta$. Therefore (59) implies that the upper bound in (63) is at most δ. $\square$

Acknowledgement

This work was supported in part by NSF IIS-1447879, DMS-1501103, CCF-1718771, DMS-1800872, and an NSF CAREER award CCF-1943201. K.L. thanks Yoram Bresler, Sjoerd Dirksen, Massimo Fornasier, Laurent Jacques, Felix Krahmer, Holger Rauhut, and Dominik Stöger for helpful discussions. The authors also thank the anonymous reviewers for their careful reading of the manuscript and their many insightful comments and suggestions.

References

[1] R. Adams and J. Fournier. *Sobolev Spaces*. Elsevier Science, 2003.

[2] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Foundations of Computational Mathematics*, 16(5):1263–1323, 2016.

[3] G. Aubert and P. Kornprobst. Can the nonlocal characterization of Sobolev spaces by Bourgain et al. be useful for solving variational problems? *SIAM Journal on Numerical Analysis*, 47(2):844–860, 2009.

[4] J.-M. Azais, Y. De Castro, and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

[5] N. Boyd, G. Schiebinger, and B. Recht. The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*, 27(2):616–639, 2017.

[6] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

[7] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[8] B. Carl. Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces. In *Annales de l'institut Fourier*, volume 35, pages 79–118, 1985.

[9] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.

[10] Y. Chen and M. J. Wainwright. Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. *arXiv preprint arXiv:1509.03025*, 2015.

[11] J. Diestel, H. Jarchow, and A. Tonge. *Absolutely summing operators*, volume 43. Cambridge University Press, 1995.

[12] S. Dirksen. Tail bounds via generic chaining. *Electronic Journal of Probability*, 20 (53), 2015.

[13] C. Dossal, V. Duval, and C. Poon. Sampling the fourier transform along radial lines. *SIAM Journal on Numerical Analysis*, 55(6):2540–2564, 2017.

[14] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. In *Geometric Aspects of Functional Analysis*, pages 84–106. Springer, 1988.

[15] U. Haagerup. The best constants in the Khintchine inequality. *Studia Mathematica*, 70(3):231–283, 1981.

[16] R. Heckel, V. I. Morgenshtern, and M. Soltanolkotabi. Super-resolution radar. *Information and Inference: A Journal of the IMA*, 5(1):22–75, 2016.

[17] C. J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.

[18] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, 2013.

[19] M. Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pages 427–435, 2013.

[20] M. Junge and K. Lee. Generalized notions of sparsity and restricted isometry property. part I: A unified framework. *Information and Inference: A Journal of the IMA*, 9(1): 157–193, 2020.

[21] B. Klartag and S. Mendelson. Empirical processes and random projections. *Journal of Functional Analysis*, 225(1):229–245, 2005.

[22] V. Koltchinskii and S. Mendelson. Bounding the smallest singular value of a random matrix without concentration. *International Mathematics Research Notices*, 2015(23): 12991–13008, 2015.

[23] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3): 1269–1281, 2011.

[24] H. J. Landau. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Mathematica*, 117(1):37–52, 1967.

[25] R. Latała and K. Oleszkiewicz. On the best constant in the khinchin-kahane inequality. *Studia Mathematica*, 109(1):101–104, 1994.

[26] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[27] Y.-K. Liu. Universal low-rank matrix recovery from Pauli measurements. In *Advances in Neural Information Processing Systems*, pages 1638–1646, 2011.

[28] B. Maurey. Type, cotype and k-convexity. In *Handbook of the Geometry of Banach Spaces*, volume 2, pages 1299–1332. Elsevier, 2003.

[29] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geometric and Functional Analysis*, 17(4):1248–1282, 2007.

[30] G. Ongie and M. Jacob. Recovery of piecewise smooth images from few Fourier samples. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 543–547. IEEE, 2015.

[31] G. Ongie and M. Jacob. Off-the-grid recovery of piecewise constant images from few Fourier samples. *SIAM journal on imaging sciences*, 9(3):1004–1041, 2016.

[32] G. Ongie, S. Biswas, and M. Jacob. Convex recovery of continuous domain piecewise constant images from nonuniform Fourier samples. *IEEE Transactions on Signal Processing*, 66(1):236–250, 2017.

[33] S. Oymak and J. A. Tropp. Universality laws for randomized dimension reduction, with applications. *Information and Inference: A Journal of the IMA*, 7(3):337–446, 2018.

[34] S. Oymak, B. Recht, and M. Soltanolkotabi. Isometric sketching of any set via the restricted isometry property. *Information and Inference: A Journal of the IMA*, 7(4): 707–726, 2018.

[35] H. Pan, T. Blu, and P. L. Dragotti. Sampling curves with finite rate of innovation. *IEEE Transactions on Signal Processing*, 62(2):458–471, 2013.

[36] H. Rauhut. Compressive sensing and structured random matrices. In *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9, pages 1–92. De Gruyter, 2010.

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

[38] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[39] M. Talagrand. Majorizing measures: the generic chaining. *The Annals of Probability*, pages 1049–1103, 1996.

[40] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013.

[41] V. Temlyakov. The Marcinkiewicz-type discretization theorems. *Constructive Approximation*, 48(2):337–369, 2018.

[42] Y. Traonmilin and R. Gribonval. Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all. *Applied and Computational Harmonic Analysis*, 45 (1):170–205, 2018.

[43] M. Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6):22–38, 1999.

[44] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

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

[46] P. Xu and M. Madiman. The norm of the Fourier series operator. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pages 750–754, 2015.

[47] Q. Zheng and J. Lafferty. Convergence analysis for rectangular matrix completion using Burer-Monteiro factorization and gradient descent. *arXiv preprint arXiv:1605.07051*, 2016.