Generalized notions of sparsity and restricted isometry property. Part I: A unified framework

MARIUS JUNGE,

Department of Mathematics,

University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

mjunge@illinois.edu

KIRYUNG LEE*

Department of Electrical and Computer Engineering,

The Ohio State University, Columbus, OH 43210, USA

*Corresponding author: lee.8763@osu.edu

August 22, 2018

Abstract

The restricted isometry property (RIP) is an integral tool in the analysis of various inverse problems with sparsity models. Motivated by the applications of compressed sensing and dimensionality reduction of low-rank tensors, we propose generalized notions of sparsity and provide a unified framework for the corresponding RIP, in particular when combined with isotropic group actions. Our results extend an approach by Rudelson and Vershynin to a much broader context including commutative and non-commutative function spaces. Moreover, our Banach space notion of sparsity applies to affine group actions. The generalized approach in particular applies to high order tensor products.

1 Introduction

The *restricted isometry property* (RIP) has been used as a universal tool in the analysis of many modern inverse problems with sparsity prior models. Indeed, the RIP implies that certain linear maps act as near isometries when restricted to “nice” (or sparse) vectors. Motivated from emerging big data applications such as compressed sensing or dimensionality reduction of massively sized data with a low-rank tensor structure, we provide a unified framework for the RIP allowing a generalized notion of sparsity and extend the existing theory to a much broader context.

Let us recall that in compressed sensing the RIP played a crucial role in providing guarantees for the recovery of sparse vectors from a small number of observations. Moreover, these guarantees were achieved by practical polynomial-time algorithms (e.g., [10, 42]). In machine learning, the RIP enabled a fast and guaranteed dimensionality reduction of data with a sparsity structure. The notion of sparsity has been shown for various models and in many cases the RIP turns out to hold near optimally in terms of the scaling of parameters for several classes of random linear operators. For example, a linear map with random subgaussian entries satisfies a near optimal RIP for the canonical sparsity model [10, 3, 28], low-rank matrix model [40, 9], low-rank tensor model [39], and manifold models [19]. Baraniuk et al. [3] provided an alternative elementary derivation that combines exponential concentration of a subgaussian quadratic form and standard geometric argument with union bounds.

Linear operators with special structures such as subsampled Fourier transform arise in practical applications. These structures are naturally given by the physics of applications (e.g., Fourier imaging) and subsampled versions of these structured linear operators can be implemented within existing physical systems. Furthermore, structured linear operators also enable scalable implementation at low computational cost, which is highly desirable for dimensionality reduction. A partial Fourier operator has been shown to satisfy a near optimal RIP for the canonical sparsity model in the context of compressed sensing

[11, 42, 38]. The linear operator that generates randomly sampled Pauli measurements in quantum tomography was also shown to satisfy a near optimal RIP for a low-rank matrix model [31]. The RIP of other structured random matrices such as block diagonal matrices and subsampled circulant matrices has been shown too [20, 41, 28].

Although the RIP under certain scenarios of sparsity models and structured random matrix as in the above examples has been studied in the literature, there are still applications whose setting does not fit in the existing theory because the classical sparsity model does not hold and/or the assumptions on the linear operator are not satisfied. To develop theory for such scenarios, in this paper, we extend the notion of sparsity and the RIP for structured linear operators in several ways described below.

1.1 Generalized notion of sparsity

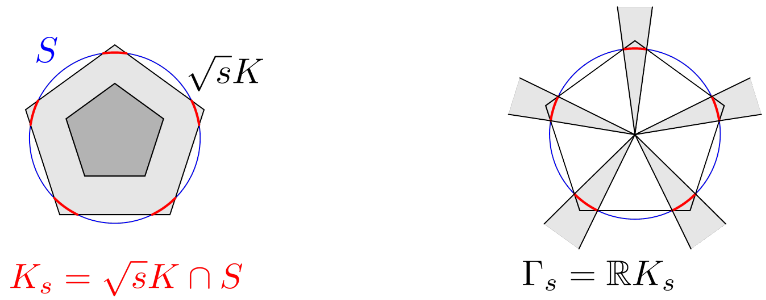
We first generalize the notion of sparsity. Let H be a Hilbert space and $K \subset H$ be a centered convex body. We will consider the Banach space $(X, \|\cdot\|_X)$ obtained by completing the linear span of K with the norm $\|\cdot\|_X$ given as the Minkowski functional $p_K(\cdot) : X \to \mathbb{R}$, which is defined by $p_K(x) := \inf\{\lambda > 0 \mid x \in \lambda K\}$. Then a generalized sparsity model is induced from X and H as follows:

Definition 1.1. *A vector $x \in H$ is (K, s)-sparse if*

$$\|x\|_X \leq \sqrt{s}\|x\|_H \ ,$$

where X is the Banach space with unit ball K.

The set of (K, s)-sparse unit vector in H, denoted by K_s, is geometrically given as the intersection of $\sqrt{s}K$ and the unit sphere $S = \{x \mid \|x\|_H = 1\}$ in H. Then the set of (K, s)-sparse vectors, denoted by Γ_s, is the star-shaped nonconvex cone given by $\mathbb{R}K_s$ (or $\mathbb{C}K_s$ if the scalar field is complex). These two sets are visualized in Figure 1. For example, if $H = \ell_2^N$ and $X = \ell_1^N$, then Γ_s corresponds to the set of approximately s-sparse vectors with respect to the canonical basis. The authors of this paper showed that



(a) (b)

Figure 1: Visualization of an abstract sparsity model using a convex set K and the unit sphere S in a Hilbert space H. (a) The set of s-sparse unit vectors (red). (b) The set of s-sparse vectors (gray-shaded).

existing near optimal RIP results extend from the exact canonical sparsity model to this *approximately* sparse model [26]. This generalized notion of sparsity covers a wider class of models beyond the classical atomic model. For example, in the companion paper [27, Section 4], we demonstrate a case where a sparse vector is not represented as a finite linear combination of atoms but defined by an infinite dimensional Banach space. It also allows a machinery that optimizes the sample complexity for the RIP of a given atomic sparsity model by choosing an appropriate Banach space (see [27, Section 2]). In a special case, where the "sparsity level" s in Definition 1.1 is set to 1, our theory covers an arbitrary set, where K is its convex hull.[1] The Minkowski functional in the generalized sparsity model is closely related to the atomic norm in the signal processing literature. Chandrasekaran et al. [16] presented a unified theory where the regularization with an atomic norm, which is the Minkowski functional given by the convex hull of an atomic set, to inverse problems induces a sparse representation of the solution in the corresponding atomic model. Here we define a generalized sparsity model directly from a convex set K, which is not necessarily derived from a specific atomic model.

[1] Note that taking the convex hull of a given set does not increase the number of measurements for RIP. Therefore, the convex set K can be considered as the convex hull of a given set of interest in this case.

1.2 Vector-valued measurements

Next we consider vector-valued measurements which generalize the conventional scalar-valued measurements. This situation arises in several practical applications. For example, in medical imaging and multi-dimensional signal acquisition, measurements are taken by sampling transform of the input not individually but in blocks. The performance of ℓ_1-norm minimization has been analyzed in this setting [37, 6] and it was shown that block sampling scheme, enforced by applications, adds a penalty to the number of measurements for the recovery. This analysis extends the noiseless part of the analogous theory for the scalar-valued measurements [8], which relies on a property called *local isometry*, which is a weaker version of the RIP. For stable recovery from noisy measurements, one essentially needs the RIP of the measurement operator but block sampling setting does not fit to existing RIP results for structured linear operators. In this paper, we will consider general vector-valued measurements in a Hilbert space and generalize the notion of incoherence and other properties accordingly. This extension, in particular combined with a generalized sparsity model, requires the use of theory of factorization of a linear operator in Banach spaces [33].

1.3 Sparsity with enough symmetries and group-structured RIP

We also generalize the theory of the RIP for a partial Fourier measurement operator to more general group-structured measurement operators, which will exploit the inherent structure in the Banach space that determines a sparsity model. Particularly, we consider Banach spaces with "enough symmetries" described below. Let G be a group and $\sigma : G \to O_N$ be an affine representation that maps an element in G to the orthogonal group O_N in $\mathbb{R}^N$. An affine representation is isotropic if averaging the conjugate actions on any linear operator becomes a scalar multiple of the identity. A convex set K has *enough symmetries* if there exists an isotropic affine representation such that $\sigma(g)K = K$ for all $g \in G$. A Banach space has enough symmetries if its unit ball does. Finite-dimensional

Banach spaces with enough symmetries have been studied extensively (see [46, 17, 33]). For example, the Banach space ℓ_1^N, which induces the "approximate" canonical sparsity model, has enough symmetries (see Section 4.2 for more details). Our original motivation for this generalization comes from studying the low-rank tensor model in $\ell_2^n \otimes_\pi \ell_2^n \otimes_\pi \ell_2^n$, where $\otimes_\pi$ denotes the projective tensor product.[2] In fact a nice feature of spaces with enough symmetries comes from their stability under tensor products.

When X has enough symmetries with respect to G with an affine representation σ, let $v : X \to \ell_\infty^m$ be a linear operator given by $v(x) = [v_1(x), \ldots, v_m(x)]^\top$, where $v_1, \ldots, v_m \in X^*$ are obtained by sampling the (adjoint) orbit $\{\sigma(g)^*\eta \mid g \in G\}$ of $\eta \in X^*$, i.e. $v_j(x) = \langle \eta, \sigma(g_j)x \rangle$ with $g_j \in G$ for $j = 1, \ldots, m$. For a certain class of group actions, the corresponding group-structured measurement operator $A = (1/\sqrt{m})v$ has fast implementation. For example, the Banach space ℓ_1^N has enough symmetries with respect to the group actions consisting of circular shifts in the canonical basis and in the Fourier basis. When $\eta = [1, \ldots, 1]^\top \in \mathbb{R}^N$, the orbit of η by circular shifts in the Fourier domain generates the Fourier basis and the resulting matrix $A : \mathbb{C}^N \to \mathbb{C}^m$ becomes a partial discrete Fourier transform. In general, A corresponds to a windowed partial Fourier transform where the window is determined by η. Fast algorithms exist for computing these transforms. We will demonstrate the RIP of these group-structured measurement operators holds with high probability when the group elements are randomly selected.

Again, the group-structured measurement operator is a natural extension of a partial Fourier operator. Unlike the other extension to subsampled bounded orthogonal system [38], the group-structured measurement operator is tightly connected to a given generalized sparsity model through the underlying Banach space.

[2] For normed spaces X and Y, the projective tensor product, denoted by $X \otimes_\pi Y$, is defined by the norm $\|T\|_\pi := \inf\{\sum_k \|x_k\|\|y_k\| \mid T = \sum_k x_k \otimes y_k,\ x_k \in X,\ y_k \in Y\}$ (see e.g. [34]).

1.4 Main results

We illustrate our main results in the general setting on concrete examples of random group-structured measurement operators and generalized sparsity models. The next two theorems commonly assume the following conditions: i) $H = \ell_2^N$; ii) The convex set K that determines the sparsity model has enough symmetries with respect to a compact group G and an isotropic affine representation $\sigma : G \to O_N$; iii) The Banach space X is given by the norm given as the Minkowski functional of K as before. A set of random measurements are obtained by randomly sampling group actions. Specifically, we assume that $g_1, \ldots, g_m$ are independent copies of a Haar-distributed random variable g in G.

The first theorem demonstrates our main result in the case where K is a polytope given as an absolute convex hull of finitely many vectors.[3]

Theorem 1.2 (Polytope)**.** *Let K be an absolute convex hull of M points in $\mathbb{C}^N$, X be the Banach space with the norm $\|\cdot\|_X$ given as the Minkowki functional of K, $u : X \to \ell_2^d$ be a linear map that satisfies $\mathrm{tr}(u^* u) = N$. Suppose that X has enough symmetries with respect to a group G and an isotropic affine representation $\sigma : G \to O_N$. Let $g_1, \ldots, g_m$ be independent copies of a Haar-distributed random variable g in G. Then*

$$\sup_{\substack{\|x\|_X \le \sqrt{s} \\ \|x\|_2 = 1}} \Big| \frac{1}{m} \sum_{j=1}^m \|u(\sigma(g_j)x)\|_2^2 - \|x\|_2^2 \Big| \;\le\; \delta \vee \delta^2$$

holds with high probability for $m = O(\delta^{-2} s \|u\|_{X \to \ell_2^d}^2 (1 + \ln m)(1 + \ln(md))^2 (1 + \ln M))$.

Theorem 1.2 generalizes the RIP result of a partial Fourier operator (e.g., [42]) in the three ways discussed above. The operator norm of u in Theorem 1.2 generalizes the notion of incoherence in existing theory. Most interestingly, combined with a clever net argument, Theorem 1.2 enables the RIP of a random group-structured measurement operator for low-rank tensors (see Section 6).

[3] The absolute convex hull of a set S is defined by $\{\sum_{k=1}^n \lambda_k x_k \mid n \in \mathbb{N},\ x_k \in S,\ \lambda_k \in \mathbb{K},\ \sum_{k=1}^n |\lambda_k| \le 1\}$, where $\mathbb{K}$ denotes the underlying scalar field.

The second theorem deals with the sparsity model with respect to a "nice" Banach space whose norm dual has type p [34]. (Details are explained in Section 3.) Here for simplicity we only demonstrate an example where $p = 2$.

Theorem 1.3 (Dual of type 2)**.** *Let X be an N-dimensional Banach space in ℓ_2^N such that i) The norm dual X^* has type 2; ii) X has enough symmetries with respect to a group G and an isotropic affine representation σ. Let $\eta \in X^*$ satisfy $\|\eta\|_2 = \sqrt{N}$ and $g_1, \ldots, g_m$ be independent copies of a Haar-distributed random variable g in G. Then*

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

holds with high probability for $m = O(\delta^{-2} s [T_2(X^*)]^6 \|\eta\|_{X^*} (1 + \ln m)^3)$, where $T_2(X^*)$ denotes the type-2 constant of X^*.

Theorem 1.3 covers many known results on the RIP of structured random linear operator and should be considered as an umbrella result for this theory. Importantly Theorem 1.3 applies to noncommutative cases such as Schatten classes. For example, Theorem 1.3 implies the previous RIP result for a partial Pauli operator applied to low-rank matrices [31] as a special case.

Theorems 1.2 and 1.3 are just exemplar of the main result in full generality given in Theorem 2.1. Indeed, the enough symmetries of the underlying Banach space of a generalized sparsity model is only a sufficient condition that guarantees that the expectation of the random measurements preserves the norm. In the companion paper [27], we demonstrate that Theorem 2.1 also provides theory for the RIP for infinite-dimensional sparsity models without relying on the enough symmetry with isotropic group actions.

1.5 Notation

In this paper, the symbols $c, c_1, c_2, \ldots$ and $C, C_1, C_2, \ldots$ will be reserved for numerical constants, which might vary from line to line. We will use notation for various Banach spaces and norms. The identity operator on Banach space X will be denoted by Id_X. For a linear map $T : X \to Y$ between Banach spaces X and Y, the operator norm of T will be denoted by $\|T\|_{X \to Y}$. (The subscript will be dropped when the corresponding Banach spaces are obvious from the context.) We will use the shorthand notation $\|\cdot\|_p$ and B_p^N respectively for the norm and unit ball in ℓ_p^N for $1 \leq p \leq \infty$ and $N \in \mathbb{N}$. For a set $\mathcal{I} \subset \mathbb{Z}$, let $(\boldsymbol{e}_j)_{j\in\mathcal{I}}$ denote the canonical basis for $\mathbb{C}^{|\mathcal{I}|}$. The index set $\mathcal{I}$ should be clear from the context. For a linear operators $v_j : X \to Y$ for $j = 1, \ldots, m$, the composition map, denoted by $(v_j)_{1\leq j\leq m}$, is defined by $[(v_j)_{1\leq j\leq m}](x) = (v_1(x), \ldots, v_m(x)) \in Y^m$ for $x \in X$, where Y^m denotes the Cartesian product of m copies of Y given by $\prod_{1\leq j\leq m} Y$.

1.6 Organization

The rest of this paper is organized as follows: The main theorem in full generality is stated and proved in Section 2. Various examples of generalized sparsity models and their complexity are discussed in Section 3, followed by the illustration of Banach spaces with enough symmetries for various affine group representations in Section 4. By collecting these results, we derive rigorous versions of Theorems 1.2 and 1.3 in Section 5. Finally, we conclude the paper with the application of the main results for a low-rank tensor model in Section 6.

2 Rudelson-Vershynin method

In this section, we derive a unified framework that identifies a sufficient number of measurements for the RIP of structured random operators in the general setting introduced in Sections 1.1 and 1.2. We will start with the statement of the property in the general setting, followed by the proof.

2.1 RIP in the general setting

Let H be a Hilbert spaces, K be a centrally symmetric convex set in H, and X be the Banach space with the norm given as the Minkowski functional of K. Let

$$\Gamma_s := \{x \in X \mid \|x\|_X \leq \sqrt{s}\|x\|_H\} \tag{1}$$

denote the set of (K, s)-sparse vectors and K_s be the intersection of Γ_s and the unit sphere in H. Let $v_1, \ldots, v_m$ be independent random linear operator from X to ℓ_2^d. For notational simplicity, we let $v : X \to \ell_\infty^m(\ell_2^d)$ denote the composite map $(v_j)_{1 \leq j \leq m}$. Then v generates a set of m vector-valued linear measurements in ℓ_2^d. We will consider the measurement operator given as the normalized map $A = (1/\sqrt{m})v$.

Our results are stated for a class of incoherent random measurement operators. We adopt the arguments by Candes and Plan [8] to describe these measurement operators. In the special case of $X = \ell_1^N$ and $d = 1$, Candes and Plan considered a class of linear operators given by measurement maps satisfying the following two key properties. i) Isotropy: $\mathbb{E}v_j^* v_j = I_N$ for all $j = 1, \ldots, m$, where I_N denotes the identity matrix of size N; ii) Incoherence: $\|v_j\|_\infty$ is upper-bounded by a numerical constant μ (deterministically or with high probability). In our setting, the isotropy extends to

$$\mathbb{E}v_j^* v_j = \mathrm{Id}\ . \tag{2}$$

But we will also consider the case where

$$\mathbb{E}v_j^* v_j = \Phi, \quad \forall j = 1, \ldots, m \tag{3}$$

holds with $\Phi : H \to H$ satisfying

$$\|\Phi\| \leq 1\ . \tag{4}$$

Obviously, the isotropy is a sufficient condition for the relaxed properties in (3) and (4).

Non-isotropic cases where the expectation of a random operator is not an isometry has been also studied in specific scenarios of compressed sensing [8] and embedding of dynamic systems [21].

We also generalize the notion of incoherence by using an 1-homogeneous function $\alpha_d : B(X, \ell_2^d) \to [0, \infty)$ that maps a bounded linear map from X to ℓ_2^d to a nonnegative number. A natural choice of α_d is the operator norm, which is consistent with the above example of $K = B_1^N$ and $d = 1$. The operator norm of v_j in this case reduces to $\|v_j\|_{X^*} = \|v_j\|_\infty$. However, in certain scenarios, there exists a better choice of α_d than the operator norm that further reduces the sample complexity that identifies a sufficient number of measurements for the RIP. One such example is demonstrated for the windowed Fourier transform in the companion paper [27, Section 2].

Under the relaxed isotropy conditions in (3) and (4), with a slight abuse of terminology, we say that A satisfies the RIP on Γ_s with constant δ if

$$\left| \|Ax\|_{\ell_2^m(\ell_2^d)}^2 - \langle x, \Phi x \rangle \right| \leq (\delta \vee \delta^2) \|x\|^2 , \quad \forall x \in \Gamma_s . \tag{5}$$

In the special case where the isotropy ($\Phi = \mathrm{Id}$) is satisfied, the deviation inequality in (5) reduces to the conventional RIP. Note that Φ is a nonnegative operator by construction. If Φ is a positive operator, then $\langle x, \Phi x \rangle$ is a weighted norm of x and (5) preserves this weighted norm through w with a small perturbation proportional to $\|x\|_H^2$.

Our main result is a far reaching generalization of the RIP of a partial Fourier operator by Rudelson and Vershynin [42]. We adapt their derivation that consists of the following two steps: The first step is to show that the expectation of the restricted isometry constant is upper-bounded by the γ_2-functional [44] of the restriction set, then by an integral of the metric entropy number by Dudley's theorem [29]. Later in this section, we show that the first step extends to the general setting with the upper bound given by

$$\int_0^\infty \sqrt{\ln N(v(K), \varepsilon B_{\ell_\infty^m(\ell_2^d)})} \, d\epsilon \leq C \sum_{l=0}^\infty \frac{e_l(v)}{\sqrt{l}} =: \mathcal{E}_{2,1}(v) , \tag{6}$$

where $N(\cdot,\cdot)$ and $e_l(\cdot)$ respectively denote the covering number and the dyadic entropy number [14]. The second step is where our theory deviates significantly from the previous work [42]. In the scalar-valued measurement case ($d = 1$), Rudelson and Vershynin used a variation of Maurey's empirical method [12] to get an upper bound on the integral in (6) for K being the unit ball in ℓ_1^N, which in turn provided a near optimal sample complexity up to a logarithmic factor. Liu [31] later extended the result by Rudelson and Vershynin [42] to the case of a partial Pauli operator applied to low-rank matrices via the dual entropy argument by Guédon et al. [24].

Our result further generalizes these results. In particular, our result provides flexibility that can address the vector-valued measurement case and optimize sample complexity over the choice of the 1-homogeneous function α_d on $L(X, \ell_2^d)$. In the general setting, we need to adopt other tools in Banach space theory to get an analogous upper bound. For this purpose, we introduce a property of the convex set K, defined as follows: Let $1 \leq p \leq 2$. We say that K is of *entropy-type* (p, α_d) if there exists a constant $M_{p,\alpha_d}(K)$ such that

$$\mathcal{E}_{2,1}(v) \;\leq\; M_{p,\alpha_d}(K) m^{1/2-1/p}(1+\ln m)^{1+\lfloor p \rfloor - 1/p} \max_{1\leq j\leq m} \alpha_d(v_j) \tag{7}$$

holds for any $m \in \mathbb{N}$ and any composite map $v = (v_j)_{1\leq j\leq m}$, where $\lfloor p \rfloor$ denotes the largest integer that is equal to or smaller than p. [4] Throughout this paper, $M_{p,\alpha_d}(K)$ will denote the smallest constant that satisfies (7). Note that α_d generalizes the notion of incoherence and M_{p,α_d} represents the complexity of a given sparsity model, which is discussed in more details in Section 3. In most examples we discuss later, M_{p,α_d} is indeed upper bounded by a logarithmic factor, which in turn provides near optimal scaling of the sample complexity.

Our main theorem below identifies a sufficient number of measurements for the RIP of a random linear operator in the general setting.

Theorem 2.1. *Let K be a symmetric centered convex body in a Hilbert space H and X be the Banach space with the norm given as the Minkowski functional of K. Let Γ_s be defined*

[4]The exponent of $(1 + \ln m)$ is chosen as $1 + \lfloor p \rfloor - 1/p$ in order to make $M_{p,\alpha_d}(K)$ independent of m in the examples in Section 3.

by (1) *from* X *and* H. *Let* $A = (1/\sqrt{m})v$, *where* $v = (v_j)_{1 \le j \le m}$ *denotes the composition map constructed from* $v_1, \ldots, v_m \in L(X, \ell_2^d)$. *Suppose that* K *is of entropy-type* (p, α_d) *with* $M_{p,\alpha_d}(K)$ *as in* (7) *and* $v_1, \ldots, v_m$ *satisfy* (3) *and* (4). *Let* $1 < p \le 2$ *and* $0 < \zeta < 1$. *Then there exists a numerical constant* c *such that if*

$$\frac{m^{1-1/p}}{(1+\ln m)^{1+\lfloor p \rfloor - 1/p}} \;\ge\; cM_{p,\alpha_d}(K)\sqrt{s}\delta^{-1} \sup_{k \in \mathbb{N}} \Big(\mathbb{E} \sup_{1 \le j \le m} \alpha_d(v_j)^{2k} \Big)^{1/2k} \tag{8}$$

and

$$m \;\ge\; c\delta^{-2} s \ln(\zeta^{-1}) \sup_{k \in \mathbb{N}} \Big(\mathbb{E} \sup_{1 \le j \le m} \|v_j\|^{2k} \Big)^{1/k}, \tag{9}$$

then $A : X \to \ell_\infty^m(\ell_2^d)$ *satisfies the RIP in* (5) *on* Γ_s *with constant* δ *with probability* $1 - \zeta$.

The moment terms in (8) and (9) are essentially probabilistic or deterministic upper bounds on $\sup_{1 \le j \le m} \alpha_d(v_j)$ and $\sup_{1 \le j \le m} \|v_j\|$, respectively. Indeed, a tail bound implies moment bounds by the Markov inequality and the converse can be shown by direct calculation with Stirling's approximation of the gamma function. (e.g., see [22, Chapter 7].) These two terms extend the notion of incoherence of measurement functionals with respect to the given sparsity model. On the other hand, $M_{p,\alpha_d}(K)$ describes the complexity of sparsity model. The number of measurements providing the RIP given in (8) is proportional to $M_{p,\alpha_d}(K)$. Indeed, this complexity is up to a logarithmic factor for the generalized sparsity models derived from some canonical sparsity models in the literature such as sparse vectors in the standard basis and low-rank matrices. The incoherence and complexity parameters are controlled by a choice of the parameter p and the 1-homogeneous function α_d.

Remark 2.2. A natural choice for the parameters in Theorem 2.1 is $p = 2$ and $\alpha_d(\cdot) = \|\cdot\|$. Then the conditions in (8) and (9) reduces to

$$m \;\ge\; c\delta^{-2} s \left[M_{2,\|\cdot\|}^2 (1+\ln m)^3 \vee \ln(\zeta^{-1}) \right] \sup_{k \in \mathbb{N}} \Big(\mathbb{E} \max_{1 \le j \le m} \|v_j\|^{2k} \Big)^{1/k}.$$

However, as shown in Sections 3.4 and 3.5 (also see [27, Section 2]), there are cases where we can further reduce the number of measurements for the RIP in (8) by optimizing over K, p, and α_d.

2.2 Proof of Theorem 2.1

Since Theorem 2.1 generalizes a special example of the partial Fourier case with the canonical sparsity model, we derive the proof of Theorem 2.1 by modifying the Rudelson-Vershynin argument [42] so that it applies to the general setting. In high level, the left-hand side of (5) denotes the deviation of sum of random variables from its expectation. In high level, the symmetrization [29, Lemma 6.3] followed by Dudley's inequality provides an upper estimate given as a function of the "entropy numbers" of the linear operator $v : X \to \ell_\infty^m(\ell_2^d)$.

Let us start with recalling the relevant notation. Let D and E be symmetric convex bodies in Banach space Y. The *covering number* $N(D, E)$ is the minimal number of translates of E to cover D, i.e.

$$N(D,E) := \min \left\{ k \mid \exists y_1, \dots, y_N \in D,\ D \subset \bigcup_{1 \le j \le k} (y_j + E) \right\} .$$

The *packing number* $M(D, E)$ is the maximal number of disjoint translates of E by elements of D, i.e.

$$M(D,E) := \max \left\{ k \mid \exists y_1, \dots, y_N \in D,\ y_j - y_l \notin E,\ \forall j \neq l \right\} .$$

Then the covering number and packing numbers are related by

$$N(D,E) \le M(D,E) \le N(D,E/2) .$$

Let $T : X \to Y$ be a linear map. The dyadic entropy number [14] is defined by

$$e_l(T,D) := \inf\{\varepsilon > 0 \mid M(T(D), \varepsilon B_Y) \ \le \ 2^{l-1}\} .$$

For $D = B_X$, we use the shorthand notation $e_l(T) = e_l(T, B_X)$. The following equivalence between metric and dyadic entropy numbers is well known (see e.g., [34]).

Lemma 2.3 ([34])**.** *There exist numerical constants $c, C > 0$ such that*

$$c \int_0^\infty \sqrt{\ln N(T(D), \varepsilon)} d\varepsilon \leq \sum_{l=1}^\infty \frac{e_l(T, D)}{\sqrt{l}} \leq C \int_0^\infty \sqrt{\ln N(T(D), \varepsilon)} d\varepsilon \, .$$

Note that since $(e_l(T))$ is a nonincreasing sequence of positive numbers, $\sum_{l=1}^\infty e_l(T)/\sqrt{l}$ coincides with the norm of $(e_l(T))$ in the Lorentz sequence space $\ell(2,1)$ [5]. Therefore, we will use the shorthand notation $\mathcal{E}_{2,1}(T)$ to denote $\sum_{l=1}^\infty e_l(T)/\sqrt{l}$.

The following lemma provides a key estimate in proving Theorem 2.1.

Lemma 2.4. *Let K be a symmetric convex set in $H = \ell_2^N$, X be the Banach determined from K as before, and $K_s = \sqrt{s} K \cap S$, where S is the unit sphere in ℓ_2^N. Let D be a subset of K_s. Let $v : X \to \ell_\infty^m(\ell_2^d)$ be the composition map $(v_j)_{1 \leq j \leq m}$ from linear operators $v_1, \ldots, v_m$ from X to ℓ_2^d. Let $\xi_1, \ldots, \xi_m$ be independent copies of $\xi \sim \mathcal{N}(0,1)$. Then for all $k \in \mathbb{N}$*

$$\left(\mathbb{E} \sup_{x \in D} \left| \sum_{j=1}^m \xi_j \|v_j(x)\|_2^2 \right|^k \right)^{1/k} \leq C\sqrt{s} \left(\sup_{x \in D} \sum_{j=1}^m \|v_j(x)\|_2^2 \right)^{1/2} \left(\mathcal{E}_{2,1}(v) + \sqrt{k} \|v\| \right) .$$

Proof. Define $\beta : \ell_2^d \to \ell_\infty^m$ by

$$\beta(x) = [\|v_1(x)\|_2^2, \ldots, \|v_m(x)\|_2^2]^\top, \quad x \in \ell_2^d \, .$$

Let $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_m]^\top$. Then $\sum_{j=1}^m \xi_j \|v_j(x)\|_2^2$ is written as $\langle \boldsymbol{\xi}, \beta(x) \rangle$, which is a subgaussian process indexed by x. By the tail bound result via generic chaining [18, Theorem 3.2] and Dudley's inequality [29], for all $k \in \mathbb{N}$, we have

$$\left(\mathbb{E} \sup_{x \in D} |\langle \boldsymbol{\xi}, \beta(x) \rangle|^k \right)^{1/k} \lesssim \int_0^\infty \sqrt{\ln N(\beta(D), \varepsilon B_{\ell_2^m})} d\varepsilon + \sup_{x \in D} \left(\mathbb{E} |\langle \boldsymbol{\xi}, \beta(x) \rangle|^k \right)^{1/k} , \qquad (10)$$

where $\beta(D)$ denotes the set $\{\beta(x) \mid x \in D\}$.

We first compute an upper bound on the first summand in the right-hand side of (10). Let

$$R := \Big(\sup_{x\in D}\sum_{j=1}^{m}\|v_j(x)\|_2^2\Big)^{1/2} . \tag{11}$$

Then for $x, x' \in D$ we have

$$\begin{aligned}
\|\beta(x)-\beta(x')\|_2 \;&=\; \Big(\sum_{j=1}^{m}\Big|\|v_j(x)\|_2^2-\|v_j(x')\|_2^2\Big|^2\Big)^{1/2} \\
&= \Big(\sum_{j=1}^{m}\Big|\langle v_j(x), v_j(x)-v_j(x')\rangle + \langle v_j(x'), v_j(x')-v_j(x)\rangle\Big|^2\Big)^{1/2} \\
&\le\; \Big(\sum_{j=1}^{m}\Big|\langle v_j(x-x'), v_j(x)\rangle\Big|^2\Big)^{1/2} + \Big(\sum_{j=1}^{m}\Big|\langle v_j(x'), v_j(x-x')\rangle\Big|^2\Big)^{1/2} \\
&\le\; \Big(\sum_{j=1}^{m}\|v_j(x)-v_j(x')\|_2^2\|v_j(x)\|_2^2\Big)^{1/2} + \Big(\sum_{j=1}^{m}\|v_j(x')\|_2^2\|v_j(x)-v_j(x')\|_2^2\Big)^{1/2} \\
&\le\; \max_{1\le j\le m}\|v_j(x)-v_j(x')\|_2\Big[\Big(\sum_{j=1}^{m}\|v_j(x)\|_2^2\Big)^{1/2} + \Big(\sum_{j=1}^{m}\|v_j(x')\|_2^2\Big)^{1/2}\Big] \\
&\le\; 2R\max_{1\le j\le m}\|v_j(x)-v_j(x')\|_2 .
\end{aligned}$$

Let T denote the maximal family of elements in D such that $\inf_{x\neq x'\in T}\|\beta(x)-\beta(x')\|_2 > \varepsilon$. Then it follows that $\inf_{x\neq x'\in T}\|v(x)-v(x')\|_{\ell_\infty^m(\ell_2^d)} > \varepsilon/(2R)$. This implies that

$$\begin{aligned}
N(\beta(D), \varepsilon B_{\ell_2^m}) \;&\le\; |T| \;\le\; N\Big(v(D), \frac{\varepsilon}{4R}B_{\ell_\infty^m(\ell_2^d)}\Big) \\
&\le\; N\Big(v(\sqrt{s}K), \frac{\varepsilon}{4R}B_{\ell_\infty^m(\ell_2^d)}\Big) \;=\; N\Big(v(K), \frac{\varepsilon}{4R\sqrt{s}}B_{\ell_\infty^m(\ell_2^d)}\Big) .
\end{aligned}$$

Using a change of variables, this implies

$$\begin{aligned}
\int_0^\infty \sqrt{\ln N(\beta(D), \varepsilon B_{\ell_2^m})}d\varepsilon \;&\le\; 4R\sqrt{s}\int_0^\infty \sqrt{\ln N(v(K), \varepsilon B_{\ell_\infty^m(\ell_2^d)})}d\varepsilon \\
&\le\; cR\sqrt{s}\,\mathcal{E}_{2,1}(v : X \to \ell_\infty^m(\ell_2^d)) .
\end{aligned}$$

Next, we compute an upper bound on the second summand in the right-hand side of

(10). By Khintchine's inequality for all $x \in D$

$$\Big(\mathbb{E}|\langle \boldsymbol{\xi}, \beta(x)\rangle|^k\Big)^{1/k} \;\leq\; \sqrt{k}\|\beta(x)\|_2 \;=\; \sqrt{k}\Big(\sum_{j=1}^m \|v_j(x)\|_2^4\Big)^{1/2}$$

$$\leq\; \sqrt{k}\Big(\sum_{j=1}^m \|v_j(x)\|_2^2\Big)^{1/2} \max_{1\leq j\leq m} \|v_j(x)\|_2 \;\leq\; \sqrt{k}R \max_{1\leq j\leq m} \|v_j(x)\|_2 \,.$$

Therefore

$$\sup_{x\in D} \Big(\mathbb{E}|\langle \boldsymbol{\xi}, \beta(x)\rangle|^k\Big)^{1/k} \;\leq\; \sup_{x\in\sqrt{s}K} \sqrt{k}R \max_{1\leq j\leq m} \|v_j(x)\|_2 \;=\; \sqrt{ks}R\,\|v\|_{X\to\ell_2^m(\ell_2^d)} \,.$$

Combining these estimates yields the assertion. □

Corollary 2.5. *Suppose the hypothesis of Lemma 2.4. Then*

$$\Big(\mathbb{E}\sup_{x,y\in D}\Big|\sum_{j=1}^m \xi_j\langle v_j(x), v_j(y)\rangle\Big|^k\Big)^{1/k} \;\leq\; C\sqrt{s}\,\Big(\sup_{x\in D}\sum_{j=1}^m \|v_j(x)\|_2^2\Big)^{1/2}\,\Big(\mathcal{E}_{2,1}(v) + \sqrt{k}\|v\|\Big)\,.$$

Proof. By the polarization identity, we have

$$\langle v_j(x), v_j(y)\rangle \;=\; \frac{1}{4}\sum_{l=0}^3 \mathfrak{i}^l\langle v_j(x+\mathfrak{i}^l y), v_j(x+\mathfrak{i}^l y)\rangle\,,$$

where $\mathfrak{i} = \sqrt{-1}$. Then we apply the argument for $\widetilde{D} = \bigcup_{l=0}^3 D + \mathfrak{i}^l D$. Note that

$$\sum_{j=1}^m \langle v_j(x+\mathfrak{i}^l y), v_j(x+\mathfrak{i}^l y)\rangle = \sum_{j=1}^m \langle v_j(x), v_j(x)\rangle + (-1)^l\sum_{j=1}^m \langle v_j(y), v_j(y)\rangle$$
$$+\, \mathfrak{i}^l\sum_{j=1}^m \langle v_j(x), v_j(y)\rangle + \mathfrak{i}^{l+1}\sum_{j=1}^m \langle v_j(y), v_j(x)\rangle\,.$$

Then by the Cauchy-Schwartz inequality for $x, y \in D$ we have

$$\Big|\sum_{j=1}^m \langle v_j(x+\mathfrak{i}^l y), v_j(x+\mathfrak{i}^l y)\rangle\Big| \;\leq\; 2\sum_{j=1}^m \|v_j(x)\|_2^2 + 2\sum_{j=1}^m \|v_j(y)\|_2^2 \;\leq\; 2R\,,$$

where R is defined in (11). Thus, the assertion follows by replacing $\sqrt{s}$ by $4\sqrt{s}$. □

Proposition 2.6. *Let H, K, and K_s be defined as before. Let $\delta > 0$, $0 < \zeta < 1$, and v be defined by $v(x) = [v_1(x), \ldots, v_m(x)]^\top$ for $x \in X$, where $v_1, \ldots, v_m$ are independent random maps from X to ℓ_2^d satisfying (3) with Φ and (4). Suppose that*

i) The linear operator Φ satisfies $\|\Phi : H \to H\| \leq 1$.

ii) The random linear operator $v : X \to \ell_\infty^m(\ell_2^d)$ satisfies

$$\sup_{k \in \mathbb{N}} \frac{c\sqrt{s}}{\sqrt{m}} \Big((\mathbb{E}\mathcal{E}_{2,1}(v)^{2k})^{1/2k} \vee \sqrt{\ln(\zeta^{-1})} (\mathbb{E}_v \|v\|^{2k})^{1/2k} \Big) \leq \delta \tag{12}$$

for a numerical constant c.

Then

$$\mathbb{P}\Bigg(\sup_{x \in K_s} \Big| \frac{1}{m} \sum_{j=1}^m \|v_j(x)\|_2^2 - \langle x, \Phi x \rangle \Big| \geq \delta \vee \delta^2 \Bigg) \leq \zeta. \tag{13}$$

Proof. Let Z denote the left-hand side of the inequality in (13). Let $(v'_j)_{1 \leq j' \leq m}$ be independent copies of $(v_j)_{1 \leq j \leq m}$. By the standard symmetrization (see e.g. [29, Lemma 6.3]), we have

$$\begin{aligned}
(\mathbb{E}Z^k)^{1/k} &\leq \Big(\mathbb{E}\mathbb{E} \sup_{x \in K_s} \Big| \frac{1}{m} \sum_{j=1}^m \|v_j(x)\|_2^2 - \|v'_j(x)\|_2^2 \Big|^k \Big)^{1/k} \\
&\leq 2\Big(\mathbb{E} \sup_{x \in K_s} \Big| \frac{1}{m} \sum_{j=1}^m \varepsilon_j \|v_j(x)\|_2^2 \Big|^k \Big)^{1/k} \leq 2\sqrt{\frac{\pi}{2}} \Big(\mathbb{E} \sup_{x \in K_s} \Big| \frac{1}{m} \sum_{j=1}^m \xi_j \|v_j(x)\|_2^2 \Big|^k \Big)^{1/k},
\end{aligned}$$

where (ε_j) is a Rademacher sequence and $\xi_1, \ldots, \xi_m$ are independent copies of $\xi \sim \mathcal{N}(0,1)$.

By conditioning on $(v_j)_{1\le j\le m}$, we deduce from Lemma 2.4 that

$$
\begin{aligned}
(\mathbb{E}Z^k)^{1/k} &\le \frac{c_1}{m}\Big(\mathbb{E}\mathbb{E}_{\boldsymbol{\xi}} \sup_{x\in K_s}\Big|\sum_{j=1}^m \xi_j \|v_j(x)\|_2^2\Big|^k\Big)^{1/k} \\
&\le \frac{c_1\sqrt{s}}{m}\left[\mathbb{E}_v\Big(\sup_{x\in K_s}\sum_{j=1}^m \|v_j(x)\|_2^2\Big)^{k/2}(\mathcal{E}_{2,1}(v)+\sqrt{k}\|v\|)^k\right]^{1/k} \\
&\le \frac{c_1\sqrt{s}}{m}\Big[\mathbb{E}_v(\mathcal{E}_{2,1}(v)+\sqrt{k}\|v\|)^{2k}\Big]^{1/2k}\left[\mathbb{E}_v\Big(\sup_{x\in K_s}\sum_{j=1}^m \|v_j(x)\|_2^2\Big)^{k}\right]^{1/2k} \\
&\le \frac{c_1\sqrt{s}}{\sqrt{m}}\Big[\mathbb{E}(\mathcal{E}_{2,1}(v)+\sqrt{k}\|v\|)^{2k}\Big]^{1/2k}\left[\mathbb{E}_v\Big(\sup_{x\in K_s}\frac{1}{m}\sum_{j=1}^m \|v_j(x)\|_2^2\Big)^{k}\right]^{1/2k} \\
&\le \frac{c_1\sqrt{s}}{\sqrt{m}}\Big[\mathbb{E}_v(\mathcal{E}_{2,1}(v)+\sqrt{k}\|v\|)^{2k}\Big]^{1/2k} \\
&\cdot\left[\mathbb{E}_v\Big(\sup_{x\in K_s}\frac{1}{m}\sum_{j=1}^m \|v_j(x)\|_2^2-\langle x,\Phi x\rangle+\langle x,\Phi x\rangle\Big)^{k}\right]^{1/2k} \\
&\le \frac{c_1\sqrt{s}}{\sqrt{m}}\Big[(\mathbb{E}_v\mathcal{E}_{2,1}(v)^{2k})^{1/2k}+\sqrt{k}(\mathbb{E}_v\|v\|^{2k})^{1/2k}\Big]\Big[1+(\mathbb{E}Z^k)^{1/k}\Big]^{1/2}.
\end{aligned}
$$

Let b be the factor before $(1+(\mathbb{E}Z^k)^{1/k})^{1/2}$, then we have $(\mathbb{E}Z^k)^{1/k}\le\sqrt{2}(b+b^2)$. Since $k\in\mathbb{Z}$ was arbitrary, a consequence of the Markov inequality [18, Lemma A.1] implies that there exists a numerical constant c_2 such that

$$
\begin{aligned}
Z \;\le\; &\frac{c_2\sqrt{s}}{\sqrt{m}}(\mathbb{E}_v\mathcal{E}_{2,1}(v)^{2k})^{1/2k}+\frac{c_2 s}{m}(\mathbb{E}_v\mathcal{E}_{2,1}(v)^{2k})^{1/k} \\
&+\frac{c_2\sqrt{s}}{\sqrt{m}}\sqrt{\ln(\zeta^{-1})}(\mathbb{E}_v\|v\|^{2k})^{1/2k}+\frac{c_2 s}{m}\ln(\zeta^{-1})(\mathbb{E}_v\|v\|^{2k})^{1/k}
\end{aligned}
$$

holds with probability $1-\zeta$. The condition in (12) implies $Z\;\le\;\delta\vee\delta^2$. □

Proof of Theorem 2.1. Since K is of entropy-type (p,α_d), for every $v: X\to\ell_\infty^m(H)$ we have

$$
\mathcal{E}_{2,1}(v)\;\le\; M_{p,\alpha_d}(K)m^{1/2-1/p}(1+\ln m)^{1+\lfloor p\rfloor-1/p}\max_{1\le j\le m}\alpha_d(v_j)\,.
$$

Then we get

$$(\mathbb{E}\mathcal{E}_{2,1}(v)^{2k})^{1/2k} \;\le\; M_{p,\alpha_d}(K)\Big(\mathbb{E}\max_{1\le j\le m}\alpha_d(v_j)^{2k}\Big)^{1/2k} m^{1/2-1/p}(1+\ln m)^{1+\lfloor p\rfloor-1/p}$$

and hence

$$\frac{c\sqrt{s}}{\sqrt{m}}(\mathbb{E}\mathcal{E}_{2,1}(v)^{2k})^{1/2k} \;\le\; \frac{c\sqrt{s}M_{p,\alpha_d}(K)(1+\ln m)^{1+\lfloor p\rfloor-1/p}}{m^{1/p}}\sup_{k\in\mathbb{N}}\Big(\mathbb{E}\max_{1\le j\le m}\alpha_d(v_j)^{2k}\Big)^{1/2k}.$$

By Proposition 2.6, it suffices to satisfy

$$\frac{m^{1/p}}{(1+\ln m)^{1+\lfloor p\rfloor-1/p}} \;\ge\; \frac{cM_{p,\alpha_d}(K)\sqrt{s}}{\delta}\sup_{k\in\mathbb{N}}\Big(\mathbb{E}\max_{1\le j\le m}\alpha_d(v_j)^{2k}\Big)^{1/2k}$$

and

$$m \;\ge\; \frac{cs\ln(\zeta^{-1})}{\delta^2}\sup_{k\in\mathbb{N}}(\mathbb{E}_v\|v\|^{2k})^{1/k}.$$

□

3 Complexity of sparsity models

Our generalized sparsity model is induced from a convex set K in a Hilbert space H. A sufficient number of measurements for the RIP is determined by the geometry of the underlying Banach space X whose norm is the Minkowski functional of K. In this section, we discuss the complexity of K given in terms of $M_{p,\alpha_d}(K)$ for various examples of the generalized sparsity model. We first recall the notion of Banach spaces of type p, which will be frequently used in the remainder of this section.

Definition 3.1. *A Banach space X has* type p *if there exists a constant C such that for all finite sequence (x_j) in X*

$$\Big(\mathbb{E}\Big\|\sum_j \varepsilon_j x_j\Big\|_X^p\Big)^{1/p} \;\le\; C\Big(\sum_j \|x_j\|_X^p\Big)^{1/p}, \tag{14}$$

where (ε_j) *is a Rademacher sequence [34]. The type-p constant of* X*, denoted by* $T_p(X)$*, is the smallest constant* C *that satisfies* (3.1)*.*

The following upper estimates of entropy numbers given by Maurey's empirical method will be used to compute the complexity of various generalized sparsity models.

Lemma 3.2 (Maurey's empirical method [12, Proposition 1])**.** *Let* X *be a Banach space of type* p *and* $v : \ell_1^n \to X$*. Then there exists a numerical constant* C *such that*

$$e_k(v) \ \le \ C T_p(X) \|v\| f(k, n, p) \ , \tag{15}$$

where

$$f(k, n, p) := 2^{-(k/n \vee 1)} \left(\frac{\log_2(1 + n/k)}{k} \vee \frac{1}{n} \right)^{1-1/p} \ . \tag{16}$$

Remark 3.3. Note that $f(k, n, p)$ in Lemma 3.2 satisfies

$$f(k, n, p) \le 2^{-1} (\log_2 n)^{1-1/p} k^{-1+1/p}, \quad \forall k \le n \ , \tag{17}$$

and

$$f(k, n, p) \le 2^{-k/n} n^{-1+1/p}, \quad \forall k > n \ . \tag{18}$$

3.1 Relaxed canonical sparsity

The sparsity level in the canonical sparsity model, which consists of exactly sparse vectors, is implicitly controlled by the ℓ_1 norm. A *relaxed* canonical sparsity model is obtained by Definition 1.1 with $H = \ell_2^N$ and $X = \ell_1^N$ and includes exactly sparse vectors and their approximation with small perturbation. The convex set K that generates this model is the unit ball B_1^N in ℓ_1^N. We derive an upper bound on $M_{2,\|\cdot\|}(B_1^N)$ by using a well known application of Maurey's empirical method, which is given in the following lemma.

Lemma 3.4. *Let* $v : \ell_1^N \to \ell_\infty^m(\ell_2^d)$*. Then*

$$\mathcal{E}_{2,1}(v) \ \le \ C \sqrt{1 + \ln N} \ (1 + \ln m)^{3/2} \|v\| \ .$$

Proof. Let $1 \le q < \infty$ be arbitrarily fixed. Let $v_q : \ell_1^N \to \ell_q^m(\ell_2^d)$ be defined as $v_q = (\mathrm{Id} \otimes \iota) v$ where $\iota : \ell_\infty^m \to \ell_q^m$ is the formal identity. Then we have

$$\|v_q\| \le \|\iota\| \cdot \|v\| \le m^{1/q} \|v\| \; .$$

Furthermore, by [12, Lemma 4], which follows from the result due to Schütt [43], the entropy number of $\iota^{-1} : \ell_q^m \to \ell_\infty^m$ is upper bounded by

$$e_k(\iota^{-1}) \le C_1 2^{-(k/m \vee 1)} \; .$$

On the other hand, the noncommutative Khintchine inequalities (see e.g. [35]) implies that $\ell_q^m(\ell_2^d)$ has type 2 and $T_2(\ell_q^m(\ell_2^d)) \le C_2 \sqrt{q}$.

Similar to the proof of [12, Proposition 3], by choosing $q = 1 + \ln m$, we obtain

$$\begin{aligned} e_{2k}(v) &\le e_k(v_q) e_k(\iota^{-1}) \\ &\le C_3 m^{1/q} \|v\| T_2(\ell_q^m(\ell_2^d)) f(k, N, 2) 2^{-(k/m \vee 1)} \\ &\le C_4 \sqrt{1 + \ln m} \|v\| f(k, N, 2) 2^{-(k/m \vee 1)} \; . \end{aligned}$$

First, let us assume that $m \le N$. Then, by plugging in (18) and (17), we have

$$e_k(v) \le C_5 \sqrt{1 + \ln m} \|v\| \sqrt{1 + \ln N} k^{-1/2}, \quad \forall k \le 2m \; , \tag{19}$$

and

$$e_k(v) \le C_5 \sqrt{1 + \ln m} \|v\| 2^{-k/(2m)} N^{-1/2}, \quad \forall k > 2m \; . \tag{20}$$

Therefore,

$$\sum_{k=1}^{\infty} \frac{e_k(v)}{\sqrt{k}} \le C_5 \sqrt{1 + \ln m} \|v\| \left(\sqrt{1 + \ln N} \underbrace{\sum_{k=1}^{2m} k^{-1}}_{(\S)} + N^{-1/2} \underbrace{\sum_{k=2m+1}^{\infty} k^{-1/2} 2^{-k/(2m)}}_{(\S\S)} \right) , \tag{21}$$

where (§) is upper-bounded by

$$(\S) \leq 1 + \int_1^{2m} t^{-1} dt \leq 1 + \ln 2 + \ln m \tag{22}$$

and (§§) is upper-bounded by

$$(\S\S) \leq \int_{2m}^{\infty} t^{-1/2} 2^{-t/(2m)} dt \leq 2 \int_{\sqrt{2m}}^{\infty} \exp\left(-\frac{\zeta^2}{2m/\ln 2}\right) d\zeta \leq \sqrt{\frac{\pi 2m}{\ln 2}} \, . \tag{23}$$

Then the assertion follows by plugging in (22) and (23) to (21).

When $m > N$, instead we have

$$e_k(v) \leq C_5 \sqrt{1 + \ln m} \|v\| \sqrt{1 + \ln N} k^{-1/2}, \quad \forall k \leq 2N \, , \tag{24}$$

and

$$e_k(v) \leq C_5 \sqrt{1 + \ln m} \|v\| 2^{-k/(2N)} N^{-1/2}, \quad \forall k > 2N \, . \tag{25}$$

The logarithmic factor in the resulting is $\sqrt{1 + \ln m}\, (1 + \ln N)^{3/2}$, which is less than $\sqrt{1 + \ln N}\, (1 + \ln m)^{3/2}$. This completes the proof. □

The following upper estimate of $M_{2,\alpha_d}(B_1^N)$ is obtained as a direct consequence of Lemma 3.4 through the definition in (7).

Proposition 3.5. *Let* $\alpha_d(u) = \|u\|_{\ell_1^N \to \ell_2^d}$. *Then*

$$M_{2,\alpha_d}(B_1^N) \; \leq \; C \sqrt{1 + \ln N} \, .$$

3.2 Relaxed atomic sparsity over finite dictionary

An immediate extension of the canonical sparsity model is the atomic model over a finite dictionary. A vector is *atomic s-sparse* if x is represented as a finite linear combination of a given set of vectors called *atoms* [16]. The set of all atoms is called a *dictionary*. This atomic model generalizes the sparsity model over a finite dictionary (see e.g. [4, 1]). Here we consider a special case where the dictionary is a finite set $\{x_k \mid 1 \le k \le M\}$ consisting of unit vectors in a Hilbert space H, i.e. $\|x_k\|_H = 1$ for $k = 1, \ldots, M$. Then a relaxed atomic sparsity model is derived from the absolute convex hull of the dictionary denoted by $K = \mathrm{absconv}\{x_k \mid 1 \le k \le M\}$. The set of "sparse" vectors is given as the cone generated by $K_s = \sqrt{s}K \cap S$ where S denotes the unit sphere in H. The normalization of the atoms implies that the dictionary is the set of all unit *1-sparse* vectors in H. The complexity of K is upper-bounded by the following corollary.

Corollary 3.6. *Let H be a Hilbert space, $K = \mathrm{absconv}\{x_j \mid 1 \le j \le M\} \subset H$, X be the Banach space with the norm given as the Minkowski functional of K, and α_d be the operator norm, i.e. $\alpha_d(u) = \|u\|$ for $u : X \to \ell_2^d$. Then*

$$M_{2,\alpha_d}(K) \ \le\ C\sqrt{1 + \ln M}\ .$$

Proof. Let $v = (v_j)_{1 \le j \le m}$ where $v_j : X \to \ell_2^d$ for $j = 1, \ldots, m$. Then we have

$$\max_{1 \le j \le M} \alpha_d(v_j) = \|v\|_{X \to \ell_\infty^m(\ell_2^d)}\ .$$

Define $Q : \ell_1^M \to X$ so that $Q(\boldsymbol{e}_k) = x_k$ for all $k = 1, \ldots, M$, where $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_M$ are standard basis vectors in $\mathbb{R}^M$. Then

$$\|Q\|_{\ell_1^M \to X} = \max_{1 \le k \le M} \|x_k\|_X = 1\ ,$$

where the last identity follows from the construction of $\|\cdot\|_X$ from K. Therefore

$$\|v_j Q\|_{\ell_1^M \to \ell_2^d} \le \|v_j\|_{X \to \ell_2^d}\ , \quad \forall j = 1, \ldots, M\ .$$

By Lemma 3.4, we have

$$\mathcal{E}_{2,1}(vQ : \ell_1^M \to \ell_\infty^m(\ell_2^d)) \ \le\ C\|v\|_{X \to \ell_\infty^m(\ell_2^d)}\sqrt{1 + \ln M}(1 + \ln m)^{3/2}\ ,$$

Thus Proposition 3.5 applies. Indeed, entropy numbers are surjective, i.e.

$$e_k(vQ : \ell_1^M \to Y) \;=\; e_k(v : X \to Y)$$

for any Banach space Y. Therefore, the estimate in Corollary 3.6 is as tight as that in Lemma 3.5. □

3.3 Norm dual of type-p Banach spaces

Previous sparsity models are constructed with the Banach space ℓ_1^M, where Maurey's empirical method applies directly. In this section, we further generalize the result to the scenario where the norm dual X^* has type p by using the entropy duality. We restrict the measurements to the scalar-valued case in this section. The extension to the vector-valued case will be discussed in the next section. The following lemma shows through Maurey's method that the unit ball K of X has type (p, α_d) if X^* has type p when α_d is the operator norm.

Lemma 3.7. *Let X be a Banach space such that X^* has type $1 < p < 2$ and $v : X \to \ell_\infty^m$. Then*

$$\sum_{k=1}^{\infty} \frac{e_k(v)}{\sqrt{k}} \;\le\; c(p) T_p(X^*) \|v\| m^{1/p-1/2} (1 + \ln m)^{1+1/p'} \;,$$

where $p' = p/(p-1)$ and $c(p)$ is a constant that depends only on p. Moreover, for $p = 2$,

$$\sum_{k=1}^{\infty} \frac{e_k(v)}{\sqrt{k}} \;\le\; c T_2(X^*) \|v\| (1 + \ln m)^{5/2}$$

for a numerical constant c.

The following corollary is a direct consequence of Lemma 3.7.

Corollary 3.8. *Suppose that X^* is of type p for $1 < p \le 2$. Then*

$$M_{p,\|\cdot\|}(K) \le c(p) T_p(X^*) m^{1/p-1/2} \;,$$

where $c(p)$ is a constant that depends only on p.

Remark 3.9. A similar estimate to Lemma 3.7 has been shown by Chafaï et al. under a stronger assumption that the Banach space X is uniformly convex [15, Chapter 5]. The entropy duality by Bourgain et al. [7] applies to this more restrictive setting too and one can obtain the same upper estimate.

Proof of Lemma 3.7. Let $v^* : \ell_1^m \to X^*$ denote the adjoint of v. Let $q \geq 2$ be arbitrary fixed and $q' = q/(q-1)$. Let $v_q : X \to \ell_q^m$ be defined by $v_q = \iota v$, where $\iota : \ell_\infty^m \to \ell_q^m$ is the identity operator. Let v_q^* be the adjoint of v_q. Then the duality of the entropy numbers [7, Proposition 4 and Lemma C] implies that there exists a numerical constant C_1 for which

$$\sum_{j=1}^{k} e_j(v) \leq \sum_{j=1}^{k} e_j(v_q) \leq [C_1 T_2(\ell_q^m)]^2 \sum_{j=1}^{k} e_j(v_q^*) \leq [C_1 T_2(\ell_q^m)]^2 \|\iota^*\| \sum_{j=1}^{k} e_j(v^*)$$

holds for all $k \in \mathbb{N}$. Since $T_2(\ell_q^m) \leq \sqrt{q}$ [12, Lemma 3] and $\|\iota^*\|_{\ell_{q'}^m \to \ell_1^m} = m^{1/q}$, by choosing $q = \ln m$ as in [13], we obtain

$$\sum_{j=1}^{k} e_j(v) \leq C_2 \log m \sum_{j=1}^{k} e_j(v^*) \,. \tag{26}$$

Since $(e_k(v))$ is a nonincreasing sequence of positive numbers, it satisfies

$$e_k(v) \leq \frac{1}{k} \sum_{j=1}^{k} e_j(v), \quad \forall k \in \mathbb{N} \,.$$

Therefore, for any $l \in \mathbb{N}$, we have

$$\begin{aligned} \sum_{k=1}^{l} \frac{e_k(v)}{\sqrt{k}} &\leq \sum_{k=1}^{l} \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \frac{e_j(v)}{k} \leq C_2 \log m \sum_{k=1}^{l} \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \frac{e_j(v^*)}{k} \\ &= C_2 \log m \sum_{j=1}^{l} \sum_{k=j}^{l} \frac{e_j(v^*)}{k\sqrt{k}} \leq C_2 \log m \sum_{j=1}^{l} e_j(v^*) \int_j^\infty t^{-3/2} dt \\ &\leq 2C_2 \log m \sum_{j=1}^{l} \frac{e_j(v^*)}{\sqrt{j}} \,, \end{aligned} \tag{27}$$

where the second inequality follows from (26).

Moreover, since v^* is an operator from ℓ_1^m to X^*, Maurey's empirical method in Lemma 3.2 implies

$$e_k(v^*) \;\leq\; C_3 T_p(X^*)\|v\| f(k,m,p)\;, \tag{28}$$

where f is defined in (16).

Therefore, by (18) and (17), we obtain

$$\sum_{k=1}^{\infty} \frac{e_k(v^*)}{\sqrt{k}} \leq C_3 T_p(X^*)\|v\| \Bigg(2^{-1}(\log_2 m)^{1/p'} \underbrace{\sum_{k=1}^{m} k^{-1/2-1/p'}}_{(\S\S\S)} + m^{-1/p'} \underbrace{\sum_{k=m+1}^{\infty} k^{-1/2} 2^{-k/m}}_{(\S\S\S\S)} \Bigg)\;. \tag{29}$$

If $p < 2$ (or equivalently $p' > 2$), then (§§§) is upper-bounded by

$$(\S\S\S) \leq 1 + \int_1^m t^{-1/2-1/p'} dt \leq 1 + \left(\frac{1}{2} - \frac{1}{p'}\right)^{-1} m^{1/2-1/p'}\;. \tag{30}$$

Otherwise, if $p = 2$ (or equivalently $p' = 2$), then (§§§) is upper-bounded by

$$(\S\S\S) \leq 1 + \int_1^m t^{-1} dt \leq 1 + \ln m\;. \tag{31}$$

On the other hand, from a tail bound on the standard Gaussian distribution, (§§§§) is upper-bounded by

$$(\S\S\S\S) \leq \int_m^\infty t^{-1/2} 2^{-t/m} dt \leq 2 \int_{\sqrt{m}}^\infty \exp\left(-\frac{u^2}{m/\ln 2}\right) du \leq \sqrt{\frac{\pi m}{\ln 2}}\;. \tag{32}$$

By plugging in (30) and (32) to (27) through (29) followed by maximizing over $l \in \mathbb{N}$, we obtain

$$\sum_{j=1}^{\infty} \frac{e_k(v)}{\sqrt{k}} \;\leq\; C_4 \left(\frac{1}{2} - \frac{1}{p'}\right)^{-1} T_p(X^*)\|v\| m^{1/2-1/p'} (1+\ln m)^{1+1/p'}, \quad \forall 1 < p < 2$$

for a numerical constant C_3. Let $c(p) = C_3(1/2 - 1/p')^{-1}$. This proves the first part. The proof for $p = 2$ is given by replacing (30) by (31) in the above argument. □

3.4 Unconditional basis and lattices

Next we discuss the generalization of the previous section to the vector-valued measurement case. When $v : X \to \ell_\infty^m(\ell_2^d)$ takes m measurements in ℓ_2^d with $d > 1$, the domain of its adjoint v^* is no longer ℓ_1^{md}. Thus the estimate by Maurey's method does not apply to this setting even when combined with the entropy duality. Here we show that imposing the lattice structure to a Banach space X with norm dual of type p provides the extension of the estimate in the scalar-valued case to the vector-valued case.

We first consider the finite-dimensional case. A Banach space X in $\mathbb{R}^N$ is a *lattice* if its unit ball K is a convex symmetric set that satisfies

$$D_\varepsilon(K) \;=\; K, \quad \forall \varepsilon \in \{-1, 1\}^N\,, \tag{33}$$

where D_ε denotes the diagonal operator that performs the element-wise multiplication with ε. In the complex case we require this condition for all $\varepsilon \in \mathbb{T}^N$, where $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$ denotes the set of unit modulus complex numbers. Equivalently, the norm $\|\cdot\|_X$ given as the Minkowski functional of K satisfies

$$\|(x_i)_{1\le i\le N}\|_X \;=\; \|(|x_i|_{1\le i\le N})\|_X\,.$$

The general definition of Banach lattice is more involved. We will consider a subset of Banach lattices given by a norm on measurable function f on $\{1, \dots, n\}$, $\mathbb{N}$, $[0, 1]$ or $[0, \infty)$ such that

$$\||f|\|_X = \|f\|_X, \tag{34}$$

where $|f|$ is defined by $|f|(t) = |f(t)|$ for all t in the domain of f. (See [30] for more details.) The arguments in this section apply to all infinite-dimensional Banach lattices satisfying (34).

Recall that the norm dual of a Banach lattice is also a lattice [30]. Indeed, the norm dual consists of functions defined on the same domain. From this property, there exists a normed space $X^*(\ell_2^d)$ of length-d sequences in X^* with the norm defined by

$$\|(x_i^*)_{1\le i\le d}\|_{X^*(\ell_2^d)} := \left\| \Big(\sum_{i=1}^d |x_i^*|^2 \Big)^{1/2} \right\|_{X^*} , \quad \forall x_i^* \in X^* , \tag{35}$$

where $(\sum_{i=1}^d |x_i^*|^2)^{1/2}$ denotes the "square root" function defined by

$$\left[\Big(\sum_{i=1}^d |x_i^*|^2 \Big)^{1/2} \right](t) := \Big(\sum_{i=1}^d |x_i^*(t)|^2 \Big)^{1/2} , \quad \forall t .$$

Then we consider a factorization of $v : X \to \ell_\infty^m(\ell_2^d)$ through ℓ_∞^{2md}. Since the norm dual of ℓ_∞^{2md} is ℓ_1^{2md}, Maurey's method applies. The lattice structure of X enables to control the norms of the factors via $X^*(\ell_2^d)$. In this way we extend the upper estimate of $M_{p,\alpha_d}(K)$ in Section 3.3 for the scalar-valued case ($d = 1$) to the vector-valued case ($d > 1$). For the analysis of the factorization of v, we will use a homogeneous function γ_∞^k on $L(X, \ell_2^d)$ defined by

$$\gamma_\infty^k(u) \; = \; \inf_{u \, = \, ab} \|b\|_{X\to\ell_\infty^k} \, \|a\|_{\ell_\infty^k \to \ell_2^d} \, . \tag{36}$$

Note that γ_∞^k is not necessarily a valid norm. The following Lemma shows a well-known upper estimate of γ_∞^k (see e.g. [33]).

Lemma 3.10. *Let X be a lattice with unit ball K as above and $u : X \to \ell_2^d$. Then there exists a numerical constant C such that*

$$\gamma_\infty^{2d}(u) \; \le \; C\|(u^*(\boldsymbol{e}_i))_{1\le i\le d}\|_{X^*(\ell_2^d)} \; \le \; 2C\ell(u(K)) ,$$

where

$$\ell(u(K)) := \mathbb{E} \sup_{z\in u(K)} \langle \xi, z\rangle$$

for $\xi \sim \mathcal{N}(0, I_d)$ in the real case (or $\xi \sim \mathcal{CN}(0, I_d)$ in the complex case).

Proof. There exists an isomorphic embedding $a : \ell_2^d \hookrightarrow \ell_1^{2d}$ due to Kashin (see e.g., [34]). Since X is a lattice, $X(\ell_2^d)$ is defined as in (35) with X^* replaced by X. Moreover, one can define $X(\ell_1^{2d})$ with

$$\|(y_k)_{1\le k\le 2d}\|_{X(\ell_1^{2d})} := \left\| \sum_{k=1}^{2d} |y_k| \right\|_X , \quad \forall y_k \in X ,$$

where

$$\left(\sum_{k=1}^{2d} |y_k| \right)(t) := \sum_{k=1}^{2d} |y_k(t)| , \quad \forall t .$$

Let $x = (x_i)_{1\le i\le d} \in X(\ell_2^d)$ and $y = (y_k)_{1\le k\le 2d} \in X(\ell_1^{2d})$ such that $y = (\mathrm{Id} \otimes a)x$, i.e.

$$y_k(t) = \sum_{i=1}^{d} a_{ki} x_i(t) , \quad \forall t ,$$

where a_{ki} denotes the (k,i)th entry of the matrix representation of a. Then we have $\|y\|_{X(\ell_1^{2d})} \le \|a\| \|x\|_{X(\ell_2^d)}$. This implies $X(\ell_2^d)$ is also embedded into $X(\ell_1^{2d})$.

Moreover, since ℓ_2^d and ℓ_1^{2d} are finite-dimensional lattices, the dual spaces of $X(\ell_2^d)$ and $X(\ell_1^{2d})$ are given by $X^*(\ell_2^d)$ and $X^*(\ell_\infty^{2d})$, respectively [30]. By the Hahn-Banach theorem, for every $x^* = (x_i^*)_{1\le i\le d} \in X^*(\ell_2^d)$ with $\|x^*\|_{X^*(\ell_2^d)} = 1$, there exists $y^* = (y_k^*)_{1\le k\le 2d} \in X^*(\ell_\infty^{2d})$ that satisfies $\|y^*\|_{X^*(\ell_\infty^{2d})} \le C_1$ for a numerical constant C_1 and $x^* = (\mathrm{Id} \otimes a^*)y^*$.

Choose $x^* = (x_i^*)_{1\le i\le d} = (u^*(\boldsymbol{e}_i))_{1\le i\le d}$. Then let $y^* = (y_k^*)_{1\le k\le 2d}$ be the extension of x^* as above. Define a linear operator $\phi : X \to \ell_\infty^{2d}$ by $\phi(x) = (\langle y_k^*, x\rangle)_{1\le k\le 2d}$ for $x \in X$. Then u is factorized as $u = a^*\phi$. Since $X^*(\ell_\infty^{2d}) \subset \ell_\infty^{2d}(X^*)$, the operator norm of ϕ is upper-bounded by

$$\|\phi\| = \max_{1\le k\le 2d} \|y_k^*\|_{X^*} = \|y^*\|_{\ell_\infty^{2d}(X^*)} \;\le\; \|y^*\|_{X^*(\ell_\infty^{2d})} \;\le\; C_1 \|x^*\|_{X^*(\ell_2^d)} ,$$

where $\|\cdot\|_{\ell_\infty^{2d}(X^*)}$ and $\|\cdot\|_{X^*(\ell_\infty^{2d})}$ are respectively defined by

$$\|y^*\|_{\ell_\infty^{2d}(X^*)} := \max_{1\le k\le 2d} \|y_k^*\|_{X^*}$$

and

$$\|y^*\|_{X^*(\ell_\infty^{2d})} := \Big\| \max_{1\le k\le 2d} |y_k^*| \Big\|_{X^*}$$

with the "max" function given by

$$\left(\max_{1\le k\le 2d} |y_k^*|\right)(t) := \sup_{1\le k\le 2d} |y_k^*(t)| \,, \quad \forall t \,.$$

Therefore,

$$\gamma_\infty^{2d}(u) \;\le\; C_1 \|x^*\|_{X^*(\ell_2^d)} \|a^*\| \,,$$

where $\|a^*\|$ is a numerical constant. This proves the first assertion.

Let $\xi_1, \dots, \xi_d$ are independent copies of $\xi \sim \mathcal{N}(0,1)$ in the real case (or $\xi \sim \mathcal{CN}(0,1)$ in the complex case). By Khintchine's inequality (see e.g. [25]),

$$\begin{aligned}
\|x^*\|_{X^*(\ell_2^d)} \;&=\; \Big\|(u^*(\boldsymbol{e}_i))_{1\le i\le d}\Big\|_{X^*(\ell_2^d)} = \sup_{x\in K} \Big\|(\langle \boldsymbol{e}_i, u(x)\rangle)_{1\le i\le d}\Big\|_2 \\
&\le \sqrt{2} \sup_{x\in K} \mathbb{E} \Big| \sum_{i=1}^d \xi_i \langle \boldsymbol{e}_i, u(x)\rangle \Big| \le \sqrt{2}\, \mathbb{E} \sup_{x\in K} \Big| \sum_{i=1}^d \xi_i \langle \boldsymbol{e}_i, u(x)\rangle \Big| \\
&=\; \sqrt{2}\, \mathbb{E} \sup_{x\in K} \mathrm{Re} \Big\langle \sum_{i=1}^d \xi_i \boldsymbol{e}_i, u(x) \Big\rangle \;=\; \sqrt{2}\ell(u(K)) \,.
\end{aligned}$$

Here we used the invariance of the distribution of ξ with respect to $\{-1,1\}$ in the real case (or with respect to $\mathbb{T}$ in the complex case) and the symmetry of K. □

Choosing the homogeneous function α_d as

$$\alpha_d(u) = \|u^*(\boldsymbol{e}_i)_{1\le i\le d}\|_{X^*(\ell_2^d)} \tag{37}$$

allows us to use Maurey's empirical method on a factorization of $(v_j)_{1\le j\le m} : X \to \ell_\infty^m(\ell_2^d)$ while the involved operator norms are controlled by Lemma 3.10. The following theorem provides the resulting upper estimate of $M_{p,\alpha_d}(K)$.

Theorem 3.11. *Let K be a convex symmetric set, X be the Banach space with the norm* *given by the Minkowski functional of K, and α_d be chosen as in (37). Suppose that the norm dual X^* has type p for $1 < p \le 2$ and K satisfies (33) in the finite-dimensional case (or (34) in the infinite-dimensional case). Then*

$$M_{p,\alpha_d}(K) \ \le\ C(p) T_p(X^*) d^{1/p-1/2}(1+\ln d)^{1+\lfloor p\rfloor-1/p}$$

for a constant $C(p)$ that depends only on p.

Proof. Let $v_1, \dots, v_m \in L(X, \ell_2^d)$ be maps with $\alpha_d(v_j) \le 1$ for $j = 1, \dots, m$. Then by Lemma 3.10 and the definition of γ_∞^{2d} in (36), for each j, there exists a factorization $v_j = a_j b_j$ by $b_j : X \to \ell_\infty^{2d}$ and $a_j : \ell_\infty^{2d} \to \ell_2^d$ such that $\|b_j\|_{X\to\ell_\infty^{2d}} \le 1$ and $\|a_j\|_{\ell_\infty^{2d}\to\ell_2^d} \le \gamma_\infty^{2d}(v_j) \ \le\ C$, where C is the constant in Lemma 3.10.

Let $b = (b_j)_{1\le j\le m} : X \to \ell_\infty^m(\ell_\infty^{2d}) = \ell_\infty^{2md}$ be the composition map given by $b(x) = (b_1(x), \dots, b_m(x))$ for $x \in X$. Since b^* is a linear operator from ℓ_1^{2md} to X^*, by Lemma 3.7, we obtain

$$\mathcal{E}_{2,1}(b) \ \le\ c(p) T_p(X^*) (2md)^{1/p-1/2}(1+\ln m+\ln d)^{1+\lfloor p\rfloor-1/p} \,. \tag{38}$$

Let $D : \ell_\infty^m(\ell_\infty^{2d}) \to \ell_\infty^m(\ell_2^d)$ be the block diagonal map defined from $(a_j)_{1\le j\le m}$ by $D((z_j)_{1\le j\le m}) = (a_1(z_1), \dots, a_m(z_m))$ for $(z_j)_{1\le j\le m} \in \ell_\infty^m(\ell_\infty^{2d})$. Then

$$\|D\|_{\ell_\infty^m(\ell_\infty^{2d})\to\ell_\infty^m(\ell_2^d)} = \max_{1\le j\le m} \|a_j\|_{\ell_\infty^{2d}\to\ell_2^d} \le C \,. \tag{39}$$

Since $v = Db$, it follows that

$$\mathcal{E}_{2,1}(v) \le \|D\|_{\ell_\infty^m(\ell_\infty^{2d})\to\ell_\infty^m(\ell_2^d)} \cdot \mathcal{E}_{2,1}(b) \,.$$

Plugging in (38) and (39) to the above upper estimate implies the assertion. □

Remark 3.12. In Section 3.2 we have shown that a convex body K given as the convex hull of M vectors induces a sparsity model. Recall that K is expressed as $K = Q(B_1^M)$ with $Q : \ell_1^M \to \ell_2^N$ that maps the vertices of B_1^M to given M vectors in ℓ_2^N. Indeed, ℓ_1^M is a Banach lattice and the above choice of $\alpha_d(u) = \|(Q^* u^*(\boldsymbol{e}_i))_{1 \le i \le d}\|_{\ell_\infty^M(\ell_2^d)}$ reduces to the operator norm $\|uQ\|_{\ell_1^M \to \ell_2^d}$ as in Section Section 3.2. In this sense, the upper estimate on $M_{p,\alpha_d}(K)$ in this section naturally extends the corresponding result for $K = Q(B_1^M)$ to the image of the unit ball in a Banach lattice.

3.5 Schatten classes

The RIP on the low-rank matrix model has been studied with the Banach space with the norm as the ℓ_1 norm of the singular values. This Banach space of n_1-by-n_2 matrices is the Schatten 1-class and denoted by $S_1^{n_1,n_2}$. In general, Schatten p-classes $S_p^{n_1,n_2}$ are examples of noncommutative L_p spaces. In this section, we present that the result for the Banach lattices in the previous section extends to analogous results for these "noncommutative lattices".

The choice of the homogeneous function α_d as the operator norm does not provide the desired decay of entropy numbers for the Schatten p-classes. The following lemma, which is analogous to Lemma 3.10 for Banach lattices, provides a factorization of a vector-valued map from a Schatten-p class through the norm dual of $\ell_1^{cd^2}$ so that one can apply Maurey's empirical method. For simplicity of notation, we state our results in the square case ($n_1 = n_2 = n$). The rectangular case can be shown in a similar way.

Lemma 3.13. *Let $1 < q < 2$ and $u : S_q^n \to \ell_2^d$. Then there exists a numerical constant c such that*

$$\gamma_\infty^{cd^2}(u) \;\le\; C\left(\left\| \Big(\sum_{j=1}^{d} [u^*(\boldsymbol{e}_j)]^* [u^*(\boldsymbol{e}_j)] \Big)^{1/2} \right\|_{S_{q'}^n} + \left\| \Big(\sum_{j=1}^{d} [u^*(\boldsymbol{e}_j)] [u^*(\boldsymbol{e}_j)]^* \Big)^{1/2} \right\|_{S_{q'}^n} \right).$$

Proof. Let $G_q^d := \{\sum_{j=1}^d \xi_j z_j \mid z_j \in \mathbb{K}\} \subset L_q(\Omega, P)$, where $\xi_1, \ldots, \xi_d$ are independent copies of $\xi \sim \mathcal{N}(0,1)$ and $\mathbb{K}$ denotes the scalar field, which is either $\mathbb{R}$ or $\mathbb{C}$. Similarly, we define $G_q^d(S_q^n) := \{\sum_{j=1}^d \xi_j x_j \mid x_j \in S_q^n\} \subset L_q(\Omega, P; S_q^n)$. Khintchine's inequality implies that there exists an isomorphism $\iota : \ell_2^d \to (G_q^d)^*$, which is an isometry up to a constant. Furthermore, it has been shown by Pisier [36, Theorem 1.13] that for $1 \leq q < 2$ the space G_q^d *completely C-isomorphic* to $\ell_q^{cd^2}$ for a numerical constant C, i.e. there exists an isomorphic map $a : G_q^d \hookrightarrow \ell_q^{cd^2}$ such that $\mathrm{Id} \otimes a : G_q^d(S_q^n) \hookrightarrow \ell_q^{cd^2}(S_q^n)$ satisfies $\|\mathrm{Id} \otimes a\| \leq C$.

By duality the norm dual of $G_q^d(S_q^n)$, denoted by $G_q^d(S_q^n)^*$, is a quotient of $\ell_{q'}^{cd^2}(S_{q'}^n)$. Therefore, for every $x^* = (x_j^*)_{1 \leq j \leq d} \in \ell_2^d(S_q^n)^*$ with $\|x^*\|_{\ell_2^d(S_q^n)^*} = 1$, there exists $y^* = (y_i^*)_{1 \leq i \leq cd^2} \in \ell_{q'}^{cd^2}(S_{q'}^n)$ that satisfies $\|y^*\|_{\ell_{q'}^{cd^2}(S_{q'}^n)} \leq C_1$ for a numerical constant C_1 and $x^* = (\mathrm{Id} \otimes \iota^{-1} a^*) y^*$.

Let $x^* = (x_j^*)_{1 \leq j \leq d}$ where $x_j^* = u^*(\boldsymbol{e}_j)$ for $j = 1, \ldots, d$. Then let y^* be the image of x^* via $\mathrm{Id} \otimes \iota^{-1} a^*$ as above. Note that the action of u on $x \in S_q^n$ is written as

$$
\begin{aligned}
u(x) &= (\langle u^*(\boldsymbol{e}_j), x \rangle)_{1 \leq j \leq d} = (\langle x_j^*, x \rangle)_{1 \leq j \leq d} = x^*(x) \\
&= [(\mathrm{Id} \otimes \iota^{-1} a^*) y^*](x) = \iota^{-1} a^* (\langle y_j^*, x \rangle)_{1 \leq j \leq cd^2} \ .
\end{aligned}
$$

Thus u is factorized as

$$
u = \iota^{-1} a^* D_\sigma b \ ,
$$

where $b : S_q^n \to \ell_\infty^{cd^2}$ is defined by

$$
b(x) = \left(\frac{\langle y_i^*, x \rangle}{\|y_i^*\|_{S_{q'}^n}} \right)_{1 \leq i \leq cd^2}, \quad \forall x \in S_q^n
$$

and $D_\sigma : \ell_\infty^{cd^2} \to \ell_{q'}^{cd^2}$ is the diagonal operator that takes the element-wise product with $\sigma = (\|y_i^*\|_{S_{q'}^n})_{1 \leq i \leq cd^2} \in \ell_{q'}^{cd^2}$. Then it follows that $\|b\| \leq 1$ and $\|D_\sigma\| = \|\sigma\|_{q'}$.

Therefore, by the definition of $\gamma_\infty^{cd^2}$, we have

$$
\gamma_\infty^{cd^2}(u) \leq \|\iota^{-1} a^* D_\sigma\| \leq \|\iota^{-1}\| \cdot \|\sigma\|_{q'} \cdot \|a^*\| \ ,
$$

where $\|\iota^{-1}\|$ and $\|a^*\|$ are numerical constants.

Finally, we note

$$\|\sigma\|_{q'} = \|y^*\|_{\ell_{q'}^{cd^2}(S_{q'}^n)} \le \|\iota^{-1}\| \cdot \left\| \sum_{j=1}^{d} \xi_j u^*(\boldsymbol{e}_j) \right\|_{G_q^d(S_q^n)^*} ,$$

where the last term is upper-bounded by the right-hand side of our assertion by noncommutative Khintchine inequalities (see e.g. [35]). □

Remark 3.14. For $q \ge 2$, we expect a similar result using polynomial Λ-cb sets. We leave the details to a future publication.

Theorem 3.15. *Suppose* $1 \le q < 2$. *Let*

$$\alpha_d(u) \;=\; \left\| \Big(\sum_{j=1}^{d} [u^*(\boldsymbol{e}_j)]^* [u^*(\boldsymbol{e}_j)] \Big)^{1/2} \right\|_{S_{q'}^n} + \left\| \Big(\sum_{j=1}^{d} [u^*(\boldsymbol{e}_j)] [u^*(\boldsymbol{e}_j)]^* \Big)^{1/2} \right\|_{S_{q'}^n} ,$$

where $q' = q/(q-1)$. *If* $q > 1$, *then*

$$M_{2,\alpha_d}(B_{S_q^n}) \;\le\; C(1+\ln d)^{3/2} (q')^{3/2} .$$

Otherwise

$$M_{2,\alpha_d}(B_{S_1^n}) \;\le\; (1+\ln d)^{3/2} (1+\ln n)^{3/2} .$$

Proof. Theorem 3.15 is analogous to Theorem 3.11 for $p = 2$ and their proofs are almost identical. The only difference is that we replace Lemma 3.10 in the proof of Theorem 3.11 by Lemma 3.13 to obtain an upper bound on $\gamma_\infty^{cd^2}$ in the noncommutative case.

For $q > 1$, we use the fact that $S_{q'}^n$ satisfies $T_2(S_{q'}^n) = \sqrt{q'}$ for all $2 \le q' < \infty$ [45].

For $q = 1$, let $r > 1$ be arbitrary. Since

$$e_k(v : S_1 \to \ell_\infty^m(\ell_2^d)) \le e_k(v : S_r \to \ell_\infty^m(\ell_2^d))), \quad \forall k \in \mathbb{N}, \ ,$$

we have

$$\mathcal{E}_{2,1}(v : S_1 \to \ell_\infty^m(\ell_2^d)) \le \mathcal{E}_{2,1}(v : S_r \to \ell_\infty^m(\ell_2^d)) \ . \tag{40}$$

Furthermore,

$$\alpha_d(v_j : S_1 \to \ell_2^d) \le n^{1/(2r')} \alpha_d(v_j : S_r \to \ell_2^d) \ , \tag{41}$$

where $r' = r/(r-1)$. We choose $r' = \ln n$. Then $n^{1/(2r')}$ is upper-bounded by a numerical constant. Therefore the assertion follows by applying the other case of $q > 1$ together with (40) and (41). □

Remark 3.16. (Noncommutative polytope). We have shown in Section 3.2 that convex hulls generated by a few points induce sparsity models. Our estimate above provides a noncommutative analogue as follows: Let $K \in \ell_2^N$ be given as $K = Q(B_{S_1^n})$ via $Q : S_1^n \to \ell_2^N$. Let X be the Banach space with the norm given as the Minkowski functional of K. Choose α_d on $L(X, \ell_2^d)$ as

$$\alpha_d(u) \ = \ \left\| \Big(\sum_{j=1}^d [Q^* u^*(\boldsymbol{e}_j)]^* [Q^* u^*(\boldsymbol{e}_j)] \Big)^{1/2} \right\|_{S_\infty^n} + \left\| \Big(\sum_{j=1}^d [Q^* u^*(\boldsymbol{e}_j)][Q^* u^*(\boldsymbol{e}_j)]^* \Big)^{1/2} \right\|_{S_\infty^n} .$$

Then we deduce from Theorem 3.15 that

$$M_{2,\alpha_d}(K) \ \le \ C(1+\ln d)^{3/2}(1+\ln n)^{3/2} \ .$$

4 Sparsity models with enough symmetries

In this section, we explore another geometric aspect of generalized sparsity models in terms of isometry group actions on the underlying Banach spaces. When a linear measurement operator is constructed with random sampling of relevant group actions, this geometric property implies the isotropy of the random measurements in (2), which is a key ingredient of the restricted isometry property. To this purpose, we introduce Banach spaces of enough symmetries and the relevant isometric group actions below.

4.1 Banach spaces with enough symmetries

A Banach space of enough symmetries is defined with relevant isometric group actions. Let us first recall relevant definitions. Let G be a group and $\sigma : G \to L(X)$ for a Banach space X. A map $\sigma : G \to L(X)$ is an *affine representation* if there exists $\phi : G \times G \to [0, 2\pi)$ such that

$$\sigma(g)\sigma(h) \;=\; e^{\mathrm{i}\phi(g,h)}\sigma(gh) \;, \quad \forall g, h \in G \;.$$

In other words, an affine representation is almost multiplicative. Affine representations are usually obtained from the representation of the Lie algebra. Moreover, an affine representation yields an honest group presentations $\pi : G \to L(X)$ via conjugation so that

$$\pi(g)(T) \;=\; \sigma(g)T\sigma(g^{-1}) \;, \quad \forall g \in G,\; T \in L(X) \;.$$

Indeed, we have

$$\begin{aligned} \pi(g)\pi(h)T \;&=\; \sigma(g)\sigma(h)T\sigma(h)^{-1}\sigma(g)^{-1} \\ &= \sigma(gh)e^{\mathrm{i}\phi(gh)}Te^{-\mathrm{i}\phi(gh)}\sigma(h^{-1}g^{-1}) \;=\; \pi(gh)T \;. \end{aligned}$$

In the Banach space literature, a space X is called to have *enough symmetries* if there is an affine representation $\sigma : G \to L(X)$ such that

$$\|\sigma(g)x\|_X = \|x\|_X, \quad \forall x \in X,\; \forall g \in G \tag{42}$$

and

$$\sigma(g)T = T\sigma(g), \quad \forall g \in G \implies T = \lambda \mathrm{Id} \;. \tag{43}$$

For a set $\mathcal{S}$ of linear operators in $L(X)$, the *commutant*, denoted by $\mathcal{S}'$, is defined as the set of linear operators those commute with all elements in $\mathcal{S}$, i.e.

$$\mathcal{S}' := \{T \in L(X) \mid TS = ST,\; \forall S \in \mathcal{S}\}.$$

With this definition, the condition in (43) is equivalently written as $(\sigma(G))' = \mathbb{C}\mathrm{Id}$, i.e. the commutant of the orbit $\sigma(G) := \{\sigma(g) \mid g \in G\}$ consists of multiples of the identity. Indeed, this is equivalent to say that σ is *irreducible*, which means $\{0\}$ and X are the only subspaces invariant under group actions [2].

When X is finite-dimensional and G is compact, then an affine representation $\sigma : G \to L(X)$ induces a Hilbert space H via Lewis' theorem so that $\sigma(g)$ is an isometry on both X and H for all $g \in G$ [46]. Furthermore, we also have an equivalent condition for (43). These results are shown in the following lemma.

Lemma 4.1. *Let X be a Banach space of dimension N with enough symmetries with respect to a compact group G. Then there exists an inner product on X, with the corresponding Hilbert space H, and an affine representation $\sigma : G \to L(X)$ such that $\sigma(g)$ is an isometry on both X and H for all $g \in G$ and*

$$\int_G \sigma(g)T\sigma(g^{-1})d\mu(g) \;=\; \frac{\mathrm{tr}(T)}{N}\mathrm{Id}\,, \tag{44}$$

where μ denotes the Haar measure on G.

Proof. Let α be an ideal norm on $L(\ell_2^N, X)$ and $u_0 : \ell_2^N \to X$ be the Lewis map with respect to α such that $\alpha(u_0) = 1$ and

$$\det(u_0) \;=\; \max_{\alpha(u)\leq 1} |\det(u)|\,. \tag{45}$$

Here we have chosen a fixed basis on $X \cong \mathbb{K}^N$ in order to calculate the determinant [34]. Since X has enough symmetries, there exists an affine representation σ such that

$$\sigma(g)(B_X) \;=\; B_X\,, \tag{46}$$

which implies $\mathrm{vol}(\sigma(g)B_X) = |\det(\sigma(g))|^N \cdot \mathrm{vol}(B_X)$ and hence $|\det(\sigma(g))| = 1$. Then it follows that $|\det(\sigma(g)u_0)| = |\det(\sigma(g))| \cdot |\det(u_0)| = |\det(u_0)|$. Furthermore (46) also implies $\alpha(\sigma(g)u_0) \leq 1$. Thus $\sigma(g)u_0$ also attains the maximum in (45). Since the ellipsoid

$\mathcal{E} = u_0(B_2^N)$ is unique for all maximizers u_0 [34], we deduce that $[\sigma(g)](\mathcal{E}) = \mathcal{E}$. This implies that $(u_0^{-1}\sigma(g)u_0)(B_2^N) \subset B_2^N$ and hence $\hat{\sigma}(g) := u_0^{-1}\sigma(g)u_0$ is a contraction in $L(\ell_2^N)$. Applying this to $\sigma(g^{-1})$, which differs from $\sigma(g)^{-1}$, up to a scalar of absolute value 1, we deduce that $\hat{\sigma}(g)$ is a unitary operator, i.e. $[\hat{\sigma}(g)](B_2^N) = B_2^N$. The Hilbert space H is now obtained from the norm $\|x\|_H := \|u_0^{-1}(x)\|_2$, and then $\sigma(g)$ simultaneously preserves the unit ball of X and is a unitary operator on H. Indeed by the definition of the norm in H, it follows that

$$\|\sigma(g)x\|_H = \|u_0^{-1}\sigma(g)u_0u_0^{-1}(x)\|_2 = \|u_0^{-1}(x)\|_2 = \|x\|_H, \quad \forall x \in X.$$

In particular, the linear map

$$\Phi(T) := \int_G \sigma(g)T\sigma(g^{-1})d\mu(g)$$

satisfies $\sigma(g_0)\Phi(T)\sigma(g_0^{-1}) = \Phi(T)$ for all $g_0 \in G$. Thus since X has enough symmetries, it follows that $\Phi(T) = \lambda(T)\mathrm{Id}$, where

$$\lambda(T)N \;=\; \mathrm{tr}(\Phi(T)) \;=\; \int_G \mathrm{tr}(\sigma(g)T\sigma(g^{-1}))d\mu(g) \;=\; \mathrm{tr}(T)\,.$$

The assertion follows by normalization. $\square$

4.2 Examples of isotropic affine representations

A Banach space X can have enough symmetries with respect to multiple isotropic group actions, i.e. there might exist pairs of a compact group G and an affine isotropic representation $\sigma : G \to L(X)$ such that the commutant of the corresponding group actions consists of multiples of the identity. For concreteness, we provide a few examples in the following.

4.2.1 Quantum Fourier transform on ℓ_p^N

Let $X = \ell_p^N$ and each vector $x \in \ell_p^N$ be represented as $[x[0], \ldots, x[N-1]]^\top \in \mathbb{C}^N$, where the indexing is zero-based and modulo N. Let Sh denote the cyclic shift operator defined by

$$\mathrm{Sh}(x)[r] = x[r+1], \quad \forall r \in \mathbb{Z}_N, \quad \forall x \, .$$

Let Λ be the diagonal operator representing the modulation with the Nth primitive root of unity such that

$$\Lambda(x)[r] = e^{\frac{\mathrm{i}2\pi r}{N}} x[r], \quad \forall r \in \mathbb{Z}_N, \quad \forall x \, .$$

Then $\sigma : \mathbb{Z}_N^2 \to L(X)$ defined by

$$\sigma(l,k) \;=\; \Lambda^l \, \mathrm{Sh}^k, \quad \forall (l,k) \in \mathbb{Z}_N^2$$

satisfies

$$\sigma(l,k)\sigma(l',k') \;=\; \Lambda^l \, \mathrm{Sh}^k \, \Lambda^{l'} \, \mathrm{Sh}^{-k} \, \mathrm{Sh}^{k+k'} \;=\; e^{\frac{-\mathrm{i}2\pi l' k}{N}} \sigma(l+l', k+k') \, .$$

Therefore σ is an affine representation.

Since Λ is a single generator of the C^*-algebra ℓ_∞^N, every matrix a commuting with Λ is a diagonal matrix. Furthermore every diagonal matrix commuting with the cyclic shift Sh has constant entries. Thus the commutant of the group actions is given by $(\sigma(\mathbb{Z}_{N^2}))' = \mathbb{C}\mathrm{Id}$. Alternatively, since the orbit of any nonzero x spans ℓ_2^N, σ is irreducible. Furthermore since ℓ_p^N is a lattice and the ℓ_p norm is symmetric, i.e. invariant under generalized permutations, $\|\sigma(l,k)x\|_p = \|x\|_p$ for all $x \in \ell_p^N$ and $(l,k) \in \mathbb{Z}_{N^2}$.

4.2.2 Sign flips and cyclic shifts on ℓ_p^N

Next we consider another affine representation for $X = \ell_p^N$. Let $\tilde{\sigma} : \widetilde{G} \to L(X)$ be defined on $\widetilde{G} = \{-1,1\}^N \rtimes \mathbb{Z}_N$ by

$$\tilde{\sigma}(\varepsilon, k) = D_\varepsilon \mathrm{Sh}^k, \quad \forall \varepsilon \in \{-1,1\}^N, \quad \forall k \in \mathbb{Z}_N \ .$$

Let us show that we deal in fact with a suitable representation. Indeed, since

$$\mathrm{Sh}^{-1} D_\varepsilon \mathrm{Sh} \ = \ D_{\varepsilon'}$$

with $\varepsilon' = \mathrm{Sh}(\varepsilon)$, it follows that

$$\tilde{\sigma}(\widetilde{G}) \ = \ \left\{ D_\varepsilon \mathrm{Sh}^k \ \middle| \ \varepsilon \in \{-1,1\}^N, k \in \mathbb{Z}_N \right\} \subset O_N$$

is a subgroup and the normalized counting measure is the Haar measure. Note that $\widetilde{G}$ is indeed the semi-direct product of $\{-1,1\}^N$ and $\mathbb{Z}_N$. Similarly to the previous example, one can verify that $(\tilde{\sigma}(\widetilde{G}))' = \mathbb{C}\mathrm{Id}$. Furthermore by the construction $\tilde{\sigma}$ also satisfies $\|\tilde{\sigma}(\varepsilon, k)(x)\|_p = \|x\|_p$ for all $x \in \ell_p^N$, $\varepsilon \in \{-1,1\}^N$, and $k \in \mathbb{Z}_N$.

4.2.3 Clifford group and Schatten class

Next we show that there are multiple affine isotropic representations for which the Schatten class $X = S_q^n$ has enough symmetries.

Example 4.2. *Let $\sigma : G \to \mathbb{C}^{n\times n}$ be an affine isotropic representation on ℓ_2^n. Then $\tilde{\sigma} : G \times G \to L(S_q^n)$ given with the left and right multiplications by*

$$[\tilde{\sigma}(g, g')](x) \ = \ \sigma(g) x \sigma(g')^*$$

defines an affine isotropic representation on S_q^n as shown below: It follows from Lemma 4.1 *that σ satisfies* (44)*. Thus*

$$\begin{aligned}
&\int_G\int_G [\sigma(g)\otimes\sigma(g')](T\otimes S)[\sigma(g^{-1})\otimes\sigma(g'^{-1})]d\mu(g)d\mu(g') \\
&=\int_G [\mathrm{Id}\otimes\sigma(g')]\left(\int_G [\sigma(g)\otimes\mathrm{Id}](T\otimes S)[\sigma(g^{-1})\otimes\mathrm{Id}]d\mu(g)\right)[\mathrm{Id}\otimes\sigma(g'^{-1})]d\mu(g') \\
&=\int_G [\mathrm{Id}\otimes\sigma(g')]\left(\int_G \sigma(g)T\sigma(g)^* d\mu(g)\otimes S\right)[\mathrm{Id}\otimes\sigma(g'^{-1})]d\mu(g') \\
&=\int_G [\mathrm{Id}\otimes\sigma(g')]\left(\frac{\mathrm{tr}(T)\mathrm{Id}}{n}\otimes S\right)[\mathrm{Id}\otimes\sigma(g'^{-1})]d\mu(g') \\
&=\frac{\mathrm{tr}(T)\mathrm{Id}}{n}\otimes\int_G \sigma(g')S\sigma(g'^{-1})d\mu(g') = \frac{\mathrm{tr}(T)\mathrm{Id}}{n}\otimes\frac{\mathrm{tr}(S)\mathrm{Id}}{n} = \frac{\mathrm{tr}(T\otimes S)\mathrm{Id}}{n^2}.
\end{aligned}$$

Since each linear operator in $L(S_q^n)$ is written as a superposition of rank-1 operators, the above identity implies that $\tilde{\sigma}$ satisfies (44)*; hence $(\tilde{\sigma}(G\times G))' = \mathbb{C}\mathrm{Id}$. On the other hand, since by Lemma 4.1 $\sigma(g)$ is unitary on ℓ_2^n for all $g\in G$ and the S_q norm is unitarily invariant [32], it follows that $\|\tilde{\sigma}(g,g')(x)\|_{S_q^n} = \|x\|_{S_q^n}$ for all $x\in S_q^n$ and $(g,g')\in G\times G$. For example, S_q^n will have enough symmetries with the tensor product of the aforementioned affine isotropic representations on ℓ_p^n.*

Example 4.3. *Let σ be the affine representation that maps $G=\mathbb{Z}_2^2$ to the left matrix multiplication with the Pauli matrices $I,\varepsilon,J,\mathrm{i}\varepsilon J$ given by*

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Let $\tilde{\sigma}$ be the the tensor product of σ that maps $\mathbb{Z}_2^{2k}$ to the left matrix multiplication with the Kronecker product of k Pauli matrices. Since $(\sigma(\mathbb{Z}_2^2))' = \mathbb{C}\mathrm{Id}$, similarly to the above example, one can show that $\tilde{\sigma}$ satisfies (43)*. Furthermore, since the Kronecker product of unitary matrices is also unitary, it also follows that $\tilde{\sigma}$ satisfies* (42)*. Thus $\tilde{\sigma}$ is an affine isotropic representation on $S_q^{2^k}$. The set of measurements generated with $u: S_q^{2^k}\to\mathbb{C}$ given by $u(x)=\mathrm{tr}(x)$ and $\tilde{\sigma}$ above appears in the compressed sensing in quantum tomography*

[31].

Example 4.4. *The above affine isotropic representation with the Pauli matrices generalizes as follows [35, p. 175]: Let $c_1, \ldots, c_k \in \mathbb{C}^{2^k \times 2^k}$ be the standard Clifford generators that satisfy*

$$c_r = c_r^*, \quad c_r^2 = 1, \quad \forall r, \qquad c_r c_s = -c_s c_r, \quad \forall r \neq s \; .$$

Let σ be an affine representation that maps $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k) \in \{0,1\}^k$ to the left matrix multiplication with $c_1^{\varepsilon_1} c_2^{\varepsilon_2} \ldots c_k^{\varepsilon_k}$. Indeed $\sigma(\varepsilon)\sigma(\varepsilon') = \pm\sigma(\varepsilon\varepsilon')$ for all $\varepsilon, \varepsilon' \in \{0,1\}^k$. One can also show that σ satisfies (42) and (43) with $S_q^{2^k}$ similarly to Example 4.3.

4.3 Smoothed sparsity model and enough symmetries

In general, a Banach space X does not have enough symmetries. We present an example that modifies a given convex set K by averaging over group actions so that the modified set has enough symmetries.

Let $K = Q(S_1^n)$ with $Q : S_1^n \to \ell_2^N$ and G be a finite group. Since $\ell_1^{|G|}$ is a lattice, $\ell_1^{|G|}(S_1^k)$ is defined with

$$\|(x_g)_{g \in G}\|_{\ell_1^{|G|}(S_1^n)} := \sum_{g \in G} \|x_g\|_{S_1} \; .$$

Let $Q^G : \ell_1^{|G|}(S_q^k) \to \ell_2^N$ be defined by

$$Q^G((x_g)_{g \in G}) \; = \; \sum_{g \in G} Q(x_g) \; .$$

Then $K^G = Q^G(B_{\ell_1^{|G|}(S_1^k)})$ be the image of the G-orbit of $B_{S_1^n}$ via Q. Then K^G has enough symmetries with any isotropic group actions with G.

On the other hand, since $\ell_1^{|G|}(S_1^k)$ is isometrically embedded into $S_1^{k|G|}$, similar to Section 3.5, it follows that

$$M_{2,\alpha_d}(K^G) \; \leq \; C(1 + \ln d)^{3/2}(1 + \ln k + \ln |G|)^{3/2} \; .$$

This implies that $M_{2,\alpha_d}(K^G)$ is larger than $M_{2,\alpha_d}(K)$ only by a poly-logarithmic factor in $|G|$.

5 RIP via group action

In this section, we illustrate implication of the general RIP result in Theorem 2.1 for prototype sparsity models when K has enough symmetries with an affine representation and the measurement operator is group-structured accordingly.

5.1 Relaxed atomic sparsity with a finite dictionary

First, we consider the case where K is a polytope given as an absolute convex hull of finitely many points. The corresponding sparsity model is is a generalization of the canonical sparsity model with the unit ℓ_1 ball.

Theorem 5.1 (Polytope)**.** *Let K be an absolute convex hull of M points in $\mathbb{C}^N$ and X be the Banach space with the norm $\|\cdot\|_X$ given as the Minkowski functional of K. Suppose that X has enough symmetries with respect to G with an isotropic affine representation $\sigma : G \to L(X)$. Let $g_1, \ldots, g_m$ be independent copies of a Haar-distributed random variable $g \in G$ and $u : X \to \ell_2^d$ be a linear map satisfying $\mathrm{tr}(u^* u) = N$. Then there exists a numerical constant c such that if*

$$m \geq c\delta^{-2} s \|u\|_{X \to \ell_2^d}^2 \left[(1 + \ln m)(1 + \ln m + \ln d)^2 (1 + \ln M) \vee \ln(\zeta^{-1})\right] ,$$

then

$$\mathbb{P}\left(\sup_{\substack{\|x\| \leq \sqrt{s} \\ \|x\|_2 = 1}} \left| \frac{1}{m} \sum_{j=1}^{m} \|u(\sigma(g_j)x)\|_2^2 - \|x\|_2^2 \right| \geq \delta \vee \delta^2 \right) \leq \zeta .$$

Proof. By Lemma 4.1, we have

$$\mathbb{E}\sigma(g)^* u^* u \sigma(g) = \frac{\mathrm{tr}(u^* u)}{N} \mathrm{Id} = \mathrm{Id} .$$

Thus the random measurements satisfy the isotropy, i.e.

$$\mathbb{E}\,\frac{1}{m}\sum_{j=1}^{m}\|u\sigma(g_j)x\|_2^2 = \|x\|_2^2\,.$$

Let $\alpha_d : L(X, \ell_2^d) \to \mathbb{R}$ be the operator norm. Then, since K is G-invariant, it follows that

$$\alpha_d(u\sigma(g_j)) = \|u\sigma(g_j)\|_{X\to\ell_2^d} = \|u\|_{X\to\ell_2^d}\,.$$

Note that $\|u\|_{X\to\ell_2^d}$ is no longer random. The assertion follows by applying the above results with the upper bound on $M_{2,\|\cdot\|}(K)$ by Corollary 3.6 to Remark 2.2. □

Remark 5.2. When K is not G invariant, one can show the RIP for

$$\widetilde{K} = \operatorname{absconv}\{\sigma(g)x_j \mid 1 \le j \le M,\ g \in G\}$$

instead of K. By construction, $\widetilde{K}$ is G invariant. Moreover, since $K \subset \widetilde{K}$, it follows that $\|x\|_{\widetilde{X}} \le \|x\|_X$ for all $x \in X$, where $\widetilde{X}$ is the Banach space with unit ball $\widetilde{K}$. Therefore the RIP on $\Gamma_s(\widetilde{K})$ implies the RIP on Γ_s. For example, if $G = \mathbb{Z}_N^2$, then this replacement of K by $\widetilde{K}$ will increase the number of measurements for the RIP in Theorem 5.1 by an additive term of $O(\ln N)$.

Let us consider the application of Theorem 5.1 to a specific case where the sparsity model is given by $K = B_1^N$ (or equivalently, $X = \ell_1^N$) and the group-structured measurements are given by the group $G = \mathbb{Z}^{N^2}$, the isotropic affine representation σ in Section 4.2.1, and the map u given as $u(x) = \langle \eta, x\rangle$ with the generator $\eta = [1, \ldots, 1]^\top \in \mathbb{R}^N$. The group actions on x produce the set of its discrete Fourier transform coefficients. Furthermore, the map u satisfies $\|u\|_{X\to\ell_2^d} = \|\eta\|_\infty$ and $\operatorname{tr}(u^*u) = \|\eta\|_2^2 = N$. Since the sparsity model here is a superset of the canonical sparsity model in the standard basis, Theorem 5.1 in this case implies the known RIP result for a partial Fourier operator [42, 38, 18].

5.2 Dual of type-p Banach spaces

Next we consider the case where the norm dual X^* has type p and the measurements are scalar-valued ($d = 1$). Let α_d be the operator norm. Then Lemma 3.7 implies that K has type (p, α_d) if X^* has type p. Therefore we obtain the following theorem.

Theorem 5.3 (Dual of type p)**.** *Let X be an N-dimensional Banach space such that i) X^* has type p where $1 < p \leq 2$; ii) the unit ball K has enough symmetries with an affine representation σ that maps a compact group G to isometries on both X and ℓ_2^N. Let H be a Hilbert space in the same vector space with X such that H is isometrically isomorphic to ℓ_2^N. Let $g_1, \ldots, g_m$ be independent copies of a Haar-distributed random variable g in G and $\eta \in X^*$ satisfy $\|\eta\|_H = \sqrt{N}$. Let $p' = p/(p-1)$. Then there exists a constant $c(p)$ depending only on p such that if*

$$\frac{m^{1-1/p}}{(1+\ln m)^{1+\lfloor p \rfloor - 1/p}} \geq c(p)\delta^{-1}\sqrt{s}T_p(X^*)\|\eta\|_{X^*}$$

and

$$m \ \geq \ c\delta^{-2} s \|\eta\|_{X^*} \ln(\zeta^{-1}) \,,$$

then

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

Proof. Let $u : X \to \mathbb{C}$ be defined by $u(x) = \langle \eta, x \rangle$ for $x \in X$. Then by Lemma 4.1 we have

$$\begin{aligned} \mathbb{E}\sigma(g)^* u^* u \sigma(g) &= \int_G \sigma(g)^* u^* u \sigma(g) d\mu(g) = \int_G \sigma(g^{-1}) u^* u \sigma(g) d\mu(g) \\ &= \frac{\mathrm{tr}(u^* u)}{N} \mathrm{Id} = \frac{\|\eta\|_H^2}{N} \mathrm{Id} = \mathrm{Id} \,. \end{aligned}$$

Furthermore, since $\sigma(g)(B_X) \subset B_X$ we deduce that

$$\mathbb{E} \sup_{1 \leq j \leq m} \sup_{x \in B_X} |\langle \eta, \sigma(g_j)x \rangle|^{2q} \leq \|\eta\|_{X^*}^{2q} \,, \quad \forall q \in \mathbb{N} \,.$$

Lastly Corollary 3.8 implies that $M_{p,\alpha_d}(K) \leq (1/p - 1/2)^{-1}[CT_p(X^*)]^{p'+1}$. Hence the assertion follows from Theorem 2.1. $\square$

Example 5.4. *Schatten class S_q^n of n-by-n square matrices has enough symmetries. Therefore, Theorem 5.3 provides an alternative proof for a near optimal RIP of partial Pauli measurements by Liu [31]. Pauli measurements are given as an orbit of the Clifford group with $\eta = \sqrt{n}\mathrm{Id}_{\mathbb{C}^n}$. Since S_∞^n is not type 2, let X be S_q^n with $q' = \ln n$ instead of S_1^n. Obviously, all rank-s matrices is (K,s)-sparse with $K = B_{S_q^n}$ in our generalized sparsity model. Then $T_2(X^*) \leq \sqrt{\ln n}$. In other words, the complexity of K is a logarithmic term. On the other hand, the incoherence satisfies $\|\eta\|_{S_{q'}^n} \leq e\|\eta\|_{S_\infty^n} = e\sqrt{n}$ and the upper bound is proportional to n. This large incoherence is the penalty for noncommutativity.*

Remark 5.5. As shown in the proof of the two theorems in this section, the isotropy of random group structured measurements is implied by enough symmetries of the underlying Banach space X. Furthermore, it simplified the incoherence condition by leveraging the isometry of all group actions. However, it is not necessary for X to have enough symmetry to get the RIP result. For example, we present an infinite-dimensional example with L_p in the companion paper [27] without relying on the enough symmetries.

6 RIP on low-rank tensors

In this section, we apply the RIP results in the previous sections to demonstrate that the group-structured measurement can be useful for dimensionality reduction of higher-order low-rank tensors. Let $N, n, d \in \mathbb{N}$ satisfy $N = n^d$. We consider the convex set $K \subset B_2^N$ given by the convex hull of rank-1 tensors, i.e.

$$K = \mathrm{absconv}\{y_1 \otimes \cdots \otimes y_d \mid y_j \in B_2^n,\ \forall j = 1, \ldots, d\}\,. \tag{47}$$

Note that K is the unit ball of $(\ell_2^n)^{\otimes_\pi^d}$, which is the projective tensor product of d copies of ℓ_2^n with respect to the largest tensor norm [34]. Let G be any compact group with an affine isotropic action, then the product G^d admits an affine isotropic action which leaves K invariant. For example, the tensor product representation $\mathbb{Z}_{n^{2d}}$ of the quantum Fourier transform shows that K has enough symmetries.

Lemma 6.1. *Let K be defined in (47). There exist rank-1 tensors $x_1, \dots, x_M \in B_2^N$ with $\ln M \ \le\ 3nd(1 + \ln d)$ such that $K \subset e\,\mathrm{absconv}\{x_j \mid 1 \le j \le M\}$.*

Proof. Let $0 < \varepsilon < 1/2$ and Δ be an ε-net for the unit ball B_2^n. Then we may assume that $|\Delta| \le (1 + 2/\varepsilon)^n$ (we consider the real scalar field). It follows that every element $y \in B_2^n$ has a representation

$$y \ =\ \sum_{j=1}^{\infty} \alpha_j z_j$$

with $z_j \in \Delta$ and $\sum_j |\alpha_j| \ \le\ \frac{1+\varepsilon}{1-\varepsilon} \ \le\ (1 + 3\varepsilon)$. This implies

$$y_1 \otimes \cdots \otimes y_d \ =\ \sum_{j_1,\dots,j_d=1}^{\infty} \bigotimes_{l=1}^{d} \alpha_{j_l} z_{j_l}\ .$$

Note that $\bigotimes_{l=1}^{d} z_{j_l} \in \Delta^{\otimes^d}$ and

$$\sum_{j_1,\dots,j_d=1}^{\infty} \left| \prod_{l=1}^{d} \alpha_{j_l} \right| \ \le\ (1 + 3\varepsilon)^d\ .$$

Thus we choose $1/(d + 1) < 3\varepsilon \le 1/d$ and deduce the assertion from

$$\ln |\Delta^{\otimes^d}| \le nd \ln \left(1 + \frac{2}{\varepsilon}\right) \ \le\ nd \ln(13d)\ . \qquad \square$$

Corollary 6.2. *Let X be the Banach space with the unit ball K defined in (47), $0 < \zeta < 1$, and $\boldsymbol{\xi} \sim \mathcal{N}(0, I_N)$. Then for all $r \in \mathbb{N}$ we have*

$$(\mathbb{E}\|\boldsymbol{\xi}\|_{X^*}^r)^{1/r} \ \le\ c \left[\sqrt{r} \vee \sqrt{3nd(1 + \ln d)}\ \right]\ . \tag{48}$$

for a numerical constant c. Furthermore,

$$\mathrm{Prob}\Big(\|\boldsymbol{\xi}\|_{X^*} \geq \sqrt{2[1+3nd(1+\ln d)+\ln(\zeta^{-1})]}\,\Big) \leq \zeta.$$

Here $\|\cdot\|_{X^*}$ *denotes the dual norm of* X*.*

Proof. By Lemma 6.1, there exists $\Delta \subset B_2^N$ such that $K \subset e\,\mathrm{absconv}\,\Delta$ and $\ln|\Delta| \leq 3nd(1+\ln d)$. Then

$$\|\boldsymbol{\xi}\|_{X^*} \;=\; \sup_{x\in K}|\langle x,\boldsymbol{\xi}\rangle| \;\leq\; e\sup_{x\in\Delta}|\langle x,\boldsymbol{\xi}\rangle|\;.$$

Let $p = 3nd\ln d$ and $r \;\geq\; p$. Since the net Δ is contained in the unit ball we deduce

$$(\mathbb{E}\|\boldsymbol{\xi}\|_{X^*}^r)^{1/r} \;\leq\; e(\mathbb{E}\sup_{x\in\Delta}|\langle\boldsymbol{\xi},x\rangle|^r)^{1/r} \;\leq\; e|\Delta|^{1/r}\sup_{x\in\Delta}(\mathbb{E}|(\boldsymbol{\xi},x)|^r)^{1/r} \;\leq\; e^2\sqrt{r}\;.$$

For $r \leq p$ we just use the L_p norm. The second assertion follows by the union bound over Δ. $\square$

Theorem 6.3. *Let* $N = n^d$*,* K *be given in (47),* $K_s = \sqrt{s}K\cap\mathbb{S}^{N-1}$*, and* $\boldsymbol{\xi}\ \mathcal{N}(0,I_N)$*. Let* G *be a compact group with an affine isotropic action,* $g_1,\ldots,g_m$ *be independent random variables in* G^d *with respect to the Haar measure, and* $\sigma : G^d \to O_N$ *be a tensor product representation. Let* $\delta > 0$ *and* $0 < \zeta < 1$*. Then there exists a numerical constant* c *such that if*

$$m \;\geq\; c\delta^{-2}s(1+\ln m)^3\left[1+3nd(1+\ln d)+\ln(\zeta^{-1})\right]^2\;,$$

then

$$\mathbb{P}\left(\sup_{x\in K_s}\left|\frac{1}{m}\sum_{j=1}^m|\langle\sigma(g_j)^*\boldsymbol{\xi},x\rangle|^2-\|x\|_2^2\right|\geq\delta\vee\delta^2\right)\leq\zeta\;.$$

Proof. As in the proof of Lemma 6.2, construct $\Delta \subset B_2^N$ from Lemma 6.1 such that $\ln|\Delta| \;\leq\; 3dn(1+\ln d)$ and $\widetilde{K} \;=\; e\,\mathrm{absconv}\,\Delta$ contains K. Let $\widetilde{X}$ be the Banach space induced from $\widetilde{K}$ such that the unit ball in $\widetilde{X}$ is $\widetilde{K}$. Since $K \subset \widetilde{K}$, it follows that $K_s \subset \sqrt{s}\widetilde{K}\cap\mathbb{S}^{N-1}$. Moreover, we have $\|x\|_{X^*} \leq \|x\|_{\widetilde{X}^*}$ for all x, where $\|\cdot\|_{\widetilde{X}^*}$ denotes the dual norm of $\widetilde{X}$. Therefore by Theorem 5.1 the assertion holds if

$$m \;\geq\; c\delta^{-2} s \|\eta\|_{\widetilde{X}^*}^2 \left[(1+\ln m)^3 (1+\ln M) \vee \ln(\zeta^{-1})\right] \;.$$

Indeed, according to the proof of Lemma 6.2, the right-hand side of the inequality in (48) is also a valid upper bound on $\|\eta\|_{\widetilde{X}^*}$. This completes the proof. □

Let us now compare the estimate in Theorem 6.3 to the Gaussian measurement operator. Let $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m$ be independent copies of $\boldsymbol{\xi} \sim \mathcal{N}(0, I_N)$. Then by Gordon's escape through the mesh [23, Corollary 1.2] and Lemma 6.2, it follows that

$$\mathbb{P}\left(\sup_{x \in K_s} \left| \frac{1}{m} \sum_{j=1}^{m} |\langle \boldsymbol{\xi}_j, x \rangle|^2 - \|x\|_2^2 \right| \geq \delta \vee \delta^2 \right) \leq \zeta$$

provided

$$m \geq c\delta^{-2} \left[nd(1+\ln d) + \ln(\zeta^{-1}) \right] \;.$$

To simplify the expressions for the number of measurements, let us choose ζ not too small so that $\ln(\zeta^{-1})$ is dominated by the other logarithmic terms and then ignore the logarithmic terms. While the Gaussian measurement operator provides the RIP with roughly snd measurements, the group-structured measurement with a Gaussian instrument provides the RIP roughly with sn^2d^2 measurements. However, the suboptimal scaling of m for the group-structured measurement can be compensated by applying a Gaussian matrix to the obtained measurements (see [27, Theorem 4.2]). Since $m \approx sn^2d^2$ is already significantly small compared to the dimension n^d of the ambient space $(\ell_2^n)^{\otimes_d^\pi}$, this two step measurement system is much more practical than the measurement system with a single big Gaussian matrix.

Moreover, besides suboptimal scaling of the number of measurements for the RIP, the group-structured measurement operator has the following advantages. The transformations $\sigma(g)$ preserve both the convex body K and the ℓ_2 norm. The incoherence in this case of group-structured measurement is determined by the instrument. Lemma 6.2 suggests that a random gaussian vector $\boldsymbol{\xi}$ can be a good choice for the instrument in the sense that it makes the incoherence parameter small. There exist fast implementations for certain group action transforms, which enable highly scalable dimensionality reduction for massively sized tensor data.

Funding

This work was supported in part by NSF under grants IIS 14-47879, DMS 15-01103, and CCF 17-18771.

Acknowledgement

M.J. and K.L. thank Yihong Wu, Yoram Bresler, Sjoerd Dirksen, and Holger Rauhut for helpful discussions. The authors also appreciate the anonymous referees for their careful reading and constructive suggestions.

References

[1] Aharon, M., Elad, M. & Bruckstein, A. (2006) K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, **54**(11), 4311.

[2] Arveson, W. (2006) *A short course on spectral theory*, vol. 209. Springer Science & Business Media.

[3] Baraniuk, R., Davenport, M., DeVore, R. & Wakin, M. (2008) A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, **28**(3), 253–263.

[4] Baraniuk, R., Foucart, S., Needell, D., Plan, Y. & Wootters, M. (2017) One-bit compressive sensing of dictionary-sparse signals. *Information and Inference: A Journal of the IMA*, **7**(1), 83–104.

[5] Berg, J. & Löfström, J. (1976) *Interpolation spaces. An introduction*. Springer.

[6] Bigot, J., Boyer, C. & Weiss, P. (2016) An analysis of block sampling strategies in compressed sensing. *IEEE Transactions on Information Theory*, **62**(4), 2125–2139.

[7] Bourgain, J., Pajor, A., Szarek, S. J. & Tomczak-Jaegermann, N. (1989) On the duality problem for entropy numbers of operators. in *Geometric aspects of functional analysis*, pp. 50–63. Springer.

[8] Candes, E. J. & Plan, Y. (2011a) A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, **57**(11), 7235–7254.

[9] ——— (2011b) Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, **57**(4), 2342–2359.

[10] Candes, E. J. & Tao, T. (2005) Decoding by linear programming. *IEEE transactions on information theory*, **51**(12), 4203–4215.

[11] ——— (2006) Near-optimal signal recovery from random projections: Universal encoding strategies?. *Information Theory, IEEE Transactions on*, **52**(12), 5406–5425.

[12] Carl, B. (1985) Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces. in *Annales de l'institut Fourier*, vol. 35, pp. 79–118.

[13] Carl, B. & Pajor, A. (1988) Gelfand numbers of operators with values in a Hilbert space. *Inventiones mathematicae*, **94**(3), 479–504.

[14] Carl, B. & Stephani, I. (1990) *Entropy, Compactness and the Approximation of Operators*, Cambridge Tracts in Mathematics. Cambridge University Press.

[15] Chafai, D., Guédon, O., Lecué, G. & Pajor, A. (2012) *Interactions between compressed sensing, random matrices, and high dimensional geometry*. Société Mathématique de France.

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

[17] Defant, A. & Floret, K. (1992) *Tensor norms and operator ideals*, vol. 176. Elsevier.

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

[19] Eftekhari, A. & Wakin, M. B. (2015) New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, **39**(1), 67–109.

[20] Eftekhari, A., Yap, H. L., Rozell, C. J. & Wakin, M. B. (2015) The restricted isometry property for random block diagonal matrices. *Applied and Computational Harmonic Analysis*, **38**(1), 1–31.

[21] Eftekhari, A., Yap, H. L., Wakin, M. B. & Rozell, C. J. (2018) Stabilizing embedology: Geometry-preserving delay-coordinate maps. *Physical Review E*, **97**(2), 022222.

[22] Foucart, S. & Rauhut, H. (2013) *A mathematical introduction to compressive sensing*, vol. 1. Birkhäuser Basel.

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

[24] Guédon, O., Mendelson, S., Pajor, A. & Tomczak-Jaegermann, N. (2008) Majorizing measures and proportional subsets of bounded orthonormal systems. *Revista matemática iberoamericana*, **24**(3), 1075–1095.

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

[26] Junge, M. & Lee, K. (2015) RIP-like properties in subsampled blind deconvolution. *arXiv preprint arXiv:1511.06146*.

[27] ——— (2017) Generalized notions of sparsity and restricted isometry property. Part II: Applications. *arXiv preprint arXiv:1706.09411*.

[28] Krahmer, F., Mendelson, S. & Rauhut, H. (2014) Suprema of chaos processes and the restricted isometry property. *Communications on Pure and Applied Mathematics*, **67**(11), 1877–1904.

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

[30] Lindenstrauss, J. & Tzafriri, L. (1996) *Classical Banach spaces I and II*, Classics in mathematics. Springer-Verlag.

[31] Liu, Y.-K. (2011) Universal low-rank matrix recovery from Pauli measurements. in *Advances in Neural Information Processing Systems*, pp. 1638–1646.

[32] Mirsky, L. (1960) Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics*, **11**(1), 50–59.

[33] Pisier, G. (1986) Factorization of linear operators and geometry of Banach spaces. Conference Board of the Mathematical Sciences Washington.

[34] ——— (1999) *The volume of convex bodies and Banach space geometry*, vol. 94. Cambridge University Press.

[35] ——— (2003) *Introduction to operator space theory*, vol. 294. Cambridge University Press.

[36] ——— (2009) Remarks on the non-commutative Khintchine inequalities for $0 < p < 2$. *Journal of Functional Analysis*, **256**(12), 4128–4161.

[37] Polak, A. C., Duarte, M. F. & Goeckel, D. L. (2015) Performance bounds for grouped incoherent measurements in compressive sensing. *IEEE Transactions on Signal Processing*, **63**(11), 2877–2887.

[38] Rauhut, H. (2010) Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, **9**, 1–92.

[39] Rauhut, H., Schneider, R. & Stojanac, Z. (2016) Low rank tensor recovery via iterative hard thresholding. *arXiv preprint arXiv:1602.05217*.

[40] Recht, B., Fazel, M. & Parrilo, P. A. (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, **52**(3), 471–501.

[41] Romberg, J. (2009) Compressive sensing by random convolution. *SIAM Journal on Imaging Sciences*, **2**(4), 1098–1128.

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

[43] Schütt, C. (1984) Entropy numbers of diagonal operators between symmetric Banach spaces. *Journal of approximation theory*, **40**(2), 121–128.

[44] Talagrand, M. (1996) Majorizing measures: the generic chaining. *The Annals of Probability*, pp. 1049–1103.

[45] Tomczak-Jaegermann, N. (1974) The moduli of smoothness and convexity and the Rademacher averages of the trace classes S_p $(1 \leq p < \infty)$. *Studia Mathematica*, **50**(2), 163–182.

[46] ——— (1989) *Banach-Mazur distances and finite-dimensional operator ideals*, vol. 38. Longman Sc & Tech.