FOUNDATIONS OF
COMPUTATIONAL
MATHEMATICS
The Journal of the Society for the Foundations of **Computational Mathematics**

CrossMark

Dimensionality Reduction with Subgaussian Matrices: A Unified Theory

Sjoerd Dirksen[1]

Abstract We present a theory for Euclidean dimensionality reduction with sub-gaussian matrices which unifies several restricted isometry property and Johnson–Lindenstrauss-type results obtained earlier for specific datasets. In particular, we recover and, in several cases, improve results for sets of sparse and structured sparse vectors, low-rank matrices and tensors, and smooth manifolds. In addition, we establish a new Johnson–Lindenstrauss embedding for datasets taking the form of an infinite union of subspaces of a Hilbert space.

Keywords Random dimensionality reduction · Johnson–Lindenstrauss embeddings · Restricted isometry properties · Compressed sensing · Union of subspaces

Mathematics Subject Classification 60F10 · 68Q87

1 Introduction

The analysis of high-dimensional data leads to various computational issues which are gathered informally under the term 'curse of dimensionality.' To circumvent such issues, several methods have been proposed to reduce the dimensionality of the data, i.e., to map the dataset in a lower-dimensional space while approximately preserving

✉ Sjoerd Dirksen
 dirksen@mathc.rwth-aachen.de

1 Hausdorff Center for Mathematics, Universität Bonn, Endenicher Allee 60,
 53115 Bonn, Germany

🍀 Springer

certain relevant properties of the set. This is often possible as many high-dimensional datasets possess some additional structure, which ensures that it has a low 'intrinsic dimension' or 'complexity.'

This paper concerns the *random* dimensionality reduction method, which seeks to embed a dataset using a random linear map. Undoubtedly the most famous known result in this direction is the classical embedding of Johnson and Lindenstrauss [29]. They showed that if Φ is the orthogonal projection onto an m-dimensional subspace of $\mathbb{R}^n$ which is chosen uniformly at random, then with high probability Φ preserves the pairwise distances in a given finite subset $\mathcal{P}$ of $\mathbb{R}^n$ up to a multiplicative (or relative) error ε, provided that $m \geq C\varepsilon^{-2} \log |\mathcal{P}|$, where $|\mathcal{P}|$ is the cardinality of $\mathcal{P}$. Simpler proofs of this result later appeared in [15,22]. Due to these historic origins, many authors refer to random dimensionality reduction as the 'random projection method.' However, it is well known that one can replace the random projection matrix Φ by a computationally more attractive subgaussian matrix. These matrices perform equally well, as stated in the following modernized version of the Johnson–Lindenstrauss embedding [27,41].

Theorem 1.1 *(Johnson–Lindenstrauss embedding) Let $\mathcal{P}$ be a set of $|\mathcal{P}|$ points in $\mathbb{R}^n$. Let $\tilde{\Phi}$ be an $m \times n$ matrix with entries $\tilde{\Phi}_{ij}$ which are independent, mean-zero, unit variance and $\sqrt{\alpha}$-subgaussian. Set $\Phi = \frac{1}{\sqrt{m}}\tilde{\Phi}$. There exists an absolute constant $C > 0$ such that for any given $0 < \varepsilon, \eta < 1$, we have*

$$(1 - \varepsilon)\|x - y\|_2^2 \leq \|\Phi(x) - \Phi(y)\|_2^2 \leq (1 + \varepsilon)\|x - y\|_2^2 \quad \text{for all } x, y \in \mathcal{P} \quad (1)$$

with probability $1 - \eta$, provided that

$$m \geq C\alpha^2\varepsilon^{-2} \max\{\log |\mathcal{P}|, \log(\eta^{-1})\}. \quad (2)$$

If $\varepsilon_{\mathcal{P},\Phi}$ denotes the smallest constant such that (1) holds, then we can formulate Theorem 1.1 compactly by saying that $\mathbb{P}(\varepsilon_{\mathcal{P},\Phi} \geq \varepsilon) \leq \eta$ if m satisfies (2). All results in this paper will be phrased in this manner. In general, the dependence on ε and the number of points $|\mathcal{P}|$ in (2) cannot be improved [4,33], and the dependence on ε and η is known to be optimal if $|\mathcal{P}| = 1$ [28].

Random dimensionality reduction is attractive for several reasons. It is easy to implement and computationally inexpensive in comparison with other dimensionality reduction methods such as principal component analysis; see, e.g., [8] for an empirical comparison on image and text data. Moreover, the method is *nonadaptive* or *oblivious* to the dataset, meaning that the method does not require any prior knowledge of the dataset as input. It is therefore particularly suitable as a preprocessing step. Methods incorporating random dimensionality reduction have been proposed for a wide range of tasks, such as approximate nearest-neighbor search [26,27], learning mixtures of Gaussians [14,30], clustering [7,50], manifold learning [25,56], matched field processing [38,40] and least squares regression [18,37]. Various other applications can be found in [55]. Sometimes these methods are coined a 'compressive' version of the original method, e.g., compressive matched field processing [40]. Several of these applications rely on extensions of Theorem 1.1 to an infinite, but structured dataset $\mathcal{P}$. In these

results, the factor $\log |\mathcal{P}|$ in (2) is replaced by a different quantity that represents the intrinsic dimension of the dataset $\mathcal{P}$. For instance, if $\mathcal{P}$ is a K-dimensional subspace, then one can show that $\mathbb{P}(\varepsilon_{\mathcal{P},\Phi} \geq \varepsilon) \leq \eta$ if $m \geq C\alpha^2\varepsilon^{-2} \max\{K, \log(\eta^{-1})\}$. Such Johnson–Lindenstrauss embeddings for subspaces were introduced in [49] for use in numerical linear algebra. We also mention the embedding results for smooth manifolds [6,13,19] which are motivated by manifold learning. In a slightly different vein, some authors have investigated lower bounds on m that guarantee that a subgaussian matrix preserves pairwise distances up to an *additive* rather than a multiplicative error [3,27].

In the signal processing literature, some closely related results appear in the form of *restricted isometry properties*. Recall that a map $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ satisfies a restricted isometry property with constant δ on a set $\mathcal{P}$ if

$$(1-\delta)\|x\|_2^2 \leq \|\Phi(x)\|_2^2 \leq (1+\delta)\|x\|_2^2 \qquad \text{for all } x \in \mathcal{P}.$$

It is a well-known result from compressed sensing that a subgaussian matrix Φ satisfies the restricted isometry property on the set of s-sparse vectors in $\mathbb{R}^n$ with probability $1 - \eta$ if $m \geq C\alpha^2\delta^{-2} \max\{s\log(n/s), \log(\eta^{-1})\}$ [5,12,17,42,43]. This property implies that with high probability, one can recover any s-sparse signal in a stable, robust and algorithmically efficient manner from $m \sim s\log(n/s)$ subgaussian measurements; see, e.g., [21, Chapter 6] and the references therein. Inspired by these developments, restricted isometry properties of subgaussian matrices have been established for various signal sets with structured sparsity [10,20,23]. Another example is the restricted isometry property of subgaussian matrices acting on low-rank matrices. This result is used as a substitute for the restricted isometry property on sparse vectors in the low-rank matrix recovery literature [11,48].

The purpose of this paper is to give a unified treatment of the aforementioned collection of restricted isometry and Johnson–Lindenstrauss-type properties for subgaussian matrices, as well as their 'additive error counterparts.' In Theorem 4.8 we formulate a 'master bound' from which one can deduce these properties for subgaussian maps (as in Definition 4.4) acting on any given dataset in a possibly infinite-dimensional Hilbert space. This result is an extension of earlier work in [24,31,42]; see the discussion after Theorem 4.8 for details. We give a transparent proof using a new tail bound for suprema of empirical processes from [16]. The main focus of our work is to make Theorem 4.8 an accessible tool for nonspecialists, by demonstrating extensively how to apply it to extract results for concrete data structures. As it turns out, in all considered applications we recover the best known results in the literature and often find an improved lower bound on the target dimension m. On several occasions, we also extend earlier results for Gaussian matrices to general subgaussian matrices. This class contains computationally more efficient matrices than Gaussian matrices; see [1] and Example 4.6. Moreover, the extension to subgaussian matrices is of interest for certain signal processing applications; see [21, Section 1.2] and [46] for examples.

To conclude, we give a brief overview of the considered applications. In Sect. 5 we consider sets with low covering dimension. This class of sets includes finite

unions of subspaces, with sets of sparse and cosparse vectors as particular examples, as well as low-rank matrices and tensors. In Sect. 6 we consider sets forming an *infinite* union of subspaces of a Hilbert space. A wide variety of models in signal processing can be expressed in this form. For instance, signals exhibiting structured sparsity, piecewise polynomials, certain finite rate of innovations models and overlapping echoes can be described in this fashion [9,10,20,36]. The main result in this section, Theorem 6.3, establishes a new Johnson–Lindenstrauss-type embedding for an infinite union of subspaces of a Hilbert space. The embedding dimension in this result depends on the maximal dimension of the subspaces and the complexity of the index set measured in terms of the Finsler distance, which is related to the largest principal angles between the subspaces. Our result significantly improves upon recent work in this direction [38,39]; see the discussion after Theorem 6.3. By combining with the main result of [9], we deduce that one can robustly reconstruct a signal in an infinite union of subspaces from a small number of subgaussian measurements using a generalized iterative hard thresholding method; see Remark 6.4. Finally, in Sect. 7 we deduce three different dimensionality reduction results for smooth submanifolds of $\mathbb{R}^n$. We first deduce a guarantee under which lengths of curves in the manifold are preserved uniformly by a subgaussian map. Further on, we give conditions under which pairwise ambient distances are preserved up to an additive error. Finally, we establish Johnson–Lindenstrauss embedding results for smooth manifolds. We first give an improvement of the embedding result for manifolds with low linearization dimension from [3]. In Theorems 7.7 and 7.9 we present embeddings in the spirit of [6,13,19]. In particular, we extend the recent result of [19] from Gaussian to subgaussian matrices and achieve optimal scaling in the error parameter ε.

2 Preliminaries and Notation

Throughout the paper, we use the following terminology. We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote a probability space and write $\mathbb{E}$ for the expected value. For a real-valued random variable X, we define its ψ_2 or *subgaussian norm* by

$$\|X\|_{\psi_2} = \inf\{C > 0 \,:\, \mathbb{E}\exp(|X|^2/C^2) \le 2\}.$$

If $\|X\|_{\psi_2} < \infty$, then we call X a *subgaussian* random variable. In particular, any centered Gaussian random variable g with variance σ^2 is subgaussian and $\|g\|_{\psi_2} \lesssim \sigma$. Also, if $|X|$ is bounded by K, then X is subgaussian and $\|X\|_{\psi_2} \lesssim K$. We call a random vector $X : \Omega \to \mathbb{R}^n$ subgaussian if

$$\sup_{\|x\|_2 \le 1} \|\langle X, x\rangle\|_{\psi_2} < \infty.$$

We say that X is *isotropic* if

$$\mathbb{E}\langle X, x\rangle^2 = \|x\|_2^2, \qquad \text{for all } x \in \mathbb{R}^n.$$

If T is a set, then $d : T \times T \to \mathbb{R}_+$ is called a *semi-metric* on T if $d(x, y) = d(y, x)$ and $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in T$. If $S \subset T$, then we use

$$\Delta_d(S) = \sup_{s,t \in S} d(s, t)$$

to denote its *diameter*.

We conclude by fixing some notation. We use $\| \cdot \|_2$ to denote the Euclidean norm on $\mathbb{R}^n$ and let d_2 denote the associated Euclidean metric. If H is a Hilbert space, then $\langle \cdot, \cdot \rangle$ denotes the inner product on H, $\| \cdot \|_H$ the induced inner product and d_H the induced metric on H. If $T \subset H$, we use $\Delta_H(T)$ to denote its diameter. If $A : H_1 \to H_2$ is a bounded linear operator between two Hilbert spaces, then $\|A\|$ denotes its operator norm. If S is a set, then we let $|S|$ denote its cardinality. Given $0 < \alpha < \infty$, we write $\log^\alpha x := (\log(x))^\alpha$ and $\log_+(x) := \max\{\log(x), 0\}$ for brevity. Finally, we write $A \lesssim B$ if $A \leq CB$ for some universal constant $C > 0$ and write $A \simeq B$ if both $A \lesssim B$ and $A \gtrsim B$ hold.

3 γ_2-Functional, Gaussian Width and Entropy

In this section we discuss the γ_2-functional of a semi-metric space (T, d), which plays a central role in the formulation of Theorem 4.8. Intuitively, one should think of $\gamma_2(T, d)$ as measuring the complexity of (T, d).

Definition 3.1 Let (T, d) be a semi-metric space. A sequence $\mathcal{T} = (T_n)_{n \geq 0}$ of T is called *admissible* if $|T_0| = 1$ and $|T_n| \leq 2^{2^n}$. The γ_2-*functional* of (T, d) is defined by

$$\gamma_2(T, d) = \inf_{(\mathcal{T}, \pi)} \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} d(t, \pi_n(t)),$$

where the infimum is taken over all admissible sequences $\mathcal{T} = (T_n)_{n \geq 0}$ in T and all sequences $\pi = (\pi_n)_{n \geq 0}$ of maps $\pi_n : T \to T_n$.

In the literature it is common to take $\pi_n(t) := \operatorname{argmin}_{s \in T_n} d(t, s)$ in Definition 3.1, i.e., to define the γ_2-functional as

$$\gamma_2(T, d) = \inf_{\mathcal{T}} \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} d(t, T_n),$$

where $d(t, T_n) = \inf_{s \in T_n} d(t, s)$. Our slightly relaxed definition will be convenient later on.

Let us recall the role of the γ_2-functional in the theory of generic chaining. We recall two results from this theory. Suppose that $(X_t)_{t \in T}$ is a real-valued stochastic process, which has subgaussian increments with respect to a semi-metric d. That is, for all $s, t \in T$,

$$\mathbb{P}(|X_t - X_s| \geq u d(t, s)) \leq 2 \exp(-u^2) \qquad (u \geq 0). \tag{3}$$

Talagrand's generic chaining method [54] yields

$$\mathbb{E} \sup_{t \in T} |X_t| \lesssim \gamma_2(T, d).$$

This bound is known to be sharp in the following interesting special case. Suppose that $(G_t)_{t \in T}$ is a centered Gaussian process, and let $d_{\mathrm{can}}(s, t) = (\mathbb{E}|G_s - G_t|^2)^{1/2}$ be the induced canonical metric on T. Then, Talagrand's celebrated majorizing measures theorem [52,53] states that

$$\mathbb{E} \sup_{t \in T} |G_t| \simeq \gamma_2(T, d_{\mathrm{can}}). \tag{4}$$

Let $g = (g_1, \ldots, g_n)$ be a vector consisting of independent standard Gaussian variables and for any $x \in \mathbb{R}^n$ define $G_x = \langle g, x \rangle$. Then $(G_x)_{x \in T}$ is a centered Gaussian process for any given subset T of $\mathbb{R}^n$. Note that the canonical metric d_{can} coincides with the usual Euclidean metric d_2 in this case. Hence, (4) translates into

$$\gamma_2(T, d_2) \simeq \mathbb{E} \sup_{x \in T} |\langle g, x \rangle|. \tag{5}$$

The quantity on the right-hand side is known as the *Gaussian width* of the set T.

The γ_2-functional can be estimated using covering numbers. For any given $u > 0$, let $N(T, d, u)$ denote the covering number of T, i.e., the smallest number of balls of radius u in (T, d) needed to cover T. Then $\log N(T, d, u)$ is called the *u-entropy* of (T, d). Let

$$I_2(T, d) = \int_0^{\Delta_d(T)} \log^{1/2} N(T, d, u) \, du$$

be the associated entropy integral. It is shown in [54, Section 1.2] that

$$\gamma_2(T, d) \lesssim I_2(T, d). \tag{6}$$

and in particular $\gamma_2(T, d) \lesssim \Delta_d(T) \log^{1/2} |T|$ if T is finite. The reverse estimate of (6) fails [54, Section 2.1]. However, if $T \subset \mathbb{R}^n$, then one can show that

$$I_2(T, d_2) \lesssim (\log n) \gamma_2(T, d_2).$$

This worst-case bound is attained by natural objects, such as ellipsoids [54, Section 2.2]. Even though (6) is not sharp, it is a very important tool to estimate the γ_2-functional in practical situations and we will use it several times below.

For our analysis, we use the following tail bound for suprema of empirical processes from [16, Theorem 5.5]. This result improves and extends two earlier results in the same direction of Klartag and Mendelson [31] and Mendelson, Pajor and Tomczak-Jaegermann [42]; see [16] for a detailed comparison.

Theorem 3.2 *Fix a probability space $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ for every $1 \leq i \leq m$. For every $t \in T$ and $1 \leq i \leq m$ let $X_{t,i} \in L^2(\Omega_i)$. Define the process of averages*

$$A_t = \frac{1}{m} \sum_{i=1}^{m} (X_{t,i}^2 - \mathbb{E}X_{t,i}^2). \tag{7}$$

Consider the semi-metric

$$d_{\psi_2}(s, t) = \max_{1 \leq i \leq m} \|X_{s,i} - X_{t,i}\|_{\psi_2} \qquad (s, t \in T)$$

and define the radius

$$\bar{\Delta}_{\psi_2}(T) = \sup_{t \in T} \max_{1 \leq i \leq m} \|X_{t,i}\|_{\psi_2}.$$

There exist constants $c, C > 0$ such that for any $u \geq 1$,

$$\mathbb{P}\left(\sup_{t \in T} |A_t| \geq C\left(\frac{1}{m}\gamma_2^2(T, d_{\psi_2}) + \frac{1}{\sqrt{m}}\bar{\Delta}_{\psi_2}(T)\gamma_2(T, d_{\psi_2})\right) \right.$$
$$\left. + c\left(\sqrt{u}\frac{\bar{\Delta}_{\psi_2}^2(T)}{\sqrt{m}} + u\frac{\bar{\Delta}_{\psi_2}^2(T)}{m}\right)\right) \leq e^{-u}.$$

In a typical application of this result, one chooses m large enough to make the right-hand side inside the probability smaller than a desired threshold. The two terms involving $\sqrt{m}$ are usually dominant in such an application.

4 Master Bound

Throughout, let H be a real Hilbert space. Let $\mathcal{P}$ be a set of points in H. We are interested in reducing the dimensionality of $\mathcal{P}$, meaning that we would like to construct a map $\Phi : H \to \mathbb{R}^m$ with the *target dimension m* as small as possible. We call the dimension of H the *original dimension*, which we think of as being very large or even infinite. All the results in this paper provide a lower bound on m under which $\mathcal{P}$ can be mapped into $\mathbb{R}^m$, while preserving certain properties of the set $\mathcal{P}$. The following definition expresses that Φ approximately preserves the *size* of the original vectors.

Definition 4.1 Let $\mathcal{P}$ be a set in H and let $\Phi : H \to \mathbb{R}^m$. The *restricted isometry constant $\delta_{\mathcal{P},\Phi}$* of Φ on $\mathcal{P}$ is the least possible constant $0 \leq \delta \leq \infty$ such that

$$(1 - \delta)\|x\|_H^2 \leq \|\Phi(x)\|_2^2 \leq (1 + \delta)\|x\|_H^2.$$

It is common parlance to loosely say that a map Φ satisfies *the restricted isometry property* on $\mathcal{P}$ if $\delta_{\mathcal{P},\Phi} < \delta_*$, where $\delta_* < 1$ is some small value.

The following definition expresses that Φ preserves *pairwise distances* between elements of $\mathcal{P}$. We distinguish between a *multiplicative* error and an *additive* error.

Definition 4.2 Let $\mathcal{P}$ be a set in H and let $\Phi : H \to \mathbb{R}^m$. We say that Φ preserves distances on $\mathcal{P}$ with *multiplicative* error $0 < \varepsilon < 1$ if

$$(1-\varepsilon)\|x-y\|_H^2 \le \|\Phi(x)-\Phi(y)\|_2^2 \le (1+\varepsilon)\|x-y\|_H^2 \qquad \text{for all } x, y \in \mathcal{P}. \quad (8)$$

The least possible constant $\varepsilon_{\mathcal{P},\Phi}$ for which this holds is called the *multiplicative precision* of Φ. We say that Φ preserves distances on $\mathcal{P}$ with *additive* error $0 \le \zeta < 1$ if

$$\|x - y\|_H^2 - \zeta \le \|\Phi(x) - \Phi(y)\|_2^2 \le \|x - y\|_H^2 + \zeta \qquad \text{for all } x, y \in \mathcal{P}. \quad (9)$$

The least possible constant $\zeta_{\mathcal{P},\Phi}$ for which this holds is called the *additive precision* of Φ.

The restricted isometry constant and the multiplicative error in Definition 4.2 are closely related. If

$$\mathcal{P}_c = \{x-y \: : \: x, y \in \mathcal{P}\} \quad (10)$$

denotes the set of chords associated with $\mathcal{P}$, then $\varepsilon_{\mathcal{P},\Phi} = \delta_{\mathcal{P}_c,\Phi}$.

Maps Φ that preserve pairwise distances up to a multiplicative error ε are the most interesting for applications as they often preserve additional properties of $\mathcal{P}$. For instance, in applications it is often used that if Φ is in addition linear and $-\mathcal{P} = \mathcal{P}$, then inner products are preserved up to additive error, i.e.,

$$|\langle \Phi(x), \Phi(y) \rangle - \langle x, y \rangle| \le \varepsilon$$

for all unit vectors $x, y \in \mathcal{P}$. This can be readily shown using a polarization argument. More can be said if $\mathcal{P}$ is a manifold; see Sect. 7 below.

Remark 4.3 In parts of the literature, it is customary to say that a map Φ approximately preserves distances on $\mathcal{P}$ with multiplicative error $0 < \hat{\varepsilon} < 1$ if
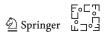
$$(1-\hat{\varepsilon})\|x-y\|_H \le \|\Phi(x)-\Phi(y)\|_2 \le (1 + \hat{\varepsilon})\|x-y\|_H \qquad \text{for all } x, y \in \mathcal{P}. \quad (11)$$

That is, one leaves out the squares in (8). Let $\hat{\varepsilon}_{\mathcal{P},\Phi}$ be the smallest possible $\hat{\varepsilon}$ in (11). One readily checks that $\hat{\varepsilon}_{\mathcal{P},\Phi} \le \hat{\varepsilon}$ if $\varepsilon_{\mathcal{P},\Phi} \le 2\hat{\varepsilon} - \hat{\varepsilon}^2$.

Below we will derive various dimensionality reduction results for the following class of random maps.

Definition 4.4 (Subgaussian map) Let $\mathcal{S}$ be any set of points in H. For every $1 \le i \le m$, let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ be a probability space. Let Ω be the corresponding product probability space. We call $\Phi : \Omega \times H \to \mathbb{R}^m$ a linear, isotropic, subgaussian map, or briefly a *subgaussian map on $\mathcal{S}$* if the following conditions hold.

(a) (Linearity) For any $\omega \in \Omega$ the map $\Phi(\omega) : H \to \mathbb{R}^m$ is linear;
(b) (Independence) For all $x \in \mathcal{S}$ and $1 \le i \le m$, $[\Phi(x)]_i \in L^2(\Omega_i)$;
(c) (Isotropy) For any $x \in \mathcal{S}$, we have $\mathbb{E}\|\Phi(x)\|_2^2 = \|x\|_H^2$;

(d) (Subgaussianity) There is an $\alpha \geq 1$ such that for all $x, y \in \mathcal{S} \cup \{0\}$,

$$\max_{1 \leq i \leq m} \|[\Phi(x) - \Phi(y)]_i\|_{\psi_2} \leq \sqrt{\frac{\alpha}{m}} \|x - y\|_H.$$

Note that the condition $\alpha \geq 1$ is forced by the assumption that Φ is isotropic.

Example 4.5 (Subgaussian matrices) Suppose that $H = \mathbb{R}^n$ for some $n \in \mathbb{N}$, equipped with the Euclidean norm. Let $\tilde{\Phi}$ be an $m \times n$ random matrix, whose rows $\tilde{\Phi}_1, \ldots, \tilde{\Phi}_m$ are independent, mean-zero, isotropic, subgaussian random vectors in $\mathbb{R}^n$. By setting $\Phi = \frac{1}{\sqrt{m}}\tilde{\Phi}$, we obtain a subgaussian map on $\mathbb{R}^n$.

Example 4.6 (Database-friendly maps [1]) As a particular instance of the previous example, we can let $\tilde{\Phi}$ be any random matrix filled with independent, mean-zero, unit variance, subgaussian (in particular, bounded) entries Φ_{ij}. In [1], Achlioptas proposed to take independent random variables satisfying

$$\mathbb{P}(\Phi_{ij} = -\sqrt{3}) = \frac{1}{6}, \qquad \mathbb{P}(\Phi_{ij} = 0) = \frac{2}{3}, \qquad \mathbb{P}(\Phi_{ij} = \sqrt{3}) = \frac{1}{6}.$$

Due to the (expected) large number of zeroes occurring in Φ, this map requires less storage space (or, as it is phrased in [1], it is 'database-friendly') and allows for faster matrix-vector multiplication than a densely populated matrix [1, Section 7]. More generally, for any $q \geq 1$ one can take

$$\mathbb{P}(\Phi_{ij} = -\sqrt{q}) = \frac{1}{2q}, \qquad \mathbb{P}(\Phi_{ij} = 0) = \frac{q-1}{q}, \qquad \mathbb{P}(\Phi_{ij} = \sqrt{q}) = \frac{1}{2q}.$$

One can readily compute that the subgaussian parameter α in part (d) of Definition 4.4 is bounded by q in this case.

Example 4.7 (Infinite-dimensional Hilbert spaces) Suppose that H is a separable Hilbert space. Let $(x_j)_{j \geq 1}$ be any orthonormal basis of H and, for $1 \leq i \leq m$, let $(g_j^{(i)})_{j \geq 1}$ be independent sequences of i.i.d. standard Gaussian random variables. For $1 \leq i \leq m$, we define $\tilde{\Phi}_i : H \to L^2(\Omega)$ by

$$\tilde{\Phi}_i x = \sum_{j=1}^{\infty} g_j^{(i)} \langle x, x_j \rangle.$$

The map $\Phi : \Omega \times H \to \mathbb{R}^m$ defined by $\Phi x = \frac{1}{\sqrt{m}}(\tilde{\Phi}_1 x, \ldots, \tilde{\Phi}_m x)$ is subgaussian on H.

To give a unified presentation of dimensionality reduction results for operators which are size and pairwise distance preserving, we define for a given set $\mathcal{S}$ in H the constant $0 \leq \kappa_{\mathcal{S}, \Phi} \leq \infty$ by

$$\kappa_{\mathcal{S}, \Phi} = \sup_{x \in \mathcal{S}} \left| \|\Phi(x)\|_2^2 - \|x\|_H^2 \right|.$$

The restricted isometry constant of Φ on $\mathcal{P}$ is exactly $\kappa_{\mathcal{P}_{nv}, \Phi}$, where

$$\mathcal{P}_{nv} = \left\{ \frac{x}{\|x\|_H} : x \in \mathcal{P} \right\} \tag{12}$$

is the set of normalized vectors in $\mathcal{P}$. For a pairwise distance preserving operator Φ, the multiplicative error on $\mathcal{P}$ is equal to $\kappa_{\mathcal{P}_{nc}, \Phi}$, where

$$\mathcal{P}_{nc} = \left\{ \frac{x - y}{\|x - y\|_H} : x, y \in \mathcal{P} \right\}$$

is the set of normalized chords corresponding to $\mathcal{P}$. The additive error on $\mathcal{P}$ is equal to $\kappa_{\mathcal{P}_c, \Phi}$, with $\mathcal{P}_c$ as in (10).

For our treatment in Sect. 6, it will be convenient to consider the situation where $\mathcal{S}$ is described by a parameter set Ξ. We will say that $\xi : \Xi \to \mathcal{S}$ is a *parametrization of $\mathcal{S}$* if ξ is a surjective map. To any parametrization ξ, we associate a semi-metric d_ξ on Ξ defined by

$$d_\xi(x, y) := \|\xi(x) - \xi(y)\|_H \qquad (x, y \in \Xi). \tag{13}$$

We can now state a 'master bound.' Every dimensionality reduction result stated below is a corollary of this theorem.

Theorem 4.8 *Let $\mathcal{S}$ be a set of points in H with radius $\bar{\Delta}_H(\mathcal{S}) = \sup_{y \in \mathcal{S}} \|y\|_H$ and let $\xi : \Xi \to \mathcal{S}$ be a parametrization of $\mathcal{S}$. Let $\Phi : \Omega \times H \to \mathbb{R}^m$ be a subgaussian map on $\mathcal{S}$. There is a constant $C > 0$ such that for any $0 < \kappa, \eta < 1$ we have $\mathbb{P}(\kappa_{\mathcal{S}, \Phi} \geq \kappa) \leq \eta$, provided that*

$$m \geq C(\alpha \kappa^{-1} + \alpha^2 \kappa^{-2} \bar{\Delta}_H^2(\mathcal{S})) \, \max\{\gamma_2^2(\Xi, d_\xi), \bar{\Delta}_H^2(\mathcal{S}) \log(\eta^{-1})\}. \tag{14}$$

Proof For any $x \in \Xi$, we write $\Phi_i(x) := \sqrt{m}[\Phi(\xi(x))]_i$. By isotropy of Φ,

$$\|\Phi(x)\|_2^2 - \|x\|_H^2 = \|\Phi(x)\|_2^2 - \mathbb{E}\|\Phi(x)\|_2^2 = \frac{1}{m} \sum_{i=1}^m (\Phi_i(x)^2 - \mathbb{E}\Phi_i(x)^2).$$

We can now set $T = \Xi$ and $X_{t,i} = \Phi_i(x)$ in Theorem 3.2 to obtain for any $u \geq 1$

$$\mathbb{P}\left(\kappa_{\mathcal{S}, \Phi} \geq C \left(\frac{\gamma_2^2(\Xi, d_{\psi_2})}{m} + \frac{\gamma_2(\Xi, d_{\psi_2})\bar{\Delta}_{\psi_2}(\Xi)}{\sqrt{m}} + \sqrt{u}\frac{\bar{\Delta}_{\psi_2}^2(\Xi)}{\sqrt{m}} + u\frac{\bar{\Delta}_{\psi_2}^2(\Xi)}{m} \right) \right)$$
$$\leq e^{-u}.$$

Since Φ is subgaussian, we have for any $x, y \in \Xi$,

$$d_{\psi_2}(x, y) = \max_{1 \leq i \leq m} \|\Phi_i(x) - \Phi_i(y)\|_{\psi_2}$$
$$= \sqrt{m} \max_{1 \leq i \leq m} \|[\Phi(\xi(x)) - \Phi(\xi(y))]_i\|_{\psi_2} \leq \sqrt{\alpha}\|\xi(x) - \xi(y)\|_H = \sqrt{\alpha}d_\xi(x, y)$$

and similarly, for any $x \in \Xi$,

$$\max_{1 \leq i \leq m} \|\Phi_i(\xi(x))\|_{\psi_2} \leq \sqrt{\alpha} \|\xi(x)\|_H \leq \sqrt{\alpha} \bar{\Delta}_H(\mathcal{S}).$$

In particular,

$$\gamma_2(\Xi, d_{\psi_2}) \leq \sqrt{\alpha} \gamma_2(\Xi, d_\xi), \qquad \bar{\Delta}_{\psi_2}(\Xi) \leq \sqrt{\alpha} \bar{\Delta}_H(\mathcal{S}).$$

We conclude that $\mathbb{P}(\kappa_{\mathcal{P},\Phi} \geq \kappa) \leq \eta$ if (14) holds.   □

In the special case that $\mathcal{S}$ is a subset of the unit sphere of H and ξ is the trivial parametrization, Theorem 4.8 corresponds to a result of Mendelson, Pajor and Tomczak-Jaegermann [42, Corollary 2.7], in which the γ_2-functional in (14) is replaced by the Gaussian width of $\mathcal{S}$ (this is equivalent by (5) in this case). They refined important earlier work of Klartag and Mendelson [31], who proved the same result with a suboptimal dependence in η. The result in [42] was obtained much earlier for Gaussian matrices by Gordon [24]. The proof given in [42] makes specific use of the assumption that $\mathcal{S}$ is contained in the unit sphere and cannot be easily modified to cover the more general case considered here.

Remark 4.9 (Anisotropy) One can relax the assumption that the subgaussian map Φ is isotropic on the set $\mathcal{S}$. Suppose that Φ satisfies (a), (b) and (d) in Definition 4.4. Set $\Psi = \mathbb{E}\Phi^*\Phi$, then

$$\mathbb{E}\|\Phi x\|_2^2 = x^* \mathbb{E}(\Phi^*\Phi)x = x^* \Psi x = \|\Psi^{1/2} x\|_H^2.$$

The proof of Theorem 4.8 shows that for any $0 < \kappa, \eta < 1$, we have

$$-\kappa + \|\Psi^{1/2} x\|_H^2 \leq \|\Phi x\|_2^2 \leq \|\Psi^{1/2} x\|_H^2 + \kappa, \qquad \text{for all } x \in \mathcal{S}$$

with probability at least $1 - \eta$, provided that (14) holds.

The following statements are immediate from Theorem 4.8 by setting $\mathcal{S} = \mathcal{P}_{\text{nv}}$, $\mathcal{S} = \mathcal{P}_{\text{nc}}$ and $\mathcal{S} = \mathcal{P}_c$, respectively, and taking the trivial parametrization $\xi(x) = x$.

Corollary 4.10 *Let $\mathcal{P}$ be a set of points and let Φ be as in Theorem 4.8. For any $0 < \delta, \eta < 1$ we have $\mathbb{P}(\delta_{\mathcal{P},\Phi} \geq \delta) \leq \eta$, provided that*
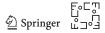
$$m \geq C\alpha^2 \delta^{-2} \max\{\gamma_2^2(\mathcal{P}_{\text{nv}}, d_H), \log(\eta^{-1})\}.$$

Moreover, for any $0 < \varepsilon, \eta < 1$ we have $\mathbb{P}(\varepsilon_{\mathcal{P},\Phi} \geq \varepsilon) \leq \eta$, whenever

$$m \geq C\alpha^2 \varepsilon^{-2} \max\{\gamma_2^2(\mathcal{P}_{\text{nc}}, d_H), \log(\eta^{-1})\}.$$

Finally, for any $0 < \zeta, \eta < 1$ we have $\mathbb{P}(\zeta_{\mathcal{P},\Phi} \geq \zeta) \leq \eta$ if

$$m \geq C(\alpha\zeta^{-1} + \alpha^2\zeta^{-2}\Delta_H^2(\mathcal{P})) \max\{\gamma_2^2(\mathcal{P}, d_H), \Delta_H^2(\mathcal{P}) \log(\eta^{-1})\}.$$

As a first illustration of Corollary 4.10, note that it implies an extension of Theorem 1.1 to general subgaussian maps as

$$\gamma_2^2(\mathcal{P}_{nc}) \lesssim \log |\mathcal{P}_{nc}| \leq 2 \log |\mathcal{P}|.$$

Since the dependence of m on ε, $|\mathcal{P}|$ and η in (2) is optimal [28,33], we see that in general one cannot expect a better dependence of m on κ, $\gamma_2^2(\Xi, d_\xi)$ and η in (14).

In the remainder of this paper, we derive dimensionality reduction results for concrete data structures from Theorem 4.8. The technical work is to derive a good estimate for the complexity parameter $\gamma_2^2(\Xi, d_\xi)$ appearing in (14).

5 Sets with Low Covering Dimension

In this section we consider dimensionality reduction for sets with low covering dimension, in particular finite unions of subspaces.

Definition 5.1 We say that a metric space $(\mathcal{X}, d)$ has *covering dimension* $K > 0$ *with parameter* $c \geq 1$ *and base covering* $N_0 > 0$ if, for all $0 < u \leq 1$,

$$N(\mathcal{X}, d, u\Delta_d(\mathcal{X})) \leq N_0 \left(\frac{c}{u}\right)^K.$$

Often c and N_0 are some small universal constants. In this situation we will loosely say that $(\mathcal{X}, d)$ has covering dimension K.

Example 5.2 (Unit ball of a finite-dimensional space) A well-known example is the unit ball $B_{\mathcal{X}}$ of a K-dimensional normed space $\mathcal{X}$. Using a standard volumetric argument (see, e.g., [21, Proposition C.3]), one shows that for any $0 < u \leq 1$,

$$N(B_{\mathcal{X}}, d_{\mathcal{X}}, u) \leq \left(1 + \frac{2}{u}\right)^K \leq \left(\frac{3}{u}\right)^K.$$

Example 5.3 (Doubling dimension) If $(\mathcal{X}, d)$ is a metric space, then the *doubling constant* $\lambda_{\mathcal{X}}$ of $\mathcal{X}$ is the smallest integer λ such that for any $x \in \mathcal{X}$ and $u > 0$, the ball $B(x, u)$ can be covered by at most λ balls of radius $u/2$. One can show that [27] for all $0 < u \leq 1$,

$$N(\mathcal{X}, d, u\Delta_d(\mathcal{X})) \leq \left(\frac{2}{u}\right)^{\log_2 \lambda_{\mathcal{X}}}.$$

That is, $(\mathcal{X}, d)$ has covering dimension $\log_2 \lambda_{\mathcal{X}}$. The latter number is also known as the *doubling dimension* of (X, d). The notion of doubling dimension was considered in the context of dimensionality reduction in, for example, [3] and [27].

We now formulate a dimensionality reduction result for sets with low covering dimension. In the proof we use that for $c, u_* > 0$,

$$\int_0^{u_*} \log^{1/2}(c/u) \, du \leq u_* \log^{1/2}\left(\frac{ec}{u_*}\right). \tag{15}$$

A short proof of this estimate can be found in [21, Lemma C.9]. The second statement in the following result was obtained by a different method in [9, Theorem 3] (note that this result was erroneously stated in [10]), but with a suboptimal dependence on δ.

Corollary 5.4 *Let $S_1, \ldots, S_k$ be subsets of a Hilbert space H and let $S = \cup_{i=1}^{k} S_i$. Set*

$$S_{i,\mathrm{nv}} = \{x/\|x\|_2 \ : \ x \in S_i\}.$$

Suppose that $S_{i,\mathrm{nv}}$ has covering dimension K_i with parameter c_i and base covering $N_{0,i}$ with respect to d_H. Set $K = \max_i K_i$, $c = \max_i c_i$ and $N_0 = \max_i N_{0,i}$. Let $\Phi : \Omega \times H \to \mathbb{R}^m$ be a subgaussian map on S_{nv}. Then, for any $0 < \delta, \eta < 1$, we have $\mathbb{P}(\delta_{S,\Phi} \geq \delta) \leq \eta$, provided that

$$m \geq C\alpha^2 \delta^{-2} \ \max\{\log k + \log N_0 + K \log(c), \log(\eta^{-1})\}.$$

In particular, if each S_i is a K_i-dimensional subspace of $\mathbb{R}^n$, then $\mathbb{P}(\delta_{S,\Phi} \geq \delta) \leq \eta$ if

$$m \geq C\alpha^2 \delta^{-2} \ \max\{\log k + K, \log(\eta^{-1})\}.$$

From Corollary 5.4, one can readily deduce that $\mathbb{P}(\varepsilon_{S,\Phi} \geq \varepsilon) \leq \eta$ if

$$m \geq C\alpha^2 \delta^{-2} \ \max\{\log k + \log N_0 + K \log(c), \log(\eta^{-1})\},$$

see the proof of Theorem 6.3 below.

Proof We use (6) to estimate

$$\gamma_2(S_{\mathrm{nv}}, d_H) \lesssim \int_0^1 \log^{1/2} N(S_{\mathrm{nv}}, d_H, u) \, du.$$

Clearly, if N_i is an u-net for $S_{i,\mathrm{nv}}$, then $\cup_{i=1}^{k} N_i$ is an u-net for S_{nv}. Therefore, using our assumption on the $S_{i,\mathrm{nv}}$, we obtain

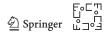$$N(S_{\mathrm{nv}}, d_H, u) \leq \sum_{i=1}^{k} N(S_{i,\mathrm{nv}}, d_H, u) \leq \sum_{i=1}^{k} \left(\frac{c_i}{u}\right)^{K_i} \leq k\left(\frac{c}{u}\right)^{K}.$$

Using (15) we arrive at

$$\gamma_2(S_{\mathrm{nv}}, d_H) \lesssim \int_0^1 \log^{1/2}(k(c/u)^K) \, du$$

$$\leq \log^{1/2}(k) + K^{1/2} \int_0^1 \log^{1/2}(c/u) \, du \lesssim \log^{1/2}(k) + K^{1/2} \log^{1/2}(c).$$

The first part of the result now follows from the first statement in Corollary 4.10, and the second part follows by the observation in Example 5.2. □

To illustrate Corollary 5.4, we consider four examples.

Example 5.5 (Sparse vectors: the 'usual' RIP) We derive the restricted isometry property on s-sparse vectors for subgaussian maps $\Phi : \Omega \times \mathbb{R}^n \to \mathbb{R}^m$, a classical result from compressed sensing [5,12,17,42,43]. For $x \in \mathbb{R}^n$ we set

$$\|x\|_0 = |\{1 \leq i \leq n \ : \ x_i \neq 0\}|.$$

A vector is called s-sparse if $\|x\|_0 \leq s$. Let

$$D_{s,n} = \{x \in \mathbb{R}^n \ : \ \|x\|_0 \leq s\}$$

be the set of s-sparse vectors. The restricted isometry constant δ_s of Φ is defined as the smallest constant δ such that

$$(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2 \qquad \text{for all } x \in D_{s,n}.$$

In our notation, $\delta_s = \delta_{D_{s,n},\Phi}$. Note that we can write

$$D_{s,n} = \bigcup_{I \subset \{1,\dots,n\}, |I|=s} S_I,$$

where S_I is the s-dimensional subspace

$$S_I = \{x \in \mathbb{R}^n \ : \ x_i = 0 \text{ if } i \in I^c\}.$$

Since the number of s-element subsets of $\{1, \dots, n\}$ is

$$\binom{n}{s} \leq \left(\frac{en}{s}\right)^s,$$

the second part of Corollary 5.4 implies that $\mathbb{P}(\delta_s \geq \delta) \leq \eta$, provided that

$$m \geq C\alpha^2\delta^{-2} \ \max\{s \log(en/s), \log(\eta^{-1})\}. \tag{16}$$

As is well known, if the restricted isometry constant δ_s of Φ is small enough, then Φ satisfies an 'exact reconstruction property' for s-sparse vectors: One can recover all s-sparse $x \in \mathbb{R}^n$ exactly from its linear measurements $y = \Phi x$ via ℓ_1-minimization. Conversely, if Φ has the latter property, then necessarily $m \gtrsim s \log(n/s)$ (see, e.g., [21, Theorem 10.11]). In particular, the scaling of the lower bound (16) in n and s is optimal. For the latter conclusion, the subgaussian assumption on Φ is important. For example, if Φ has independent, isotropic, *log-concave* rows, then there is a constant $c(\delta) > 0$ depending on δ such that $\delta_s \leq \delta$ with high probability if

$$m \geq c(\delta)s \log^2(en/s),$$

and the square in the log-factor cannot be removed [2]. On the other hand, this matrix still satisfies, with high probability, the exact reconstruction property for s-sparse

vectors via ℓ_1-minimization in the optimal regime $m \simeq s \log(n/s)$ [32]. Near-matching necessary and sufficient conditions for a matrix to satisfy the exact reconstruction property in this regime were recently derived in [34].

Example 5.6 (Cosparse vectors: the Ψ-RIP) In many signal processing applications, signals of interest are not sparse themselves in the standard basis, but can rather be represented as a sparse vector. Let $\Upsilon : \mathbb{R}^n \rightarrow \mathbb{R}^p$ be a linear operator, which is usually called the 'analysis operator' in the literature. We are interested in elements $x \in \mathbb{R}^n$ such that Υx is sparse. It has become customary to count the number of zero components of Υx, rather than the number of nonzero ones. Accordingly, a vector $x \in \mathbb{R}^n$ is called *l-cosparse with respect to* Υ if there is a set $\Lambda \subset \{1, \ldots, p\}$ with $|\Lambda| = l$ such that $\Upsilon_\Lambda x = 0$, where $\Upsilon_\Lambda : \mathbb{R}^n \rightarrow \mathbb{R}^p$ is the operator obtained by setting the rows of Υ indexed by Λ^c equal to zero. Let N_Λ be the null space of Υ_Λ, then we can write the set of l-cosparse vectors as

$$C_{\Upsilon,l,p} = \bigcup_{|\Lambda|=l} N_\Lambda.$$

The Υ-*restricted isometry constant* δ_l of Φ is defined as the smallest possible $\delta > 0$ such that

$$(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2 \quad \text{for all } x \in C_{\Upsilon,l,p}.$$

In our notation, $\delta_l = \delta_{C_{\Upsilon,l,p},\Phi}$. Observe that $\dim(N_\Lambda) \leq n - l$ and the number of l-element subsets of $\{1, \ldots, p\}$ is

$$\binom{p}{l} \leq \left(\frac{ep}{l}\right)^l.$$

The second part of Corollary 5.4 now implies that $\mathbb{P}(\delta_l \geq \delta) \leq \eta$ if

$$m \geq C\alpha^2 \delta^{-2} \ \max\{l \log(ep/l) + (n - l), \log(\eta^{-1})\}.$$

This result improves upon the RIP result in [23, Theorem 3.8]. In fact, as is heuristically explained in [44, Section 6.1], one cannot expect a better lower bound for m. An upper bound on δ_l leads to performance guarantees for greedy-like recovery algorithms for cosparse vectors from subgaussian measurements; see [23] for some results in this direction.

Example 5.7 (Matrix RIP) Another direct consequence of Corollary 5.4 is a new proof of the restricted isometry property for subgaussian matrix maps. This property plays the same role in low-rank matrix recovery as the 'usual' restricted isometry property discussed in Example 5.5 plays in compressed sensing; see, e.g., [11,48] for further information.

We use the following notation. Given two matrices $X, Y \in \mathbb{R}^{n_1 \times n_2}$, we consider the Frobenius inner product

$$\langle X, Y \rangle = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{ij} Y_{ij}.$$

Let $\|X\|_F = \langle X, X \rangle^{1/2}$ be the corresponding norm and $d_F(X, Y) = \|X - Y\|_F$ be the induced metric. Also, we use $\mathrm{Rank}(X)$ to denote the rank of X. For $1 \leq r \leq \min\{n_1, n_2\}$, we define the restricted isometry constant δ_r of a map $\Phi : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ as the smallest constant $\delta > 0$ such that

$$(1 - \delta)\|X\|_F^2 \leq \|\Phi X\|_F^2 \leq (1 + \delta)\|X\|_F^2 \quad \text{for all } X \in \mathbb{R}^{n_1 \times n_2} \text{ with } \mathrm{Rank}(X) \leq r.$$

If we set

$$D_r = \{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_F = 1, \mathrm{Rank}(X) \leq r\},$$

then $\delta_r = \delta_{D_r, \Phi}$ in our notation. The covering number estimate [11, Lemma 3.1]

$$N(D_r, d_F, u) \leq (9/u)^{r(n_1 + n_2 + 1)} \quad (0 < u \leq 1),$$

shows that D_r has covering dimension $r(n_1 + n_2 + 1)$ in $(\mathbb{R}^{n_1 \times n_2}, d_F)$. The first part of Corollary 5.4 implies for any subgaussian map Φ that $\mathbb{P}(\delta_r \geq \delta) \leq \eta$ if

$$m \geq C\alpha^2 \delta^{-2} \max\{r(n_1 + n_2 + 1), \log(\eta^{-1})\}.$$

This result was obtained in a different way in [11, Theorem 2.3]; see also [48, Theorem 4.2] for a slightly worse result.

Example 5.8 (Tensor RIP) The previous example can be extended to higher-order tensors. Let $d \geq 2$ and set $n = (n_1, \ldots, n_d) \in \mathbb{N}^d$. Given two tensors $X, Y \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, we consider their Frobenius inner product

$$\langle X, Y \rangle = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} X(i_1, \ldots, i_d) Y(i_1, \ldots, i_d).$$

Let $\|X\|_F = \langle X, X \rangle^{1/2}$ be the corresponding norm and $d_F(X, Y) = \|X - Y\|_F$ be the induced metric. Let $\mathrm{Rank}(X)$ denote the rank of X associated with its HOSVD decomposition; see [47] for more information.

Given $r = (r_1, \ldots, r_d)$, $0 \leq r_i \leq n_i$, we define the restricted isometry constant δ_r of a map $\Phi : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^m$ as the smallest constant $\delta > 0$ such that

$$(1 - \delta)\|X\|_F^2 \leq \|\Phi X\|_F^2 \leq (1 + \delta)\|X\|_F^2 \quad \text{for all } X \in \mathbb{R}^{n_1 \times \cdots \times n_d} \text{ with } \mathrm{Rank}(X) \leq r.$$

If we set

$$D_r = \{X \in \mathbb{R}^{n_1 \times \cdots \times n_d} \ : \ \|X\|_F = 1, \text{Rank}(X) \le r\},$$

then $\delta_r = \delta_{D_r, \Phi}$ in our notation. It is shown in [47] that for any $0 < u \le 1$

$$N(D_r, d_F, u) \le (3(d+1)/u)^{r_1 \cdots r_d + \sum_{i=1}^d n_i r_i}.$$

In other words, D_r has covering dimension $r_1 \cdots r_d + \sum_{i=1}^d n_i r_i$ with parameter $3(d+1)$. Corollary 5.4 implies that for any subgaussian map Φ and any $0 < \delta, \eta < 1$, we have $\mathbb{P}(\delta_r \ge \delta) \le \eta$, provided that

$$m \ge C\alpha^2 \delta^{-2} \ \max \left\{ \left(r_1 \cdots r_d + \sum_{i=1}^d n_i r_i \right) \log(d), \log(\eta^{-1}) \right\}.$$

This result was obtained originally for a more restricted class of subgaussian maps in [47].

The results presented in the four examples above can be derived in a different, more elementary fashion using the ε-net technique; see [5,11,23,43,48], and [47], respectively. In fact, this is already true for the statement in Corollary 5.4. The results in the following two sections cannot be achieved using the ε-net technique, however, and therefore generic chaining methods, which are at the basis of Theorem 3.2, become necessary to achieve the best results.

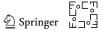## 6 Infinite Union of Subspaces

Many sets of signals relevant to signal processing can be expressed as a possibly infinite union of finite-dimensional subspaces of a Hilbert space. For example, signals with various forms of structured sparsity (e.g., sparse, cosparse, block sparse, simultaneously sparse data), piecewise polynomials, certain finite rate of innovations models and overlapping echoes can be described in this fashion [9,10,20,36]. In this section we prove a Johnson–Lindenstrauss embedding result for an infinite union of subspaces.

We consider the following setup. Let H be a Hilbert space, and let B_H denote its unit ball. Let Θ be a parameter set, and suppose that for every $\theta \in \Theta$, we are given a finite-dimensional subspace S_θ of H. We use P_θ to denote the orthogonal projection onto S_θ. It will be natural to consider the Finsler metric on Θ, which is defined by

$$d_{\text{Fin}}(\theta, \theta') := \|P_\theta - P_{\theta'}\| \qquad (\theta, \theta' \in \Theta).$$

If two subspaces S_θ, $S_{\theta'}$ have the same dimension, then the Finsler distance satisfies

$$d_{\text{Fin}}(\theta, \theta') = \sin(\gamma(\theta, \theta')),$$

where $\gamma(\theta, \theta')$ is the largest canonical angle (or largest principal angle) between the subspaces S_θ and $S_{\theta'}$ [51, Corollary 2.6]. Define the union

$$\mathcal{U} = \bigcup_{\theta \in \Theta} S_\theta. \tag{17}$$

We are interested in reducing the dimensionality of $\mathcal{U}$ using a subgaussian map. To achieve this, we apply Theorem 4.8 with a suitable parametrization of $\mathcal{U} \cap B_H$. We estimate the relevant γ_2-functional in the following lemma.

Lemma 6.1 *Set $K = \sup_{\theta \in \Theta} \dim(S_\theta)$. Let $\xi : \Theta \times B_H \to \mathcal{U} \cap B_H$ be the parametrization defined by $\xi(\theta, x) = P_\theta x$ and let d_ξ be as in (13). Then,*

$$\gamma_2(\Theta \times B_H, d_\xi) \lesssim \sqrt{K} + \gamma_2(\Theta, d_{\mathrm{Fin}}).$$

Proof Let $(\Theta_n)_{n \geq 0}$ be any admissible sequence in Θ and let $\rho = (\rho_n)_{n \geq 0}$ be an associated sequence of maps $\rho_n : \Theta \to \Theta_n$. For any given $\theta \in \Theta$ and $n \geq 0$, we define a semi-metric on B_H by

$$d_{n,\theta}(x, y) = \|P_{\rho_n(\theta)}(x - y)\|_H.$$

Next, for every $\theta \in \Theta$, we define an admissible sequence $\mathcal{H}_\theta = (H_{n,\theta})_{n \geq 0}$ of B_H by

$$H_{n,\theta} := \mathrm{argmin}_A \sup_{x \in B_H} d_{n,\theta}(x, A),$$

where the minimization is over all subsets A of B_H with $|A| \leq 2^{2^n}$. We use

$$e_{n,\theta} = \inf_A \sup_{x \in B_H} d_{n,\theta}(x, A) = \sup_{x \in B_H} d_{n,\theta}(x, H_{n,\theta})$$

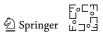to denote the associated entropy numbers. Finally, we define

$$\sigma_{n,\theta}(x) = \mathrm{argmin}_{y \in H_{n,\theta}} d_{n,\theta}(x, y).$$

For completeness we set Θ_{-1} equal to Θ_0, $d_{-1,\theta}$ equal to $d_{0,\theta}$ and $H_{-1,\theta}$ equal to $H_{0,\theta}$. Now we define

$$T_n = \{(\rho_{n-1}(\theta), \sigma_{n-1,\theta}(x)) \in \Theta \times B_H \ : \ \theta \in \Theta, x \in B_H\}.$$

Note that $\sigma_{n,\theta}(x)$ depends only on θ through $\rho_n(\theta)$. It follows that

$$|T_n| \leq |\Theta_{n-1}| \, 2^{2^{n-1}} \leq 2^{2^n}$$

and therefore $\mathcal{T} = (T_n)_{n \geq 0}$ is an admissible sequence for $\Theta \times B_H$. For $(\theta, x) \in \Theta \times B_H$, we define $\pi_n(\theta, x) = (\rho_{n-1}(\theta), \sigma_{n-1,\theta}(x))$. By the triangle inequality,

$$
\begin{aligned}
d_\xi((\theta, x), \pi_n(\theta, x)) &= \|P_\theta x - P_{\rho_{n-1}(\theta)} \sigma_{n-1,\theta}(x)\|_H \\
&\leq \|(P_\theta - P_{\rho_{n-1}(\theta)})x\|_H + \|P_{\rho_{n-1}(\theta)}(x - \sigma_{n-1,\theta}(x))\|_H \\
&\leq \|P_\theta - P_{\rho_{n-1}(\theta)}\| + d_{n-1,\theta}(x, \sigma_{n-1,\theta}(x)) \\
&\leq \|P_\theta - P_{\rho_{n-1}(\theta)}\| + e_{n-1,\theta},
\end{aligned}
$$

where we used in the final estimate that

$$
d_{n,\theta}(x, \sigma_{n,\theta}(x)) = d_{n,\theta}(x, H_{n,\theta}) \leq \sup_{x \in B_H} d_{n,\theta}(x, H_{n,\theta}) = e_{n,\theta}.
$$

Using these observations, we obtain

$$
\begin{aligned}
\gamma_2(\Theta \times B_H, d_\xi) &\leq \sum_{n \geq 0} 2^{n/2} d_\xi((\theta, x), \pi_n(\theta, x)) \\
&\leq \sum_{n \geq 0} 2^{n/2} d_{\text{Fin}}(\theta, \rho_{n-1}(\theta)) + 2^{n/2} e_{n-1,\theta} \\
&\leq (1 + \sqrt{2})\left(\sum_{n \geq 0} 2^{n/2} d_{\text{Fin}}(\theta, \rho_n(\theta)) + \sum_{n \geq 0} 2^{n/2} e_{n,\theta} \right).
\end{aligned}
$$

It remains to bound the second term on the right-hand side. Observe that

$$
e_{n,\theta} = \inf\{u \ : \ N(B_H, d_{n,\theta}, u) \leq 2^{2^n}\}.
$$

If (a_α) is a u-net for the unit ball in $S_{\rho_n(\theta)}$ with respect to d_H and we pick x_α such that $a_\alpha = P_{\rho_n(\theta)} x_\alpha$, then (x_α) is a u-net for B_H with respect to $d_{n,\theta}$. Since $S_{\rho_n(\theta)}$ is at most K-dimensional, we find for all $u > 0$,

$$
N(B_H, d_{n,\theta}, u) \leq N(B_{\mathbb{R}^K}, d_2, u).
$$

Thus, we can conclude that $e_{n,\theta} \leq e_n$, where

$$
e_n = \inf\{u \ : \ N(B_{\mathbb{R}^K}, d_2, u) \leq 2^{2^n}\}.
$$

Now, if $u < e_n$ then $N(B_{\mathbb{R}^K}, d_2, u) \geq 2^{2^n} + 1$ and hence we can estimate

$$
\begin{aligned}
\left(1 - \frac{1}{\sqrt{2}}\right) \sum_{n \geq 0} 2^{n/2} e_n &\leq \sum_{n \geq 0} 2^{n/2} e_n - \sum_{n \geq 1} 2^{(n-1)/2} e_n \\
&= \sum_{n \geq 0} 2^{n/2}(e_n - e_{n+1}) \\
&\leq \frac{1}{\log^{1/2}(2)} \sum_{n \geq 0} \log^{1/2}(1 + 2^{2^n}) \, (e_n - e_{n+1})
\end{aligned}
$$

$$\leq \frac{1}{\log^{1/2}(2)} \sum_{n \geq 0} \int_{e_{n+1}}^{e_n} \log^{1/2} N(B_{\mathbb{R}^K}, d_2, u) \, du$$

$$= \frac{1}{\log^{1/2}(2)} \int_0^1 \log^{1/2} N(B_{\mathbb{R}^K}, d_2, u) \, du.$$

As observed in Example 5.2,

$$N(B_{\mathbb{R}^K}, d_2, u) \leq (1 + 2u^{-1})^K.$$

Putting these estimates together, we conclude using (15) that

$$\sum_{n \geq 0} 2^{n/2} e_{n,\theta} \leq \sqrt{K} \left(\log^{1/2}(2) - \frac{\log^{1/2}(2)}{\sqrt{2}} \right)^{-1} \int_0^1 \log^{1/2}(1 + 2u^{-1}) \, du$$

$$\leq \sqrt{K} \left(\log^{1/2}(2) - \frac{\log^{1/2}(2)}{\sqrt{2}} \right)^{-1} \log^{1/2}(3e).$$

This completes the proof.   □

Theorem 4.8 and Lemma 6.1 together imply the following result.

Theorem 6.2 *Let $\mathcal{U}$ be the union of subspaces defined in (17) and let $K = \sup_{\theta \in \Theta} \dim(S_\theta)$. Let $\Phi : \Omega \times H \to \mathbb{R}^m$ be a subgaussian map on $\mathcal{U}$. Then there is a constant $C > 0$ such that for any $0 < \delta, \eta < 1$ we have $\mathbb{P}(\delta_{\mathcal{U}, \Phi} \geq \delta) \leq \eta$, provided that*

$$m \geq C\alpha^2 \delta^{-2} \, \max\{K + \gamma_2^2(\Theta, d_{\mathrm{Fin}}), \log(\eta^{-1})\}.$$

Proof Recall from (12) that $\delta_{\mathcal{U}, \Phi} = \kappa_{\mathcal{U}_{\mathrm{nv}}, \Phi}$ and clearly $\kappa_{\mathcal{U}_{\mathrm{nv}}, \Phi} \leq \kappa_{\mathcal{U} \cap B_H, \Phi}$. Let ξ be the parametrization of $\mathcal{U} \cap B_H$ defined in Lemma 6.1. By Theorem 4.8 we have $\mathbb{P}(\kappa_{\mathcal{U} \cap B_H, \Phi} \geq \delta) \leq \eta$ if

$$m \geq C\alpha^2 \delta^{-2} \, \max\{\gamma_2^2(\Theta \times B_H, d_\xi), \log(\eta^{-1})\}.$$

The assertion now follows from Lemma 6.1.   □

Theorem 6.2 improves upon the second part of Corollary 5.4 even if Θ is a finite set. Indeed, this follows from the bound $\gamma_2^2(\Theta, d_{\mathrm{Fin}}) \lesssim \log |\Theta|$.

Let us now derive a condition under which Φ preserves pairwise distances in $\mathcal{U}$. For $\theta, \theta' \in \Theta$, let $S_{(\theta, \theta')}$ be the subspace spanned by S_θ and $S_{\theta'}$ and let $P_{(\theta, \theta')}$ be the projection onto this subspace.

Theorem 6.3 *Let $\Phi : \Omega \times H \to \mathbb{R}^m$ be a subgaussian map on $\mathcal{U}$. Set $K = \sup_{\theta, \theta'} \dim(S_{(\theta, \theta')})$. On the set $\Theta \times \Theta$ consider the metric*

$$d_{\mathrm{Fin}}((\theta, \theta'), (\tau, \tau')) = \|P_{(\theta, \theta')} - P_{(\tau, \tau')}\|.$$

Then, there is a constant $C > 0$ such that for any $0 < \varepsilon, \eta < 1$ we have $\mathbb{P}(\varepsilon_{\mathcal{U},\Phi} \geq \varepsilon) \leq \eta$, provided that

$$m \geq C\alpha^2 \varepsilon^{-2} \max\{K + \gamma_2^2(\Theta \times \Theta, d_{\mathrm{Fin}}), \log(\eta^{-1})\}. \tag{18}$$

Proof Recall that $\varepsilon_{\mathcal{U},\Phi} = \delta_{\mathcal{U}-\mathcal{U},\Phi}$. Since

$$\mathcal{U} - \mathcal{U} \subset \mathcal{U}_* := \cup_{(\theta,\theta') \in \Theta \times \Theta} S_{(\theta,\theta')},$$

we have $\delta_{\mathcal{U}-\mathcal{U},\Phi} \leq \delta_{\mathcal{U}_*,\Phi}$. The result follows by applying Theorem 6.2 to $\mathcal{U}_*$, noting that $\dim(S_{(\theta,\theta')}) \leq 2K$ for all $(\theta, \theta') \in S_{(\theta,\theta')}$. $\qquad\square$

Clearly, if there exists a one-to-one map from $\mathcal{U}$ into $\mathbb{R}^m$, then we must have $m \geq K$. In particular, the scaling of m in K in (18) cannot be improved.

Remark 6.4 Together with the main result of [9], Theorem 6.3 implies the following very general uniform signal recovery result. Suppose that we wish to recover a vector $x \in \mathcal{U}$ from m noisy measurements $y \in \mathbb{R}^m$ given by

$$y = \Phi x + e, \tag{19}$$

where $e \in H$ represents the measurement error. If m satisfies (18), then in the terminology of [9] the subgaussian map Φ is with probability $1 - \eta$ a bilipschitz map on $\mathcal{U}$ with constants $1 - \varepsilon$ and $1 + \varepsilon$. Therefore, if $(1 + \varepsilon)/(1 - \varepsilon) < 3/2$, then with probability $1 - \eta$ we can recover *any* $x \in \mathcal{U}$ robustly from the m measurements y in (19) using a projective Landweber algorithm. We refer to [9, Theorem 2] for details and a quantitative statement.

In [38,39], Mantzel and Romberg prove a version of Theorem 6.2 for a matrix Φ populated with i.i.d. standard Gaussian entries and discuss some interesting applications. They assume that Θ has covering dimension K_{Fin} with respect to d_{Fin}, with base covering N_0. Their result in [39] states (in our terminology) that $\mathbb{P}(\delta_{\mathcal{U},\Phi} \geq \delta) \leq \eta$, provided that

$$m \geq C\delta^{-2} \max\{K(K_{\mathrm{Fin}} + \log K + \log N_0 + \log(\eta^{-1})),$$
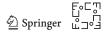$$(K_{\mathrm{Fin}} + \log K + \log N_0 + \log(\eta^{-1}))^2\}. \tag{20}$$

Note that in this setup Theorem 6.2 implies that (cf. the argument in the proof of Corollary 5.4)

$$m \geq C\delta^{-2} \max\{K + \log N_0 + K_{\mathrm{Fin}}, \log(\eta^{-1})\} \tag{21}$$

is already sufficient. Moreover, this statement extends to any subgaussian map, in particular the database-friendly map discussed in Example 4.6.

The approach in [38,39] is very different from ours. The idea is to write

$$\delta_{\mathcal{U},\Phi} = \sup_{\theta \in \Theta} \|P_\theta \Phi^* \Phi P_\theta - P_\theta\|$$

and to estimate the expected value of the right-hand side using a (classical) chaining argument in the operator norm, based on the noncommutative Bernstein inequality. Note that this approach cannot yield the improved condition (21). For example, the factor $K \log(K)$ in (20) is incurred through the use of the noncommutative Bernstein inequality, and therefore an artifact of the used method.

7 Manifolds

Let $\mathcal{M}$ be a K-dimensional C^1-submanifold of $\mathbb{R}^n$, equipped with the Riemannian metric induced by the Euclidean inner product on $\mathbb{R}^n$. We use the following standard notation and terminology. For any $x \in \mathcal{M}$, we let $T_x\mathcal{M}$ denote the tangent space of $\mathcal{M}$ at x and let $P_x : \mathbb{R}^n \to T_x\mathcal{M}$ be the associated projection onto $T_x\mathcal{M}$. We use

$$T\mathcal{M} = \bigcup_{x \in \mathcal{M}} T_x\mathcal{M}$$

to denote the tangent bundle of $\mathcal{M}$. If $\gamma : [a, b] \to \mathcal{M}$ is a piecewise C^1 curve in $\mathcal{M}$, then its length is defined as

$$L(\gamma) = \int_a^b \|\gamma'(t)\|_2 \, dt.$$

For $x, y \in \mathcal{M}$, let $d_{\mathcal{M}}(x, y)$ be the geodesic distance between x and y, which can be described as

$$d_{\mathcal{M}}(x, y) = \inf\{L(\gamma) \; : \; \gamma : [a, b] \to \mathcal{M} \text{ piecewise } C^1, \; a, b \in \mathbb{R},$$
$$\gamma(a) = x, \; \gamma(b) = y\}.$$

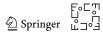For more information on Riemannian (sub)manifolds, we refer to [35].

Below we prove three different types of dimensionality reduction results for a subgaussian map Φ. We derive a sufficient condition under which Φ uniformly preserves the lengths of all curves in $\mathcal{M}$ up to a specified multiplicative error and conditions under which ambient distances are preserved up to an additive error and a multiplicative error, respectively.

7.1 Preservation of Curve Lengths

We can immediately apply Theorem 6.2 to derive a condition under which Φ *uniformly* preserves the lengths of all curves in $\mathcal{M}$ up to a specified multiplicative error.

Theorem 7.1 *Let $\mathcal{M}$ be a K-dimensional C^1-submanifold of $\mathbb{R}^n$. Let $\Phi : \Omega \times \mathbb{R}^n \to \mathbb{R}^m$ be a subgaussian map. There is a constant $C > 0$ such that for any $0 < \varepsilon, \eta < 1$ we have with probability at least $1 - \eta$ for any piecewise C^1-curve γ in $\mathcal{M}$,*

$$(1 - \varepsilon)L(\gamma) \le L(\Phi\gamma) \le (1 + \varepsilon)L(\gamma) \tag{22}$$

provided that

$$m \geq C\alpha^2(2\varepsilon - \varepsilon^2)^{-2} \ \max\{K + \gamma_2^2(\mathcal{M}, d_{\text{Fin}}), \log(\eta^{-1})\}. \tag{23}$$

Proof Let $\gamma : [a, b] \to \mathcal{M}$ be any piecewise C^1-curve in $\mathcal{M}$, then $\Phi\gamma$ is a piecewise C^1-curve and $(\Phi\gamma)'(t) = \Phi\gamma'(t)$, whenever γ is differentiable at t. Therefore,

$$(1 - \delta_{T\mathcal{M},\Phi})\|\gamma'(t)\|_2^2 \leq \|(\Phi\gamma)'(t)\|_2^2 \leq (1 + \delta_{T\mathcal{M},\Phi})\|\gamma'(t)\|_2^2.$$

Note that if $\delta_{T\mathcal{M},\Phi} \leq 2\varepsilon - \varepsilon^2$, then (see Remark 4.3 for a similar observation)

$$(1 - \varepsilon)\|\gamma'(t)\|_2 \leq \|(\Phi\gamma)'(t)\|_2 \leq (1 + \varepsilon)\|\gamma'(t)\|_2.$$

Integrating on both sides over $[a, b]$ yields

$$(1 - \varepsilon)L(\gamma) \leq L(\Phi\gamma) \leq (1 + \varepsilon)L(\gamma).$$

By Theorem 6.2, applied with $\Theta = \mathcal{M}$ and $S_\theta = T_x\mathcal{M}$, we have $\mathbb{P}(\delta_{T\mathcal{M},\Phi} \geq 2\varepsilon - \varepsilon^2) \leq \eta$ under condition (23) and this implies the result.   □

Remark 7.2 If the map Φ in Theorem 7.1 also happens to be a manifold embedding, i.e., an immersion that is homeomorphic onto its image, then it preserves geodesic distances. That is, if for a given $0 < \varepsilon < 1$, (22) holds for all piecewise C^1-curves in $\mathcal{M}$, then

$$(1 - \varepsilon)d_\mathcal{M}(x, y) \leq d_{\Phi\mathcal{M}}(\Phi x, \Phi y) \leq (1 + \varepsilon)d_\mathcal{M}(x, y), \qquad \text{for all } x, y \in \mathcal{M}.$$

Indeed, for given $x, y \in \mathcal{M}$, let $\gamma_{g,\Phi}$ be a geodesic between $\Phi(x)$ and $\Phi(y)$, and let γ be the preimage of $\gamma_{g,\Phi}$. By (22),

$$(1 - \varepsilon)d_\mathcal{M}(x, y) \leq (1 - \varepsilon)L(\gamma) \leq L(\gamma_{g,\Phi}) = d_{\Phi\mathcal{M}}(\Phi(x), \Phi(y)).$$

Similarly, if γ_g is a geodesic between x and y in $\mathcal{M}$, then

$$d_{\Phi\mathcal{M}}(\Phi(x), \Phi(y)) \leq L(\Phi\gamma_g) \leq (1 + \varepsilon)L(\gamma_g) = (1 + \varepsilon)d_\mathcal{M}(x, y).$$

7.2 Preservation of Ambient Distances: Additive Error

We briefly consider maps that preserve pairwise ambient distances up to a specified additive error. The following result is similar to a result established for random projections in [3, Theorem 9].

Proposition 7.3 *Let $\mathcal{M}$ be a C^1-manifold with doubling dimension $D_\mathcal{M}$ in the geodesic distance $d_\mathcal{M}$ and let $\Delta_\mathcal{M}$ be its diameter in $d_\mathcal{M}$. Let $\Phi : \Omega \times \mathbb{R}^n \to \mathbb{R}^m$ be*

a subgaussian map. Then, there is a constant $C > 0$ such that $\mathbb{P}(\zeta_{\mathcal{M},\Phi} \geq \zeta) \leq \eta$, provided that

$$m \geq C(\alpha\zeta^{-1} + \alpha^2\zeta^{-2}\Delta_{\mathcal{M}}^2)\Delta_{\mathcal{M}}^2 \ \max\{D_{\mathcal{M}}, \log(\eta^{-1})\}.$$

Proof Since $d_2 \leq d_{\mathcal{M}}$, we find using (6)

$$\gamma_2(\mathcal{M}, d_2) \lesssim \int_0^{\Delta_{\mathcal{M}}} \log^{1/2} N(\mathcal{M}, d_{\mathcal{M}}, \varepsilon) \, d\varepsilon$$

$$= \Delta_{\mathcal{M}} \int_0^1 \log^{1/2} N(\mathcal{M}, d_{\mathcal{M}}, \varepsilon\Delta_{\mathcal{M}}) \, d\varepsilon$$

$$\leq \Delta_{\mathcal{M}} D_{\mathcal{M}}^{1/2} \int_0^1 \log^{1/2}(c/\varepsilon) \, d\varepsilon \lesssim \Delta_{\mathcal{M}} D_{\mathcal{M}}^{1/2},$$

where in the final step we used (15). The result is now immediate from the third statement in Corollary 4.10. □

If γ is a C^1-curve in $\mathbb{R}^n$, then it has doubling dimension 2 with respect to the geodesic distance. Therefore, Proposition 7.3 implies in this case that with probability $1 - \eta$

$$\|x - y\|_2^2 - \zeta \leq \|\Phi(x - y)\|_2^2 \leq \|x - y\|_2^2 + \zeta, \qquad \text{for all } x, y \in \gamma,$$

whenever

$$m \geq C(\alpha\zeta^{-1} + \alpha^2\zeta^{-2}\Delta_\gamma^2)\Delta_\gamma^2 \max\{2, \log(\eta^{-1})\}.$$

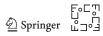### 7.3 Preservation of Ambient Distances: Multiplicative Error

We will now investigate under which conditions a subgaussian map Φ on $\mathcal{M}$ preserves pairwise ambient distances up to a small multiplicative error. These maps also approximately preserve several other properties of the manifold, such as its volume and the length and curvature of curves in the manifold (see [6, Section 4.2] for a discussion). Results in this direction were first obtained in [6] and improved upon in [13,19].

Let us first observe an embedding result for manifolds with a low linearization dimension, which substantially improves [3, Theorem 8].

Corollary 7.4 *For every $1 \leq i \leq k$ let $\mathcal{M}_i$ be smooth submanifold of $\mathbb{R}^n$ with linearization dimension K_i. Set $\mathcal{M} = \cup_{i=1}^k \mathcal{M}_i$ and $K = \max_i K_i$. Let $\Phi : \Omega \times \mathbb{R}^n \to \mathbb{R}^m$ be a subgaussian map. Then there is a constant $C > 0$ such that for every $\varepsilon, \eta > 0$ we have $\mathbb{P}(\varepsilon_{\mathcal{M},\Phi} \geq \varepsilon) \leq \eta$ if*

$$m \geq C\alpha^2\varepsilon^{-2} \ \max\{\log k + K, \log(\eta^{-1})\}.$$

Proof By [3, Lemma 3], $\mathcal{M}_i$ is contained in an affine subspace of dimension K_i and therefore in a linear subspace of dimension $K_i + 1$. The result is now immediate from the second statement in Corollary 5.4. □

To derive the main results of this section, Theorems 7.7 and 7.9, we apply Corollary 4.10 and estimate the γ_2-functional of the set $\mathcal{M}_{\mathrm{nc}}$ of normalized chords. We use (6), i.e.,

$$\gamma_2(\mathcal{M}_{\mathrm{nc}}, d_2) \lesssim \int_0^1 \log^{1/2} N(\mathcal{M}_{\mathrm{nc}}, d_2, u) \, du$$

and estimate the covering numbers of $\mathcal{M}_{\mathrm{nc}}$. The idea behind the covering number estimates, which is already implicit in [13], is to divide the set of normalized chords into two categories. Firstly, one considers normalized chords corresponding to $x, y \in \mathcal{M}$ which are 'close' in the Euclidean metric. In [13], these chords are called the 'short chords,' which should be taken as shorthand for 'the normalized chords corresponding to short chords.' Let

$$\mathrm{Ch}(x, y) = \frac{y - x}{\|y - x\|_2}$$

denote the normalized chord from x to y. Since $\mathrm{Ch}(x, y)$ converges to a unit tangent vector in $T_x\mathcal{M}$ as y approaches x, it is clear that this part of the covering number estimate requires good control of the 'intrinsic dimension' of the tangent bundle of $\mathcal{M}$. Secondly, one considers normalized chords corresponding to x and y which are 'far apart' in Euclidean distance (the 'long chords' in the terminology of [13]). These chords can be approximated well by chords $\mathrm{Ch}(a, b)$, where a and b are taken from a covering of $\mathcal{M}$ itself; see Lemma 7.6 below. To be able to decide whether two points are 'close' or 'far apart,' we need to quantify how well we can approximate a normalized chord by a tangent vector. For this purpose we introduce the following parameter.

Definition 7.5 If $\mathcal{M}$ is a C^1-submanifold of $\mathbb{R}^n$, then we let $\iota(\mathcal{M})$ be the smallest constant $0 < \iota \le \infty$ satisfying

$$\|\mathrm{Ch}(x_1, x_2) - P_{x_1}\mathrm{Ch}(x_1, x_2)\|_2 \le \iota \|x_1 - x_2\|_2, \qquad \text{for all } x_1, x_2 \in \mathcal{M}.$$
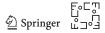
In the proof of Theorem 7.7, we use the following observation, which is implicitly used in [13]. It is readily proven using the triangle and reverse triangle inequalities.

Lemma 7.6 If $x_1, x_2 \in \mathbb{R}^n$ satisfy $\|x_1 - x_2\|_2 \ge t > 0$, then

$$\|\mathrm{Ch}(x_1, x_2) - \mathrm{Ch}(y_1, y_2)\|_2 \le 2t^{-1}(\|x_1 - y_1\|_2 + \|x_2 - y_2\|_2).$$

Theorem 7.7 Let $\mathcal{M}$ be a K-dimensional C^1-submanifold of $\mathbb{R}^n$. Let $\Phi : \Omega \times \mathbb{R}^n \to \mathbb{R}^m$ be a subgaussian map. Suppose that $\mathcal{M}$ has covering dimensions K_2 and K_{Fin} with respect to d_2 and d_{Fin}, respectively. Then there is a constant $C > 0$ such that for any $0 < \varepsilon, \eta < 1$ we have $\mathbb{P}(\varepsilon_{\mathcal{M}, \Phi} \ge \varepsilon) \le \eta$, provided that

$$m \ge C\alpha^2\varepsilon^{-2} \, \max\{K_2 \log_+(\iota(\mathcal{M})\Delta_{d_2}(\mathcal{M})) + K_{\mathrm{Fin}} + K, \log(\eta^{-1})\}.$$

Proof Let $0 < a, b, c, t < \infty$ be parameters to be determined later. Let N_2 be an a-net of $\mathcal{M}$ with respect to the Euclidean distance d_2 and let N_{Fin} be a b-net for $\mathcal{M}$ with respect to d_{Fin}. Finally, for any $y \in N_{\mathrm{Fin}}$ let N_y be a c-net for the unit sphere $\mathcal{S}_{\mathbb{R}^n}$ in $\mathbb{R}^n$ with respect to the induced semi-metric $d_y(z_1, z_2) := \|P_y(z_1 - z_2)\|_2$.

Suppose first that $\|x_1 - x_2\|_2 > t$. Let $y_1, y_2 \in N_2$ be such that $\|x_1 - y_1\|_2 < a$ and $\|x_2 - y_2\|_2 < a$. By Lemma 7.6,

$$\|\mathrm{Ch}(x_1, x_2) - \mathrm{Ch}(y_1, y_2)\|_2 \leq 2t^{-1}(\|x_1 - y_2\|_2 + \|x_2 - y_2\|_2) \leq 4t^{-1}a.$$

Suppose now that $\|x_1 - x_2\|_2 \leq t$. Pick $y \in N_{\mathrm{Fin}}$ such that $\|P_{x_1} - P_y\| < b$ and subsequently $z \in N_y$ such that $\|P_y(\mathrm{Ch}(x_1, x_2) - z)\|_2 < c$. Letting ι be as in Definition 7.5, we find

$$
\begin{aligned}
\|\mathrm{Ch}(x_1, &x_2) - P_y z\|_2 \\
&\leq \|\mathrm{Ch}(x_1, x_2) - P_{x_1}\mathrm{Ch}(x_1, x_2)\|_2 + \|P_{x_1}\mathrm{Ch}(x_1, x_2) - P_y\mathrm{Ch}(x_1, x_2)\|_2 \\
&\quad + \|P_y(\mathrm{Ch}(x_1, x_2) - z)\|_2 \\
&\leq \iota t + b + c.
\end{aligned}
$$

Now let $0 < u \leq 1$. From our estimates, we see that if we pick $t = u/(3\iota)$, $a = u^2/(12\iota)$, $b = u/3$ and $c = u/3$, then

$$\{\mathrm{Ch}(y_1, y_2) \ : \ y_1, y_2 \in N_2\} \cup \{P_y z \ : \ y \in N_{\mathrm{Fin}}, \ z \in N_y\}$$

yields a u-net for $\mathcal{M}_{\mathrm{nc}}$ with respect to d_2. Since for every $y \in N_{\mathrm{Fin}}$ and $v > 0$,

$$N(\mathcal{S}_{\mathbb{R}^n}, d_y, v) \leq N(B_{\mathbb{R}^K}, d_2, v) \leq (1 + 2v^{-1})^K,$$

we obtain

$$N(\mathcal{M}_{\mathrm{nc}}, d_2, u) \leq N^2\left(\mathcal{M}, d_2, \frac{u^2}{12\iota}\right) + N\left(\mathcal{M}, d_{\mathrm{Fin}}, \frac{u}{3}\right)\left(1 + \frac{6}{u}\right)^K. \tag{24}$$

By (6),

$$
\begin{aligned}
\gamma_2(\mathcal{M}_{\mathrm{nc}}, d_2) &\lesssim \int_0^1 \log^{1/2} N(\mathcal{M}_{\mathrm{nc}}, d_2, u) \, du \\
&\leq 2\sqrt{2} \int_0^1 \log^{1/2} N\left(\mathcal{M}, d_2, \frac{u^2}{12\iota}\right) du \\
&\quad + 2 \int_0^1 \log^{1/2} N\left(\mathcal{M}, d_{\mathrm{Fin}}, \frac{u}{3}\right) du + 2\sqrt{K} \int_0^1 \log^{1/2}\left(1 + \frac{6}{u}\right) du \\
&\leq 2\sqrt{2K_2} \int_0^1 \log_+^{1/2}\left(\frac{12\iota\Delta_{d_2}(\mathcal{M})}{u^2}\right) du \\
&\quad + 2\sqrt{K_{\mathrm{Fin}}} \int_0^1 \log^{1/2}\left(\frac{3}{u}\right) du + \sqrt{K} \int_0^1 \log^{1/2}\left(1 + \frac{6}{u}\right) du \\
&\lesssim \sqrt{K_2} \log_+^{1/2}(\iota\Delta_{d_2}(\mathcal{M})) + \sqrt{K_{\mathrm{Fin}}} + \sqrt{K}, \tag{25}
\end{aligned}
$$

where in the final step we used (15). The result now follows from the second statement in Corollary 4.10.   □

We conclude by proving a result related to [6,19] using some tools from [19], which can in turn be traced back to [45]. Recall that the *reach* $\tau(\mathcal{M})$ of a smooth submanifold $\mathcal{M}$ of $\mathbb{R}^n$ is the smallest $\tau > 0$ such that some point of $\mathbb{R}^n$ at distance τ from $\mathcal{M}$ has two distinct points of $\mathcal{M}$ as closest points in $\mathcal{M}$.

Lemma 7.8 *If $\mathcal{M}$ has reach τ, then $\iota(\mathcal{M}) \leq 2\tau^{-1}$. Moreover, for any $x_1, x_2 \in \mathcal{M}$,*

$$d_{\mathrm{Fin}}(x_1, x_2) \leq 2\sqrt{2}\tau^{-1/2}\|x_1 - x_2\|_2^{1/2}.$$

Proof Suppose first that $\|x_1 - x_2\|_2 \leq \tau/2$. Then,

$$\|\mathrm{Ch}(x_1, x_2) - P_{x_1}\mathrm{Ch}(x_1, x_2)\|_2 = \sin(\angle(\mathrm{Ch}(x_1, x_2), P_{x_1}\mathrm{Ch}(x_1, x_2)))$$
$$= \sin(\angle(x_2 - x_1, P_{x_1}(x_2 - x_1))) \leq \frac{\|x_1 - x_2\|_2}{2\tau},$$

where the final inequality follows from [19, Lemma 3]. On the other hand, if $\|x_1 - x_2\|_2 > \tau/2$, then trivially,

$$\|\mathrm{Ch}(x_1, x_2) - P_{x_1}\mathrm{Ch}(x_1, x_2)\|_2 \leq 1 \leq 2\tau^{-1}\|x_1 - x_2\|_2.$$

The second statement for x_1, x_2 satisfying $\|x_1 - x_2\|_2 < \tau/2$ follows from [19, Lemma 10] and is trivial if $\|x_1 - x_2\|_2 \geq \tau/2$.   □

Theorem 7.9 *Let $\mathcal{M}$ be a K-dimensional C^∞-submanifold of $\mathbb{R}^n$ with reach τ and covering dimension K_2 with respect to d_2. Let $\Phi : \Omega \times \mathbb{R}^n \to \mathbb{R}^m$ be a subgaussian map. Then there is a constant $C > 0$ such that for any $0 < \varepsilon, \eta < 1$ we have $\mathbb{P}(\varepsilon_{\mathcal{M},\Phi} \geq \varepsilon) \leq \eta$, provided that*

$$m \geq C\alpha^2\varepsilon^{-2} \, \max\{K_2 \log_+(\tau^{-1}\Delta_{d_2}(\mathcal{M})) + K, \log(\eta^{-1})\}.$$

If $\mathcal{M}$ has volume $V_{\mathcal{M}}$, then $\mathbb{P}(\varepsilon_{\mathcal{M},\Phi} \geq \varepsilon) \leq \eta$ if

$$m \geq C\alpha^2\varepsilon^{-2} \, \max\{K \log_+(K\tau^{-1}) + K + \log_+(V_{\mathcal{M}}), \log(\eta^{-1})\}.$$

Proof By the second part of Lemma 7.8, if N is a v-net of $\mathcal{M}$ with respect to d_2, then it is also a $2\sqrt{2}\tau^{-1/2}\sqrt{v}$-net with respect to d_{Fin}. Hence,

$$N(\mathcal{M}, d_{\mathrm{Fin}}, 2\sqrt{2}\tau^{-1/2}\sqrt{v}) \leq N(\mathcal{M}, d_2, v),$$

which implies that for any $u > 0$,

$$N(\mathcal{M}, d_{\mathrm{Fin}}, u) \leq N\left(\mathcal{M}, d_2, \frac{\tau}{8}u^2\right).$$

Combining this with our estimate (24) in the proof of Theorem 7.7, we obtain

$$N(\mathcal{M}_{\mathrm{nc}}, d_2, u) \le N\Big(\mathcal{M}, d_2, \frac{u^2}{12\iota}\Big) + N\Big(\mathcal{M}, d_{\mathrm{Fin}}, \frac{u}{3}\Big)\Big(1 + \frac{6}{u}\Big)^K$$
$$\le N\Big(\mathcal{M}, d_2, \frac{\tau u^2}{24}\Big) + N\Big(\mathcal{M}, d_2, \frac{\tau u^2}{72}\Big)\Big(1 + \frac{6}{u}\Big)^K, \quad (26)$$

where we applied Lemma 7.8. By a computation similar to (25), we find

$$\gamma_2(\mathcal{M}_{\mathrm{nc}}, d_2) \lesssim \sqrt{K_2} \log_+^{1/2}(\tau^{-1} \Delta_{d_2}(\mathcal{M})) + \sqrt{K}.$$

The first statement now follows from Corollary 4.10.

For the second result, we use that for any $v \le \tau/2$ (cf. [19, Lemma 14])

$$N(\mathcal{M}, d_2, v) \le \Big(\frac{v^2}{4} - \frac{v^4}{64\tau^2}\Big)^{-K/2} V_{\mathcal{M}} V_{B_{\mathbb{R}^K}}^{-1}$$
$$\le \Big(\frac{v^2}{4} - \frac{v^4}{64\tau^2}\Big)^{-K/2} \Big(\frac{K+2}{4\pi}\Big)^{K/2} V_{\mathcal{M}}.$$

We apply this bound to the terms on the far right-hand side of (26) to find for some absolute constants $c, \tilde{c} > 0$ and $0 < u \le 1$,

$$N(\mathcal{M}_{\mathrm{nc}}, d_2, u) \le V_{\mathcal{M}}\Big(\frac{K+2}{4\pi}\Big)^{K/2}\Big(\Big(\frac{c}{\tau u^2}\Big)^K + \Big(\frac{\tilde{c}}{\tau u^2}\Big)^K \Big(1 + \frac{6}{u}\Big)^K\Big).$$

A computation similar to (25) shows that

$$\gamma_2(\mathcal{M}_{\mathrm{nc}}, d_2) \lesssim \sqrt{K}(\log^{1/2}(K) + \log_+^{1/2}(\tau^{-1}) + 1) + \log_+^{1/2}(V_{\mathcal{M}}).$$
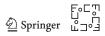
The claim now follows from Corollary 4.10.   □

The first statement in Theorem 7.9 improves upon [6, Theorem 3.1]. The second statement extends the result in [19, Theorem 2] from Gaussian matrices to general subgaussian maps and removes an additional $\mathcal{O}(\log(\varepsilon^{-1}))$ dependence of m on ε. The superfluous factor $\mathcal{O}(\log(\varepsilon^{-1}))$ seems to be an inherent construct of the proof in [19], as it occurs in several other papers which use essentially the same method [3,13,27].

References

1. D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. System Sci.*, 66(4):671–687, 2003.

2. R. Adamczak, R. Latała, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Geometry of log-concave ensembles of random matrices and approximate reconstruction. *C. R. Math. Acad. Sci. Paris*, 349(13-14):783–786, 2011.

3. P. Agarwal, S. Har-Peled, and H. Yu. Embeddings of surfaces, curves, and moving points in Euclidean space. *SIAM J. Comput.*, 42(2):442–458, 2013.
4. N. Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003.
5. R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.
6. R. Baraniuk and M. Wakin. Random projections of smooth manifolds. *Found. Comput. Math.*, 9(1):51–77, 2009.
7. G. Biau, L. Devroye, and G. Lugosi. On the performance of clustering in Hilbert spaces. *IEEE Trans. Inform. Theory*, 54(2):781–790, 2008.
8. E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 245–250. ACM, 2001.
9. T. Blumensath. Sampling and reconstructing signals from a union of linear subspaces. *IEEE Trans. Inform. Theory*, 57(7):4660–4671, 2011.
10. T. Blumensath and M. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Trans. Inform. Theory*, 55(4):1872–1882, 2009.
11. E. Candès and Y. Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inform. Theory*, 57(4):2342–2359, 2011.
12. E. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
13. K. Clarkson. Tighter bounds for random projections of manifolds. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 39–48. ACM, 2008.
14. S. Dasgupta. Learning mixtures of Gaussians. In *40th Annual Symposium on Foundations of Computer Science (New York, 1999)*, pages 634–644. IEEE Computer Soc., Los Alamitos, CA, 1999.
15. S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures Algorithms*, 22(1):60–65, 2003.
16. S. Dirksen. Tail bounds via generic chaining. *Electron. J. Probab.*, 20:no. 53, 1–29, 2015.
17. D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
18. P. Drineas, M. Mahoney, S. Muthukrishnan, and T. Sarlós. Faster least squares approximation. *Numer. Math.*, 117(2):219–249, 2011.
19. A. Eftekhari and M. Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, 39(1):67 – 109, 2015.
20. Y. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inform. Theory*, 55(11):5302–5316, 2009.
21. S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhaüser, Boston, 2013.
22. P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(3):355–362, 1988.
23. R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies. Greedy-like algorithms for the cosparse analysis model. *Linear Algebra Appl.*, 441:22–60, 2014.
24. Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 84–106. Springer, Berlin, 1988.
25. C. Hegde, M. Wakin, and R. Baraniuk. Random projections for manifold learning. In *Advances in neural information processing systems*, pages 641–648, 2007.
26. P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pages 604–613, New York, NY, USA, 1998. ACM.
27. P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):Art. 31, 12, 2007.
28. T. Jayram and D. Woodruff. Optimal bounds for Johnson-Lindenstrauss transforms and streaming problems with subconstant error. *ACM Trans. Algorithms*, 9(3):Art. 26, 17, 2013.
29. W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.
30. A. Kalai, A. Moitra, and G. Valiant. Disentangling gaussians. *Commun. ACM*, 55(2):113–120, February 2012.

31. B. Klartag and S. Mendelson. Empirical processes and random projections. *J. Funct. Anal.*, 225(1):229–245, 2005.
32. V. Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. Springer, Berlin, 2011.
33. K. Larsen and J. Nelson. The Johnson-Lindenstrauss lemma is optimal for linear dimensionality reduction. ArXiv:1411.2404, 2014.
34. G. Lecué and S. Mendelson. Sparse recovery under weak moment assumptions. ArXiv:1401.2188, 2014.
35. J. Lee. *Riemannian manifolds*. Springer-Verlag, New York, 1997.
36. Y. Lu and M. Do. A theory for sampling signals from a union of subspaces. *IEEE Trans. Signal Process.*, 56(6):2334–2345, 2008.
37. O. Maillard and R. Munos. Linear regression with random projections. *J. Mach. Learn. Res.*, 13:2735–2772, 2012.
38. W. Mantzel. *Parametric estimation of randomly compressed functions*. PhD. Thesis, Georgia Institute of Technology, 2013.
39. W. Mantzel and J. Romberg. Compressed subspace matching on the continuum. *Information and Inference*, 4(2):79–107, 2015.
40. W. Mantzel, J. Romberg, and K. Sabra. Compressive matched-field processing. *The Journal of the Acoustical Society of America*, 132(1), 2012.
41. J. Matoušek. On variants of the Johnson-Lindenstrauss lemma. *Random Structures Algorithms*, 33(2):142–156, 2008.
42. S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.*, 17(4):1248–1282, 2007.
43. S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.*, 28(3):277–289, 2008.
44. S. Nam, M. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Appl. Comput. Harmon. Anal.*, 34(1):30–56, 2013.
45. P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008.
46. M. Raginsky, R. Willett, Z. Harmany, and R. Marcia. Compressed sensing performance bounds under Poisson noise. *IEEE Trans. Signal Process.*, 58(8):3990–4002, 2010.
47. H. Rauhut, R. Schneider, and Z. Stojanac. Low rank tensor recovery via iterative hard thresholding. In *Proc. 10th International Conference on Sampling Theory and Applications*, 2013.
48. B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.
49. T. Sarlos. Improved approximation algorithms for large matrices via random projections. In *Foundations of Computer Science, 2006. FOCS '06. 47th Annual IEEE Symposium on*, pages 143–152, Oct 2006.
50. L. Schulman. Clustering for edge-cost minimization. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, STOC '00, pages 547–555, New York, NY, USA, 2000. ACM.
51. G. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev.*, 15:727–764, 1973.
52. M. Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.
53. M. Talagrand. Majorizing measures without measures. *Ann. Probab.*, 29(1):411–417, 2001.
54. M. Talagrand. *The generic chaining*. Springer-Verlag, Berlin, 2005.
55. S. Vempala. *The random projection method*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 65. American Mathematical Society, Providence, RI, 2004.
56. N. Verma. Distance preserving embeddings for general n-dimensional manifolds. *J. Mach. Learn. Res.*, 14:2415–2448, 2013.