View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Edinburgh Research Explorer

THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Recipes for stable linear embeddings from Hilbert spaces to R^m

Citation for published version:
Puy, G, Davies, M & Gribonval, R 2017, 'Recipes for stable linear embeddings from Hilbert spaces to R^m' IEEE Transactions on Information Theory. DOI: 10.1109/TIT.2017.2664858

Digital Object Identifier (DOI):
10.1109/TIT.2017.2664858

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
IEEE Transactions on Information Theory

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 09. May. 2018

Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$

Gilles Puy, Mike E. Davies, and Rémi Gribonval

***Abstract*—We consider the problem of constructing a linear map from a Hilbert space $\mathcal{H}$ (possibly infinite dimensional) to $\mathbb{R}^m$ that satisfies a restricted isometry property (RIP) on an arbitrary signal model, i.e., a subset of $\mathcal{H}$. We present a generic framework that handles a large class of low-dimensional subsets but also *unstructured* and *structured* linear maps. We provide a simple recipe to prove that a random linear map satisfies a general RIP with high probability. We also describe a generic technique to construct linear maps that satisfy the RIP. Finally, we detail how to use our results in several examples, which allow us to recover and extend many known compressive sampling results.**

***Index Terms*—Compressed sensing, restricted isometry property, box-counting dimension.**

I. Introduction

THE restricted isometry property (RIP) is at the core of many theoretical developments in compressive sensing (CS). It allows ones to show that sparse signals can be "captured" by few linear and non-adaptive measurements and recovered by non-linear decoders [2]. In a finite dimensional space, a matrix $\mathsf{A} \in \mathbb{R}^{m\times n}$ satisfies the RIP on a general set $\mathcal{S} \subset \mathbb{R}^n$, if there exists a constant $\delta \in (0,1)$, such that for all $x \in \mathcal{S}$,

$$(1-\delta)\left\|x\right\|_2^2 \leqslant \left\|\mathsf{A}x\right\|_2^2 \leqslant (1+\delta)\left\|x\right\|_2^2. \tag{1}$$

Random matrices with independent entries drawn from the centered Gaussian distribution with variance m^{-1} are examples of matrices that satisfy the RIP with high probability for many low-dimensional sets such as $\mathcal{S} = \Sigma_{2k}$, associated to k-sparse signals [2], $\mathcal{S} = \Sigma_{2r}$, associated to rank-$r$ matrices [3], or $\mathcal{S} = \{x_1 - x_2 | x_1, x_2 \in \Sigma\}$, associated to a compact Riemannian manifold Σ [4]. In these scenarios, the RIP holds for a number of measurements m essentially proportional to a measure of intrinsic dimension of $\mathcal{S}$.

In this work, we extend the construction of linear maps that satisfy the RIP in a finite-dimensional ambient space to linear maps that satisfy the RIP on subsets of a possibly *infinite*-dimensional space. This extension of the CS theory to infinite-dimensional spaces is important to properly apply CS in an analog setting [5], explore connections with the sampling of signals with finite rate of innovation [6], or also in machine learning to develop efficient methods to compute information-preserving sketches of probability distributions [7], [8]. As an example of the application of our results, we will explain how to build a stable linear embedding of sparse signals in the Haar wavelet basis of $L_2([0,1])$ with a sampling in the Fourier basis in Section V-A. Note that this type of signal model and sampling is often used to study the theoretical aspects of CS-MRI acquisition.

A. The normalised secant set

The RIP is a very convenient tool when one wants to prove that a signal $x \in \mathcal{H}$ can be reconstructed from its compressed measurements $\mathsf{A}x$ when it belongs to a given model set, *i.e.*, $x \in \Sigma \subset \mathcal{H}$. If A satisfies the RIP on the set of $2k$-sparse vectors then every k-sparse vector x can be accurately and stably recovered from its noisy measurements $\mathsf{A}x + \boldsymbol{n}$ by solving the Basis Pursuit problem [2]. For more general low-dimensional signal models Σ, one can prove that a signal $x \in \Sigma$ can be stably recovered from its compressed measurements if the matrix A satisfies the RIP on the secant set $\mathcal{S} = \Sigma - \Sigma$, [9], [10], *i.e.*, if there exists a constant $\delta \in (0,1)$ such that

$$(1-\delta)\left\|x_1 - x_2\right\|_2^2 \leqslant \left\|\mathsf{A}(x_1 - x_2)\right\|_2^2 \leqslant (1+\delta)\left\|x_1 - x_2\right\|_2^2.$$

for all $x_1, x_2 \in \Sigma$. When this condition is satisfied, we say that the matrix A stably embeds the set Σ in $\mathbb{R}^m$.

Let us recall that the *secant set* of Σ is defined as

$$\Sigma - \Sigma := \left\{ y = x_1 - x_2 \;\middle|\; x_1, x_2 \in \Sigma \right\}.$$

One can remark that the above condition for stable recovery is equivalent to

$$\sup_{z \in \mathcal{S}(\Sigma)} \left| \left\|\mathsf{A}z\right\|_2^2 - 1 \right| \leqslant \delta, \tag{2}$$

where

$$\mathcal{S}(\Sigma) := \left\{ z = y/\left\|y\right\|_2 \;\middle|\; y \in (\Sigma - \Sigma) \setminus \{\mathbf{0}\} \right\}$$

is the *normalised secant set* of Σ. This indicates that the set of interest to prove that a matrix A stably embeds the set Σ is not directly Σ, but rather its normalised secant set $\mathcal{S}(\Sigma)$.

From now on, we concentrate on proving a generalised version of (2) for an arbitrary set $\mathcal{S}$ lying on the unit sphere in the ambient space. If one wants to prove that a matrix or, more generally, a linear map stably embeds a set Σ, one just needs to substitute $\mathcal{S}(\Sigma)$ for $\mathcal{S}$ in the following results.

G. Puy is with Technicolor, 975 Avenue des Champs Blancs, 35576 Cesson-Sévigné, FR. He was with INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, FR, when the first verison of this work was submitted.

M. E. Davies is with the institute for Digital Communications (IDCom), The University of Edinburgh, EH9 3JL, UK.

R. Gribonval is with with INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, FR.

This work was partly funded by the European Research Council, PLEASE project (ERC-StG-2011-277906). M. E. Davies would like to acknowledge the support of EPSRC grant EP/J015180/1.

This work is an extension of the paper [1] presented at the European Signal Processing Conference (EUSIPCO), 2015.

In this paper, we consider that the ambient space is an infinite-dimensional Hilbert space denoted by $\mathcal{H}$. The inner product in $\mathcal{H}$ is denoted by $\langle \cdot, \cdot \rangle$ and the associated norm by $\|\cdot\|$. Vectors in $\mathcal{H}$ are given a bold letter while vectors in $\mathbb{R}^n$ are given a regular letter. We denote by $\mathbb{S}$ the unit sphere in $\mathcal{H}$. We assume everywhere $\mathcal{S} \subset \mathbb{S}$: $\|\boldsymbol{x}\| = 1$ for all $\boldsymbol{x} \in \mathcal{S}$.

B. Measuring the dimension of $\mathcal{S}$

All the developments in this work are based on the assumption that $\mathcal{S}$ has a small intrinsic dimension. In the literature, several definitions of dimension exist. The reader can refer to, *e.g.*, the monograph of Robinson [11], for an exhaustive list of definitions. In an infinite-dimensional space, one should be careful with the definition of dimension used. Indeed, as described in [11], there are examples of sets for which no stable linear embedding to a finite dimensional space exists even though their dimension is finite (according to some definition). Therefore, there is no hope to construct a linear map that satisfies the RIP for these sets.

In this paper, we use the *upper box-counting dimension* as our measure of dimension. The upper box-counting dimension is also at the centre of most of the developments in [11].

Definition I.1 (Covering number)**.** *Let $\epsilon > 0$. The covering number $N_{\mathcal{S}}(\epsilon)$ of $\mathcal{S}$ is the minimum number of closed balls of radius ϵ, with respect to the norm $\|\cdot\|$, with centres in $\mathcal{S}$ needed to cover $\mathcal{S}$. The set of centres of these balls is a minimal ϵ-net for $\mathcal{S}$.*

Definition I.2 (Upper box-counting dimension)**.** *The upper box-counting dimension of $\mathcal{S}$ is*

$$\dim(\mathcal{S}) := \limsup_{\epsilon \to 0} \ \log[N_{\mathcal{S}}(\epsilon)]/\log[1/\epsilon].$$

From the above definition, one can remark that if $d > \dim(\mathcal{S})$ then there exists $\epsilon_{\mathcal{S}} > 0$ such that $N_{\mathcal{S}}(\epsilon) \leqslant \epsilon^{-d}$ for all $\epsilon \leqslant \epsilon_{\mathcal{S}}$. In this paper, we make the following assumption on $\mathcal{S}$.

Assumption A. *The set $\mathcal{S} \subset \mathbb{S}$ has a finite upper box-counting dimension $\dim(\mathcal{S})$ which is <u>strictly</u> bounded by $s \geqslant 1$: $\dim(\mathcal{S}) < s$.*

Therefore, there exists a model-set dependent constant $\epsilon_{\mathcal{S}} \in (0, 1/2)$ such that $N_{\mathcal{S}}(\epsilon) \leqslant \epsilon^{-s}$ for all $\epsilon \leqslant \epsilon_{\mathcal{S}}$.

In many well-behaved cases, the upper-box counting dimension of the normalised secant set $\mathcal{S} = \mathcal{S}(\Sigma)$ of Σ is simply twice the upper-box counting dimension of Σ (or $\Sigma \cap \mathbb{S}$ if Σ is not bounded). For example, the set of normalised k-sparse vectors in $\mathbb{R}^n$ has dimension k and its secant set has dimension $2k$. The same results holds for low-rank matrices and low-dimensional manifolds. However, this is not always the case as we will see in Section V-B.

C. Contributions and organisation of the paper

This work has four main contributions.

- First, given a set $\mathcal{S}$ that satisfies Assumption A and a random linear map $L : \mathcal{H} \to \mathbb{R}^m$, we factorise the proof of the RIP for L on $\mathcal{S}$ into a small number of steps. We separate the steps that are problem-specific from the ones that are common to most problems and do not need to be repeated every time. This allows us to provide a simple recipe for the proof of the RIP, with techniques similar to those of, *e.g.*, [4], [12]–[15]. This is the subject of Section II.
- Second, given an *arbitrary* set $\mathcal{S}$ that satisfies Assumption A, we propose a *generic construction* of linear maps that satisfy the RIP on $\mathcal{S}$. This construction is made of two steps. In the first step, we show that one can always find a finite-dimensional subspace of large but finite dimension that accurately approximates $\mathcal{S}$. We then use this subspace to build a linear functional that maps the vectors in $\mathcal{S}$ into finite but potentially large dimension while preserving their norm to a prescribed accuracy. The second step consists in reducing the embedding dimension further with a random matrix. This construction is detailed in Section III.
- Third, we show how to use the developed techniques to recover and extend many known CS results. In particular, we show that our recipe is general enough to handle *structured* measurement strategies such as the ones proposed in [16], [17]. These examples are presented in Section IV.
- Fourth, we show that while it is sufficient to have a set $\mathcal{S}$ of finite upper-box counting dimension to guarantee the existence of linear maps that have the RIP on $\mathcal{S}$, this condition is not *necessary*. This fact is discussed in Section V-B.

We discuss related works in Section V and conclude in Section VI. All technical proofs can be found in the appendices.

II. Recipe to prove that the RIP holds

In this section, we give a generic recipe to prove that a random linear map $L : \mathcal{H} \to \mathbb{R}^m$ preserves the norm of all vectors in the set $\mathcal{S} \subset \mathbb{S}$.

We suppose that L is built by drawing at random m vectors $(\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m)$ in $\mathcal{H}^m$ using a probability measure μ on $\mathcal{H}^m$:

$$\begin{aligned} L \colon \mathcal{H} &\longrightarrow \mathbb{R}^m \\ \boldsymbol{x} &\longmapsto (\langle \boldsymbol{l}_i, \boldsymbol{x} \rangle)_{1 \leqslant i \leqslant m}. \end{aligned} \tag{3}$$

Notice that L is a continuous linear map.

In the following, we start by introducing a generalised form of RIP. Then we detail generic properties of the probability measure μ which are sufficient to prove that L preserves the norm of all vectors in $\mathcal{S}$ with high probability, provided that m is sufficiently large. Finally, we give a recipe to show that a random linear map preserves the norm of all vectors in $\mathcal{S}$ with high probability.

A. Generalised form of the RIP

Let us come back to the usual form of the RIP (1), which involves the Euclidian norms in the ambient spaces. First, we notice that it can be rewritten in the following equivalent form:

$$\left| \|\mathsf{A}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| \leqslant \delta \, \|\boldsymbol{x}\|_2^2 \,. \tag{4}$$

Second, we remark that for typical random matrices A satisfying this RIP, one has $\mathbb{E}\,\|\mathsf{A}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2$. This property is satisfied, *e.g.*, for random matrices $\mathsf{A} \in \mathbb{R}^{m\times n}$ whose entries are independent Gaussian variables with zero-mean and variance $1/m$, for $\pm 1/\sqrt{m}$ Bernoulli random matrices, or for (rescaled) sensing matrices constructed by selecting independently m vectors from an orthonormal basis using the uniform distribution. For all these cases, inequality (4) becomes

$$\left|\|\mathsf{A}\boldsymbol{x}\|_2^2 - \mathbb{E}\,\|\mathsf{A}\boldsymbol{x}\|_2^2\right| \leqslant \delta\,\|\boldsymbol{x}\|_2^2\,. \tag{5}$$

In this form, we see that the right-hand side of (5) characterises how much $\|\mathsf{A}\boldsymbol{x}\|_2^2$ deviates from its mean. In (5), this deviation is proportional to $\|\boldsymbol{x}\|_2^2$, the square of the Euclidean norm in the ambient space. We replace here the Euclidean norm by the norm $\|\cdot\|$ in the considered Hilbert space. Finally, instead of concentrating only on the ℓ_2-norm in the measurement space, we will consider arbitrary ℓ_p-norms, $p \geqslant 1$.

We define the following semi-norm for all $\boldsymbol{x} \in \mathcal{H}$,

$$\|\boldsymbol{x}\|_{\mu,p} := \left(\mathbb{E}_\mu\,\|L(\boldsymbol{x})\|_p^p\right)^{1/p}.$$

Gathering all the above remarks and adapting them to the case of random linear maps from $\mathcal{H}$ to $\mathbb{R}^m$, we consider the following general form for the desired RIP:

$$\left|\|L(\boldsymbol{x})\|_p^p - \|\boldsymbol{x}\|_{\mu,p}^p\right| \leqslant \delta, \quad \forall \boldsymbol{x} \in \mathcal{S} \subset \mathbb{S}.$$

Definition II.1 (RIP)**.** *Let $L : \mathcal{H} \to \mathbb{R}^m$ be a linear map. Define*

$$\delta_{\mathcal{S},\mu,p} := \sup_{\boldsymbol{x}\in\mathcal{S}} \left|\|L(\boldsymbol{x})\|_p^p - \|\boldsymbol{x}\|_{\mu,p}^p\right| \tag{6}$$

and

$$\underline{\delta}_{\mathcal{S},\mu,p} := \inf_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{x}\|_{\mu,p}^p\,.$$

The linear map L satisfies the RIP on $\mathcal{S} \subset \mathbb{S}$ with constant δ if $\delta_{\mathcal{S},\mu,p} \leqslant \delta < \underline{\delta}_{\mathcal{S},\mu,p}$.

It will also be convenient to define

$$\bar{\delta}_{\mathcal{S},\mu,p} := \sup_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{x}\|_{\mu,p}^p\,.$$

To simplify notations, we substitute δ_p, $\underline{\delta}_p$, and $\bar{\delta}_p$ for $\delta_{\mathcal{S},\mu,p}$, $\underline{\delta}_{\mathcal{S},\mu,p}$, and $\bar{\delta}_{\mathcal{S},\mu,p}$, respectively, but one should keep in mind that these quantities depend on $\mathcal{S}$, p and μ.

We remark that if L satisfies the RIP on $\mathcal{S}$ with constant δ then

$$\underline{\delta}_p - \delta \leqslant \|L(\boldsymbol{x})\|_p^p \leqslant \bar{\delta}_p - \delta,$$

for all $\boldsymbol{x} \in \mathcal{S}$. The condition $\delta < \underline{\delta}_p$ ensures that no vector in $\boldsymbol{x} \in \mathcal{S}$ is in the null space of L. Indeed, we have

$$\|L(\boldsymbol{x})\|_p^p \geqslant \underline{\delta}_p - \delta > 0,$$

for all $\boldsymbol{x} \in \mathcal{S}$.

In the remaining part of this section, we give generic sufficient conditions to ensure that the RIP holds with high probability. We will see how to recover classical RIP results in finite dimensions with $p = 2$ in Section IV-A.

B. Concentration inequalities and main result

If Assumption A holds, the only other ingredients needed to prove that a random linear map L of the form of (3) satisfies the RIP with high probability are concentration inequalities. The choice of the probability distribution μ is thus important to ensure a preservation of the norm of the vectors in $\mathcal{S}$ by L. In this section, we assume that the following general concentration inequalities hold.

Assumption B. *Define the function*

$$\begin{aligned} h_{L,\mu,p} \colon \mathcal{H} &\longrightarrow \mathbb{R} \\ \boldsymbol{x} &\longmapsto \|L(\boldsymbol{x})\|_p^p - \|\boldsymbol{x}\|_{\mu,p}^p\,. \end{aligned}$$

There exist two constants $c_1, c_2 \in (0,\infty]$ such that for any fixed $\boldsymbol{y}, \boldsymbol{z} \in \mathcal{S} \cup \{\boldsymbol{0}\}$,

$$\mathbb{P}\left\{|h_{L,\mu,p}(\boldsymbol{y}) - h_{L,\mu,p}(\boldsymbol{z})| \geqslant \lambda\,\|\boldsymbol{y}-\boldsymbol{z}\|\right\} \leqslant 2\mathrm{e}^{-c_1 m\lambda^2}, \tag{7}$$

for every $0 \leqslant \lambda \leqslant c_2/c_1$, and

$$\mathbb{P}\left\{|h_{L,\mu,p}(\boldsymbol{y}) - h_{L,\mu,p}(\boldsymbol{z})| \geqslant \lambda\,\|\boldsymbol{y}-\boldsymbol{z}\|\right\} \leqslant 2\mathrm{e}^{-c_2 m\lambda}, \tag{8}$$

for every $\lambda \geqslant c_2/c_1$.

To simplify notations, we substitute h_p for $h_{L,\mu,p}$ in the remaining part of the paper.

We remark that we can have $c_1 = \infty$ or $c_2 = \infty$. This is to handle the case where one of the above bounds hold for all $\lambda \geqslant 0$. If $c_1 = \infty$ then (8) holds for all $\lambda \geqslant 0$. Similarly, if $c_2 = \infty$ then (7) holds for all $\lambda \geqslant 0$.

We can now state our main theorem from which all the following results are derived.

Theorem II.2. *Let $L : \mathcal{H} \to \mathbb{R}^m$ be a random linear map constructed as in (3) using a probability distribution μ. If Assumption A and Assumption B hold, then for any $\xi, \delta \in (0,1)$, we have: $\delta_p \leqslant \delta$ with probability at least $1-\xi$ provided that*

$$m \geqslant \frac{C}{\min(c_1,c_2)\,\delta^2}\,\max\left\{s\log\left(\frac{1}{\epsilon_{\mathcal{S}}}\right),\ \log\left(\frac{6}{\xi}\right)\right\}, \tag{9}$$

where $C > 0$ is an absolute constant.

Inequality (9) above involves the *natural logarithm*. The proof of this theorem is available in Appendix A. The proof is based on a chaining argument which is a powerful technique to obtain sharp bounds for the supremum of random processes, see, *e.g.*, [4], [12]–[15]. This technique can be viewed as a refinement of the classical ϵ-net argument such as used in, *e.g.*, [18], [19].

C. Proof recipe for the RIP

We can now deduce the following recipe to show that a random linear map L of the form of (3) satisfies the RIP on $\mathcal{S}$.

The result of this recipe is that the random linear map L satisfies the RIP on the set $\mathcal{S}$ with constant $\delta \in (0, \underline{\delta}_p)$ and probability at least $1-\xi$ provided that (9) holds.

In many practical scenarios, such as the ones discussed in next sections, the concentration inequalities (7) and (8) in Step 3 are easy to obtain using well-known concentration

Recipe 1 Recipe to prove that L satisfies the RIP on $\mathcal{S}$

1: Prove that the set $\mathcal{S}$ has finite upper box-counting dimension (Assumption A).
2: Compute $\underline{\delta}_p := \inf_{\boldsymbol{x}\in\mathcal{S}} \mathbb{E}_\mu \|L(\boldsymbol{x})\|_p^p$.
3: Prove that the concentration inequalities (7) and (8) hold (Assumption B).
4: Choose m such that (9) holds.

inequalities for the sum of independent random variables, see, *e.g.*, [2], [18]. We will also see that the estimation of $\underline{\delta}_p$ simplifies in these cases. Therefore, the main difficult step in the recipe is most often the computation of the upper box-counting dimension of $\mathcal{S}$.

III. A GENERIC CONSTRUCTION OF L WITH THE RIP

In the previous section, we provided a recipe to prove that a random linear map satisfies the RIP once the probability distribution μ is given. In this section, we give a generic way of constructing a random linear map which has the RIP on a set $\mathcal{S}$ that satisfies Assumption A, *i.e.*, we give a generic construction of μ.

We divide our construction into two steps. In the first step, we build a continuous linear functional b that maps the vectors in $\mathcal{S}$ to finite but potentially large dimension while essentially preserving their norm. In the second step, we further reduce the embedding dimension by multiplication with a random matrix.

A. Mapping to a finite-dimensional subspace

In this section, we prove that it is always possible to design a continuous linear map $b\colon \mathcal{H} \to \mathbb{R}^d$, where d is potentially large but finite, such that

$$\|b(\boldsymbol{x})\|_b \leqslant \|\boldsymbol{x}\|\,,\ \forall \boldsymbol{x} \in \mathcal{H}, \tag{10}$$

and

$$\|b(\boldsymbol{x})\|_b \geqslant 1-\epsilon_*,\ \forall \boldsymbol{x} \in \mathcal{S}, \tag{11}$$

for some $\epsilon_* \in (0,1)$ and a well-chosen norm $\|\cdot\|_b$. One can remark that the above properties ensure that b already satisfies a RIP but for a potentially large dimension d. We will see how to further reduce the dimension in Section III-B.

In the following subsections, we detail a generic method to construct b. This construction uses the fact that we can cover $\mathcal{S}$ with precision ϵ_* as Assumption A holds. Note that this construction can be computationally expensive in practice. Yet, we would like to highlight that one does not have to use such a covering in practice. Indeed, the only required property in the following results is that b satisfies (10) and (11). The way b is constructed is not important as long as these properties hold. In practice, one can use other properties of $\mathcal{S}$ to construct b efficiently. For example, we will see in Section V-A that, for signals sparse in the Haar wavelet basis of $L_2([0,1])$, such a linear map can be built easily using properties of these wavelets in the Fourier basis.

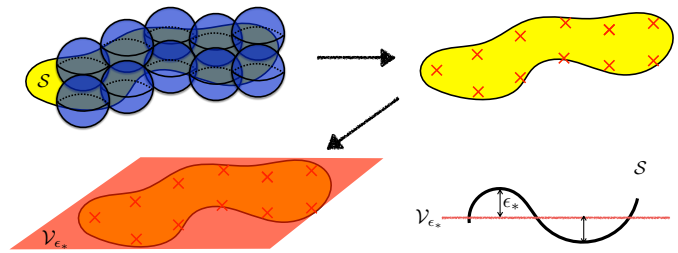

Figure 1. Construction of $\mathcal{V}_{\epsilon_*}$. *Top left*: cover of $\mathcal{S}$ with $N(\epsilon_*)$ balls of radius ϵ_*. *Top right*: the centres of the balls, indicated by the red crosses, form a ϵ_*-net, denoted by $\mathcal{C}(\epsilon_*)$, for $\mathcal{S}$. *Bottom left*: the linear span of the vectors in $\mathcal{C}(\epsilon_*)$ is $\mathcal{V}_{\epsilon_*}$. *Bottom right*: $\mathcal{V}_{\epsilon_*}$ approximates $\mathcal{S}$ with precision ϵ_*.

1) Using an orthonormal basis of a well-chosen subspace: If Assumption A holds, then we can cover $\mathcal{S}$ with a finite number of closed balls. We fix a resolution ϵ_*, find a minimum ϵ_*-net for $\mathcal{S}$, and denote the set of points in this net by $\mathcal{C}(\epsilon_*)$. The cardinality of $\mathcal{C}(\epsilon_*)$ is $N_{\mathcal{S}}(\epsilon_*)$. Let $\mathcal{V}_{\epsilon_*} \subset \mathcal{H}$ be the finite-dimensional linear subspace of $\mathcal{H}$ spanned by $\mathcal{C}(\epsilon_*)$, and $P_{\mathcal{V}_{\epsilon_*}} : \mathcal{H} \to \mathcal{H}$ be the orthogonal projection onto $\mathcal{V}_{\epsilon_*}$. By construction of $\mathcal{V}_{\epsilon_*}$, we have

$$\sup_{\boldsymbol{x}\in\mathcal{S}} \left\|\boldsymbol{x} - P_{\mathcal{V}_{\epsilon_*}}(\boldsymbol{x})\right\| \leqslant \sup_{\boldsymbol{x}\in\mathcal{S}} \min_{\boldsymbol{x}_0\in\mathcal{C}(\epsilon_*)} \|\boldsymbol{x}-\boldsymbol{x}_0\| \leqslant \epsilon_*. \tag{12}$$

The orthogonal projection onto $\mathcal{V}_{\epsilon_*}$ thus preserves the norm of all vectors in $\mathcal{S}$ with an error at most ϵ_*, as illustrated in Fig. 1.

Let $(\boldsymbol{b}_1,\dots,\boldsymbol{b}_d)$ be an orthonormal basis for $\mathcal{V}_{\epsilon_*}$ and b be the linear map

$$\begin{aligned} b\colon \mathcal{H} &\longrightarrow \mathbb{R}^d \\ \boldsymbol{x} &\longmapsto (\langle \boldsymbol{b}_i, \boldsymbol{x}\rangle)_{1\leqslant i\leqslant d}. \end{aligned}$$

We remark that

$$\|b(\boldsymbol{x})\|_2 = \left\|P_{\mathcal{V}_{\epsilon_*}}(\boldsymbol{x})\right\| \leqslant \|\boldsymbol{x}\|\,,\ \forall \boldsymbol{x} \in \mathcal{H},$$

and that

$$\|b(\boldsymbol{x})\|_2 \geqslant \|\boldsymbol{x}\| - \left\|\boldsymbol{x} - P_{\mathcal{V}_{\epsilon_*}}(\boldsymbol{x})\right\| \geqslant 1-\epsilon_*,\ \forall \boldsymbol{x} \in \mathcal{S}.$$

To obtain the last inequality, we used inequality (12) and the fact that $\mathcal{S} \subset \mathbb{S}$. We thus proved the desired result with $\|\cdot\|_b = \|\cdot\|_2$.

2) Using an arbitrary basis of $\mathcal{V}_{\epsilon_*}$: The linear map b in the last section was constructed using an orthonormal basis of $\mathcal{V}_{\epsilon_*}$. To provide more flexibility in the design of the linear map L, we propose here to use an arbitrary basis of $\mathcal{V}_{\epsilon_*}$. Let $(\boldsymbol{b}_1,\dots,\boldsymbol{b}_d)$ be such a basis and define

$$\begin{aligned} b\colon \mathcal{H} &\longrightarrow \mathbb{R}^d \\ \boldsymbol{x} &\longmapsto (\langle \boldsymbol{b}_i, \boldsymbol{x}\rangle)_{1\leqslant i\leqslant d}. \end{aligned}$$

We also define the following norm

$$\|y\|_b = \inf_{\boldsymbol{z}\in\mathcal{H}} \{\|\boldsymbol{z}\| \mid b(\boldsymbol{z}) = y\}\,,$$

for all $y \in \mathbb{R}^d$. We next show that (10) and (11) hold with this choice of norm.

We denote by $\boldsymbol{y} \in \mathcal{V}_{\epsilon_*}$ the vector with coordinates y in the basis $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_d)$. Let $\mathcal{V}_{\epsilon_*}^{\perp}$ be the orthogonal complement to $\mathcal{V}_{\epsilon_*}$ in $\mathcal{H}$. It is easy to notice that the set of vectors that satisfy $b(\boldsymbol{z}) = y$ is

$$\{\boldsymbol{z} \ = \ \boldsymbol{y} + \boldsymbol{y}^{\perp} \mid \boldsymbol{y}^{\perp} \in \mathcal{V}_{\epsilon^*}^{\perp}\},$$

i.e., the set of vectors whose orthogonal projection on $\mathcal{V}_{\epsilon_*}$ is $\boldsymbol{y}$. Then, by orthogonality, we have $\|\boldsymbol{z}\|^2 = \|\boldsymbol{y}\|^2 + \left\|\boldsymbol{y}^{\perp}\right\|^2$, whose minimum value is $\|\boldsymbol{y}\|^2$. Therefore, $\|y\|_b = \|\boldsymbol{y}\|$ and we deduce that $\|b(\boldsymbol{x})\|_b = \left\|P_{\mathcal{V}_{\epsilon_*}}(\boldsymbol{x})\right\|$ for all $\boldsymbol{x} \in \mathcal{H}$. As we have $\left\|P_{\mathcal{V}_{\epsilon_*}}(\boldsymbol{x})\right\| \leqslant \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in \mathcal{H}$, (10) holds. Then the fact that $\left\|P_{\mathcal{V}_{\epsilon_*}}(\boldsymbol{x})\right\| \geqslant \|\boldsymbol{x}\| - \left\|\boldsymbol{x} - P_{\mathcal{V}_{\epsilon_*}}(\boldsymbol{x})\right\| \geqslant 1 - \epsilon_*$ for all $\boldsymbol{x} \in \mathcal{S}$ shows that (11) holds.

We remark that when $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_d)$ is an orthonormal basis, we have $\|y\|_b = \|\boldsymbol{y}\| = \|y\|_2$, *i.e.*, the norm we naturally chose in the previous section. The choice of $\|\cdot\|_b$ as norm in $\mathbb{R}^d$ allows us to recover the result of the last section but also to generalise it. Indeed, the result is now independent of the choice of the basis $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_d)$. As explained in the remark before Recipe 2 below, this result can be useful when the choice of $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_d)$ is for example fixed by the application.

B. Dimensionality reduction with a random matrix

The linear map b in the previous section maps the vectors in $\mathcal{S}$ into finite, but potentially large, dimension with an error at most ϵ_* on the norm of the vectors. The goal is now to reduce the dimension further without degrading much the norm of the vectors in $b(\mathcal{S})$.

Before continuing, we introduce the following definitions (definitions 5.7 and 5.13 in [18]).

Definition III.1 (Subexponential random variable)**.** *A subexponential random variable X is a random variable that satisfies*

$$(\mathbb{E}\,|X|^q)^{1/q} \leqslant Cq \text{ for all } q \geqslant 1,$$

with $C > 0$. The subexponential norm of X, denoted by $\|X\|_{\Psi_1}$, is the smallest constant C for which the last property holds, i.e.,

$$\|X\|_{\Psi_1} := \sup_{q \geqslant 1} \left\{ q^{-1} \left(\mathbb{E}\,|X|^q\right)^{1/q} \right\}.$$

Definition III.2 (Subgaussian random variable)**.** *A subgaussian random variable X is a random variable that satisfies*

$$(\mathbb{E}\,|X|^q)^{1/q} \leqslant C\sqrt{q} \text{ for all } q \geqslant 1,$$

with $C > 0$. The subgaussian norm of X, denoted by $\|X\|_{\Psi_2}$, is the smallest constant C for which the last property holds, i.e.,

$$\|X\|_{\Psi_2} := \sup_{q \geqslant 1} \left\{ q^{-1/2} \left(\mathbb{E}\,|X|^q\right)^{1/q} \right\}.$$

To reduce the dimension, we draw m independent random vectors $a_1, \dots, a_m \in \mathbb{R}^d$ according to a probability distribution ν in $\mathbb{R}^d$. The measurements $a_i^{\intercal} b(\boldsymbol{x})$, $i = 1, \dots, m$ are thus independent and identically distributed random variables. The choice of b and ν defines the probability distribution μ in $\mathcal{H}^m$.

To characterise the number of measurements needed to preserve the norm of all vectors in $\mathcal{S}$, we need the following quantities

$$\Lambda_1 := \sup_{\substack{\boldsymbol{x} \in \{\mathcal{S}-\mathcal{S}\} \cup \mathcal{S} \\ \boldsymbol{x} \neq \boldsymbol{0}}} \left\{ \|a_i^{\intercal} b(\boldsymbol{x})\|_{\Psi_1} / \|b(\boldsymbol{x})\|_b \right\} \text{ and} \tag{13}$$

$$\Lambda_2 := \sup_{\substack{\boldsymbol{x} \in \{\mathcal{S}-\mathcal{S}\} \cup \mathcal{S} \\ \boldsymbol{x} \neq \boldsymbol{0}}} \left\{ \|a_i^{\intercal} b(\boldsymbol{x})\|_{\Psi_2} / \|b(\boldsymbol{x})\|_b \right\}. \tag{14}$$

Note that these quantities may be infinite. However, we can ensure that Λ_1 and Λ_2 are finite, for example, by drawing $a_1, \dots, a_m$ in the ball of radius 1 with respect to the dual norm of $\|\cdot\|_b$ defined as

$$\|a\|_{b^*} := \sup \left\{ |a^{\intercal} x| \mid x \in \mathbb{R}^d, \|x\|_b \leqslant 1 \right\}.$$

Indeed, we have $|a^{\intercal} b(\boldsymbol{x})| \leqslant \|b(\boldsymbol{x})\|_b$ for any $\boldsymbol{x} \in \mathcal{H}$ in this case, which ensures that $\Lambda_2 \leqslant 1$ and $\Lambda_1 \leqslant 1$.

We are now ready to give the two main results of this section, which follow from Theorem II.2.

Theorem III.3. *Assume that $\Lambda_1 < +\infty$ and define*

$$\begin{aligned} L \colon \mathcal{H} &\longrightarrow \mathbb{R}^m \\ \boldsymbol{x} &\longmapsto \left(\frac{a_i^{\intercal} b(\boldsymbol{x})}{m} \right)_{1 \leqslant i \leqslant d}. \end{aligned} \tag{15}$$

There exists an absolute constant $C > 0$ such that if Assumption A holds then, for any $\xi \in (0,1)$ and $\delta \in (0,1)$, with probability at least $1 - \xi$,

$$\|\boldsymbol{x}\|_{\mu,1} - \delta \leqslant \|L(\boldsymbol{x})\|_1 \leqslant \|\boldsymbol{x}\|_{\mu,1} + \delta,$$

holds for all $\boldsymbol{x} \in \mathcal{S}$ provided that

$$m \geqslant \frac{C}{\delta^2} \max(2\Lambda_1^2, \Lambda_1) \max\left\{ s \log\left(\frac{1}{\epsilon_{\mathcal{S}}}\right),\ \log\left(\frac{6}{\xi}\right) \right\}. \tag{16}$$

Theorem III.4. *Assume that $\Lambda_2 < +\infty$ and define*

$$\begin{aligned} L \colon \mathcal{H} &\longrightarrow \mathbb{R}^m \\ \boldsymbol{x} &\longmapsto \left(\frac{a_i^{\intercal} b(\boldsymbol{x})}{\sqrt{m}} \right)_{1 \leqslant i \leqslant d}. \end{aligned} \tag{17}$$

There exists an absolute constant $C > 0$ such that if Assumption A holds, then, for any $\xi \in (0,1)$ and $\delta \in (0,1)$, with probability at least $1 - \xi$,

$$\|\boldsymbol{x}\|_{\mu,2}^2 - \delta \leqslant \|L(\boldsymbol{x})\|_2^2 \leqslant \|\boldsymbol{x}\|_{\mu,2}^2 + \delta,$$

holds for all $\boldsymbol{x} \in \mathcal{S}$ provided that

$$m \geqslant \frac{C}{\delta^2} \max(8\Lambda_2^4, \Lambda_2^2) \max\left\{ s \log\left(\frac{1}{\epsilon_{\mathcal{S}}}\right),\ \log\left(\frac{6}{\xi}\right) \right\}. \tag{18}$$

Both theorems are proved in Appendix C. Let us comment on these results and highlight the situations where they are useful. We will study several examples in the next section.

- First, we remark that

$$\underline{\delta}_p = \inf_{\boldsymbol{x} \in \mathcal{S}} \mathbb{E}_{\nu} \left|a_i^{\intercal} b(\boldsymbol{x})\right|^p \ \leqslant \ \sup_{\boldsymbol{x} \in \mathcal{S}} \mathbb{E}_{\nu} \left|a_i^{\intercal} b(\boldsymbol{x})\right|^p = \bar{\delta}_p$$

with $p = 1$ in Theorem III.3 and $p = 2$ in Theorem III.4. By definition of Λ_1 and Λ_2, we have

$$\mathbb{E}_\nu \left|a_i^\intercal b(\boldsymbol{x})\right|^p \leqslant p\,\Lambda_p^p \left\|b(\boldsymbol{x})\right\|_b^p \leqslant p\,\Lambda_p^p, \quad \forall \boldsymbol{x} \in \mathcal{S},$$

for $p = 1$ and 2. The last inequality follows from (10). Therefore,

$$\underline{\delta}_p \leqslant \bar{\delta}_p \leqslant p\,\Lambda_p^p$$

with $p = 1$ in Theorem III.3 and $p = 2$ in Theorem III.4.

- Second, we notice that the RIP is satisfied for $\delta < \underline{\delta}_1$ in the case of Theorem III.3 and for $\delta < \underline{\delta}_2$ in the case of Theorem III.4. Consequently, if L is constructed as in this section, then Recipe 1 can be modified to Recipe 2.

Recipe 2 Recipe to prove that L in (15) or (17) satisfies the RIP on $\mathcal{S}$

1: Prove that the set $\mathcal{S}$ has finite upper box-counting dimension (Assumption A).
2: Compute $\underline{\delta}_p = \inf_{\boldsymbol{x}\in\mathcal{S}} \mathbb{E}_\nu \left|a_i^\intercal b(\boldsymbol{x})\right|^p$, with $p = 1$ for L defined in (15) or $p = 2$ for L defined in (17).
3: Compute Λ_1 or Λ_2.
4: Choose m such that (16) holds if $p = 1$, or such that (18) holds if $p = 2$.

- Third, the number of measurements essentially scales with $s \max(2\Lambda_1^2, \Lambda_1)/\delta^2$ in the first case and $s \max(8\Lambda_2^4, \Lambda_2^2)/\delta^2$ in the second case. As we should have $\delta < \underline{\delta}_p$ (with $p = 1$ or $p = 2$) to satisfy the RIP, the results have a practical interest when $\max(2\Lambda_1^2, \Lambda_1)/\underline{\delta}_1^2$ or $\max(8\Lambda_2^4, \Lambda_2^2)/\underline{\delta}_2^2$ is small. These ratios are the quantities to optimise when designing L, as illustrated in Section IV-B.
- Finally, it is important to notice that the choice of the basis $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ and the choice of the distribution ν for the a_i's interact together in the value of the ratio $\max(2\Lambda_1^2, \Lambda_1)/\underline{\delta}_1^2$ or $\max(8\Lambda_2^4, \Lambda_2^2)/\underline{\delta}_2^2$. If one has the flexibility to choose both $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ and ν, then one should seek to minimise these ratios in order to minimise m. Even if the choice of b is fixed by the application of interest, one still has the flexibility to optimise the distribution ν to minimise m.

In Recipe 2 introduced above, when $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ is an orthonormal basis and a_i is an isotropic random vector, then $\underline{\delta}_2$ can be directly estimated using (19) as done below in Section IV-A1 for the second step of the recipe. We recall that $a_i \in \mathbb{R}^d$ is isotropic if $\mathbb{E}\left|a_i^\intercal x\right|^2 = \|x\|_2^2$ for all $x \in \mathbb{R}^d$. To estimate $\underline{\delta}_1$, one can use Lemma D.2 as also done below in Section IV-B2 and Section IV-B4 for the second step of the recipe.

IV. Examples

In this section, we show how to use our generic recipe on different examples.

A. A linear embedding from $\mathcal{H}$ to $\mathbb{R}^m$

1) The infinite-dimensional case: In Section III-A, we built a linear map $b : \mathcal{H} \to \mathbb{R}^d$ which preserves the norm of all vectors in $\mathcal{S}$. For simplicity, we consider the case where $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ is an orthonormal basis. We have

$$(1-\epsilon_*)^2 \|\boldsymbol{x}\|^2 \leqslant \|b(\boldsymbol{x})\|_b^2 = \|b(\boldsymbol{x})\|_2^2 \leqslant \|\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x} \in \mathcal{S}. \tag{19}$$

Let $a_1, \ldots, a_m \in \mathbb{R}^d$ be m vectors whose entries are independent Gaussian random variables with mean 0 and variance 1, and define L as in Theorem III.4. We follow Recipe 2 to prove that L has the RIP.

1: The set $\mathcal{S}$ has a finite upper-box counting dimension by assumption.
2: To estimate $\underline{\delta}_2$, we first notice that

$$\mathbb{E}\left|a_i^\intercal b(\boldsymbol{x})\right|^2 = \|b(\boldsymbol{x})\|_2^2.$$

Inequality (19) then yields

$$(1-\epsilon_*)^2 \|\boldsymbol{x}\|^2 \leqslant \mathbb{E}\left|a_i^\intercal b(\boldsymbol{x})\right|^2 \leqslant \|\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x} \in \mathcal{S}.$$

Therefore, $\underline{\delta}_2 \geqslant (1-\epsilon_*)^2$ and $\bar{\delta}_2 \leqslant 1$.
3: To estimate Λ_2, we use the fact that $\|a_i^\intercal x\|_{\Psi_2} \leqslant D\,\|x\|_2$, for all $x \in \mathbb{R}^d$, where $D > 0$ is an absolute constant (see [18]). Therefore,

$$\|a_i^\intercal b(\boldsymbol{x})\|_{\Psi_2} \leqslant D\,\|b(\boldsymbol{x})\|_2, \quad \forall \boldsymbol{x} \in \mathcal{H},$$

and $\Lambda_2 \leqslant D$.
4: The last step of the recipe proves that L satisfies the RIP with constant $\delta \in (0, (1-\epsilon_*)^2)$ and probability at least $1-\xi$, *i.e.*,

$$(1-\epsilon_*)^2 - \delta \leqslant \|L(\boldsymbol{x})\|_2^2 \leqslant 1+\delta, \quad \forall \boldsymbol{x} \in \mathcal{S},$$

provided that

$$m \geqslant \frac{C}{\delta^2} \max\left\{ s \log\left(\frac{1}{\epsilon_{\mathcal{S}}}\right),\ \log\left(\frac{6}{\xi}\right) \right\},$$

where C is an absolute constant.

With this example, one can remark that the only difference in the construction of the linear map L between an infinite-dimensional setting and a finite-dimensional setting is the presence of the intermediate mapping b. This mapping is built from the projection onto a well-chosen subspace that preserves the norm of all vectors in $\mathcal{S}$. Finally, the number of measurements m is essentially proportional to the intrinsic dimension of $\mathcal{S}$, as usual in CS results.

2) Recovery of known results in finite dimension: The above results holds for any set $\mathcal{S}$ that satisfies Assumption A in a Hilbert space $\mathcal{H}$, so it also holds for signal models in $\mathbb{R}^n$ provided that they have a finite upper box-counting dimension. Let us take one example in finite ambient space.

Consider the set $\mathcal{S}$ of $2k$-sparse signals with unit ℓ_2-norm in $\mathbb{R}^n$. This set can be covered by at most $[3en/(2k\epsilon)]^{2k}$ balls of radius $\epsilon \in (0,1)$ [2]. Its upper box-counting dimension is thus $2k$. Let us compute $\epsilon_{\mathcal{S}}$ and s which appear in Assumption A. We remind that $\epsilon_{\mathcal{S}}$ is a constant such that $[3en/(2k\epsilon)]^{2k} \leqslant \epsilon^{-s}$ for all $\epsilon \leqslant \epsilon_{\mathcal{S}}$ with $s > 2k$. Writing $s = 2k + \eta$ with $\eta > 0$, we should have $(3en/(2k))^{2k} \leqslant \epsilon^{-\eta}$ for all $\epsilon \leqslant \epsilon_{\mathcal{S}}$, or, equivalently, $(3en/(2k))^{2k} \leqslant \epsilon_{\mathcal{S}}^{-\eta}$. We take $\epsilon_{\mathcal{S}} = 2k/(3en)$ and $\eta = 2k$. Notice that $\epsilon_{\mathcal{S}} \leqslant 1/2$ and that $s = 4k$. As we are in finite ambient dimension, we can take $\mathcal{V}_{\epsilon_*} = \mathbb{R}^n$

(which implies that $\epsilon_* = 0$ and $d = n$) and the canonical basis for $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ (which implies that b is the identity). The complete linear map L thus reduces to the matrix $\mathsf{A}/\sqrt{m}$ which satisfies, with probability at least $1-\xi$,

$$(1-\delta)\left\|\boldsymbol{x}\right\|_2^2 \leqslant \frac{\left\|\mathsf{A}\boldsymbol{x}\right\|_2^2}{m} \leqslant (1+\delta)\left\|\boldsymbol{x}\right\|_2^2, \tag{20}$$

for all $2k$-sparse vectors $\boldsymbol{x}$ provided that m satisfies

$$m \geqslant \frac{C}{\delta^2}\max\left\{4k\log\left(\frac{3\mathrm{e}n}{2k}\right),\ \log\left(\frac{6}{\xi}\right)\right\}.$$

In particular, this result shows that the set of k-sparse vectors is stably embedded into $\mathbb{R}^m$ for m satisfying the above inequality.

In comparison, Theorem 9.2 in [2] shows that $\mathsf{A}/\sqrt{m}$ satisfies (20) with probability at least $1-\xi$ provided that

$$m \geqslant \frac{D}{\delta^2}\ \left(2k\log\left(\frac{\mathrm{e}n}{2k}\right) + \log\left(\frac{2}{\xi}\right)\right),$$

where $D > 0$ is an absolute constant. This condition on m is similar to ours. This shows that we recover results similar to known ones in CS.

Note that we could have chosen any orthonormal basis of $\mathbb{R}^n$ for $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$. This illustrates the known universality of subgaussian measurement matrices relative to the sparsity basis [2].

Following the same procedure, one can easily recover many other similar results for, *e.g.*, the set of low-rank matrices [3], low-dimensional smooth manifolds [4], or the set of group-sparse signals [20]. In all these cases, the only remaining difficulty is estimating the upper box-counting dimension of the set of interest.

B. Embedding of matrices with rank-one projections

As a second example, we propose to use our results to show that one can embed a low dimensional set of matrices using rank-one projections. This scheme was proposed and studied in, *e.g.*, [16], [17], [21], [22] to embed certain low-dimensional set of matrices: low-rank matrices, sparse matrices, low-rank+sparse matrices, *etc*. We extend here these results to any set with a low-dimensional normalised secant set that satisfies Assumption A, which, *e.g.*, includes matrix manifolds.

In this section, the ambient space is the space of real matrices of size $n_1 \times n_2$, $\mathbb{R}^{n_1\times n_2}$, equipped with the Frobenius norm, $\left\|\cdot\right\|_{\mathrm{Frob}}$. We consider that $\mathcal{S} \subset \mathbb{R}^{n_1\times n_2}$ is a low-dimensional subset of matrices that satisfies Assumption A, with $\left\|\cdot\right\|_{\mathrm{Frob}}$ instead of $\left\|\cdot\right\|$.

1) Measurement strategy: As done in [16], [17], [21], [22], we propose to measure a matrix $\mathsf{M} \in \mathbb{R}^{n_1\times n_2}$ using m rank-one projections:

$$\begin{aligned} L\colon \mathbb{R}^{n_1\times n_2} &\longrightarrow \mathbb{R}^m \\ \mathsf{M} &\longmapsto \left(\frac{a_i^\intercal \mathsf{M} b_i}{m}\right)_{i=1,\ldots,m}, \end{aligned} \tag{21}$$

where $a_1, \ldots, a_m \in \mathbb{R}^{n_1}$ and $b_1, \ldots, b_m \in \mathbb{R}^{n_2}$ are independent random vectors. Note that the main advantage of this measurement strategy is its memory efficiency. Indeed, we only need to store $m\,(n_1+n_2)$ coefficients to compute the measurements while the measurement strategy proposed in [3] requires to store $m\,n_1 n_2$ coefficients.

We will pay attention to two types of random vectors. First, we will focus on the case where the entries of the vectors a_i and b_i, $i = 1, \ldots, m$, are independent draws of a random variable $N \in \mathbb{R}$ that has a standard normal distribution. Second, we will discuss the case where the entries of the vectors a_i and b_i, $i = 1, \ldots, m$, are independent draws of a random variable $P \in \{0, \pm\sqrt{q}\}$ that satisfies

$$\mathbb{P}\left(P = 0\right) = \frac{q-1}{q} \quad \text{and} \quad \mathbb{P}\left(P = \pm\sqrt{q}\right) = \frac{1}{2q}, \tag{22}$$

where $q \geqslant 1$. The larger q is, the sparser (on average) the measurement vectors are, hence improving the computational efficiency of the measurement strategy. In this case, the average sparsity of the vectors a_i and b_i is n_1/q and n_2/q, respectively, and computing one measurement can be done in $O(n_1 n_2/q^2)$ operations. This type of measurement strategy with sparse vectors was proposed in [23] for Johnson-Lindenstrauss embeddings and can also be used for compressive sensing [19]. We will see that the number of measurements needed to preserve the norm of all vectors in $\mathcal{S}$ depends on the parameter q.

2) RIP for rank-one projections with gaussian vectors: In this section, we study the case where the entries of the vectors a_i and b_i are independent draws of the random variable N. As the linear map L has the form presented in Theorem III.3 (with b the identity and $\left\|\cdot\right\|_b = \left\|\cdot\right\|_{\mathrm{Frob}}$), we can follow Recipe 2 to prove that L has the RIP.

1: The set $\mathcal{S}$ has a finite upper-box counting dimension by assumption.

2: To estimate $\underline{\delta}_1$, we use the result presented after Lemma D.2 which shows that there exists an absolute constant $D > 0$ such that

$$\underline{\delta}_1 := \inf_{\mathsf{M}\in\mathcal{S}} \mathbb{E}\left|a_i^\intercal \mathsf{M} b_i\right| \geqslant D.$$

Then, using the fact that

$$\mathbb{E}\left|a_i^\intercal \mathsf{M} b_i\right| \leqslant \left(\mathbb{E}\left|a_i^\intercal \mathsf{M} b_i\right|^2\right)^{1/2} = \left\|\mathsf{M}\right\|_{\mathrm{Frob}}, \quad \forall\, \mathsf{M} \in \mathbb{R}^{n_1\times n_2},$$

we also deduce that $\bar{\delta}_1 \leqslant 1$.

3: To estimate Λ_1, we use Lemma D.1 which shows that there exists an absolute constant $E > 0$ such that

$$\left\|a_i^\intercal \mathsf{M} b_i\right\|_{\Psi_1} \leqslant E\left\|\mathsf{M}\right\|_{\mathrm{Frob}} \quad \forall\, \mathsf{M} \in \mathbb{R}^{n_1\times n_2}.$$

Therefore $\Lambda_1 \leqslant E$.

4: The last step of the recipe proves that L satisfies the RIP with constant $\delta \in (0, D)$ and probability at least $1-\xi$, *i.e.*,

$$D - \delta \leqslant \left\|L(\mathsf{M})\right\|_1 \leqslant 1 + \delta \quad \forall\, \mathsf{M} \in \mathcal{S},$$

provided that

$$m \geqslant \frac{C}{\delta^2}\max\left\{s\log\left(\frac{1}{\epsilon_{\mathcal{S}}}\right),\ \log\left(\frac{6}{\xi}\right)\right\},$$

where $C > 0$ is an absolute constant. The number of measurements m thus scales linearly with s. This result

is true for *any set $\mathcal{S}$ that satisfies Assumption A* and is not restricted to the set of low-rank matrices.

Let us highlight that the result presented in this section involves an ℓ_1-norm in the measurement domain, instead of the ℓ_2-norm as in most CS results. The authors of [16] and [17] also chose the ℓ_1-norm in the measurement domain to prove the RIP. In addition, the authors in [16] proved that if one chooses the ℓ_2-norm in the measurement domain for the RIP, then at least $O(n_1 n_2)$ measurements are needed to ensure that L satisfies a RIP condition that guarantees the recovery of rank-1 matrices (see Lemma 2.1 in [16] and the discussion thereafter), which is much larger than n_1+n_2-1, the number of degrees of freedom for rank-1 matrices. Choosing the ℓ_1-norm in the measurement domain solves this issue.

3) Recovery of a known result: Let now $\mathcal{S}$ be the set of rank-$2r$ matrices with unit Frobenius norm. Lemma 3.1 in [3] shows that $N_{\mathcal{S}}(\epsilon) \leqslant (9/\epsilon)^{2r(n_1+n_2+1)}$ for any $\epsilon > 0$. The upper box-counting dimension of $\mathcal{S}$ is thus $2r(n_1+n_2+1)$. Using the procedure of Section IV-A2, we can satisfy Assumption A by taking $\epsilon_{\mathcal{S}} = 1/9$ and $s = 4r(n_1+n_2+1)$. The above result then shows that for $\delta \in (0, E)$, with probability at least $1-\xi$,

$$E - \delta \leqslant \|L(\mathsf{M})\|_1 \leqslant 1 + \delta, \quad \forall \mathsf{M} \in \mathcal{S}$$

provided that

$$m \geqslant \frac{C}{\delta^2} \max\left\{ 2r(n_1+n_2+1)\log(9),\ \log\left(\frac{6}{\xi}\right) \right\},$$

where $C > 0$ is an absolute constant. In particular, the set of rank-r matrices is stably embedded into $\mathbb{R}^m$ for m satisfying the above inequality.

Let us compare this result with Theorem 2.2 in [16].

Theorem IV.1 (Theorem 2.2, [16])**.** *Let $L : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be the linear map defined in (21). For positive numbers*[1] *$C_1 < 1/3$ and $C_2 > 1$, there exist constants C and δ, not depending on n_1, n_2, and r, such that if $m \geqslant Cr(n_1+n_2)$, then with probability at least $1 - \mathrm{e}^{-m\delta}$, L satisfies*

$$C_1 \|\mathsf{M}\|_{\mathrm{Frob}} \leqslant \|L(\mathsf{M})\|_1 \leqslant C_2 \|\mathsf{M}\|_{\mathrm{Frob}},$$

for all rank-$2r$ matrices $\mathsf{M} \in \mathbb{R}^{n_1 \times n_2}$.

Comparing our result and the theorem above, we notice that we recover the same type of result with a number of measurements that needs to be essentially proportional to the upper box-counting dimension of the set of rank-r matrices in order to guarantee that the Frobenius norm of these matrices is preserved.

4) RIP for rank-one projections with sparse vectors: To improve the computational efficiency of the measurement procedure, one can think of using sparse vectors a_i and b_i. Let us thus discuss the case where the entries of the vectors a_i and b_i are independent draws of the sparse random variable P according to (22). Recipe 2 still works to prove that L has the RIP.

[1]These numbers depend on the subgaussian norm of the measurement vectors a_i and b_i.

1: The set $\mathcal{S}$ has a finite upper-box counting dimension by assumption.
2: To estimate $\underline{\delta}_1$, we use the result presented after Lemma D.2 which shows that there exists an absolute constant $D > 0$ such that

$$\underline{\delta}_1 := \inf_{\mathsf{M} \in \mathcal{S}} \mathbb{E}\, |a_i^{\mathsf{T}} \mathsf{M} b_i| \geqslant \frac{D}{q\,(1+\log q)},$$

for $q \geqslant 2$. We also have $\bar{\delta}_1 \leqslant 1$, as with Gaussian random measurement vectors.
3: To estimate Λ_1, we use Lemma D.1 which shows that $\Lambda_1 \leqslant E\,q$, where $E > 0$ is an absolute contant.
4: The last step of Recipe 2 proves that L satisfies the RIP with constant $\delta \in (0, D/(q + q\log q))$ and probability at least $1-\xi$, *i.e.*,

$$\frac{D}{q} - \delta \leqslant \|L(\mathsf{M})\|_1 \leqslant 1 + \delta \quad \forall \mathsf{M} \in \mathcal{S},$$

provided that

$$m \geqslant \frac{C\,q^2}{\delta^2} \max\left\{ s\log\left(\frac{1}{\epsilon_{\mathcal{S}}}\right),\ \log\left(\frac{6}{\xi}\right) \right\},$$

where $C > 0$ is an absolute constant.

The number of measurements m thus still scales linearly with s but also with $q^4(1+\log q)^2$, when $\delta < D/(q+q\log q)$. As the (average) cost of computing one measurement is $O(n_1 n_2/q^2)$, this result is too weak in general to prove any benefit of using sparse measurement vectors.

5) Discussion: Let us highlight that the above result is derived without using any properties of $\mathcal{S}$ beyond its box counting dimension. In particular, this result also applies for sets $\mathcal{S}$ of sparse matrices for which we guess that using very sparse measurement vectors is not an adequate measurement strategy. Indeed, we need the support of the measurement vectors and the support of the vectors in the sparse matrix to both overlap in order to have non-zero measurements. As the measurement vectors get sparser, we increase the chance of having only null measurements, and thus decrease the chance of embedding the original matrix. This fact is reminiscent of the fundamental incoherence property between the sparsity domain and the sampling domain required in compressed sensing: one cannot sample a signal in the domain where it is sparse. More optimistic results might be obtained by exploiting additional structures in $\mathcal{S}$. In that perspective, the work [24] on sparse dimensionality reduction in Euclidean space might be particularly useful to obtain more optimistic results. The authors of [24] prove that, with high probability, some random sparse measurement matrices satisfy the RIP for a wide class of subsets of the unit sphere. They provide sufficient conditions on the number of measurements and on the sparsity of the measurement matrix to ensure that the RIP is satisfied. These bounds depend on additional properties of $\mathcal{S}$ beyond its box-counting dimension. For example, for sparse signals in an orthonormal basis, they further require an incoherence property to ensure that the RIP holds (Section 5.1, [24]).

This example illustrates the importance of the ratio $\max(2\Lambda_1^2, \Lambda_1)/\underline{\delta}_1^2$ in the design of the linear map L. Even

though the measurement vectors a_i, b_i are subgaussian random vectors as in Section IV-B2, the ratio $\max(2\Lambda_1^2, \Lambda_1)/\underline{\delta}_1^2$ is too large and the result does not predict any gain in the computational efficiency compared to Gaussian rank-one measurements.

V. Related works

A. Infinite dimensional CS and generalised sampling

An extension of the CS theory to an infinite dimensional setting is also studied in [5]. In their work, the ambient space is a separable Hilbert space $\mathcal{H}$ and the authors concentrate on the particular case of sparse signals in an orthonormal basis of $\mathcal{H}$. Their result does not involve a RIP but the linear map used to sense sparse signals shares some similarities with our construction of L.

Let us take the example of s-sparse signals in an orthonormal basis $(\boldsymbol{\psi}_j)_{j\in\mathbb{N}}$ of $\mathcal{H}$, *e.g.*, one can think of a wavelet basis. For simplicity, let us consider that the support of the signals in $\mathcal{S}$ belongs to $\{0, \dots, n-1\}$, hence $\mathcal{S}$ has a finite upper box-counting dimension. In Section IV-A, we measure the signals by first projecting them onto a finite dimensional subspace of large dimension d that approximates the set $\mathcal{S}$. To construct this subspace, we propose here to use another orthonormal basis $(\boldsymbol{\phi}_i)_{i\in\mathbb{N}}$ of $\mathcal{H}$, *e.g.*, one can think of the Fourier basis. We choose to construct this finite dimensional subspace by using only the first d basis vectors $\boldsymbol{\phi}_i$. Let us define the matrix $\mathsf{U} = (u_{ij} := \langle \boldsymbol{\phi}_i, \boldsymbol{\psi}_j\rangle)_{i,j\in\mathbb{N}}$. Let α denote the coordinates of $\boldsymbol{x} \in \mathcal{H}$ in $(\boldsymbol{\psi}_j)_{j\in\mathbb{N}}$. We have $b(\boldsymbol{x}) = \mathsf{P}_d\mathsf{U}\alpha \in \mathbb{R}^d$, where P_d is the matrix that selects the first d rows of U. According to our recipe, we need to choose d such that

$$(1-\epsilon_*)\,\|\boldsymbol{x}\| \leqslant \|b(\boldsymbol{x})\|_2 \leqslant \|\boldsymbol{x}\|\,,\ \forall \boldsymbol{x} \in \mathcal{S},$$

in order to have the RIP. Here, this condition is equivalent to

$$(1-\epsilon_*)\,\|\mathsf{U}\mathsf{P}_n\alpha\|_2 \leqslant \|\mathsf{P}_d\mathsf{U}\mathsf{P}_n\alpha\|_2 \leqslant \|\mathsf{U}\mathsf{P}_n\alpha\|_2\,, \tag{23}$$

for all $\mathsf{U}\mathsf{P}_n\alpha \in \mathcal{S} \subset \mathbb{S}$, where P_n is the matrix that keeps the first n entries of α and set all the other ones to 0. Using the facts that $\|\mathsf{U}\mathsf{P}_n\alpha\|_2 = \|\mathsf{P}_n\alpha\|_2$ and $\mathsf{P}_n^\intercal\mathsf{P}_n = \mathsf{P}_n\mathsf{P}_n = \mathsf{P}_n$, taking the square of (23) and rearranging the terms, one can notice that the condition

$$\|\mathsf{P}_n\mathsf{U}^\intercal\mathsf{P}_d^\intercal\mathsf{P}_d\mathsf{U}\mathsf{P}_n - \mathsf{P}_n\|_2 \leqslant \epsilon_*, \tag{24}$$

is sufficient to ensure (23).

It is interesting to notice that this condition also appears in the generalised sampling theorem of [25]. Indeed, when U is an isometry as in the present example, the condition for stable reconstruction presented in [25] amounts to the control of $\|\mathsf{P}_n\mathsf{U}^\intercal\mathsf{P}_d^\intercal\mathsf{P}_d\mathsf{U}\mathsf{P}_n - \mathsf{P}_n\|_2$ (see Section 5 in [25] and the proofs therein). It is also proved in [25] that this quantity tends to 0 as d tends to infinity. This shows that to choose d in practice, one just needs to increase it until (24) is satisfied. In the particular case where $\mathcal{H} = L_2([0,1])$, $(\boldsymbol{\psi}_j)_{j\in\mathbb{N}}$ is the Haar wavelet basis (ordered from coarse to fine scale), and $(\boldsymbol{\phi}_j)_{j\in\mathbb{N}}$ is the Fourier basis (ordered from low to high frequencies), the authors of [25] further prove that (24) is satisfied for $d = O(n^2)$, and their numerical experiments suggest that it is already sufficient to take $d = O(n)$ (see Section 5.4 in [25]). This indicates that $\mathcal{V}_{\epsilon_*}$ does not always need to be a very large dimensional subspace of $\mathcal{H}$ to ensure that the RIP holds.

Finally, we note that condition (24) also appears in the extension of the CS theory to an infinite dimensional setting presented in [5], involving however another operator norm. This property is called the weak balancing property in [5].

B. Other measures of dimension?

In his monograph [11], Robinson studies the problem of embedding a *compact* subset Σ of an infinite-dimensional (Hilbert or Banach) space into $\mathbb{R}^m$. He considers different definitions of dimension, in particular, the Hausdorff dimension (denoted dim_H hereafter) and the upper box-counting dimension. He shows that embedding the set Σ into $\mathbb{R}^m$ is possible as soon as $\dim_H(\Sigma - \Sigma) < m$ but that no modulus of continuity is possible for L^{-1}, *i.e.*, the embedding is not necessarily stable. One should further assume that the upper box-counting dimension of Σ is finite in order to find *linear* embeddings with Hölder continuous inverses.

In our work, we show that a stable embedding of a set Σ into finite dimensions exists when the upper box-counting dimension on the *normalised* secant set of Σ is finite. Compared to [11] which uses the dimension of the secant set $\Sigma - \Sigma$, we are thus adding another constraint by normalising the secant set. An interesting question is thus whether we can change our measure of dimension and still guarantee the existence of stable linear embeddings or not. A first step in answering this question is checking if a finite upper box-counting dimension of $\mathcal{S}$ is a necessary condition to the existence of linear maps that satisfies the RIP. The following example shows that it is only a sufficient condition, which leaves room for a refinement of our results. This example is inspired by the construction of the "orthogonal sequence" in [11].

Let $\mathcal{H}$ be the space of square-summable sequences, ℓ^2, and define the correlated sequence $\Sigma = \{\boldsymbol{x}_i\}_{i\geqslant 1}$ with $\boldsymbol{x}_i = r^i\,(\boldsymbol{e}_i + b\,\boldsymbol{e}_0)$, where $(\boldsymbol{e}_i)_{i\in\mathbb{N}}$ is the standard orthonormal basis of $\mathcal{H}$, $r \in (0,1)$ and $b > 0$. Now consider the linear map $L : \mathcal{H} \to \mathbb{R}$ defined by $L(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{e}_0\rangle$. Let us show that the linear map L stably embeds the set Σ. Clearly,

$$|\langle \boldsymbol{x}_i - \boldsymbol{x}_j, \boldsymbol{e}_0\rangle| \leqslant \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \quad \text{for all } \boldsymbol{x}_i, \boldsymbol{x}_j \in \Sigma,$$

so let us consider the lower isometry bound. We want to show the existence of an $\alpha > 0$ such that

$$|\langle \boldsymbol{x}_i - \boldsymbol{x}_j, \boldsymbol{e}_0\rangle| \geqslant \alpha\ \|\boldsymbol{x}_i - \boldsymbol{x}_j\|\,,\ \forall \boldsymbol{x}_i, \boldsymbol{x}_j \in \Sigma, j > i \geqslant 1,$$

or equivalently,

$$b^2(r^i - r^j)^2 \geqslant \alpha^2\ (r^{2i} + r^{2j} + b^2(r^i - r^j)^2),\ \forall j > i \geqslant 1.$$

Re-arranging the terms gives

$$b^2 r^{2i}(1 - r^{j-i})^2 \geqslant \alpha^2\ r^{2i}(1 + r^{2(j-i)} + b^2((1 - r^{j-i})^2),$$

for all $j > i \geqslant 1$. We can thus choose

$$\begin{aligned}\alpha^2 &= \min_{j>i\geqslant 1} \frac{b^2\,(1 - r^{j-i})^2}{1 + r^{2(j-i)} + b^2(1 - r^{j-i})^2} \\ &> \frac{b^2(1-r)^2}{(1 + r^2 + b^2)} > 0.\end{aligned}$$

Hence, L satisfies the RIP on $\mathcal{S}$, the normalised secant set of Σ:

$$\alpha \leqslant |\langle \boldsymbol{y}, \boldsymbol{e}_0 \rangle| \leqslant 1, \quad \text{for all } \boldsymbol{y} \in \mathcal{S}.$$

Consequently, L stably embeds the set Σ. Let us now calculate the upper box-counting dimension of $\mathcal{S}$. Consider the following infinite sequence of vectors from $\mathcal{S}$,

$$\boldsymbol{v}_k := \frac{\boldsymbol{x}_{2k} - \boldsymbol{x}_{2k+1}}{\|\boldsymbol{x}_{2k} - \boldsymbol{x}_{2k+1}\|} = \frac{\boldsymbol{e}_{2k} - r\boldsymbol{e}_{2k+1} + b(1-r)\boldsymbol{e}_0}{\sqrt{1 + r^2 + b^2(1-r)^2}},$$

with $k \geqslant 1$. Note that each $\boldsymbol{v}_k$ is the only point in the sequence that has a non-zero component in the direction $\boldsymbol{e}_{2k}$. Therefore,

$$\|\boldsymbol{v}_k - \boldsymbol{v}_{k'}\| \geqslant \frac{1}{\sqrt{1 + r^2 + b^2(1-r)^2}}, \ \forall k \neq k'.$$

Thus, for all $\epsilon < 2^{-1}[1 + r^2 + b^2(1-r)^2]^{-1/2}$, it is impossible to cover the set $\{\boldsymbol{v}_k\}_{k \geqslant 1} \subset \mathcal{S}$ with a finite ϵ-cover. Hence, $\dim(\mathcal{S}) = \infty$. We thus have a set $\mathcal{S}$ of infinite upper box-counting dimension for which there exists a linear map that satisfies the RIP. *A finite upper box-counting dimension of the normalised secant set is thus not necessary for the existence of a stable linear embedding.*

Let us terminate this section by highlighting an important fact. We mentioned earlier that the upper box-counting dimension of $\mathcal{S}$ is often twice the upper box-counting of Σ (or $\Sigma \cap \mathbb{S}$ if Σ is not bounded). It is however not always the case. Indeed, let us calculate the upper box-counting dimension of Σ. Let $\epsilon > 0$. We note that

$$\|\boldsymbol{x}_i\| = r^i (1 + b^2)^{1/2}.$$

Let $i^*(\epsilon)$ be the smallest integer such that $r^{i^*(\epsilon)}(1+b^2)^{1/2} \leqslant \epsilon$. For ϵ small enough, we have $i^*(\epsilon) \geqslant 2$. Hence, we can cover $\{x_i\}_{i \geqslant i^*(\epsilon)}$ with a single ϵ-ball at 0. Then, the remaining $i^*(\epsilon) - 1$ points, $\{x_i\}_{i=1}^{i^*(\epsilon)-1}$ can be separately covered by at most $i^*(\epsilon) - 1$ balls of radius ϵ centered at each x_i. This gives $N_\Sigma(\epsilon) \leqslant i^*(\epsilon)$. Then,

$$\frac{\log N_\Sigma(\epsilon)}{-\log \epsilon} \leqslant \frac{\log i^*(\epsilon)}{-(i^*(\epsilon) - 1)\log r - \frac{1}{2}\log(1 + b^2)}.$$

Hence,

$$\begin{aligned} \dim(\Sigma) &= \limsup_{\epsilon \to 0} \frac{\log N_\Sigma(\epsilon)}{-\log \epsilon} \\ &\leqslant \limsup_{i^* \to \infty} \frac{\log i^*}{-(i^* - 1)\log r - \frac{1}{2}\log(1 + b^2)} \\ &= 0. \end{aligned}$$

Since $\dim(\Sigma)$ cannot be negative, we have $\dim(\Sigma) = 0$. We thus have an example of a set Σ with a null upper box-counting dimension but whose normalised secant set has an infinite upper box-counting dimension. In general, one cannot directly deduce the dimension of $\mathcal{S}$ from the dimension of Σ.

C. Related works with similar proof techniques

A closely related work is the one of Dirksen [26]. In this work, the ambient space is a (possibly infinite) Hilbert space and a generic theory for dimensionality reduction with subgaussian maps is presented. This work allows one to derive embedding results *once a random linear map* L is given. Theorem III.4 can for example be derived using the generic theory developed in [26]. In our work, we provide in addition a recipe for the construction of the linear map L itself. Let us also note that one cannot directly recover the results presented in Section IV-B with the result presented in [26].

Another related work of Dirksen is [15], that extends results presented in [27], [28] and where a technique to obtain tail bounds of the supremum of a stochastic process is presented. Let us highlight that these results can be another approach to obtain our generic theorem (Theorem II.2). From a more technical point of view, [15] provides bounds for all p-th moment of the supremum of a stochastic process. Deviation inequalities immediately follows from these bounds. The proofs use the so-called generic chaining argument and the γ-functionals. Although the approximation of these functionals might not be easy, this method can produce sharper bounds than ours. In our proofs, we use more basic tools of probability theory with a another type of chaining argument to directly obtain the deviation inequalities.

Related works also exist for embeddings of manifolds in finite ambient dimension [4], [13], [29]. In particular, the result presented in Section IV-A generalises the result presented in [4] to any model set with finite upper box-counting dimension. The proof technique used in [4] is also based on a chaining argument. In [4], one of the main contribution is also the identification of the manifold properties that allow one to control the covering dimension. In [30], the same type of embedding result also appears but for signals that lie on a collection of continuously parameterised low-dimensional subspaces in $\mathbb{R}^n$.

Even though the following result applies in a finite dimensional Euclidean space, let us mention that Oymak *et al.* proved in [31] that certain structured random matrices can embed any low-dimensional set of $\mathbb{R}^n$. This is a particularly useful result for practical applications as it shows that matrices for which fast matrix-vector multiplication algorithms exist can be used to embed any set, as long as it has a small intrinsic dimension. It would be interesting in a future work to study how such computationally efficient matrices can be used in the construction of our linear map L.

VI. Conclusion

We presented a generic recipe to prove that a random linear map satisfies the RIP on arbitrary low dimensional sets $\mathcal{S}$ which lives in a Hilbert space. The proposed framework is general enough to take into account a large class of subsets $\mathcal{S}$ as well as structured and unstructured measurement processes. We also explained how to *construct* random linear maps that satisfy the RIP with high probability in this general setting.

The linear map presented in Section IV-A is built in two steps. The first step consists in a projection onto a well-chosen finite-dimensional subspace $\mathcal{V}_{\epsilon_*}$ and the second step uses a random subgaussian matrix to reduce the dimension. In order to obtain linear maps which are computationally more efficient, one can consider replacing the random matrix in the

second step by a more structured one. One possibility is to use a variable density sampling technique, as we considered in [1]. Note that the condition on m presented in [1] is not optimal. Indeed for k-sparse signals in $\mathbb{R}^n$, the condition requires a number of measurements essentially proportional to k^2 to ensure that the RIP is satisfied. To understand the performance of the variable density sampling technique in practical applications, it will be important to determine under which additional structures this condition on m can be improved.

We also saw that a finite upper box-counting dimension of the normalised secant set $\mathcal{S}(\Sigma)$ of Σ is not necessary for the existence of a stable dimension reducing linear embedding of Σ. An interesting question is thus what appropriate measure of dimension of $\mathcal{S}(\Sigma)$ is both necessary and sufficient for the existence of a stable dimension reducing linear embedding of Σ.

Appendix A
Proof of Theorem II.2

The proof of Theorem II.2 is based on a chaining argument which is a powerful technique to obtain sharp bounds for the supremum of random processes [4], [12]–[14]. Before proving Theorem II.2, we need some preparations.

First, we notice that if Assumption A holds then we can cover the set $\mathcal{S}$ with a finite number of balls. Furthermore, a bound on the number of balls sufficient to cover $\mathcal{S}$ is available for all radius $\epsilon \leqslant \epsilon_{\mathcal{S}}$. In the proof below, for each $j \geqslant 0$,

- $\mathcal{C}_j \subset \mathcal{S}$, denotes a minimal $(2^{-j}\epsilon_{\mathcal{S}})$-net for $\mathcal{S}$;
- π_j denotes the mapping $\pi_j(\boldsymbol{x}) \in \operatorname{argmin}_{\boldsymbol{z}\in\mathcal{C}_j} \|\boldsymbol{x}-\boldsymbol{z}\|$;
- $\mathcal{D}_j$ denotes the finite set $\{(\pi_{j+1}(\boldsymbol{x}), \pi_j(\boldsymbol{x})) \mid \boldsymbol{x} \in \mathcal{S}\} \subset \mathcal{C}_{j+1} \times \mathcal{C}_j$.

One can remark that the cardinality of $\mathcal{C}_j$ is $N_{\mathcal{S}}(2^{-j}\epsilon_{\mathcal{S}})$, that the cardinality of $\mathcal{D}_j$ is bounded above by $N_{\mathcal{S}}^2(2^{-j-1}\epsilon_{\mathcal{S}})$, and that $\sup_{(\boldsymbol{y},\boldsymbol{z})\in\mathcal{D}_j} \|\boldsymbol{y}-\boldsymbol{z}\| \leqslant 2^{-j+1}\epsilon_{\mathcal{S}}$.

Second, we remark that Assumption B implies that for any fixed $\boldsymbol{x} \in \mathcal{S} \subset \mathbb{S}$

$$\mathbb{P}\left\{|h_p(\boldsymbol{x})| \geqslant \lambda\right\} \leqslant \begin{cases} 2\mathrm{e}^{-c_1 m\lambda^2}, \text{if } 0 \leqslant \lambda \leqslant \frac{c_2}{c_1}, \\ 2\mathrm{e}^{-c_1 m\lambda}, \text{if } \lambda \geqslant \frac{c_2}{c_1}. \end{cases} \tag{25}$$

Indeed, it suffices to take $\boldsymbol{y} = \boldsymbol{x}$ and $\boldsymbol{z} = \boldsymbol{0}$ in (7) and (8).

With these tools in hand, we can prove the following intermediate result from which Theorem II.2 follows.

Lemma A.1. *Let $(\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m)$ be m random vectors drawn from $\mathcal{H}^m$ according to a probability distribution μ, $L : \mathcal{H} \to \mathbb{R}^m$ be the linear map defined in (3), δ_p be the quantity defined in (6) and $\xi \in (0,1)$. Define*

$$S_1(\epsilon_{\mathcal{S}}, \xi) := \sqrt{\log\left(\frac{2}{\xi} \cdot N_{\mathcal{S}}\left(\epsilon_{\mathcal{S}}\right)\right)},$$

$$S_2(\epsilon_{\mathcal{S}}, \xi) := \sum_{j\in\mathbb{N}} 2^{-j+1}\sqrt{\log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right)},$$

$$S_3(\epsilon_{\mathcal{S}}, \xi) := \sum_{j\in\mathbb{N}} 2^{-j+1}\log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right).$$

If Assumption B holds, then

$$\delta_p \leqslant \frac{S_1(\epsilon_{\mathcal{S}}, \xi) + S_2(\epsilon_{\mathcal{S}}, \xi)\epsilon_{\mathcal{S}}}{\sqrt{c_1 m}} + \frac{S_1^2(\epsilon_{\mathcal{S}}, \xi) + S_3(\epsilon_{\mathcal{S}}, \xi)\epsilon_{\mathcal{S}}}{c_2 m}$$

with probability at least $1 - 3\xi$.

Proof: We begin by establishing the telescopic sum expression:

$$h_p(\boldsymbol{x}) = h_p(\pi_0(\boldsymbol{x})) + \sum_{j=0}^{\infty}\left[h_p(\pi_{j+1}(\boldsymbol{x})) - h_p(\pi_j(\boldsymbol{x}))\right].$$

The above equality holds because the linear maps of the form of L are continuous with respect to $\|\cdot\|$. Therefore, $\|L(\cdot)\|_p^p$, $\mathbb{E}\,\|L(\cdot)\|_p^p$, and thus h_p are also continuous with respect to $\|\cdot\|$. Then,

$$\begin{aligned} h_p(\boldsymbol{x}) - h_p(\pi_0(\boldsymbol{x})) - \sum_{j=0}^{N}\left[h_p(\pi_{j+1}(\boldsymbol{x})) - h_p(\pi_j(\boldsymbol{x}))\right] \\ = h_p(\boldsymbol{x}) - h_p(\pi_{N+1}(\boldsymbol{x})). \end{aligned}$$

As $\lim_{N\to\infty}\|\pi_{N+1}(\boldsymbol{x}) - \boldsymbol{x}\| = 0$, we have $\lim_{N\to\infty}|h_p(\boldsymbol{x}) - h_p(\pi_{N+1}(\boldsymbol{x}))| = 0$ by continuity.

We continue with the triangle inequality which yields

$$\begin{aligned} \delta_p &= \sup_{\boldsymbol{x}\in\mathcal{S}}|h_p(\boldsymbol{x})| \\ &\leqslant \sup_{\boldsymbol{x}\in\mathcal{S}}|h_p(\pi_0(\boldsymbol{x}))| + \sum_{j=0}^{\infty}\sup_{\boldsymbol{x}\in\mathcal{S}}|h_p(\pi_{j+1}(\boldsymbol{x})) - h_p(\pi_j(\boldsymbol{x}))| \\ &= \max_{\boldsymbol{x}_0\in\mathcal{C}_0}|h_p(\boldsymbol{x}_0)| + \sum_{j=0}^{\infty}\max_{(\boldsymbol{y},\boldsymbol{z})\in\mathcal{D}_j}|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})|. \end{aligned}$$

Let $a_j, b > 0$, with $j \in \mathbb{N}$, be parameters whose values will be chosen later on. The union bound yields

$$\begin{aligned} &\mathbb{P}\left\{\delta_p \geqslant b + \sum_{j\geqslant 0} a_j\right\} \\ &\leqslant \mathbb{P}\left\{\max_{\boldsymbol{x}_0\in\mathcal{C}_0}|h_p(\boldsymbol{x}_0)| \geqslant b\right\} \\ &\qquad + \sum_{j\geqslant 0}\mathbb{P}\left\{\max_{(\boldsymbol{y},\boldsymbol{z})\in\mathcal{D}_j}|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})| \geqslant a_j\right\} \\ &\leqslant N_{\mathcal{S}}\left(\epsilon_{\mathcal{S}}\right) \cdot \max_{\boldsymbol{x}_0\in\mathcal{C}_0}\mathbb{P}\left\{|h_p(\boldsymbol{x}_0)| \geqslant b\right\} \\ &+ \sum_{j\geqslant 0}\left[N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right) \cdot \max_{(\boldsymbol{y},\boldsymbol{z})\in\mathcal{D}_j}\mathbb{P}\left\{|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})| \geqslant a_j\right\}\right]. \end{aligned} \tag{26}$$

In the last step, we used the facts that the cardinality of $\mathcal{C}_0$ is $N_{\mathcal{S}}\left(\epsilon_{\mathcal{S}}\right)$, and that the cardinality of $\mathcal{D}_j$ is bounded above by $N_{\mathcal{S}}^2\left(2^{-j-1}\epsilon_{\mathcal{S}}\right)$.

We now bound the first term on the right-hand side (rhs) of (26). We recall that $\mathcal{C}_0 \subset \mathcal{S}$. Inequality (25) thus yields

$$\max_{\boldsymbol{x}_0\in\mathcal{C}_0}\mathbb{P}\left\{|h_p(\boldsymbol{x}_0)| \geqslant b\right\} \leqslant \begin{cases} 2\mathrm{e}^{-c_1 m b^2}, \text{if } b \leqslant \frac{c_2}{c_1}, \\ 2\mathrm{e}^{-c_2 m b}, \text{if } b \geqslant \frac{c_2}{c_1}. \end{cases}$$

If

$$\frac{S_1^2(\epsilon_{\mathcal{S}}, \xi)}{m} \leqslant \frac{c_2^2}{c_1},$$

we choose

$$b := \frac{S_1(\epsilon_{\mathcal{S}}, \xi)}{\sqrt{c_1 m}} = \sqrt{\frac{\log(2N_{\mathcal{S}}(\epsilon_{\mathcal{S}})/\xi)}{c_1 m}},$$

and

$$b := \frac{S_1^2(\epsilon_{\mathcal{S}}, \xi)}{c_2 m} = \frac{\log(2N_{\mathcal{S}}(\epsilon_{\mathcal{S}})/\xi)}{c_2 m},$$

otherwise. This choice yields

$$N_{\mathcal{S}}(\epsilon_{\mathcal{S}}) \cdot \max_{\boldsymbol{x}_0 \in \mathcal{C}_0} \mathbb{P}\{|h_p(\boldsymbol{x}_0)| \geqslant b\} \leqslant \xi,$$

and

$$b \leqslant \frac{S_1(\epsilon_{\mathcal{S}}, \xi)}{\sqrt{c_1 m}} + \frac{S_1^2(\epsilon_{\mathcal{S}}, \xi)}{c_2 m},$$

in both cases.

We continue by bounding the infinite sum on the rhs of (26). We notice that for $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{D}_j$, we have $\|\boldsymbol{y} - \boldsymbol{z}\| \leqslant 2^{-j+1}\epsilon_{\mathcal{S}}$. Assumption B shows that

$$\begin{aligned}
&\mathbb{P}\{|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})| \geqslant a_j\} \\
&\quad = \mathbb{P}\left\{|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})| \geqslant \frac{a_j 2^{j-1}}{\epsilon_{\mathcal{S}}} \cdot \frac{\epsilon_{\mathcal{S}}}{2^{j-1}}\right\} \\
&\quad \leqslant \mathbb{P}\left\{|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})| \geqslant \frac{a_j 2^{j-1}}{\epsilon_{\mathcal{S}}} \|\boldsymbol{y} - \boldsymbol{z}\|\right\} \\
&\quad \leqslant \begin{cases} 2\,\mathrm{e}^{-c_1\, m\, (2^{j-1} a_j/\epsilon_{\mathcal{S}})^2}, \text{ if } \frac{a_j}{2^{-j+1}\epsilon_{\mathcal{S}}} \leqslant c_2/c_1, \\ 2\,\mathrm{e}^{-c_2\, m\, 2^{j-1} a_j/\epsilon_{\mathcal{S}}}, \text{ if } \frac{a_j}{2^{-j+1}\epsilon_{\mathcal{S}}} \geqslant c_2/c_1. \end{cases}
\end{aligned} \tag{27}$$

Define

$$\mathcal{J} := \left\{ j \in \mathbb{N} \,\middle|\, \frac{1}{m} \log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right) \leqslant \frac{c_2^2}{c_1} \right\}.$$

We choose

$$a_j := \frac{2^{-j+1}\epsilon_{\mathcal{S}}}{\sqrt{c_1 m}} \sqrt{\log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right)},$$

for all $j \in \mathcal{J}$, and

$$a_j := \frac{2^{-j+1}\epsilon_{\mathcal{S}}}{c_2 m} \log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right),$$

for all $j \in \mathbb{N} \setminus \mathcal{J}$. Notice that, by definition of $\mathcal{J}$, we have

$$\frac{a_j}{2^{-j+1}\epsilon_{\mathcal{S}}} \leqslant \frac{c_2}{c_1}$$

for all $j \in \mathcal{J}$, and

$$\frac{a_j}{2^{-j+1}\epsilon_{\mathcal{S}}} \geqslant \frac{c_2}{c_1},$$

otherwise. Replacing a_j by its expression in (27) shows that

$$N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right) \max_{(\boldsymbol{y},\boldsymbol{z}) \in \mathcal{D}_j} \mathbb{P}\{|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})| \geqslant a_j\} \leqslant 2^{-j}\xi$$

for all $j \in \mathbb{N}$. Therefore,

$$\sum_{j \geqslant 0} N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right) \max_{(\boldsymbol{y},\boldsymbol{z}) \in \mathcal{D}_j} \mathbb{P}\{|h_p(\boldsymbol{y}) - h_p(\boldsymbol{z})| \geqslant a_j\} \leqslant 2\xi.$$

In total, the rhs of (26) is thus bounded by 3ξ. Finally, we have

$$\begin{aligned}
\sum_{j \geqslant 0} a_j &= \frac{\epsilon_{\mathcal{S}}}{\sqrt{c_1 m}} \sum_{j \in \mathcal{J}} 2^{-j+1} \sqrt{\log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right)} \\
&\quad + \frac{\epsilon_{\mathcal{S}}}{c_2 m} \sum_{j \in \mathbb{N} \setminus \mathcal{J}} 2^{-j+1} \log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right) \\
&\leqslant \frac{\epsilon_{\mathcal{S}}}{\sqrt{c_1 m}} \sum_{j \in \mathbb{N}} 2^{-j+1} \sqrt{\log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right)} \\
&\quad + \frac{\epsilon_{\mathcal{S}}}{c_2 m} \sum_{j \in \mathbb{N}} 2^{-j+1} \log\left(\frac{2^{j+1}}{\xi} \cdot N_{\mathcal{S}}^2\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right)\right) \\
&= \epsilon_{\mathcal{S}}\, \frac{S_2(\epsilon_{\mathcal{S}}, \xi)}{\sqrt{c_1 m}} + \epsilon_{\mathcal{S}}\, \frac{S_3(\epsilon_{\mathcal{S}}, \xi)}{c_2 m}.
\end{aligned}$$

In summary, we have shown that

$$\delta_p \leqslant b + \sum_{j \geqslant 0} a_j$$

with probability at least $1 - 3\xi$. Using the bounds on b and $\sum_{j \geqslant 0} a_j$, we have

$$\delta_p \leqslant \frac{S_1(\epsilon_{\mathcal{S}}, \xi) + S_2(\epsilon_{\mathcal{S}}, \xi)\epsilon_{\mathcal{S}}}{\sqrt{c_1 m}} + \frac{S_1^2(\epsilon_{\mathcal{S}}, \xi) + S_3(\epsilon_{\mathcal{S}}, \xi)\epsilon_{\mathcal{S}}}{c_2 m}$$

with at least the same probability. ■

We can now prove Theorem II.2. First, we need to compute the sums S_1, S_2, S_3 in the above lemma. It involves uninteresting computations which, for completeness, we detail in Appendix B. For $\xi \in (0,1)$ and $\epsilon_{\mathcal{S}} \in (0, 1/2)$, if Assumption A holds, these computations show that

$$\begin{aligned}
&S_1(\epsilon_{\mathcal{S}}, \xi) \leqslant \sqrt{\log(2/\xi)} + \sqrt{s \log(1/\epsilon_{\mathcal{S}})}, \\
&S_2(\epsilon_{\mathcal{S}}, \xi) \leqslant 8\sqrt{\log(2/\xi)} + 8\sqrt{2s\log(2)} + 4\sqrt{2s\log(1/\epsilon_{\mathcal{S}})}, \\
&S_1^2(\epsilon_{\mathcal{S}}, \xi) \leqslant \log(2/\xi) + s\log(1/\epsilon_{\mathcal{S}}), \\
&S_3(\epsilon_{\mathcal{S}}, \xi) \leqslant 8\log(2/\xi) + 16s\log(2) + 8s\log(1/\epsilon_{\mathcal{S}}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\delta_p &\leqslant \sqrt{\frac{\log(2/\xi)}{c_1 m}} + \sqrt{\frac{s\log(1/\epsilon_{\mathcal{S}})}{c_1 m}} \\
&\quad + 8\epsilon_{\mathcal{S}}\sqrt{\frac{\log(2/\xi)}{c_1 m}} + 8\epsilon_{\mathcal{S}}\sqrt{\frac{2s\log(2)}{c_1 m}} + 4\epsilon_{\mathcal{S}}\sqrt{\frac{2s\log(1/\epsilon_{\mathcal{S}})}{c_1 m}} \\
&\quad + \frac{\log(2/\xi)}{c_2 m} + \frac{s\log(1/\epsilon_{\mathcal{S}})}{c_2 m} \\
&\quad + \frac{8\epsilon_{\mathcal{S}}\log(2/\xi)}{c_2 m} + \frac{16\epsilon_{\mathcal{S}}\, s\log(2)}{c_2 m} + \frac{8\epsilon_{\mathcal{S}}\, s\log(1/\epsilon_{\mathcal{S}})}{c_2 m}.
\end{aligned}$$

with probability at least $1 - 3\xi$.

Let $\delta \in (0,1)$ and define $D = 1/\min(c_1, c_2)$. We remark that if

$$m \geqslant \frac{16D}{\delta^2} \log(2/\xi), \tag{28}$$

then, since $\epsilon_{\mathcal{S}} < 1/2$,

$$\sqrt{\frac{\log(2/\xi)}{c_1 m}} \leqslant \delta,\ 8\epsilon_{\mathcal{S}}\sqrt{\frac{\log(2/\xi)}{c_1 m}} \leqslant \delta,$$
$$\frac{\log(2/\xi)}{c_2 m} \leqslant \delta^2 \leqslant \delta,\ \text{and}\ \ 8\epsilon_{\mathcal{S}}\ \frac{\log(2/\xi)}{c_2 m} \leqslant \delta^2 \leqslant \delta.$$

Similarly, if

$$m \geqslant \frac{8D}{\delta^2}\ s\ \log(1/\epsilon_{\mathcal{S}}), \tag{29}$$

then, since $\epsilon_{\mathcal{S}} < 1/2$,

$$\sqrt{\frac{s\log(1/\epsilon_{\mathcal{S}})}{c_1 m}} \leqslant \delta,\ 4\epsilon_{\mathcal{S}}\sqrt{\frac{2s\log(1/\epsilon_{\mathcal{S}})}{c_1 m}} \leqslant \delta,$$
$$\frac{s\log(1/\epsilon_{\mathcal{S}})}{c_2 m} \leqslant \delta^2 \leqslant \delta,\ \text{and}\ \ \frac{8\epsilon_{\mathcal{S}} s\log(1/\epsilon_{\mathcal{S}})}{c_2 m} \leqslant \delta^2 \leqslant \delta.$$

Finally, if

$$m \geqslant \frac{32D}{\delta^2}\ s\ \log(2), \tag{30}$$

then, since $\epsilon_{\mathcal{S}} < 1/2$,

$$8\epsilon_{\mathcal{S}}\sqrt{\frac{2s\log(2)}{c_1 m}} \leqslant \delta\ \ \text{and}\ \ \frac{16\epsilon_{\mathcal{S}} s\log(2)}{c_2 m} \leqslant \delta^2 \leqslant \delta.$$

To terminate the proof, we notice that if

$$m \geqslant \frac{D}{\delta^2}\max\left\{s\log\left(\frac{1}{\epsilon_{\mathcal{S}}}\right),\ \log\left(\frac{2}{\xi}\right)\right\},$$

with $D := 32/\min(c_1, c_2)$, holds then (28), (29), (30) all hold. Under this condition, we have $\delta_p \leqslant 10\delta$, with probability at least $1 - 3\xi$. Two change of variables ($\xi' = 3\,\xi$, $\delta' = 10\,\delta$) prove the theorem with $C = 3200$.

Appendix B
Evaluation of S_1, S_2, S_3 of Lemma A.1

To estimates the sums S_1, S_2, S_3, we start by noticing that $\sum_{j\geqslant 0} j2^{-j} = 2$ (see, *e.g.*, Section 4.2.3 in [32]). We precompute, for $\xi \in (0,1)$,

$$\begin{aligned}\sum_{j=0}^{\infty} 2^{-j}\sqrt{\log(2^{j+1}/\xi)} &\leqslant \sum_{j=0}^{\infty} 2^{-j}\sqrt{j\log(2) + \log(2/\xi)}\\ &\leqslant \sum_{j=0}^{\infty} 2^{-j}\sqrt{j\log(2)} + \sum_{j=0}^{\infty} 2^{-j}\sqrt{\log(2/\xi)}\\ &\leqslant \sqrt{\log(2)}\sum_{j=0}^{\infty} 2^{-j} j + 2\sqrt{\log(2/\xi)}\\ &= 2\left[\sqrt{\log(2)} + \sqrt{\log(2/\xi)}\right] \leqslant 4\sqrt{\log(2/\xi)}.\end{aligned}$$

Similarly,

$$\sum_{j=0}^{\infty} 2^{-j}\log\left(2^{j+1}/\xi\right) \leqslant 2\left[\log(2) + \log(2/\xi)\right] \leqslant 4\log(2/\xi).$$

We also have, for $\epsilon_{\mathcal{S}} \in (0, 1/2)$,

$$\begin{aligned}&\sum_{j=0}^{\infty} 2^{-j}\sqrt{\log\left(2^{2(j+1)s}\epsilon_{\mathcal{S}}^{-2s}\right)}\\ &\quad= \sum_{j=0}^{\infty} 2^{-j}\sqrt{2(j+1)s\log(2) + 2s\log(1/\epsilon_{\mathcal{S}}))}\\ &\quad\leqslant \sqrt{2s\log(2)}\sum_{j=0}^{\infty} 2^{-j}\sqrt{j+1} + 2\sqrt{2s\log(1/\epsilon_{\mathcal{S}})}\\ &\quad\leqslant \sqrt{2s\log(2)}\sum_{j=0}^{\infty} 2^{-j}(j+1) + 2\sqrt{2s\log(1/\epsilon_{\mathcal{S}})}\\ &\quad= 4\sqrt{2s\log(2)} + 2\sqrt{2s\log(1/\epsilon_{\mathcal{S}})}.\end{aligned}$$

Similarly,

$$\sum_{j=0}^{\infty} 2^{-j}\log\left(2^{2(j+1)s}\epsilon_{\mathcal{S}}^{-2s}\right) \leqslant 8s\log(2) + 4s\log(1/\epsilon_{\mathcal{S}}).$$

If Assumption A holds, we have

$$N_{\mathcal{S}}\left(\frac{\epsilon_{\mathcal{S}}}{2^{j+1}}\right) \leqslant 2^{(j+1)s}\epsilon_{\mathcal{S}}^{-s},$$

for all $j \in \mathbb{N}$. Therefore, using the pre-computations above, we obtain

$$S_2(\epsilon_{\mathcal{S}}, \xi) \leqslant 8\sqrt{\log(2/\xi)} + 8\sqrt{2s\log(2)} + 4\sqrt{2s\log(1/\epsilon_{\mathcal{S}})},$$

and

$$S_3(\epsilon_{\mathcal{S}}, \xi) \leqslant 8\log(2/\xi) + 16s\log(2) + 8s\log(1/\epsilon_{\mathcal{S}}).$$

We also have

$$S_1^2(\epsilon_{\mathcal{S}}, \xi) \leqslant \log(2/\xi) + s\log(1/\epsilon_{\mathcal{S}}),$$

and

$$S_1(\epsilon_{\mathcal{S}}, \xi) \leqslant \sqrt{\log(2/\xi)} + \sqrt{s\log(1/\epsilon_{\mathcal{S}})}.$$

Appendix C
Proof of Theorem III.3 and Theorem III.4

A. Basic tools

In this section, we use several properties of subgaussian and subexponential random vectors/variables (see Definition III.1 and Definition III.2). We let the reader refer to, *e.g.*, [18] for more information about them. We recall here one useful property.

- If X is a subexponential random variable then so is $X - \mathbb{E}X$, and we have $\|X - \mathbb{E}X\|_{\Psi_1} \leqslant 2\,\|X\|_{\Psi_1}$ (Remark 5.18, [18]).

We will also use the following Bernstein-type inequality for subexponential random variables.

Lemma C.1 (Corollary 5.17, [18])**.** *There exists an absolute constant $c > 0$ such that for independent centered subexponential random variables $X_1, \ldots, X_m$ with subexponential norm bounded by $K > 0$, we have, for every $0 \leqslant t \leqslant K$,*

$$\mathbb{P}\left\{\frac{1}{m}\left|\sum_{i=1}^{m} X_i\right| \geqslant t\right\} \leqslant 2\ \mathrm{e}^{-cmt^2/K^2},$$

and, for every $t \geqslant K$,

$$\mathbb{P}\left\{\frac{1}{m}\left|\sum_{i=1}^{m} X_i\right| \geqslant t\right\} \leqslant 2\ \mathrm{e}^{-cmt/K}.$$

B. Proof of Theorem III.3

The independent random vectors $a_1, \ldots, a_m \in \mathbb{R}^d$ are assumed to be such that

$$\Lambda_1 := \sup_{\substack{\boldsymbol{x}\in\{\mathcal{S}-\mathcal{S}\}\cup\mathcal{S}\\ \boldsymbol{x}\neq\boldsymbol{0}}} \left\{\|a_i^\intercal b(\boldsymbol{x})\|_{\Psi_1}/\|b(\boldsymbol{x})\|_b\right\} < +\infty.$$

Therefore, for all $p \geqslant 1$, we have

$$\mathbb{E}\left(|a_i^\intercal b(\boldsymbol{x})|^p\right)^{1/p} \leqslant \Lambda_1\ p\ \|b(\boldsymbol{x})\|_b\,, \tag{31}$$

for any $\boldsymbol{x}$ in $\{\mathcal{S}-\mathcal{S}\}\cup\mathcal{S}$.

Let $\boldsymbol{y}, \boldsymbol{z}$ be two fixed vectors in $\mathcal{S}\cup\{\boldsymbol{0}\}$. Define $X := \sum_{i=1}^m X_i$ where

$$X_i := |a_i^\intercal b(\boldsymbol{y})| - |a_i^\intercal b(\boldsymbol{z})| - \mathbb{E}(|a_i^\intercal b(\boldsymbol{y})| - |a_i^\intercal b(\boldsymbol{z})|).$$

It is clear that X is a sum of m independent centered random variables. To use Lemma C.1, we need to show that these variables are subexponential and bound their subexponential norm. We have

$$\begin{aligned}\left(\mathbb{E}\big||a_i^\intercal b(\boldsymbol{y})| - |a_i^\intercal b(\boldsymbol{z})|\big|^p\right)^{1/p} &\leqslant \left(\mathbb{E}\ |a_i^\intercal(b(\boldsymbol{y}) - b(\boldsymbol{z}))|^p\right)^{1/p}\\ &= \left(\mathbb{E}\ |a_i^\intercal(b(\boldsymbol{y}-\boldsymbol{z}))|^p\right)^{1/p}\\ &\leqslant \Lambda_1\ p\ \|b(\boldsymbol{y}-\boldsymbol{z})\|_b\\ &\leqslant \Lambda_1\ p\ \|\boldsymbol{y}-\boldsymbol{z}\|\,.\end{aligned}$$

The second step follows from the linearity of b. The third step follows from (31) and the fact that $\boldsymbol{y}-\boldsymbol{z} \in \{\mathcal{S}-\mathcal{S}\}\cup\mathcal{S}$. The last step follows from inequality (10).

We deduce that $X_1, \ldots, X_m$ are independent centered subexponential random variables with subexponential norm bounded by $2\Lambda_1\|\boldsymbol{y}-\boldsymbol{z}\|$. Observe that

$$h_1(\boldsymbol{y}) - h_1(\boldsymbol{z}) = \frac{1}{m}\ \sum_{i=1}^m X_i.$$

Lemma C.1 with $K = 2\Lambda_1\|\boldsymbol{y}-\boldsymbol{z}\|$ yields

$$\mathbb{P}\left\{|h_1(\boldsymbol{y}) - h_1(\boldsymbol{z})| \geqslant \lambda\|\boldsymbol{y}-\boldsymbol{z}\|\right\} \leqslant 2\mathrm{e}^{-c_1 m\lambda^2},$$

for every $0 \leqslant \lambda \leqslant 2\Lambda_1$, and

$$\mathbb{P}\left\{|h_1(\boldsymbol{y}) - h_1(\boldsymbol{z})| \geqslant \lambda\|\boldsymbol{y}-\boldsymbol{z}\|\right\} \leqslant 2\mathrm{e}^{-c_2 m\lambda},$$

for every $\lambda \geqslant 2\Lambda_1$, with $c_1 = c/(4\Lambda_1^2)$ and $c_2 = c/(2\Lambda_1)$. Note that $c_2/c_1 = 2\Lambda_1$. Hence, Assumption B is satisfied. Theorem II.2 terminates the proof.

C. Proof of Theorem III.4

The independent random vectors $a_1, \ldots, a_m \in \mathbb{R}^d$ are assumed to be such that

$$\Lambda_2 := \sup_{\substack{\boldsymbol{x}\in\{\mathcal{S}-\mathcal{S}\}\cup\mathcal{S}\\ \boldsymbol{x}\neq\boldsymbol{0}}} \left\{\|a_i^\intercal b(\boldsymbol{x})\|_{\Psi_2}/\|b(\boldsymbol{x})\|_b\right\} < +\infty$$

Therefore, for all $p \geqslant 1$, we have

$$\mathbb{E}\left(|a_i^\intercal b(\boldsymbol{x})|^p\right)^{1/p} \leqslant \Lambda_2\ \sqrt{p}\ \|b(\boldsymbol{x})\|_b\,, \tag{32}$$

for any $\boldsymbol{x}$ in $\{\mathcal{S}-\mathcal{S}\}\cup\mathcal{S}$.

Let $\boldsymbol{y}, \boldsymbol{z}$ be two fixed vectors in $\mathcal{S}\cup\{\boldsymbol{0}\}$. Define $X := \sum_{i=1}^m X_i$ where

$$X_i := |a_i^\intercal b(\boldsymbol{y})|^2 - |a_i^\intercal b(\boldsymbol{z})|^2 - \mathbb{E}(|a_i^\intercal b(\boldsymbol{y})|^2 - |a_i^\intercal b(\boldsymbol{z})|^2).$$

It is clear that X is a sum of m independent centered random variables. To use Lemma C.1, we need to show that these variables are subexponential and bound their subexponential norm. We have

$$\begin{aligned}&\left(\mathbb{E}\left||a_i^\intercal b(\boldsymbol{y})|^2 - |a_i^\intercal b(\boldsymbol{z})|^2\right|^p\right)^{1/p}\\ &= \left(\mathbb{E}\left(|\boldsymbol{a}_i^\intercal(b(\boldsymbol{y}) + b(\boldsymbol{z}))|^p\,|\boldsymbol{a}_i^\intercal(b(\boldsymbol{y}) - b(\boldsymbol{z}))|^p\right)\right)^{1/p}\\ &\leqslant \left(\mathbb{E}\,|\boldsymbol{a}_i^\intercal b(\boldsymbol{y}+\boldsymbol{z})|^{2p}\right)^{\frac{1}{2p}}\left(\mathbb{E}\,|\boldsymbol{a}_i^\intercal b(\boldsymbol{y}-\boldsymbol{z})|^{2p}\right)^{\frac{1}{2p}}.\end{aligned}$$

In the last step, we used the Cauchy-Schwarz inequality and the linearity of b. Using (32) and the facts that[2] $(\boldsymbol{y}\pm\boldsymbol{z}) \in \{\mathcal{S}-\mathcal{S}\}\cup\mathcal{S}$, we obtain

$$\begin{aligned}\left(\mathbb{E}\,|\boldsymbol{a}_i^\intercal b(\boldsymbol{y}\pm\boldsymbol{z})|^{2p}\right)^{\frac{1}{2p}} &\leqslant \Lambda_2\ \sqrt{2p}\ \|b(\boldsymbol{y}\pm\boldsymbol{z})\|_b\\ &\leqslant \Lambda_2\ \sqrt{2p}\ \|\boldsymbol{y}\pm\boldsymbol{z}\|\\ &\leqslant \begin{cases} 2\sqrt{2}\,\Lambda_2\ \sqrt{p}, & \text{for } \boldsymbol{y}+\boldsymbol{z},\\ \sqrt{2}\,\Lambda_2\ \sqrt{p}\ \|\boldsymbol{y}-\boldsymbol{z}\|\,, & \text{for } \boldsymbol{y}-\boldsymbol{z}.\end{cases}\end{aligned}$$

In the second step, we used inequality (10). In the last step, we used the fact that $\|\boldsymbol{y}\| \leqslant 1$ and $\|\boldsymbol{z}\| \leqslant 1$. We thus have

$$\left(\mathbb{E}\left||a_i^\intercal \boldsymbol{y}|^2 - |a_i^\intercal \boldsymbol{z}|^2\right|^p\right)^{1/p} \leqslant 4\Lambda_2^2\ p\ \|\boldsymbol{y}-\boldsymbol{z}\|\,.$$

Therefore, $X_1, \ldots, X_m$ are independent centered subexponential random variables with subexponential norm bounded by $8\Lambda_2^2\|\boldsymbol{y}-\boldsymbol{z}\|$. Observe that

$$h_2(\boldsymbol{y}) - h_2(\boldsymbol{z}) = \frac{1}{m}\ \sum_{i=1}^m X_i.$$

Lemma C.1 with $K = 8\Lambda_2^2\|\boldsymbol{y}-\boldsymbol{z}\|$ yields

$$\mathbb{P}\left\{|h_2(\boldsymbol{y}) - h_2(\boldsymbol{z})| \geqslant \lambda\|\boldsymbol{y}-\boldsymbol{z}\|\right\} \leqslant 2\mathrm{e}^{-c_1 m\lambda^2},$$

for every $0 \leqslant \lambda \leqslant 8\Lambda_2^2$, and

$$\mathbb{P}\left\{|h_2(\boldsymbol{y}) - h_2(\boldsymbol{z})| \geqslant \lambda\|\boldsymbol{y}-\boldsymbol{z}\|\right\} \leqslant 2\mathrm{e}^{-c_2 m\lambda},$$

for every $\lambda \geqslant 8\Lambda_2^2$ with $c_1 = c/(64\Lambda_2^4)$ and $c_2 = c/(8\Lambda_2^2)$. Note that $8\Lambda_2^2 = c_2/c_1$. Hence, Assumption B is satisfied. Theorem II.2 terminates the proof.

[2]Remark that $\mathcal{S}$ is symmetric. Therefore $\mathcal{S}+\mathcal{S} = \mathcal{S}-\mathcal{S}$.

APPENDIX D
RANDOM RANK-ONE PROJECTIONS

A. Estimation of Λ_1 from Section IV-B

Let M be a fixed matrix in $\mathbb{R}^{n_1\times n_2}$ and $a \in \mathbb{R}^{n_1}$, $b \in \mathbb{R}^{n_2}$ be vectors whose entries are independent copies of the random variable N or P, defined in (22). In [16], the authors show that $a^\intercal \mathsf{M} b$ is a subexponential random variable. In this section, we estimate the parameter Λ_1 needed in Recipe 2. This estimation follows from the lemma below, whose proof starts from intermediate results presented in [16].

Before presenting the lemma, we recall the definition of the double factorial for odd numbers: $(2k-1)!! = (2k)!/(2^k\, k!)$ for all $k \geqslant 1$. Using Stirling's formula for the factorial, we obtain

$$(2k-1)!! = \frac{\sqrt{2\pi 2k}\,(2k/\mathrm{e})^{2k}\,\mathrm{e}^{\lambda_{2k}}}{2^k\sqrt{2\pi k}\,(k/\mathrm{e})^k\,\mathrm{e}^{\lambda_k}} = \sqrt{2}\,2^k (k/\mathrm{e})^k\,\mathrm{e}^{\lambda_{2k}-\lambda_k}, \tag{33}$$

where $1/(12k+1) \leqslant \lambda_k \leqslant 1/(12k)$.

Lemma D.1. *Let $a \in \mathbb{R}^{n_1}$ and $b \in \mathbb{R}^{n_2}$ be vectors whose entries are independent copies of a random variable $X \in \mathbb{R}$. Define*

$$\alpha_X := \sup_{k\geqslant 1}\left[\frac{\mathbb{E}\,X^{2k}}{(2k-1)!!}\right]^{1/(2k)}.$$

Assume that $1 \leqslant \alpha_X < \infty$. For any fixed matrix M in $\mathbb{R}^{n_1\times n_2}$ and all $p \geqslant 1$,

$$\left(\mathbb{E}\,|a^\intercal \mathsf{M} b|^p\right)^{1/p} \leqslant 2^{3/(2p)}\,\mathrm{e}^{-1}\,p\,\alpha_X^2\,\|\mathsf{M}\|_{\mathrm{Frob}}.$$

As a consequence, $a^\intercal \mathsf{M} b$ is a subexponential random variable with subexponential norm bounded by $C\,\alpha_X^2\,\|\mathsf{M}\|_{\mathrm{Frob}}$, where $C = 2^{3/2}\,\mathrm{e}^{-1} \approx 1.04 > 1$.

The proof of this lemma is given below. From the above result, we see that we only need to estimate α_X with $X = N$ and $X = P$ to estimate the constant Λ_1 in Section IV-B.

- **[Case where $X = N$]** As N is a standard random variable, we have $\mathbb{E}\,N^{2k} = (2k-1)!!$. Therefore, $\alpha_N = 1$ and $\Lambda_1 \leqslant C$ where $C > 1$ is an absolute constant.
- **[Case where $X = P$]** It is easy to notice that $\mathbb{E}\,P^{2k} = q^k/q$. Using (33) and the fact that $\mathrm{e}^{\lambda_{2k}-\lambda_k} \geqslant 1/2$, we obtain

$$\left[\frac{\mathbb{E}P^{2k}}{(2k-1)!!}\right]^{1/(2k)} \leqslant \left[\sqrt{2}\,2^{-k}\,(k/\mathrm{e})^{-k}\;q^k/q\right]^{1/(2k)} \leqslant \left[\sqrt{2}\,q^{-1}\,2^{-k}\,(k/\mathrm{e})^{-k}\right]^{1/(2k)}\;q^{1/2}.$$

  Therefore, $\alpha_P \leqslant 1.39\sqrt{q}$ and $\Lambda_1 \leqslant C\sqrt{q}$ where $C > 1$ is an absolute constant.

Proof: We start with inequality (0.20) in the supplementary material of [16], which shows that, for all $k \geqslant 1$,

$$\mathbb{E}\,|a^\intercal \mathsf{M} b|^{2k} \leqslant \alpha_X^{4k}\,[(2k-1)!!]^2\,\|\mathsf{M}\|_{\mathrm{Frob}}^{2k}.$$

Using (33) and the fact that $\mathrm{e}^{\lambda_{2k}-\lambda_k} \leqslant 1$, we obtain $(2k-1)!! \leqslant \sqrt{2}\,2^k(k/\mathrm{e})^k$. Therefore,

$$\mathbb{E}\,|a^\intercal \mathsf{M} b|^{2k} \leqslant 2\,(2\alpha_X^2/\mathrm{e})^{2k}\,k^{2k}\,\|\mathsf{M}\|_{\mathrm{Frob}}^{2k},$$

for all $k \geqslant 1$. To shorten notations, we define

$$C_{\alpha_X,\mathsf{M}} := (2\alpha_X^2/\mathrm{e})\,\|\mathsf{M}\|_{\mathrm{Frob}}.$$

We thus have

$$\mathbb{E}\,|a^\intercal \mathsf{M} b|^{2k} \leqslant 2\,C_{\alpha_X,\mathsf{M}}^{2k}\,k^{2k}.$$

We now follow exactly the interpolation technique used in the proof of Corollary 8.7 in [2]. Let $\theta \in [0,1]$. Then,

$$\begin{aligned}
&\mathbb{E}\,|a^\intercal \mathsf{M} b|^{2(k+\theta)}\\
&\quad\leqslant \left(\mathbb{E}\,|a^\intercal \mathsf{M} b|^{2k}\right)^{1-\theta}\left(\mathbb{E}\,|a^\intercal \mathsf{M} b|^{2k+2}\right)^{\theta}\\
&\quad\leqslant \left(2\,C_{\alpha_X,\mathsf{M}}^{2k}\,k^{2k}\right)^{1-\theta}\left(2\,C_{\alpha_X,\mathsf{M}}^{2(k+1)}\,(k+1)^{2(k+1)}\right)^{\theta}\\
&\quad= 2\,C_{\alpha_X,\mathsf{M}}^{2(k+\theta)}\,k^{2k(1-\theta)}\,(k+1)^{2(k+1)\theta}\\
&\quad= 2\,C_{\alpha_X,\mathsf{M}}^{2(k+\theta)}\,\left[k^{1-\theta}\,(k+1)^{\theta}\right]^{2(k+\theta)}\left(\frac{k+1}{k}\right)^{2\theta(1-\theta)}\\
&\quad\leqslant 2\,C_{\alpha_X,\mathsf{M}}^{2(k+\theta)}\,[k+\theta]^{2(k+\theta)}\left(\frac{k+1}{k}\right)^{2\theta(1-\theta)}\\
&\quad\leqslant 2\sqrt{2}\,C_{\alpha_X,\mathsf{M}}^{2(k+\theta)}\,[k+\theta]^{2(k+\theta)}
\end{aligned}$$

Hölder's inequality yielded the first step. In the second step from below, we used the fact that

$$(1-\theta)\log(k) + \theta\log(k+1) \leqslant \log\left((1-\theta)k + \theta(k+1)\right),$$

which follows from the concavity of the logarithm. In the last step, we used the facts that $(k+1)/k \leqslant 2$ and that $2\theta(1-\theta) \leqslant 1/2$. Taking $p = 2(k+\theta)$ proves the lemma for $p \geqslant 2$.

For $0 < p \leqslant 2$, we have

$$\begin{aligned}
\left(\mathbb{E}\,|a^\intercal \mathsf{M} b|^p\right)^{1/p} &\leqslant \left(\mathbb{E}\,|a^\intercal \mathsf{M} b|^2\right)^{1/2} = \|\mathsf{M}\|_{\mathrm{Frob}}\\
&\leqslant 2^{3/(2p)}\,\mathrm{e}^{-1}\,p\,\alpha_X^2\,\|\mathsf{M}\|_{\mathrm{Frob}}.
\end{aligned}$$

In the first step, we used the fact that for any random variable X, $(\mathbb{E}\,|X|^p)^{1/p} \leqslant (\mathbb{E}\,|X|^q)^{1/q}$ for $0 < p \leqslant q$. The last inequality follows from the fact that the minimum of $f(p) = 2^{3/(2p)}\mathrm{e}^{-1}\,p$ is attained at $p_0 = 3\log(2)/2$ and that $\alpha_X^2 f(p_0) \approx 1.0397\,\alpha_X^2 \geqslant 1$. The lemma is thus proved for all $p > 0$. ■

B. Bound on $\underline{\delta}_1$

Let us first introduce the following lemma which bounds $\mathbb{E}|a^\intercal x|$ for isotropic subexponential random vectors $a \in \mathbb{R}^n$ and any fixed vector $x \in \mathbb{R}^n$.

Lemma D.2. *Let $a \in \mathbb{R}^n$ be a random vector such that for any fixed vector $x \in \mathbb{R}^n$,*

$$\mathbb{E}\,|a^\intercal x|^2 = \|x\|_2^2 \quad \textit{and} \quad \|a^\intercal x\|_{\Psi_1} \leqslant C\,\|x\|_2,$$

where $C \geqslant 2$ is a constant. Then,

$$\frac{\|x\|_2}{2\mathrm{e}^3\,C\,(1+\log(C))} \leqslant \mathbb{E}\,|a^\intercal x| \leqslant \|x\|_2.$$

The proof of this lemma is given below. To estimate $\underline{\delta}_1$, we thus notice that we only need to estimate the constant C in the above lemma. For rank-one projections, Lemma D.1 indicates that $C \leqslant O(\alpha_X^2)$.

Proof: The upper bound follows from Jensen's inequality,

$$\mathbb{E}\left|a^{\intercal}x\right| \leqslant \sqrt{\mathbb{E}\left|a^{\intercal}x\right|^2} = \|x\|_2 .$$

For the lower bound, we start with Hölder's inequality. Let $\theta \in (0,1)$, we obtain

$$\begin{aligned}\mathbb{E}\left|a^{\intercal}x\right|^2 &= \mathbb{E}\left(\left|a^{\intercal}x\right|^{1-\theta}\left|a^{\intercal}x\right|^{1+\theta}\right)\\ &\leqslant \left(\mathbb{E}\left|a^{\intercal}x\right|\right)^{1-\theta}\left(\mathbb{E}\left|a^{\intercal}x\right|^{(1+\theta)/\theta}\right)^{\theta}.\end{aligned}$$

Therefore,

$$\mathbb{E}\left|a^{\intercal}x\right| \geqslant \frac{\left(\mathbb{E}\left|a^{\intercal}x\right|^2\right)^{1/(1-\theta)}}{\left(\mathbb{E}\left|a^{\intercal}x\right|^{(1+\theta)/\theta}\right)^{\theta/(1-\theta)}}$$

Using the fact that $a^{\intercal}x$ has a subexponential norm bounded by $C\|x\|_2$, we obtain

$$\mathbb{E}\left|a^{\intercal}x\right|^{(1+\theta)/\theta} \leqslant \left(\frac{1+\theta}{\theta}\right)^{(1+\theta)/\theta} C^{(1+\theta)/\theta}\|x\|_2^{(1+\theta)/\theta}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left|a^{\intercal}x\right| &\geqslant \frac{\|x\|_2^{2/(1-\theta)}}{\left[(1+\theta)/\theta\cdot C\cdot\|x\|_2\right]^{(1+\theta)/(1-\theta)}}\\ &\geqslant \frac{\|x\|_2}{\left[C\cdot(1+\theta)/\theta\right]^{(1+\theta)/(1-\theta)}}.\end{aligned}$$

The lower bound is valid for any $\theta \in (0,1)$. Therefore,

$$\mathbb{E}\left|a^{\intercal}x\right| \geqslant \frac{\|x\|_2}{\beta},$$

where

$$\beta := \inf_{\theta\in(0,1)}\left[C\cdot(1+\theta)/\theta\right]^{(1+\theta)/(1-\theta)}.$$

Let g be the function defined by

$$g(\theta) = \frac{1+\theta}{1-\theta}\log(C) + \frac{1+\theta}{1-\theta}\log\left(1+\frac{1}{\theta}\right).$$

We have $\log(\beta) = \inf_{\theta\in(0,1)} g(\theta)$. One can check that the minimum of g is attained at θ^* satisfying

$$\theta_1 := \frac{1}{5+3\log(C)} \;\leqslant\; \theta^* \;\leqslant\; \theta_2 := \frac{1}{1+2\log(C)}.$$

This gives

$$g(\theta_2) = \left[1+\frac{1}{\log(C)}\right]\left[\log(C)+\log\left(2+2\log(C)\right)\right] \geqslant g(\theta^*).$$

For $C \geqslant 2$, we have

$$3 \geqslant \frac{\log(C)+\log\left(2+2\log(C)\right)}{\log(C)}.$$

Therefore,

$$2\mathrm{e}^3\, C\,\left(1+\log(C)\right) \geqslant \beta.$$

■

REFERENCES

[1] G. Puy, M. Davies, and R. Gribonval, "Linear embeddings of low-dimensional subsets of a hilbert space to $\mathbb{R}^m$," *https://hal.inria.fr/hal-01116153*, 2015.

[2] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. Springer New York, 2013.

[3] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.

[4] A. Eftekhari and M. Wakin, "New analysis of manifold embeddings and signal recovery from compressive measurements," *Appl. Comput. Harmon. Anal.*, 2014.

[5] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: a new theory for compressed sensing," *arXiv:1302.0561*, 2013.

[6] M. Vetterli and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.

[7] N. Thaper, S. Guha, P. Indyk, and N. Koudas, "Dynamic multidimensional histograms," in *Proc. ACM SIGMOD Int. Conf. Manag. Data*, 2002, pp. 428–439.

[8] A. Bourrier, R. Gribonval, and P. Pérez, "Compressive gaussian mixture estimation," in *Proc. IEEE Conf. Acoustics, Speech and Signal Processing*, 2013, pp. 6024–6028.

[9] T. Blumensath, "Sampling and reconstructing signals from a union of linear subspaces," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4660–4671, 2011.

[10] A. Bourrier, M. E. Davies, T. Peleg, and P. Pérez, "Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7928–7946, 2014.

[11] J. C. Robinson, *Dimensions, embeddings, and attractors*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 2011.

[12] M. Talagrand, "Majorizing measures: the generic chaining," *Ann. Probab.*, vol. 24, no. 3, pp. 1049–1103, 1996.

[13] K. L. Clarkson, "Tighter bounds for random projections of manifolds," in *Proc. Symposium on Computational geometry*, 2008, pp. 39–48.

[14] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, ser. Radon Series on Computational and Applied Mathematics, M. Fornasier, Ed. De Gruyter, 2010, vol. 9, pp. 1–92.

[15] S. Dirksen, "Tail bounds via generic chaining," *Electron. J. Probab.*, vol. 20, no. 53, pp. 1–29, 2015.

[16] T. T. Cai and A. Zhang, "Rop: Matrix recovery via rank-one projections," *Ann. Statist.*, vol. 43, no. 1, pp. 102–138, 2015.

[17] Y. Chen, Y. Chi, and A. J. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 4034–4059, 2015.

[18] R. Vershynin, *Compressed Sensing, Theory and Applications*. Cambridge University Press, 2012, ch. Introduction to the non-asymptotic analysis of random matrices, pp. 210–268.

[19] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[20] U. Ayaz, S. Dirksen, and H. Rauhut, "Uniform recovery of fusion frame structured sparse signals," *Appl. Comput. Harmon. Anal.*, vol. 41, no. 2, pp. 341–361, 2016.

[21] R. Kueng, H. Rauhut, and U. Terstiege, "Low rank matrix recovery from rank one measurements," *Appl. Comput. Harmon. Anal.*, 2015.

[22] S. Bahmani and J. Romberg, "Sketching for simultaneously sparse and low-rank covariance matrices," in *Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2015, pp. 357–360.

[23] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, 2003.

[24] J. Bourgain, S. Dirksen, and J. Nelson, "Toward a unified theory of sparse dimensionality reduction in euclidean space," *Geom. Funct. Anal.*, vol. 25, no. 4, pp. 1009–1088, 2015.

[25] B. Adcock and A. C. Hansen, "A generalised sampling theorem for stable rreconstruction in arbitrary bases," *J. Fourier Anal. Appl.*, vol. 18, no. 4, pp. 685–716, 2012.

[26] S. Dirksen, "Dimensionality reduction with subgaussian matrices: a unified theory," *Found. Comput. Math.*, pp. 1–30, 2015.

[27] B. Klartag and S. Mendelson, "Empirical processes and random projections," *J. Funct. Anal.*, vol. 225, no. 1, pp. 229–245, 2005.

[28] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geom. Funct. Anal.*, vol. 17, no. 4, pp. 1248–1282, 2007.
[29] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.
[30] W. Mantzel and J. Romberg, "Compressed subspace matching on the continuum," *Inf. Inference*, 2015.
[31] S. Oymak, B. Recht, and M. Soltanolkotabi, "Isometric sketching of any set via the restricted isometry property," *arXiv:1506.03521*, 2015.
[32] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical methods for physics and engineering*. Cambridge University Press, 2006.

Gilles Puy is a researcher at Technicolor R&D France. Before, he was at INRIA in the PANAMA Research team. He obtained the Ph.D. degree in Electrical Engineering from Ecole Polytechnique Fdrale de Lausanne, Switzerland, in January 2014. He also obtained the M.Sc. degree in Electrical and Electronics Engineering from EPFL and the Engineering Diploma (M.Sc.) from Supélec, France, in 2009. He received the EPFL Chorafas Foundation award in 2013 and the EPFL Doctorate award in 2015.

His research interests are in the field of signal processing, image processing and machine learning. He works on the design of efficient sampling methods for high dimensional data (compressive sampling), low-dimensional representations, and non-linear reconstruction methods using convex and non-convex optimization. His work concerns both theory and applications, mainly in imaging problems.

Mike E. Davies (M'00SM'12F'15) holds the Jeffrey Collins Chair in Signal and Image Processing at UoE, where he also leads the Edinburgh Compressed Sensing Research Group, and is Head of the Institute for Digital Communications (IDCOM). He received an M.A. in engineering from Cambridge University in 1989 where he was awarded a Foundation Scholarship (1987), and a Ph.D. degree in nonlinear dynamics and signal processing from University College London (UCL) in 1993. He was awarded a Royal Society University Research Fellowship in 1993 and was appointed a Texas Instruments Distinguished Visiting Professor at Rice University in 2012. Currently he also leads the University Defence Research Collaboration (UDRC), a U.K. programme of signal processing research in defence in collaboration with the U.K. Defence Science and Technology Laboratory (DSTL).

His research has focused on nonlinear time series, source separation, sparse representations and compressed sensing. He has made key contributions to these fields including: the development of signal embedding theorems for nonlinear dynamical time series and the development of new theoretical and algorithmic results in Independent Component Analysis (ICA). Most recently he has pioneered the use of sparse representations as a fundamental tool in signal processing, source separation and compressed sensing. This work includes: the proposal and analysis of the highly popular Iterative Hard Thresholding algorithm for sparse reconstruction; the development of a theory for the sampling of structured sparse signal models as well as the development of a variety of new practical sampling and reconstruction techniques that exploit structured sparsity. He has also explored the application of these ideas to advanced medical imaging and RF based sensing applications. He also has an active interest in the related topics of machine learning, high-dimensional statistics and information theory.

Rémi Gribonval (FM'14) is a Senior Researcher with Inria (Rennes, France), and the scientific leader of the PANAMA research group on sparse audio processing. A former student at Ecole Normale Superieure (Paris, France), he received the Ph. D. degree in applied mathematics from Universite de Paris-IX Dauphine (Paris, France) in 1999, and his Habilitation a Diriger des Recherches in applied mathematics from Universite de Rennes I (Rennes, France) in 2007. His research focuses on mathematical signal processing, machine learning, approximation theory and statistics, with an emphasis on sparse approximation, audio source separation, dictionary learning and compressed sensing. He founded the series of international workshops SPARS on Signal Processing with Adaptive/Sparse Representations. In 2011, he was awarded the Blaise Pascal Award in Applied Mathematics and Scientific Engineering from the SMAI by the French National Academy of Sciences, and a starting investigator grant from the European Research Council.