Linear embeddings of low-dimensional subsets of a Hilbert space to R^m

Gilles Puy, Mike E. Davies, Rémi Gribonval

Linear embeddings of low-dimensional subsets of a Hilbert space to $\mathbb{R}^m$

Gilles Puy *, *Mike Davies* †, *Rémi Gribonval* *

* INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, FR.
† Institute for Digital Communications (IDCom), The University of Edinburgh, EH9 3JL, UK.

ABSTRACT

We consider the problem of embedding a low-dimensional set, $\mathcal{M}$, from an infinite-dimensional Hilbert space, $\mathcal{H}$, to a finite-dimensional space. Defining appropriate random linear projections, we propose two constructions of linear maps that have the restricted isometry property (RIP) on the secant set of $\mathcal{M}$ with high probability. The first one is optimal in the sense that it only needs a number of projections essentially proportional to the intrinsic dimension of $\mathcal{M}$ to satisfy the RIP. The second one, which is based on a variable density sampling technique, is computationally more efficient, while potentially requiring more measurements.

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

1. INTRODUCTION

The compressed sensing (CS) theory shows that the information contained in objects with a small intrinsic dimension can be "captured" by few linear and non-adaptive measurements and recovered by non-linear decoders [1]. The restricted isometry property (RIP) is at the core of many of the theoretical developments in CS. 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 $\boldsymbol{x} \in \mathcal{S}$,

$$(1-\delta)\,\|\boldsymbol{x}\|_2^2 \leqslant \|\mathsf{A}\boldsymbol{x}\|_2^2 \leqslant (1+\delta)\,\|\boldsymbol{x}\|_2^2 . \qquad (1)$$

For example, if A satisfies the RIP on the set $\mathcal{S}$ of $2s$-sparse with a sufficiently small constant δ then every s-sparse vector $\boldsymbol{x}$ is accurately and stably recovered from its noisy measurements $\boldsymbol{z} = \mathsf{A}\boldsymbol{x} + \boldsymbol{n}$ by solving the Basis Pursuit problem [1]. For a more general low-dimensional model $\Sigma \subset \mathbb{R}^n$, one needs to show that the matrix A satisfies the RIP on the secant set $\mathcal{S} = \Sigma - \Sigma$ to ensure stable recovery [2].

In this finite dimensional setting, 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 different low-dimensional models Σ in $\mathbb{R}^n$: sparse signals [1], compact Riemannian manifold [3], *etc.* In these scenarios, the RIP holds for a number of measurements m essentially proportional to the dimension of Σ.

In this work, we are interested in extending this theory to an *infinite dimensional* Hilbert space $\mathcal{H}$. We consider a low-dimensional subset, $\mathcal{M}$, of $\mathcal{H}$ and our goal is to construct a linear map L from $\mathcal{H}$ to $\mathbb{R}^m$ that stably embeds $\mathcal{M}$ in $\mathbb{R}^m$ for a number of measurements m which is: 1) essentially proportional to the intrinsic dimension of $\mathcal{M}$, and, 2) obviously independent of the (infinite) ambient dimension. These developments are important in CS to extend the theory to an analog setting [4], explore connections with the sampling of signals with finite rate of innovation [5], and also in machine learning to develop efficient methods to compute information-preserving sketches of probability distributions [6, 7].

In Section 2, we give the definition of dimension used and our assumption on the dimension of $\mathcal{M}$. In Section 3, we detail a construction of a linear map that satisfies the RIP for a number of measurement m essentially proportional to the dimension of $\mathcal{M}$. Even though this linear map has a number of measurements m reduced to its minimum, its interest can be computationally limited in certain settings. In Section 4, we propose a strategy more interesting in terms of computational cost. This strategy is based on a variable density sampling technique [4, 8] and its performance is driven by a generalised notion of coherence. Finally, in Section 5, we highlight the main key ingredients needed to construct a linear map L which satisfies the RIP. All proofs can be found in the appendices.

Notations: $\mathcal{H}$ denotes a real Hilbert space with scalar product $\langle\cdot,\cdot\rangle$ and associated norm denoted by $\|\cdot\|$. The euclidean norm in $\mathbb{R}^d$, for any $d > 0$, is denoted by $\|\cdot\|_2$.

2. THE DIMENSION OF $\mathcal{M}$

2.1. The normalised secant set

Our goal is to construct a linear map $L\colon \mathcal{H} \to \mathbb{R}^m$ that satisfies[1]

$$(1-\delta) \leqslant \|L(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2 / \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \leqslant (1+\delta) \qquad (2)$$

This work was partly funded by the European Research Council, PLEASE project (ERC-StG-2011-277906).

[1] Notice that the non-squared RIP (2) implies the squared RIP (1) with a RIP constant multiplied by 3.

for all pairs of distinct vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{M}$. As L is linear, this is equivalent to show that $\sup_{\boldsymbol{y}\in\mathcal{S}(\mathcal{M})} |\|L(\boldsymbol{y})\|_2 - 1| \leqslant \delta$, where $\mathcal{S}(\mathcal{M})$ is the normalised secant set of $\mathcal{M}$:

$$\mathcal{S}(\mathcal{M}) = \left\{ \boldsymbol{y} = \frac{\boldsymbol{x}_1 - \boldsymbol{x}_2}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|} \;\middle|\; \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{M}, \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \right\}.$$

Therefore, $\mathcal{S}(\mathcal{M})$ is our object of interest. It is the object that needs to have a small dimension. In the next section, we precise how we measure the dimension of $\mathcal{S}(\mathcal{M})$. In this work, whenever we talk about the intrinsic dimension of $\mathcal{M}$, we implicitly refer to the dimension of $\mathcal{S}(\mathcal{M})$.

We recall that the vectors $\boldsymbol{x}_1 - \boldsymbol{x}_2$ with $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{M}$ are called the chords of $\mathcal{M}$. From now on, we substitute $\mathcal{S}$ for $\mathcal{S}(\mathcal{M})$ to simplify notations.

2.2. The upper box-counting dimension

We would like to highlight that several definitions of dimension exist. The reader can refer to, *e.g.*, the monograph of Robinson [9], 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 [9], there are examples of sets for which no stable linear embedding 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 work, we use the upper box-counting dimension as our measure of dimension. This definition of the dimension is also at the centre of most of the developments in [9].

Definition 1 (Upper box-counting dimension)**.** *Let $\mathcal{X}$ be a metric space with metric denoted by ϱ and $\mathcal{A} \subseteq \mathcal{X}$. Let $N_{\mathcal{A}}(\epsilon)$ be the minimum number of closed balls of radius $\epsilon > 0$ (with respect to the metric ϱ) with centres in $\mathcal{A}$ needed to cover $\mathcal{A}$. The upper box-counting dimension of $\mathcal{A}$ is*

$$d(\mathcal{A}) := \limsup_{\epsilon\to 0} -\log(N_{\mathcal{A}}(\epsilon))/\log(\epsilon).$$

From the definition, one can remark that if $d > d(\mathcal{A})$ then there exists $\epsilon_0 > 0$ such that $N_{\mathcal{A}}(\epsilon) < \epsilon^{-d}$ for all $\epsilon < \epsilon_0^{-d}$. In this paper, $\mathcal{A}$ is always a subset of a (finite or infinite-dimensional) Hilbert space and the metric used for the definition of the upper box-counting dimension is always the norm induced by the scalar product in the ambient space.

We make the following assumption on the dimension of the normalised secant set $\mathcal{S}$ of $\mathcal{M}$.

Assumption 2. *The normalised secant set $\mathcal{S}$ of $\mathcal{M}$ has a finite upper box-counting dimension which is strictly bounded by $s > 0$: $d(\mathcal{S}) < s$. Therefore, there exist a constant $\epsilon_0 > 0$ such that $N_{\mathcal{S}}(\epsilon) < \epsilon^{-s}$ for all $\epsilon < \epsilon_0$.*

3. A TWO-STEP CONSTRUCTION OF L

We divide our first construction of the linear map L into two simple steps. The first step is an orthogonal projection onto a well-chosen subspace $\mathcal{V}$ of potentially large but finite dimension $d > 0$. This projection must preserve the norm of the vectors in $\mathcal{S}$ as well as possible. Here, we use the fact $\mathcal{S}$ has finite upper box-counting dimension to construct $\mathcal{V}$. Once this projection is performed, we are coming back to a usual embedding problem in finite ambient dimension. In this setting, we show that the embedding dimension can be reduced by multiplication with a random matrix of size $m \times d$ with m of the order of s. We highlight that the sufficient condition on m to ensure a stable embedding is independent of the potentially very large dimension d of $\mathcal{V}$.

3.1. Projection onto a finite-dimensional subspace

We fix a resolution $\epsilon \in (0, 1)$ and find a minimum cover of $\mathcal{S}$ with closed balls of radius ϵ and centres in $\mathcal{S}$. Let $T(\epsilon)$ be the set of centres of these balls. The cardinality of $T(\epsilon)$ is at most $N_{\mathcal{S}}(\epsilon)$. We denote $\mathcal{V}_\epsilon \subset \mathcal{H}$ the finite-dimensional linear subspace spanned by $T(\epsilon)$, and $P_{\mathcal{V}_\epsilon} : \mathcal{H} \to \mathcal{H}$ the orthogonal projection onto $\mathcal{V}_\epsilon$. By construction of $\mathcal{V}_\epsilon$, we have

$$\sup_{\boldsymbol{y}\in\mathcal{S}} \|\boldsymbol{y} - P_{\mathcal{V}_\epsilon}(\boldsymbol{y})\| \leqslant \sup_{\boldsymbol{y}\in\mathcal{S}} \inf_{\boldsymbol{y}_0\in T(\epsilon)} \|\boldsymbol{y} - \boldsymbol{y}_0\| \leqslant \epsilon. \quad (3)$$

The orthogonal projection onto $\mathcal{V}_\epsilon$ thus preserves the norm of all vectors in $\mathcal{S}$ with an error at most ϵ. This projection allows us to transfer our problem from infinite ambient dimension to a problem in finite ambient dimension at the cost of an error at most ϵ on the norm of the normalised chords.

Let $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$ be an orthonormal basis for $\mathcal{V}_\epsilon$ and $P_\epsilon \colon \mathcal{H} \to \mathbb{R}^d$ be the linear map $P_\epsilon(\boldsymbol{x}) = (\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle)_{1\leqslant i\leqslant d}$. We remark that

$$\|P_\epsilon(\boldsymbol{x})\|_2 = \|P_{\mathcal{V}_\epsilon}(\boldsymbol{x})\| \leqslant \|\boldsymbol{x}\|, \quad (4)$$

for every $\boldsymbol{x} \in \mathcal{H}$, and that

$$\|P_\epsilon(\boldsymbol{y})\|_2 = \|P_{\mathcal{V}_\epsilon}(\boldsymbol{y})\| \geqslant \|\boldsymbol{y}\| - \|\boldsymbol{y} - P_{\mathcal{V}_\epsilon}(\boldsymbol{y})\| \geqslant 1 - \epsilon, \quad (5)$$

for every $\boldsymbol{y} \in \mathcal{S}$. To obtain the last inequality, we used inequality (3) and the fact that $\|\boldsymbol{y}\| = 1$ for any $\boldsymbol{y} \in \mathcal{S}$. The inequalities (4) and (5) shows that P_ϵ is bi-Lipschitz on the set $\mathcal{M}$. However, the embedding dimension d is still potentially very large. Indeed, we have $d \leqslant N_{\mathcal{S}}(\epsilon) \sim \epsilon^{-s}$ while we would like an embedding dimension of the order of s. We thus need to reduce drastically the dimension of the embedding space. This reduction is achieved in the next section by a multiplication with a random flat matrix.

Let us highlight at this stage that we constructed the linear mapping P_ϵ using a covering of $\mathcal{S}$. However, the only properties that we require on P_ϵ hereafter is that its operator norm is lower and upper bounded on $\mathcal{S}$ (properties (4) and (5)). One may thus construct P_ϵ differently using other properties of $\mathcal{S}$.

3.2. Dimensionality reduction by matrix multiplication

Let $P_\epsilon(\mathcal{S}) \subset \mathbb{R}^d$ be the image of $\mathcal{S}$ under P_ϵ. Our goal in this section is to map the vectors in $P_\epsilon(\mathcal{S})$ to a space of dimension

m much smaller than ϵ^{-s}, ideally of the order of s, while preserving their norm. To show that this reduction is possible, we first prove that $P_\epsilon(\mathcal{S})$ has a small dimension.

Lemma 3. *Let $P_\epsilon(\mathcal{S})$ be the image of $\mathcal{S}$ under a linear mapping $P_\epsilon : \mathcal{H} \to \mathbb{R}^d$ satisfying (4). If Assumption 2 holds, then the upper box-counting dimension of $P_\epsilon(\mathcal{S})$ is strictly bounded by s, and $N_{P_\epsilon(\mathcal{S})}(\epsilon') < \epsilon'^{-s}$ for all $\epsilon' < \epsilon_0$.*

Proof. From (4), we see that

$$\|P_\epsilon(\boldsymbol{x}_1) - P_\epsilon(\boldsymbol{x}_2)\|_2 \leqslant \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \;\forall\, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{H}. \tag{6}$$

We proceed as in the proof of Lemma 3.3.4 in [9]. Cover $\mathcal{S}$ with $N_{\mathcal{S}}(\epsilon') < \epsilon'^{-s}$ balls of radius $\epsilon' < \epsilon_0$. The image of this cover covers $P_\epsilon(\mathcal{S})$. Consider one of these balls and let $\boldsymbol{c}$ be its centre. We denote this ball $B(\boldsymbol{c}, \epsilon')$. Using (6), we notice that $\|P_\epsilon(\boldsymbol{c}) - P_\epsilon(\boldsymbol{x})\|_2 \leqslant \epsilon'$ for all $\boldsymbol{x} \in B(\boldsymbol{c}, \epsilon')$. Therefore, the image of $B(\boldsymbol{c}, \epsilon')$ is contained in a closed ball of radius ϵ' with centre $P_\epsilon(\boldsymbol{c})$. We conclude that $N_{P_\epsilon(\mathcal{S})}(\epsilon') \leqslant N_{\mathcal{S}}(\epsilon')$ and that the upper box-counting dimension of $P_\epsilon(\mathcal{S})$ is strictly bounded by s. □

As $P_\epsilon(\mathcal{S})$ has a small dimension, it is tempting to reduce the dimension using a random matrix as done in CS. The next theorem shows that it is indeed possible with a random matrix of size $m \times d$, where m is essentially proportional to s.

Theorem 4. *Let $P_\epsilon : \mathcal{H} \to \mathbb{R}^d$ be a linear mapping satisfying (4) and (5), $\mathsf{A} \in \mathbb{R}^{m\times d}$ be a matrix whose entries are independent normal random variables with mean 0 and variance $1/m$, and $\rho \in (0,1)$.*

There exist absolute constants $D_1, D_2, D_3 > 0$ with $D_1 < 1$ such that if Assumption 2 holds, then for any $0 < \delta < \min(D_1, \epsilon_0)$, we have, with probability at least $1-\rho$,

$$1 - \epsilon - \delta \leqslant \|\mathsf{A}P_\epsilon(\boldsymbol{y})\|_2 \leqslant 1 + \delta,$$

uniformly for all $\boldsymbol{y} \in \mathcal{S}$ provided that

$$m \geqslant D_2\delta^{-2}\, \max\left\{s \log\left(D_3/\delta\right), \log\left(6/\rho\right)\right\}.$$

The same result applies (with different absolute constants D_1, D_2, D_3) if A is a matrix whose entries are independent $\pm 1/\sqrt{m}$ Bernoulli random variables, or if its rows are independent random vectors drawn from the surface of the unit sphere using the uniform distribution.

Proof. The proof, available in Appendix C, is based on a technique used by Eftekhari *et al.* in [3] to show that random Gaussian matrices satisfy the RIP on compact Riemannian manifolds of $\mathbb{R}^d$. The sufficient condition on m that they obtain is independent of the ambient dimension d. It turns out that their result generalises to any set with finite upper box-counting dimension in $\mathbb{R}^d$. □

In view of Theorem 4, we remark that $AP_\epsilon(\cdot)\colon \mathcal{H} \to \mathbb{R}^m$ has the RIP on $\mathcal{M}$ if m and ϵ are appropriately chosen.

Corollary 5. *Let $\mathsf{A} \in \mathbb{R}^{m\times d}$ be as in Theorem 4, $P_\epsilon : \mathcal{H} \to \mathbb{R}^d$ be a linear mapping satisfying (4) and (5), and $\rho \in (0,1)$.*

There exist absolute constants $D_1, D_2, D_3 > 0$ with $D_1 < 1$ such that if Assumption 2 holds, then for any $0 < \delta < \min(D_1, \epsilon_0)$, if $\epsilon \leqslant \delta/2$ and

$$m \geqslant D_2\delta^{-2}\, \max\left\{s \log\left(D_3/\delta\right), \log\left(6/\rho\right)\right\}, \tag{7}$$

then $\mathsf{A}P_\epsilon : \mathcal{H} \to \mathbb{R}^m$ satisfies the RIP (2) on $\mathcal{M}$ with constant at most δ and probability at least $1-\rho$.

3.3. Comparison with a known result in CS

Corollary 5 holds for low-dimensional signal models that satisfy Assumption 2 in an infinite-dimensional Hilbert space $\mathcal{H}$, so it also holds for signal models in $\mathbb{R}^n$ provided that their normalised secant set has finite upper box-counting dimension. Let us take one example in finite ambient space.

Consider the set $\mathcal{M}$ of k-sparse signals in $\mathbb{R}^n$. The normalised secant set $\mathcal{S}$ is the set of $2k$-sparse signals with ℓ_2-norm equal to 1. This set can be covered by at most $[3\mathrm{e}n/(2k\epsilon)]^{2k}$ balls of radius $\epsilon < 1$ [1]. Its box-counting dimension is thus $2k$. As we are in finite dimension, we can take $\mathcal{V}_\epsilon = \mathbb{R}^n$ (which implies $\epsilon = 0$) and $P_\epsilon = \mathsf{Id}$. The complete linear map L thus reduces to the matrix A. Corollary 5 shows that A has the RIP if m satisfies (7) with $s > 2k$.

In comparison, Theorem 9.2 in [1] shows that A has the RIP on $\mathcal{S}$ with constant δ and probability at least $1-\rho$ provided that

$$m \geqslant D\delta^{-2}\, \left(k \log\left(\mathrm{e}n/k\right) + \log\left(2/\rho\right)\right), \tag{8}$$

where $D > 0$ is an absolute constant. A fundamental difference seems to exist between (7) and (8). Our result seems to be independent of $\log(n/k)$ while (8) is not. This dependence does actually exist in our result but is “hidden” in the variable ϵ_0 in the statement of Corollary 5. We remind that ϵ_0 is a constant such that $[3\mathrm{e}n/(2k\epsilon)]^{2k} < \epsilon^{-s}$ for all $\epsilon < \epsilon_0$ with $s > 2k$. Let us compute ϵ_0. Writing $s = 2k + \eta$ with $\eta > 0$, we should have $(3\mathrm{e}n/(2k))^{2k} < \epsilon^{-\eta}$ for all $\epsilon < \epsilon_0$, or, equivalently, $(3\mathrm{e}n/(2k))^{2k} \leqslant \epsilon_0^{-\eta}$. We take $\epsilon_0 = 2k/(3\mathrm{e}n)$ and $\eta = 2k$. For $\delta < \epsilon_0$, we have $\log(1/\delta) > \log(3\mathrm{e}n/(2k))$. In this setting, we notice that we can slightly strengthen (8) to

$$m \geqslant D\delta^{-2}\, \left(s \log\left(1/\delta\right) + \log\left(2/\rho\right)\right).$$

This stronger condition on m is similar to ours. This shows that we recover results similar to known ones in CS.

4. AN EFFICIENT LINEAR EMBEDDING OF $\mathcal{M}$

In cases where the dimension d is very large (we recall that $d \leqslant \epsilon^{-s}$), the linear map AP_ϵ proposed in the last section is *a priori* costly in terms of computations. Indeed, one needs to compute the scalar products between the signal of interest and all the vectors of the basis $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$ before multiplication with an unstructured matrix A. In this section, we propose an computationally lighter solution where one computes the scalar products with only a small subset of the basis vectors. The method is related to the variable density sampling technique in CS [4, 8].

4.1. Construction of the linear map

To select a small subset of the basis vectors, we use a probability distribution on the discrete set $\{1, \ldots, d\}$. We represent this probability distribution by a vector $p = (p_i)_{1 \leqslant i \leqslant d} \in \mathbb{R}^d$, and the linear map L is created as follows.

1. We draw independently m indices $\Omega = \{\omega_1, \ldots, \omega_m\}$ from $\{1, \ldots, d\}$ according to p: $\mathbb{P}(\omega_k = j) = p_j$, for all $k \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, d\}$.
2. We create the sparse normalized subsampling matrix $\mathsf{R} \in \mathbb{R}^{m \times d}$ with $\mathsf{R}_{k\omega_k} = 1/\sqrt{m \|p\|_\infty}$, for all $k \in \{1, \ldots, m\}$, and 0 everywhere else.
3. We set $L := \mathsf{R}P_\epsilon$.

In order to compute $L(\boldsymbol{x})$, one just needs to compute the m scalar products $\langle \boldsymbol{a}_{\omega_k}, \boldsymbol{x} \rangle$, $k = 1, \ldots, m$. We will next see that if p is appropriately designed then L satisfies the RIP with high probability.

Before continuing, we define two quantities that characterise the interaction between p, P_ϵ, and $\mathcal{M}$. Let $\mathsf{P} \in \mathbb{R}^{d \times d}$ be the diagonal matrix with diagonal entries satisfying $\mathsf{P}_{ii} = (p_i / \|p\|_\infty)^{1/2}$. Remark that one has $\|\mathsf{P}y\|_2 \leqslant \|y\|_2$ for all $y \in \mathbb{R}^d$. The first quantity, ϵ_p, characterises how well the matrix P preserves the norm of the vectors in $P_\epsilon(\mathcal{S})$:

$$\epsilon_p := 1 - \inf_{\boldsymbol{y} \in \mathcal{S}} \left\{ \frac{\|\mathsf{P}P_\epsilon(\boldsymbol{y})\|_2}{\|P_\epsilon(\boldsymbol{y})\|_2} \right\}.$$

The second quantity, μ_p, can be viewed as a generalised definition of coherence:

$$\mu_p = \|p\|_\infty^{-1/2} \sup_{\boldsymbol{y} \in \mathcal{S}} \left\{ \frac{\|P_\epsilon(\boldsymbol{y})\|_\infty}{\|\mathsf{P}P_\epsilon(\boldsymbol{y})\|_2} \right\}. \quad (9)$$

We are now ready to state our main result.

Theorem 6. *Let $P_\epsilon : \mathcal{H} \to \mathbb{R}^d$ be a linear mapping satisfying (4) and (5), $\mathsf{R} \in \mathbb{R}^{m \times d}$ be the subsampling matrix created above and $\rho \in (0, 1)$.*

There exist absolute constants $D_1, D_2 > 0$ such that if Assumption 2 holds, then for any $0 < \delta < \min(1, \epsilon_0)$, we have, with probability at least $1 - \rho$,

$$1 - \epsilon - \epsilon_p - \delta \leqslant \|\mathsf{R}P_\epsilon(\boldsymbol{y})\|_2 \leqslant 1 + \delta,$$

uniformly for all $\boldsymbol{y} \in \mathcal{S}$ provided that

$$m \geqslant D_1 \ \max\left\{ s\mu_p^2 \delta^{-2},\ s \right\} \cdot \log\left[D_2 / (\delta \|p\|_\infty \rho) \right].$$

The proof, available in Appendix D, is similar to the proof of Theorem 4. We can now deduce sufficient conditions ensuring that $RP_\epsilon(\cdot) \colon \mathcal{H} \to \mathbb{R}^m$ has the RIP on $\mathcal{M}$.

Corollary 7. *Let $\mathsf{R} \in \mathbb{R}^{m \times d}$ be as in Theorem 4, $P_\epsilon : \mathcal{H} \to \mathbb{R}^d$ be a linear mapping satisfying (4) and (5), and $\rho \in (0, 1)$.*

There exist absolute constants $D_1, D_2 > 0$ such that if Assumption 2 holds, then for any $0 < \delta < \min(1, \epsilon_0)$, if $\epsilon \leqslant \delta/3$, $\epsilon_p \leqslant \delta/3$ and

$$m \geqslant D_1 \ \max\left\{ s\mu_p^2 \delta^{-2},\ s \right\} \cdot \log\left[D_2 / (\delta \|p\|_\infty \rho) \right], \quad (10)$$

then $\mathsf{R}P_\epsilon : \mathcal{H} \to \mathbb{R}^m$ satisfies the RIP (2) on $\mathcal{M}$ with constant at most δ and probability at least $1 - \rho$.

Condition (10) shows that one should seek to reduce the coherence parameter μ_p to optimise m. To reduce the value of μ_p, the choice of p is obviously important. However, we want to highlight that the choice of the orthonormal basis $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$ used to define P_ϵ is also very important. Indeed, this choice will influence the value of $\|P_\epsilon(\boldsymbol{y})\|_\infty$ appearing in the definition of μ_p. Therefore, if possible, one should choose an orthonormal basis that reduces $\|P_\epsilon(\boldsymbol{y})\|_\infty$ to reduce μ_p.

4.2. Comparison with a known result in CS

As in Section 3.3, we consider the set of k-sparse signals in $\mathcal{H} = \mathbb{R}^n$. We recall that $\mathcal{S}$ is the set of $2k$-sparse signals with ℓ_2-norm equal to 1. We take $\mathcal{V}_\epsilon = \mathbb{R}^n$ and $P_\epsilon = \mathsf{H} \in \mathbb{R}^{n \times n}$ the Hadamard transform. The linear map $L = \mathsf{RH} \in \mathbb{R}^{m \times n}$, is thus a random selection of m vectors of H. This case is well-known in CS. The matrix H is optimally incoherent with the identity matrix, with coherence $\|h_i\|_\infty = n^{-1/2}$ (h_i is the i^{th} row-vector of H), and the RIP is satisfied for m essentially proportional to k when the measurements are selected according to the uniform probability measure ($p_i = 1/n$) [1].

In this setting, we have $\mathsf{P} = \mathsf{Id}$ and thus $\epsilon_p = 0$. It is obvious from (10) that m should be at least proportional to s. Let us now compute μ_p. We have $\|\mathsf{PH}y\|_2 = \|y\|_2 = 1$ for all $y \in \mathcal{S}$. Therefore,

$$\begin{aligned} \mu_p = \|p\|_\infty^{-1/2} \sup_{y \in \mathcal{S}} \|\mathsf{H}y\|_\infty &= n^{1/2} \sup_{y \in \mathcal{S}} \max_{1 \leqslant i \leqslant n} |\langle h_i, y \rangle| \\ &\leqslant n^{1/2} \sup_{y \in \mathcal{S}} \max_{1 \leqslant i \leqslant n} \|h_i\|_\infty \|y\|_1 \leqslant \sup_{y \in \mathcal{S}} \|y\|_1 \leqslant \sqrt{2s}, \end{aligned}$$

and Corollary 7 requires a number of measurements essentially proportional to s^2. We do not recover an optimal result.

A careful inspection of the proof of the RIP for RH in [1] suggests that one needs to exploit additional properties on $\mathcal{S}$ than just its box-counting dimension to obtain an optimal result. Indeed, the proof is based on a clever evaluation of the covering number of $\mathcal{S}$ with an appropriate metric. The optimal result is obtained using two estimates of $N_\mathcal{S}$: a first one accurate for small radius of the balls (similar to the one used in this paper); a second one accurate for large radius. The combination of both bounds is essential to obtain m essentially proportional to s instead of s^2. It will be important in the future to determine what additional structure is needed in $\mathcal{M}$ or $\mathcal{S}$ to obtain optimal results with the variable density sampling technique.

5. KEY INGREDIENTS FOR THE DESIGN OF L

In this last section, we summarise the key steps that one should concentrate on for the design of L. For this discussion, we consider a particular case where the signal model $\mathcal{M}$ is a subset of $L_2([0,1])$. One can think of $\mathcal{M}$ as, *e.g.*, the set of k-sparse signals of the form $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{\psi}_i$ where $\|x\|_0 \leqslant k$ and $\{\boldsymbol{\psi}_i\}_{i\in\mathbb{N}}$ is the Haar wavelet basis in $L_2([0,1])$.

The first step consists in identifying a finite orthonormal basis $\Phi = \{\boldsymbol{\phi}_i\}_{1\leqslant i\leqslant d}$ (with d potentially large) that stably embeds $\mathcal{M}$. More precisely, we want (5) to hold. Intuitively, if one knows that the chords of $\mathcal{M}$ have a fast decaying spectrum, the Fourier basis $\{\sqrt{2}\cos(2\pi nt)\}_{n\in\mathbb{N}^*} \cup \{\sqrt{2}\sin(2\pi nt)\}_{n\in\mathbb{N}^*} \cup \{1\}$ seems a good candidate. The choice of d, *i.e.*, the maximum frequency that can be probed, will be determined by the decay of the spectrum of the chords.

The second step consists in constructing a good probability distribution p to select a subset of the basis vectors in Φ. In CS for MRI, a common argument is that p should favour the selection of low-frequency measurements because the energy of MR images is concentrated in this region. Let us confront this intuitive idea to our result.

According to Corollary 7, we see that p must "capture" most of the energy of the chords, *i.e.*, ϵ_p must be small. Obviously, taking the uniform discrete probability distribution, $p_i = 1/d$, ensures $\epsilon_p = 0$. However, $p_i = 1/d$ might not be the optimal choice to reduce m. The first reason is that $\|p\|_\infty$ appears in the $\log$ term of Condition (10). In the case where $d \simeq \epsilon^{-s}$, Condition (10) imposes a number of measurements at least s^2 with $p_i = 1/d$. It could be wiser to have $\epsilon_p \neq 0$ and increase the value of $\|p\|_\infty$. The second reason is that it is also essential to minimise μ_p. As the matrix P should be such that $\|\mathsf{P}P_\epsilon(\boldsymbol{y})\|_2 \simeq \|P_\epsilon(\boldsymbol{y})\|_2 \simeq 1$ for $\boldsymbol{y} \in \mathcal{S}$, we have $\mu_p^2 \simeq \sup_{\boldsymbol{y}\in\mathcal{S}} \max_i \; |\langle \boldsymbol{\phi}_i, \boldsymbol{y}\rangle|^2 / \|p\|_\infty$. Therefore, to minimise μ_p, the maximum of p should be as close as possible to $\max_i \; |\langle \boldsymbol{\phi}_i, \boldsymbol{y}\rangle|^2$. To make sure that both ϵ_p and μ_p are small, we should have p_i large whenever $|\langle \boldsymbol{\phi}_i, \boldsymbol{y}\rangle|$ is large and p_i small whenever $|\langle \boldsymbol{\phi}_i, \boldsymbol{y}\rangle|$ is small. Let us highlight that the shape of p is thus governed by the shape of the projection of the chords $\boldsymbol{x}_1 - \boldsymbol{x}_2$ on Φ and not the projection of the vectors in $\mathcal{M}$ themselves. *The measurements need to be concentrated where the energy of the normalised difference between two signals in $\mathcal{M}$ is concentrated.* In CS for MRI, one should not just look at where the energy of the images is concentrated but identify where the energy of the difference between two images in $\mathcal{M}$ is concentrated.

6. CONCLUSION

We presented two different strategies to stably embed a low-dimensional signal models from an infinite ambient dimension to $\mathbb{R}^m$. While one gives optimal results in terms of number of measurements, the other one is more appealing in terms of computational cost.

To conclude this paper, let us mention the related work of Dirksen [10] on a unified theory for dimensionality reduction with subgaussian matrices. His results also apply in infinite-dimensional Hilbert spaces. In particular, for separable Hilbert spaces, he gives an example of a linear map that satisfies the RIP with high probability but which requires the evaluation of an infinite number of scalar products. We also remark that Theorem 4 can be derived from the generic results presented in [10].

REFERENCES

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

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

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

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

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

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

[7] 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.

[8] G. Puy, P. Vandergheynst, and Y. Wiaux, "On variable density compressive sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 595–598, 2011.

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

[10] S. Dirksen, "Dimensionality reduction with subgaussian matrices: a unified theory," *arXiv:1402.3973*, 2014.

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

[12] D. P. Dubhashi and A. Panconesi, *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press, 2009.

[13] M. Talagrand, "Majorizing measures: the generic

chaining," *Ann. Probab.*, vol. 24, no. 3, pp. 1049–1103, 1996.

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

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

[16] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical methods for physics and engineering*, Cambridge University Press, 2006.

A. CONCENTRATION INEQUALITIES

In this section, we use the notions of subgaussian and subexponential random vectors/variables. We let the reader refer to, *e.g.*, [11] for more information about them. We recall here a few definitions and properties.

1. A subgaussian random variable X is a random variable that satisfies $(\mathbb{E}\,|X|^p)^{1/p} \leqslant K\sqrt{p}$ for all $p \geqslant 1$ with $K > 0$. The subgaussian norm of X, denoted by $\|X\|_{\Psi_2}$, is the smallest constant K for which the last property holds, *i.e.*, $\|X\|_{\Psi_2} = \sup_{p\geqslant 1} p^{-1/2}\,(\mathbb{E}\,|X|^p)^{1/p}$ (Definition 5.7, [11]).
2. A subexponential random variable X is a random variable that satisfies $(\mathbb{E}\,|X|^p)^{1/p} \leqslant Kp$ for all $p \geqslant 1$ with $K > 0$. The subexponential norm of X, denoted by $\|X\|_{\Psi_1}$, is the smallest constant K for which the last property holds, *i.e.*, $\|X\|_{\Psi_1} = \sup_{p\geqslant 1} p^{-1}\,(\mathbb{E}\,|X|^p)^{1/p}$ (Definition 5.13, [11]).
3. A random variable X is subgaussian if and only if X^2 is subexponential and $\|X^2\|_{\Psi_1} \leqslant 2\,\|X\|^2_{\Psi_2}$ (Lemma 5.14, [11]).
4. If X is subexponential then so is $X - \mathbb{E}X$, and we have $\|X - \mathbb{E}X\|_{\Psi_1} \leqslant 2\,\|X\|_{\Psi_1}$ (Remark 5.18, [11]).
5. A random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ is subgaussian if the one-dimensional marginals $\boldsymbol{x}^\intercal\boldsymbol{X}$ are subgaussian random variables for all $\boldsymbol{x} \in \mathbb{R}^d$. The subgaussian norm of $\boldsymbol{X}$ is defined as $\|\boldsymbol{X}\|_{\Psi_2} = \sup_{\boldsymbol{x}\in S^{d-1}} \|\boldsymbol{x}^\intercal\boldsymbol{X}\|_{\Psi_2}$, where S^{d-1} is the unit sphere in $\mathbb{R}^d$ (Definition 5.22, [11]).

We start from generic concentration inequalities that we will use in specific cases after.

Lemma 8 (Bernstein-type inequality)**.** *Let $X_1, \dots, X_m$ be independent centered subexponential random variables with subexponential norm bounded by $K > 0$. Then, for every $t \in (0, K)$, we have*

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

and for every $t \geqslant K$,

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

where $c > 0$ is an absolute constant.

The lemma above is a consequence of Proposition 5.16 in [11] and intermediate results in its proof.

Lemma 9 ([12], Theorem 1.1)**.** *Let $X = X_1 + \ldots + X_m$ be the sum of m random variables independently distributed in*

$[0, 1]$. *Then, for every $t \in (0, 1)$, we have*

$$\mathbb{P}\{X \geqslant (1+t)\mathbb{E}(X)\} \leqslant \mathrm{e}^{-ct^2\mathbb{E}(X)},$$
$$\mathbb{P}\{X \leqslant (1-t)\mathbb{E}(X)\} \leqslant \mathrm{e}^{-ct^2\mathbb{E}(X)},$$

and for every $t > 0$,

$$\mathbb{P}\{X \geqslant \mathbb{E}(X) + t\} \leqslant \mathrm{e}^{-ct^2/m},$$

where $c > 0$ is an absolute constant.

We now use Lemma 8 and Lemma 9 in respectively two scenarios of interest for dimensionality reduction. The first case below involves a multiplication with a dense random matrix as in Section 3.

Lemma 10. *Let $\mathsf{A} \in \mathbb{R}^{m\times d}$ be a matrix whose entries are independent centered normal random variables with variance $1/m$, and $x \in \mathbb{R}^d$ be a fixed vector.*

For every $\lambda \in (0, \lambda_0)$, we have

$$\mathbb{P}\{\|\mathsf{A}x\|_2 \geqslant (1+\lambda)\|x\|_2\} \leqslant \mathrm{e}^{-c_0 m\lambda^2}, \tag{11}$$
$$\mathbb{P}\{\|\mathsf{A}x\|_2 \leqslant (1-\lambda)\|x\|_2\} \leqslant \mathrm{e}^{-c_0 m\lambda^2}, \tag{12}$$

and for every $\lambda \geqslant \lambda_0$,

$$\mathbb{P}\{\|\mathsf{A}x\|_2 \geqslant (1+\lambda)\|x\|_2\} \leqslant \mathrm{e}^{-c_1 m\lambda}, \tag{13}$$

where $c_0, c_1, \lambda_0 > 0$ are absolute constants.

The same results apply (with different constants c_0, c_1, λ_0) if, e.g., A is a matrix whose entries are independent $\pm 1/\sqrt{m}$ Bernoulli random variables, or if its rows are independent random vectors drawn from the surface of the unit sphere using the uniform distribution.

Proof. Denote by $a_1, \dots, a_m \in \mathbb{R}^d$ the rows of A. We have

$$\|\mathsf{A}x\|_2^2 = \sum_{i=1}^m |a_i^\intercal x|^2,$$

where $a_1^\intercal x, \dots, a_m^\intercal x$ are independent centered subgaussian random variables. Their subgaussian norm is bounded by $C\|x\|_2/\sqrt{m}$, where $C > 0$ is an absolute constant. We also have $\mathbb{E}|a_i^\intercal x|^2 = \|x\|_2^2/m$. Therefore, $|a_1^\intercal x|^2 - \|x\|_2^2/m, \dots, |a_m^\intercal x|^2 - \|x\|_2^2/m$ are independent centered subexponential random variables with subexponential norm bounded by $C'\|x\|_2^2/m$, where $C' > 0$ is an absolute constant. Using Lemma 8 with $X_i = |a_i^\intercal x|^2 - \|x\|_2^2/m$ and $K = C'\|x\|_2^2/m$ shows that there exist absolute constants $c_0, c_1, \lambda_0 > 0$ such that, for every $\lambda \in (0, \lambda_0)$,

$$\mathbb{P}\left\{\|\mathsf{A}x\|_2^2 \geqslant (1+\lambda)\|x\|_2^2\right\} \leqslant \mathrm{e}^{-c_0 m\lambda^2},$$
$$\mathbb{P}\left\{\|\mathsf{A}x\|_2^2 \leqslant (1-\lambda)\|x\|_2^2\right\} \leqslant \mathrm{e}^{-c_0 m\lambda^2},$$

and for every $\lambda \geqslant \lambda_0$,

$$\mathbb{P}\left\{\|\mathsf{A}x\|_2^2 \geqslant (1+\lambda)\|x\|_2^2\right\} \leqslant \mathrm{e}^{-c_1 m\lambda}.$$

To finish the proof, notice that

$$\mathbb{P}\{\|\mathsf{A}x\|_2 \geqslant (1+\lambda)\|x\|_2\} \leqslant \mathbb{P}\left\{\|\mathsf{A}x\|_2^2 \geqslant (1+\lambda)\|x\|_2^2\right\}$$

and

$$\mathbb{P}\{\|\mathsf{A}x\|_2 \leqslant (1-\lambda)\|x\|_2\} \leqslant \mathbb{P}\left\{\|\mathsf{A}x\|_2^2 \leqslant (1-\lambda)\|x\|_2^2\right\}.$$

□

We now transform the bounds above in a form directly usable in Lemma 14 for the proof of Theorem 4.

Corollary 11. *Let $\mathsf{A} \in \mathbb{R}^{m\times d}$ be one of the matrices considered in Lemma 10.*

1. *For any fixed $x \in \mathbb{R}^d$ with $\|x\|_2 \leqslant 1$ and every $\lambda \in (0, \lambda_0)$, we have*
$$\mathbb{P}\{\|\mathsf{A}x\|_2 \geqslant 1+\lambda\} \leqslant \mathrm{e}^{-c_0 m\lambda^2}.$$
2. *Let $\epsilon \in (0, 1)$. For any fixed $x \in \mathbb{R}^d$ with $\|x\|_2 \geqslant 1-\epsilon$ and every $\lambda \in (0, \lambda_0)$, we have*
$$\mathbb{P}\{\|\mathsf{A}x\|_2 \leqslant 1-\lambda-\epsilon\} \leqslant \mathrm{e}^{-c_0 m\lambda^2}. \tag{14}$$
3. *Let $\epsilon' \in (0, 1)$ and $j \in \mathbb{N}$. For any fixed $x \in \mathbb{R}^d$ with $\|x\|_2 \leqslant 2^{-j+1}\epsilon'$ and every $\lambda \geqslant \lambda_1$, we have*
$$\mathbb{P}\left\{\|\mathsf{A}x\|_2 \geqslant \frac{j+1}{2^{j+2}}\lambda\epsilon'\right\} \leqslant \mathrm{e}^{-c_1 m[8^{-1}(j+1)\lambda-1]}. \tag{15}$$

The constants $\lambda_0, \lambda_1, c_0, c_1$ are absolute and $\lambda_1 > 8$.

Proof. Inequality (11) in Lemma 10 yields

$$\begin{aligned}\mathbb{P}\{\|\mathsf{A}x\|_2 \geqslant 1+\lambda\} &\leqslant \mathbb{P}\{\|\mathsf{A}x\|_2 \geqslant (1+\lambda)\|x\|_2\} \\ &\leqslant \mathrm{e}^{-c_0 m\lambda^2},\end{aligned}$$

for any $\lambda \in (0, \lambda_0)$. The first inequality is obtained by using the fact that $1 \geqslant \|x\|_2$.

Inequality (12) yields

$$\begin{aligned}\mathbb{P}\{\|\mathsf{A}x\|_2 \leqslant 1-\lambda-\epsilon\} &\leqslant \mathbb{P}\{\|\mathsf{A}x\|_2 \leqslant (1-\lambda)(1-\epsilon)\} \\ &\leqslant \mathbb{P}\{\|\mathsf{A}x\|_2 \leqslant (1-\lambda)\|x\|_2\} \\ &\leqslant \mathrm{e}^{-c_0 m\lambda^2},\end{aligned}$$

for any $\lambda \in (0, \lambda_0)$. The first inequality is obtained by noticing that $1-\lambda-\epsilon \leqslant (1-\lambda)(1-\epsilon)$, the second one by using the fact that $1-\epsilon \leqslant \|x\|_2$.

Inequality (13) yields

$$\begin{aligned}\mathbb{P}\left\{\|\mathsf{A}x\|_2 \geqslant \frac{j+1}{2^{j+2}}\lambda\epsilon'\right\} &\leqslant \mathbb{P}\left\{\|\mathsf{A}x\|_2 \geqslant \frac{j+1}{8}\lambda\|x\|_2\right\} \\ &\leqslant \mathrm{e}^{-c_1 m[8^{-1}(j+1)\lambda-1]}.\end{aligned}$$

for any $\lambda \geqslant \lambda_1 := 8\lambda_0 + 8$. In the first line, we used the fact that $2^{-j+1}\epsilon' \geqslant \|x\|_2$. □

We consider now a second scenario for dimensionality reduction. It consists in randomly selecting few entries of a vector.

As in Section 4, the vector $p \in \mathbb{R}^d$ represents a discrete probability distribution on $\{1, \ldots, d\}$ and $\mathsf{P} \in \mathbb{R}^{d \times d}$ is the diagonal matrix with diagonal entries $\mathsf{P}_{ii} = (p_i / \|p\|_\infty)^{1/2}$. The set $\Omega = \{\omega_1, \ldots, \omega_m\}$ is created by drawing independently m indices from $\{1, \ldots, d\}$ according to p. The sparse subsampling matrix $\mathsf{R} \in \mathbb{R}^{m \times d}$ has one non-zero entry on each line: $\mathsf{R}_{k\omega_k} = 1/\sqrt{m \|p\|_\infty}$, for all $k \in \{1, \ldots, m\}$.

Lemma 12. *Let $x \in \mathbb{R}^d$ be a fixed vector and define $\mu_x = \|p\|_\infty^{-1/2} \|x\|_\infty / \|\mathsf{P}x\|_2$. For every $\lambda \in (0, 1)$, we have*

$$\mathbb{P}\left\{\|\mathsf{R}x\|_2 \geqslant (1+\lambda)\|\mathsf{P}x\|_2\right\} \leqslant \mathrm{e}^{-cm\lambda^2/\mu_x^2}, \tag{16}$$

$$\mathbb{P}\left\{\|\mathsf{R}x\|_2 \leqslant (1-\lambda)\|\mathsf{P}x\|_2\right\} \leqslant \mathrm{e}^{-cm\lambda^2/\mu_x^2}, \tag{17}$$

and, for every $\lambda \geqslant 1/\|p\|_\infty$,

$$\mathbb{P}\left\{\|\mathsf{R}x\|_2 \geqslant (1+\lambda)\|x\|_2\right\} \leqslant \mathrm{e}^{-cm\lambda\|p\|_\infty}, \tag{18}$$

where $c > 0$ is an absolute constant.

Proof. We have $x_j^2 / \|x\|_\infty^2 = x_j^2 / (\mu_x^2 \|p\|_\infty \|\mathsf{P}x\|_2^2) \leqslant 1$ for all $j = 1, \ldots, d$. Therefore,

$$S = \sum_{i=1}^{m} \frac{x_{\omega_i}^2}{\mu_x^2 \|p\|_\infty \|\mathsf{P}x\|_2^2}$$

is a sum of m independent random variables bounded by 1. We have $\mathbb{E}(S) = m/\mu_x^2$ (recall that $P_{ii}^2 = p_i / \|p\|_\infty$). The first two inequalities in Lemma 9 yield

$$\mathbb{P}\left\{S \geqslant (1+\lambda)\frac{m}{\mu_x^2}\right\} \leqslant \mathrm{e}^{-c\lambda^2 m/\mu_x^2}, \text{ and}$$
$$\mathbb{P}\left\{S \leqslant (1-\lambda)\frac{m}{\mu_x^2}\right\} \leqslant \mathrm{e}^{-c\lambda^2 m/\mu_x^2}.$$

To obtain (16), we observe that

$$\begin{aligned}
&\mathbb{P}\left\{S \geqslant (1+\lambda)\frac{m}{\mu_x^2}\right\} \\
&\qquad = \mathbb{P}\left\{\frac{\mu_x^2}{m} S \geqslant (1+\lambda)\right\} \\
&\qquad = \mathbb{P}\left\{\sum_{i=1}^{m} \frac{x_{\omega_i}^2}{m\|p\|_\infty \|\mathsf{P}x\|_2^2} \geqslant (1+\lambda)\right\} \\
&\qquad = \mathbb{P}\left\{\sum_{i=1}^{m} \frac{x_{\omega_i}^2}{m\|p\|_\infty} \geqslant (1+\lambda)\|\mathsf{P}x\|_2^2\right\} \\
&\qquad = \mathbb{P}\left\{\|\mathsf{R}x\|_2^2 \geqslant (1+\lambda)\|\mathsf{P}x\|_2^2\right\},
\end{aligned}$$

and that

$$\begin{aligned}
&\mathbb{P}\left\{\|\mathsf{R}x\|_2 \geqslant (1+\lambda)\|\mathsf{P}x\|_2\right\} \\
&\qquad \leqslant \mathbb{P}\left\{\|\mathsf{R}x\|_2^2 \geqslant (1+\lambda)\|\mathsf{P}x\|_2^2\right\}.
\end{aligned}$$

The same procedure with $\mathbb{P}\{S \leqslant (1-\lambda)m/\mu_x^2\}$ yields (17).

To obtain (18), we notice that $x_j^2 \leqslant \|x\|_2^2$ for all $j = 1, \ldots, m$. Therefore,

$$\tilde{S} = \sum_{i=1}^{m} \frac{x_{\omega_i}^2}{\|x\|_2^2}$$

is a sum of m independent random variables bounded by 1. We have $\mathbb{E}(\tilde{S}) = m\|p\|_\infty \|\mathsf{P}x\|_2^2 / \|x\|_2^2$. The last inequality in Lemma 9 gives

$$\mathbb{P}\left\{\tilde{S} \geqslant \frac{m\|p\|_\infty}{\|x\|_2^2}\|\mathsf{P}x\|_2^2 + t\right\} \leqslant \mathrm{e}^{-ct^2/m}.$$

Then, we notice that

$$\begin{aligned}
&\mathbb{P}\left\{\tilde{S} \geqslant \frac{m\|p\|_\infty}{\|x\|_2^2}\|\mathsf{P}x\|_2^2 + t\right\} \\
&\qquad = \mathbb{P}\left\{\frac{\|x\|_2^2}{m\|p\|_\infty}\tilde{S} \geqslant \|\mathsf{P}x\|_2^2 + \frac{\|x\|_2^2\, t}{m\|p\|_\infty}\right\} \\
&\qquad = \mathbb{P}\left\{\sum_{i=1}^{m} \frac{x_{\omega_i}^2}{m\|p\|_\infty} \geqslant \|\mathsf{P}x\|_2^2 + \frac{\|x\|_2^2\, t}{m\|p\|_\infty}\right\} \\
&\qquad = \mathbb{P}\left\{\|\mathsf{R}x\|_2^2 \geqslant \|\mathsf{P}x\|_2^2 + \frac{\|x\|_2^2\, t}{m\|p\|_\infty}\right\}.
\end{aligned}$$

Therefore, for every $t > 0$,

$$\mathbb{P}\left\{\|\mathsf{R}x\|_2^2 \geqslant \|\mathsf{P}x\|_2^2 + \frac{\|x\|_2^2\, t}{m\|p\|_\infty}\right\} \leqslant \mathrm{e}^{-ct^2/m}$$

or, equivalently,

$$\mathbb{P}\left\{\|\mathsf{R}x\|_2^2 \geqslant \|\mathsf{P}x\|_2^2 + u\|x\|_2^2\right\} \leqslant \mathrm{e}^{-cm\|p\|_\infty^2 u^2},$$

for every $u > 0$. Finally, we use the facts that $\mathrm{e}^{-cm\|p\|_\infty^2 u^2} \leqslant \mathrm{e}^{-cm\|p\|_\infty u}$ for every $u \geqslant 1/\|p\|_\infty$ and that

$$\begin{aligned}
&\mathbb{P}\left\{\|\mathsf{R}x\|_2 \geqslant (1+u)\|x\|_2\right\} \\
&\qquad \leqslant \mathbb{P}\left\{\|\mathsf{R}x\|_2^2 \geqslant \|x\|_2^2 + u\|x\|_2^2\right\} \\
&\qquad \leqslant \mathbb{P}\left\{\|\mathsf{R}x\|_2^2 \geqslant \|\mathsf{P}x\|_2^2 + u\|x\|_2^2\right\},
\end{aligned}$$

as $\|x\|_2^2 \geqslant \|\mathsf{P}x\|_2^2$. □

We now transform the bounds above in a form directly usable in Lemma 14 for the proof of Theorem 6.

Corollary 13. *For every $x \in \mathbb{R}^d$, we associate the quantity $\mu_x = \|p\|_\infty^{-1/2} \|x\|_\infty / \|\mathsf{P}x\|_2$.*

1. *For any fixed $x \in \mathbb{R}^d$ with $\|x\|_2 \leqslant 1$ and every $\lambda \in (0, 1)$, we have*

$$\mathbb{P}\left\{\|\mathsf{R}x\|_2 \geqslant 1+\lambda\right\} \leqslant \mathrm{e}^{-cm\lambda^2/\mu_x^2}. \tag{19}$$

2. *Let $\epsilon, \epsilon_p \in (0,1)$ and assume that $\|\mathsf{P}x\|_2 \geqslant (1-\epsilon_p)\|x\|_2$. For any fixed $x \in \mathbb{R}^d$ with $\|x\|_2 \geqslant 1-\epsilon$ and every $\lambda \in (0,1)$, we have*
$$\mathbb{P}\left\{\|\mathsf{R}x\|_2 \leqslant 1-\lambda-\epsilon-\epsilon_p\right\} \leqslant \mathrm{e}^{-cm\lambda^2/\mu_x^2}. \quad (20)$$
3. *Let $\epsilon' \in (0,1)$. For any fixed $x \in \mathbb{R}^d$ with $\|x\|_2 \leqslant 2^{-j+1}\epsilon'$ and every $\lambda \geqslant 8/\|p\|_\infty + 8$, we have*
$$\mathbb{P}\left\{\|\mathsf{R}x\|_2 \geqslant \frac{j+1}{2^{j+2}}\lambda\epsilon'\right\} \leqslant \mathrm{e}^{-cm[8^{-1}(j+1)\lambda-1]\|p\|_\infty}. \quad (21)$$

The constant $c > 0$ is an absolute constant.

Proof. We follow the same procedure as for the proof of Corollary 11.

For any $x \in \mathbb{R}^d$ such that $1 \geqslant \|x\|_2$, inequality (16) in Lemma 12 yields (19) for any $\lambda \in (0,1)$ (using the fact that $\|x\|_2 \geqslant \|\mathsf{P}x\|_2$).

As $1-\epsilon-\epsilon_p \leqslant (1-\epsilon_p)\|x\|_2 \leqslant \|\mathsf{P}x\|_2$, inequality (17) yields (20) for any $\lambda \in (0,1)$.

For any $x \in \mathbb{R}^d$ such that $2^{-j+1}\epsilon' \geqslant \|x\|_2$, inequality (18) yields (21). □

B. CHAINING ARGUMENT

This section contains the main result on which the proofs of Theorem 4 and Theorem 6 are based. In the lemma below,

1. $\mathcal{A}$ is a subset of $\mathbb{R}^d$ with finite upper box-counting dimension strictly bounded by $s > 0$. Therefore, there exists $\epsilon_0' > 0$ such that its covering number satisfies $N_{\mathcal{A}}(\alpha) < \alpha^{-s}$ for all $\alpha < \epsilon_0'$;
2. ϵ' is a fixed parameter in $(0, \epsilon_0')$ (its value will chosen later on in Appendix C and Appendix D);
3. $B_j \subset \mathcal{A}$, with $j \in \mathbb{N}$, is a set of centres of closed balls of radius $2^{-j}\epsilon'$ that covers $\mathcal{A}$ with cardinality less than $2^{js}\epsilon'^{-s}$;
4. π_j is the mapping $\pi_j(y) \in \operatorname{argmin}_{z\in B_j}\|y-z\|_2$;
5. C_j is the finite set $\{(\pi_{j+1}(y), \pi_j(y)) \mid y \in \mathcal{A}\}$.

Note that $\sup_{(w,z)\in C_j}\|w-z\|_2 \leqslant 2^{-j+1}\epsilon'$.

The lemma below appears (in a slightly different form) in [3]. The proof is based on a chaining argument which is a powerful technique to obtain sharp bounds for the supremum of random processes [13–15].

Lemma 14. *Let $\mathsf{B} \in \mathbb{R}^{m\times d}$ be one of the random matrix considered in Theorem 4 or Theorem 6 and t_1, t_2, u_1, u_2 be any fixed values in $(0,\infty)$. Define $t := t_1 + t_2$ and $u := u_1 + u_2$. We have*
$$\mathbb{P}\left\{\inf_{y\in\mathcal{A}}\|\mathsf{B}y\|_2 \leqslant 1-t\right\} \leqslant \epsilon'^{-s}\max_{y_0\in B_0}\mathbb{P}\left\{\|\mathsf{B}y_0\|_2 \leqslant 1-t_1\right\} + \sum_{j\geqslant 0}\frac{2^{2(j+1)s}}{\epsilon'^{2s}}\max_{(w,z)\in C_j}\mathbb{P}\left\{\|\mathsf{B}(w-z)\|_2 \geqslant \frac{j+1}{2^{j+2}}t_2\right\}, \quad (22)$$
and
$$\mathbb{P}\left\{\sup_{y\in\mathcal{A}}\|\mathsf{B}y\|_2 \geqslant 1+u\right\} \leqslant \epsilon'^{-s}\max_{y_0\in B_0}\mathbb{P}\left\{\|\mathsf{B}y_0\|_2 \geqslant 1+u_1\right\} + \sum_{j\geqslant 0}\frac{2^{2(j+1)s}}{\epsilon'^{2s}}\max_{(w,z)\in C_j}\mathbb{P}\left\{\|\mathsf{B}(w-z)\|_2 \geqslant \frac{j+1}{2^{j+2}}u_2\right\}. \quad (23)$$

Proof. We follow the same procedure as in [3].

We remark that any vector $y \in \mathcal{A}$ can be written
$$y = \pi_0(y) + S(y) \text{ where } S(y) = \sum_{j\geqslant 0}(\pi_{j+1}(y) - \pi_j(y)).$$

We start by proving (22). Noticing that
$$\inf_{y\in\mathcal{A}}\|\mathsf{B}\pi_0(y)\|_2 - \sup_{y\in\mathcal{A}}\|\mathsf{B}S(y)\|_2 \leqslant \inf_{y\in\mathcal{A}}\|\mathsf{B}y\|_2,$$
we obtain
$$\begin{aligned}&\mathbb{P}\left\{\inf_{y\in\mathcal{A}}\|\mathsf{B}y\|_2 \leqslant 1-t\right\}\\ &\leqslant \mathbb{P}\left\{\inf_{y\in\mathcal{A}}\|\mathsf{B}\pi_0(y)\|_2 - \sup_{y\in\mathcal{A}}\|\mathsf{B}S(y)\|_2 \leqslant 1-t\right\}\\ &= \mathbb{P}\left\{\inf_{y\in\mathcal{A}}\|\mathsf{B}\pi_0(y)\|_2 - \sup_{y\in\mathcal{A}}\|\mathsf{B}S(y)\|_2 \leqslant 1-t_1-t_2\right\}.\end{aligned}$$

The union bound then yields
$$\begin{aligned}&\mathbb{P}\left\{\inf_{y\in\mathcal{A}}\|\mathsf{B}\pi_0(y)\|_2 - \sup_{y\in\mathcal{A}}\|\mathsf{B}S(y)\|_2 \leqslant 1-t_1-t_2\right\}\\ &\leqslant \mathbb{P}\left\{\inf_{y\in\mathcal{A}}\|\mathsf{B}\pi_0(y)\|_2 \leqslant 1-t_1\right\} + \mathbb{P}\left\{\sup_{y\in\mathcal{A}}\|\mathsf{B}S(y)\|_2 \geqslant t_2\right\}\\ &\leqslant \mathbb{P}\left\{\min_{y_0\in B_0}\|\mathsf{B}y_0\|_2 \leqslant 1-t_1\right\} + \mathbb{P}\left\{\sup_{y\in\mathcal{A}}\|\mathsf{B}S(y)\|_2 \geqslant t_2\right\}. \quad (24)\end{aligned}$$

In the last step we used the fact that
$$\inf_{y\in\mathcal{A}}\|\mathsf{B}\pi_0(y)\|_2 = \min_{y_0\in B_0}\|\mathsf{B}y_0\|_2.$$

Using again the union bound, we see that
$$\mathbb{P}\left\{\min_{y_0\in B_0}\|\mathsf{B}y_0\|_2 \leqslant 1-t_1\right\} \leqslant \epsilon'^{-s}\max_{y_0\in B_0}\mathbb{P}\left\{\|\mathsf{B}y_0\|_2 \leqslant 1-t_1\right\}.$$

We continue by bounding the second term on the right-hand side of (24) to get a probability easier to evaluate. The triangle inequality yields

$$\mathbb{P}\left\{\sup_{y\in\mathcal{A}} \|\mathsf{B}S(y)\|_2 \geqslant t_2\right\}$$
$$\leqslant \mathbb{P}\left\{\sum_{j\geqslant 0}\sup_{y\in\mathcal{A}} \|\mathsf{B}\left(\pi_{j+1}(y)-\pi_j(y)\right)\|_2 \geqslant t_2\right\}.$$

Then, noticing that $\sum_{j\geqslant 0} 2^{-j-2}(j+1) = 1$ (see, *e.g.*, Section 4.2.3 in [16]), we obtain

$$\mathbb{P}\left\{\sum_{j\geqslant 0}\sup_{y\in\mathcal{A}} \|\mathsf{B}\left(\pi_{j+1}(y)-\pi_j(y)\right)\|_2 \geqslant t_2\right\}$$
$$= \mathbb{P}\left\{\sum_{j\geqslant 0}\sup_{y\in\mathcal{A}} \|\mathsf{B}\left(\pi_{j+1}(y)-\pi_j(y)\right)\|_2 \geqslant \sum_{j\geqslant 0}\frac{j+1}{2^{j+2}}\ t_2\right\}$$
$$\leqslant \sum_{j\geqslant 0}\mathbb{P}\left\{\sup_{y\in\mathcal{A}} \|\mathsf{B}\left(\pi_{j+1}(y)-\pi_j(y)\right)\|_2 \geqslant \frac{j+1}{2^{j+2}}t_2\right\}$$

In the last step we used the union bound. We continue by observing that

$$\sup_{y\in\mathcal{A}} \|\mathsf{B}\left(\pi_{j+1}(y)-\pi_j(y)\right)\|_2 = \max_{(w,z)\in C_j} \|\mathsf{B}\left(w-z\right)\|_2.$$

The cardinality of C_j is bounded by $2^{2(j+1)s}\epsilon'^{-2s}$. Using once more the union bound, we have

$$\mathbb{P}\left\{\max_{(w,z)\in C_j} \|\mathsf{B}\left(w-z\right)\|_2 \geqslant \frac{j+1}{2^{j+2}}t_2\right\}$$
$$\leqslant \frac{2^{2(j+1)s}}{\epsilon'^{2s}}\max_{(w,z)\in C_j}\mathbb{P}\left\{\|\mathsf{B}\left(w-z\right)\|_2 \geqslant \frac{j+1}{2^{j+2}}t_2\right\}.$$

In total, we proved that

$$\mathbb{P}\left\{\inf_{y\in\mathcal{A}} \|\mathsf{B}y\|_2 \leqslant 1-t\right\}$$
$$\leqslant \epsilon'^{-s}\max_{y_0\in B_0}\mathbb{P}\left\{\|\mathsf{B}y_0\|_2 \leqslant 1-t_1\right\}$$
$$+\sum_{j\geqslant 0}\frac{2^{2(j+1)s}}{\epsilon'^{2s}}\max_{(w,z)\in C_j}\mathbb{P}\left\{\|\mathsf{B}\left(w-z\right)\|_2 \geqslant \frac{j+1}{2^{j+2}}t_2\right\}.$$

Repeating the same procedure for $\sup_{y\in\mathcal{A}}\|\mathsf{B}y\|_2$, one obtains (23).

□

C. PROOF OF THEOREM 4

In this section, we prove Theorem 4.

Lemma 3 shows that $N_{P_\epsilon(\mathcal{S})}(\alpha) < \alpha^{-s}$ for all $\alpha < \epsilon_0$. As A is a random matrix, we are thus in the setting of Lemma 14 with $\mathcal{A} = P_\epsilon(\mathcal{S})$, $\mathsf{B} = \mathsf{A}$, and $\epsilon_0' = \epsilon_0$. Let ϵ', B_0, C_j with $j \in \mathbb{N}$ be as in Lemma 14. We will fix the value of ϵ' later on. We start by bounding the probabilities appearing in the right-hand side (rhs) of (22).

Let $\delta \in (0,1)$ and take $t_1 = \epsilon + \delta/2$ and $t_2 = \delta/2$ in Lemma 14. We recall that

$$\sup_{y\in P_\epsilon(\mathcal{S})}\|y\|_2 \leqslant 1 \quad \text{and} \quad \inf_{y\in P_\epsilon(\mathcal{S})}\|y\|_2 \geqslant 1-\epsilon,$$

(see (4) and (5)) and that $B_0 \subset P_\epsilon(\mathcal{S})$. If $\delta < \min(2\lambda_0, 1)$, then (14) in Lemma 11 shows that

$$\max_{y_0\in B_0}\mathbb{P}\left\{\|\mathsf{A}y_0\| \leqslant 1-\epsilon-\delta/2\right\} \leqslant \mathrm{e}^{-c_0 m\delta^2/4}. \tag{25}$$

We have thus a bound on the first probability appearing in the rhs of (22). To bound the second one, we recall that $\sup_{(w,z)\in C_j}\|w-z\|_2 \leqslant 2^{-j+1}\epsilon'$. We can thus use the bound (15) of Lemma 11. In this lemma, we take $\lambda = 8\lambda_1$ and $\epsilon' = \delta/(16\lambda_1)$ so that $\lambda\epsilon' = \delta/2 = t_2$. It is obvious that $\lambda > \lambda_1$. If $\delta < \epsilon_0$ then we have[2] $\epsilon' < \epsilon_0$ as required in Lemma 14. Therefore,

$$\max_{(w,z)\in C_j}\mathbb{P}\left\{\|\mathsf{A}\left(w-z\right)\| \geqslant \frac{j+1}{2^{j+2}}\frac{\delta}{2}\right\}$$
$$\leqslant \mathrm{e}^{-c_1 m[(j+1)\lambda_1-1]}. \tag{26}$$

Combining (25) and (26), we arrive at

$$\mathbb{P}\left\{\inf_{y\in P_\epsilon(\mathcal{S})}\|\mathsf{A}y\|_2 \leqslant 1-t\right\} \leqslant \left(\frac{16\lambda_1}{\delta}\right)^s \mathrm{e}^{-c_0 m\delta^2/4}$$
$$+\left(\frac{32\lambda_1}{\delta}\right)^{2s}\mathrm{e}^{-c_1 m(\lambda_1-1)}\sum_{j\geqslant 0}2^{2js}\,\mathrm{e}^{-jc_1 m\lambda_1}.$$

We observe that if

$$m \geqslant \frac{\log(2)}{c_1\lambda_1}(2s+1), \tag{27}$$

then $\sum_{j\geqslant 0}2^{2js}\,\mathrm{e}^{-jc_1 m\lambda_1} \leqslant 2$. In addition, if

$$m \geqslant \frac{4s}{c_1(\lambda_1-1)}\log\left(\frac{32\lambda_1}{\delta}\right), \tag{28}$$

then

$$\left(\frac{32\lambda_1}{\delta}\right)^{2s}\mathrm{e}^{-c_1 m(\lambda_1-1)} \leqslant \mathrm{e}^{-c_1 m(\lambda_1-1)/2}.$$

Finally, further imposing that

$$m \geqslant \frac{8s}{c_0\delta^2}\log\left(\frac{16\lambda_1}{\delta}\right) \tag{29}$$

yields

$$\mathbb{P}\left\{\inf_{y\in P_\epsilon(\mathcal{S})}\|\mathsf{A}y\|_2 \leqslant 1-\epsilon-\delta\right\}$$
$$\leqslant \mathrm{e}^{-c_0 m\delta^2/8} + 2\mathrm{e}^{-c_1 m(\lambda_1-1)/2}.$$

[2]Recall that $\lambda_1 > 8$.

Define $c_2 = \min(c_0/8, c_1(\lambda_1 - 1)/2)$, then

$$\mathbb{P}\left\{\inf_{y \in P_\epsilon(\mathcal{S})} \|\mathsf{A}y\|_2 \leqslant 1 - \epsilon - \delta\right\} \leqslant\ 3\mathrm{e}^{-c_2 m \delta^2}.$$

The right-hand side is less than $\rho/2$ provided that

$$m \geqslant \frac{1}{c_2\delta^2} \log\left(\frac{6}{\rho}\right). \tag{30}$$

To conclude, we remark that there exists absolute constants $D_1, D_2 > 0$ such that if

$$m \geqslant \frac{D_1}{\delta^2}\ \max\left\{s \log\left(\frac{D_2}{\delta}\right), \log\left(\frac{6}{\rho}\right)\right\},$$

then (27), (28), (29), and (30) hold. Notice that the final constraint on δ is $\delta < \min(1, 2\lambda_0, \epsilon_0)$.

Repeating the same procedure to bound

$$\mathbb{P}\left\{\sup_{y \in P_\epsilon(\mathcal{S})} \|\mathsf{A}y\|_2 \geqslant 1 + u_1 + u_2\right\},$$

with $u_1 = \delta/2$ and $u_2 = \delta/2$, terminates the proof.

D. PROOF OF THEOREM 6

In this section, we prove Theorem 6.

We use Lemma 14 with $\mathcal{A} = P_\epsilon(\mathcal{S})$, $\mathsf{B} = \mathsf{R}$, and $\epsilon_0' = \epsilon_0$. Let ϵ', B_0, C_j with $j \in \mathbb{N}$ be as in Lemma 14. We start by bounding the probabilities appearing on the rhs of (22).

Let $\delta \in (0, 1)$ and take $t_1 = \epsilon + \epsilon_p + \delta/2$ and $t_2 = \delta/2$ in Lemma 14. Inequality (20) yields

$$\max_{y_0 \in B_0} \mathbb{P}\left\{\|\mathsf{R}y_0\| \leqslant 1 - \epsilon - \epsilon_p - \delta/2\right\} \leqslant \mathrm{e}^{-cm\delta^2/(4\mu_p^2)}, \tag{31}$$

where μ_p is defined in (9). We have thus a bound on the first probability appearing on the rhs of (22). To bound the second one, we recall that $\sup_{(w,z)\in C_j} \|w - z\|_2 \leqslant 2^{-j+1}\epsilon'$. We can thus use the bound (21) of Lemma 11. In this lemma, we take $\lambda = 16/\|p\|_\infty$ and $\epsilon' = \delta\|p\|_\infty/32$ so that $\lambda\epsilon' = \delta/2 = t_2$. One can check that $\lambda > 8/\|p\|_\infty + 8$. If $\delta < \epsilon_0$ then we have $\epsilon' < \epsilon_0$ as required in Lemma 14. Therefore,

$$\max_{(w,z)\in C_j} \mathbb{P}\left\{\|\mathsf{R}\,(w - z)\| \geqslant \frac{j+1}{2^{j+2}}\frac{\delta}{2}\right\} \leqslant\ \mathrm{e}^{-cm[2(j+1)/\|p\|_\infty - 1]\|p\|_\infty}. \tag{32}$$

Gathering (31) and (32), we arrive at

$$\mathbb{P}\left\{\inf_{y \in P_\epsilon(\mathcal{S})} \|\mathsf{R}y\|_2 \leqslant 1 - t\right\} \leqslant\ \left(\frac{32}{\delta\|p\|_\infty}\right)^s\ \mathrm{e}^{-cm\delta^2/(4\mu_p^2)} + \left(\frac{64}{\delta\|p\|_\infty}\right)^{2s}\ \mathrm{e}^{-cm(2-\|p\|_\infty)} \sum_{j\geqslant 0} 2^{2js}\ \mathrm{e}^{-2cmj}.$$

Then we observe that if

$$m \geqslant \frac{\log(2)}{2c}(2s + 1), \tag{33}$$

then $\sum_{j\geqslant 0} 2^{2js}\ \mathrm{e}^{-2jcm} \leqslant 2$. In addition, if

$$m \geqslant \frac{1}{c}\left(2s \log\left(\frac{64}{\delta\|p\|_\infty}\right) + \log\left(\frac{1}{\rho}\right)\right), \tag{34}$$

then

$$\left(\frac{64}{\delta\|p\|_\infty}\right)^{2s}\ \mathrm{e}^{-cm(2-\|p\|_\infty)} \leqslant \rho.$$

We used the fact that $\|p\|_\infty \leqslant 1$. Finally, also imposing that

$$m \geqslant \frac{4\mu_p^2}{c\delta^2}\left(s \log\left(\frac{32}{\delta\|p\|_\infty}\right) + \log\left(\frac{1}{\rho}\right)\right) \tag{35}$$

yields

$$\mathbb{P}\left\{\inf_{y \in P_\epsilon(\mathcal{S})} \|\mathsf{A}y\|_2 \leqslant 1 - \epsilon - \epsilon_p - \delta\right\} \leqslant\ 3\rho.$$

To conclude, we remark that there exists absolute constants $D_1, D_2 > 0$ such that if

$$m \geqslant D_1\ \max\left\{\frac{s\mu_p^2}{\delta^2}, s\right\} \log\left(\frac{D_2}{\delta\|p\|_\infty\rho}\right),$$

then (33), (34), and (35) hold. Notice that the final constraint on δ is $\delta < \min(1, \epsilon_0)$.

Repeating the same procedure for

$$\mathbb{P}\left\{\sup_{y \in P_\epsilon(\mathcal{S})} \|\mathsf{R}y\|_2 \geqslant 1 + u\right\},$$

with $u_1 = \delta/2$ and $u_2 = \delta/2$, terminates the proof.