

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

Gilles Puy <sup>\*</sup>, Mike Davies <sup>†</sup>, Rémi Gribonval <sup>\*</sup>

<sup>\*</sup> INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, FR.

<sup>†</sup> 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  $A \in \mathbb{R}^{m \times n}$  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  $\mathbf{x} \in \mathcal{S}$ ,

$$(1 - \delta) \|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta) \|\mathbf{x}\|_2^2. \quad (1)$$

For example, if  $A$  satisfies the RIP on the set  $\mathcal{S}$  of  $2s$ -sparse with a sufficiently small constant  $\delta$  then every  $s$ -sparse vector  $\mathbf{x}$  is accurately and stably recovered from its noisy measurements  $\mathbf{z} = A\mathbf{x} + \mathbf{n}$  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  $\Sigma$  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  $\Sigma$ .

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 our technical report [9].

**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  $L: \mathcal{H} \rightarrow \mathbb{R}^m$  that satisfies<sup>1</sup>

$$(1 - \delta) \leq \|L(\mathbf{x}_1 - \mathbf{x}_2)\|_2 / \|\mathbf{x}_1 - \mathbf{x}_2\| \leq (1 + \delta) \quad (2)$$

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

<sup>1</sup>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  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}$ . As  $L$  is linear, this is equivalent to show that  $\sup_{\mathbf{y} \in \mathcal{S}(\mathcal{M})} \left| \|L(\mathbf{y})\|_2 - 1 \right| \leq \delta$ , where  $\mathcal{S}(\mathcal{M})$  is the normalised secant set of  $\mathcal{M}$ :

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

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  $\mathbf{x}_1 - \mathbf{x}_2$  with  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}$  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 [10], 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 [10], 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 [10].

**Definition 1** (Upper box-counting dimension). *Let  $\mathcal{X}$  be a metric space with metric denoted by  $\varrho$  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  $\varrho$ ) with centres in  $\mathcal{A}$  needed to cover  $\mathcal{A}$ . The upper box-counting dimension of  $\mathcal{A}$  is*

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

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  $\epsilon$  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  $P_{\mathcal{V}_{\epsilon}} : \mathcal{H} \rightarrow \mathcal{H}$  the orthogonal projection onto  $\mathcal{V}_{\epsilon}$ . By construction of  $\mathcal{V}_{\epsilon}$ , we have

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

The orthogonal projection onto  $\mathcal{V}_{\epsilon}$  thus preserves the norm of all vectors in  $\mathcal{S}$  with an error at most  $\epsilon$ . 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  $\epsilon$  on the norm of the normalised chords.

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

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

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

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

for every  $\mathbf{y} \in \mathcal{S}$ . To obtain the last inequality, we used inequality (3) and the fact that  $\|\mathbf{y}\| = 1$  for any  $\mathbf{y} \in \mathcal{S}$ . The inequalities (4) and (5) shows that  $P_{\epsilon}$  is bi-Lipschitz on the set  $\mathcal{M}$ . However, the embedding dimension  $d$  is still potentially very large. Indeed, we have  $d \leq N_{\mathcal{S}}(\epsilon) \sim \epsilon^{-s}$  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  $P_{\epsilon}$  using a covering of  $\mathcal{S}$ . However, the only properties that we require on  $P_{\epsilon}$  hereafter is that its operator norm is lower and upper bounded on  $\mathcal{S}$  (properties (4) and (5)). One may thus construct  $P_{\epsilon}$  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  $P_{\epsilon}$ . Our goal in this section is to map the vectors in  $P_{\epsilon}(\mathcal{S})$  to a space of dimension

$m$  much smaller than  $\epsilon^{-s}$ , 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  $P_\epsilon : \mathcal{H} \rightarrow \mathbb{R}^d$  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

$$\|P_\epsilon(\mathbf{x}_1) - P_\epsilon(\mathbf{x}_2)\|_2 \leq \|\mathbf{x}_1 - \mathbf{x}_2\| \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{H}. \quad (6)$$

We proceed as in the proof of Lemma 3.3.4 in [10]. Cover  $\mathcal{S}$  with  $N_{\mathcal{S}}(\epsilon')$  balls of radius  $\epsilon' < \epsilon_0$ . The image of this cover covers  $P_\epsilon(\mathcal{S})$ . Consider one of these balls and let  $\mathbf{c}$  be its centre. We denote this ball  $B(\mathbf{c}, \epsilon')$ . Using (6), we notice that  $\|P_\epsilon(\mathbf{c}) - P_\epsilon(\mathbf{x})\|_2 \leq \epsilon'$  for all  $\mathbf{x} \in B(\mathbf{c}, \epsilon')$ . Therefore, the image of  $B(\mathbf{c}, \epsilon')$  is contained in a closed ball of radius  $\epsilon'$  with centre  $P_\epsilon(\mathbf{c})$ . We conclude that  $N_{P_\epsilon(\mathcal{S})}(\epsilon') \leq N_{\mathcal{S}}(\epsilon')$  and that the upper box-counting dimension of  $P_\epsilon(\mathcal{S})$  is strictly bounded by  $s$ .  $\square$

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  $P_\epsilon : \mathcal{H} \rightarrow \mathbb{R}^d$  be a linear mapping satisfying (4) and (5),  $\mathbf{A} \in \mathbb{R}^{m \times d}$  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  $1 - \rho$ ,*

$$1 - \epsilon - \delta \leq \|\mathbf{A}P_\epsilon(\mathbf{y})\|_2 \leq 1 + \delta,$$

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

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

*The same result applies (with different absolute constants  $D_1, D_2, D_3$ ) if  $\mathbf{A}$  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 the appendix of our technical report [9], 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  $\mathbb{R}^d$ .  $\square$

In view of Theorem 4, we remark that  $\mathbf{A}P_\epsilon(\cdot) : \mathcal{H} \rightarrow \mathbb{R}^m$  has the RIP on  $\mathcal{M}$  if  $m$  and  $\epsilon$  are appropriately chosen.

**Corollary 5.** *Let  $\mathbf{A} \in \mathbb{R}^{m \times d}$  be as in Theorem 4,  $P_\epsilon : \mathcal{H} \rightarrow \mathbb{R}^d$  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  $\epsilon \leq \delta/2$  and*

$$m \geq D_2 \delta^{-2} \max\{s \log(D_3/\delta), \log(6/\rho)\}, \quad (7)$$

*then  $\mathbf{A}P_\epsilon : \mathcal{H} \rightarrow \mathbb{R}^m$  satisfies the RIP (2) on  $\mathcal{M}$  with constant at most  $\delta$  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  $\ell_2$ -norm equal to 1. This set can be covered by at most  $[3en/(2k\epsilon)]^{2k}$  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  $P_\epsilon = \text{Id}$ . The complete linear map  $L$  thus reduces to the matrix  $\mathbf{A}$ . Corollary 5 shows that  $\mathbf{A}$  has the RIP if  $m$  satisfies (7) with  $s > 2k$ .

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

$$m \geq D \delta^{-2} (k \log(en/k) + \log(2/\rho)), \quad (8)$$

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  $\epsilon_0$  in the statement of Corollary 5. We remind that  $\epsilon_0$  is a constant such that  $[3en/(2k\epsilon)]^{2k} < \epsilon^{-s}$  for all  $\epsilon < \epsilon_0$  with  $s > 2k$ . Let us compute  $\epsilon_0$ . Writing  $s = 2k + \eta$  with  $\eta > 0$ , we should have  $(3en/(2k))^{2k} < \epsilon^{-\eta}$  for all  $\epsilon < \epsilon_0$ , or, equivalently,  $(3en/(2k))^{2k} \leq \epsilon_0^{-\eta}$ . We take  $\epsilon_0 = 2k/(3en)$  and  $\eta = 2k$ . For  $\delta < \epsilon_0$ , we have  $\log(1/\delta) > \log(3en/(2k))$ . In this setting, we notice that we can slightly strengthen (8) to

$$m \geq D \delta^{-2} (s \log(1/\delta) + \log(2/\rho)).$$

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  $d \leq \epsilon^{-s}$ ), the linear map  $\mathbf{A}P_\epsilon$  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  $(\mathbf{a}_1, \dots, \mathbf{a}_d)$  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  $\{1, \dots, d\}$ . We represent this probability distribution by a vector  $p = (p_i)_{1 \leq i \leq d} \in \mathbb{R}^d$ , and the linear map  $L$  is created as follows.

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

In order to compute  $L(\mathbf{x})$ , one just needs to compute the  $m$  scalar products  $\langle \mathbf{a}_{\omega_k}, \mathbf{x} \rangle$ ,  $k = 1, \dots, m$ . 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  $p$ ,  $P_\epsilon$ , and  $\mathcal{M}$ . Let  $P \in \mathbb{R}^{d \times d}$  be the diagonal matrix with diagonal entries satisfying  $P_{ii} = (p_i/\|p\|_\infty)^{1/2}$ . Remark that one has  $\|Py\|_2 \leq \|y\|_2$  for all  $y \in \mathbb{R}^d$ . The first quantity,  $\epsilon_p$ , characterises how well the matrix  $P$  preserves the norm of the vectors in  $P_\epsilon(\mathcal{S})$ :

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

The second quantity,  $\mu_p$ , can be viewed as a generalised definition of coherence:

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

We are now ready to state our main result.

**Theorem 6.** *Let  $P_\epsilon : \mathcal{H} \rightarrow \mathbb{R}^d$  be a linear mapping satisfying (4) and (5),  $R \in \mathbb{R}^{m \times d}$  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  $1 - \rho$ ,*

$$1 - \epsilon - \epsilon_p - \delta \leq \|RP_\epsilon(\mathbf{y})\|_2 \leq 1 + \delta,$$

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

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

The proof, available in the appendix of our technical report [9], is similar to the proof of Theorem 4. We can now deduce sufficient conditions ensuring that  $RP_\epsilon(\cdot) : \mathcal{H} \rightarrow \mathbb{R}^m$  has the RIP on  $\mathcal{M}$ .

**Corollary 7.** *Let  $R \in \mathbb{R}^{m \times d}$  be as in Theorem 4,  $P_\epsilon : \mathcal{H} \rightarrow \mathbb{R}^d$  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  $\epsilon \leq \delta/3$ ,  $\epsilon_p \leq \delta/3$  and*

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

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

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

#### 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  $\ell_2$ -norm equal to 1. We take  $\mathcal{V}_\epsilon = \mathbb{R}^n$  and  $P_\epsilon = H \in \mathbb{R}^{n \times n}$  the Hadamard transform. The linear map  $L = RH \in \mathbb{R}^{m \times n}$ , 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  $i^{\text{th}}$  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  $P = \text{Id}$  and thus  $\epsilon_p = 0$ . It is obvious from (10) that  $m$  should be at least proportional to  $s$ . Let us now compute  $\mu_p$ . We have  $\|PHy\|_2 = \|y\|_2 = 1$  for all  $y \in \mathcal{S}$ . Therefore,

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

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  $\mathbf{x} = \sum_{i=1}^n x_i \psi_i$  where  $\|\mathbf{x}\|_0 \leq k$  and  $\{\psi_i\}_{i \in \mathbb{N}}$  is the Haar wavelet basis in  $L_2([0, 1])$ .

The first step consists in identifying a finite orthonormal basis  $\Phi = \{\phi_i\}_{1 \leq i \leq d}$  (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  $\{\sqrt{2} \cos(2\pi n t)\}_{n \in \mathbb{N}^*} \cup \{\sqrt{2} \sin(2\pi n t)\}_{n \in \mathbb{N}^*} \cup \{1\}$  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  $\Phi$ . 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.,  $\epsilon_p$  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  $\mu_p$ . As the matrix  $P$  should be such that  $\|PP_\epsilon(\mathbf{y})\|_2 \simeq \|P_\epsilon(\mathbf{y})\|_2 \simeq 1$  for  $\mathbf{y} \in \mathcal{S}$ , we have  $\mu_p^2 \simeq \sup_{\mathbf{y} \in \mathcal{S}} \max_i |\langle \phi_i, \mathbf{y} \rangle|^2 / \|p\|_\infty$ . Therefore, to minimise  $\mu_p$ , the maximum of  $p$  should be as close as possible to  $\max_i |\langle \phi_i, \mathbf{y} \rangle|^2$ . To make sure that both  $\epsilon_p$  and  $\mu_p$  are small, we should have  $p_i$  large whenever  $|\langle \phi_i, \mathbf{y} \rangle|$  is large and  $p_i$  small whenever  $|\langle \phi_i, \mathbf{y} \rangle|$  is small. Let us highlight that the shape of  $p$  is thus governed by the shape of the projection of the chords  $\mathbf{x}_1 - \mathbf{x}_2$  on  $\Phi$  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 [11] 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 [11].

## 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] 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.
- [10] J. C. Robinson, *Dimensions, embeddings, and attractors*, Cambridge Tracts in Mathematics. Cambridge University Press, 2011.
- [11] S. Dirksen, “Dimensionality reduction with subgaussian matrices: a unified theory,” *arXiv:1402.3973*, 2014.