UNIVERSITÀ DEGLI STUDI DI UDINE

Dipartimento di Scienze Matematiche, Informatiche e Fisiche

Corso di Laurea Magistrale in Matematica

Tesi di Laurea Magistrale

ON EXTENDING THE JOHNSON-LINDENSTRAUSS LEMMA TO HILBERT SPACES

Relatore:
DIMITRI BREDA

Laureanda:
MICHELA PALERMO

Correlatore:
ENRICO BOZZO

ANNO ACCADEMICO 2023-2024

Contents

Abstract **v**

Sommario **vii**

Introduction **ix**

1 Notations and preliminary notions of probability theory **1**
1.1 Notations . . . 1
1.2 Preliminary notions of probability theory . . . 2
1.2.1 Basic notions . . . 2
1.2.2 Expected value and variance . . . 4
1.2.3 Gaussian random variables . . . 5
1.2.4 Chi-squared random variables . . . 6

2 The Johnson-Lindenstrauss Lemma **7**
2.1 Dimensional reduction *almost* preserving distances . . . 7
2.2 The proof . . . 8
2.3 Experiments . . . 13

3 Extending the Johnson-Lindenstrauss Lemma **15**
3.1 *CS*, *RIP* and more . . . 15
3.2 Preliminaries . . . 16
3.3 Infinite-dimensional and probabilistic RIP . . . 19
3.4 A random linear map that satisfies the RIP . . . 29
3.4.1 First step: from infinite to finite dimension . . . 29
3.4.2 Second step: into a lower dimension . . . 32
3.5 A finite-dimensional example: the RIP on *sparse* vectors . . . 39
3.6 Finite cardinality case . . . 42

4 A first test in infinite dimension **51**
4.1 The *JL-projection* . . . 51
4.2 The *F-projection* . . . 55

4.3 Numerical experiments 55

Conclusions and future developments **61**

A Useful definitions and results **65**
A.1 Auxiliary results 65
A.2 Legendre polynomials and Fourier-Legendre series 69

B MATLAB codes **71**
B.1 JL 71
B.2 JL_inf_Leg 72
B.3 Comp_JL_F_Leg 72

Bibliography **73**

Abstract

In this thesis we explore both the theoretical and experimental aspects of the Johnson-Lindenstrauss Lemma, a foundational result in dimensionality reduction, with a particular focus on its extension to infinite-dimensional spaces. Our work includes experiments on dimensionality reduction with *almost* preservation of distances, starting from both finite and infinite-dimensional initial spaces. We show that it is possible to *randomly project* a set of elements from an arbitrary Hilbert space into $\mathbb{R}^m$, with m significantly smaller than the lower bound prescribed by the theory, *almost* preserving the distances between these elements.

Key contributions include the comparison between the finite and infinite-dimensional versions of the Johnson-Lindenstrauss Lemma and the development of MATLAB-based experiments. We also present a confrontation between the infinite-dimensional Johnson-Lindenstrauss *projection* and the classical *projection* of Fourier-Legendre.

The thesis is organised as follows: Chapter 1 introduces the necessary probabilistic concepts; Chapter 2 provides a probabilistic proof of the Johnson-Lindenstrauss Lemma and presents experimental results about dimensionality reduction from finite-dimensional initial space; Chapter 3 extends the Lemma to infinite dimension and compares the two versions; Chapter 4 discusses experiments on dimensionality reduction with *almost* distance preservation from the infinite-dimensional Hilbert space $L^2([0,1],\mathbb{R})$. The appendices include auxiliary definitions, results, and MATLAB codes used in the experiments, available for further exploration.

Sommario

In questa tesi esploriamo gli aspetti teorici e sperimentali del Lemma di Johnson-Lindenstrauss, un risultato fondamentale nel campo della riduzione dimensionale, con un particolare focus sulla sua estensione agli spazi di dimensione infinita. Il nostro lavoro include esperimenti sulla riduzione dimensionale con *quasi* conservazione delle distanze, partendo da spazi iniziali di dimensione sia finita che infinita. Con questi, mostriamo che è possibile *proiettare casualmente* un insieme di elementi da uno spazio di Hilbert arbitrario in $\mathbb{R}^m$, con m significativamente minore del *lower bound* imposto dalla teoria, *quasi* conservando le distanze tra questi elementi.

I principali contributi di questo lavoro includono il confronto tra le versioni finita e infinita-dimensionale del Lemma di Johnson-Lindenstrauss e lo sviluppo di esperimenti eseguiti con il software MATLAB. Presentiamo inoltre un confronto tra la *proiezione* di Johnson-Lindenstrauss applicata a elementi di dimensione infinita e la *proiezione* classica di Fourier-Legendre.

La tesi è organizzata come segue: il Capitolo 1 introduce i concetti probabilistici necessari alla comprensione di ciò che segue; il Capitolo 2 fornisce una dimostrazione probabilistica del Lemma di Johnson-Lindenstrauss e presenta risultati sperimentali sulla riduzione dimensionale con *quasi* conservazione delle distanze da spazi iniziali di *grande* dimensione, ma finita; il Capitolo 3 estende il Lemma al caso di spazi iniziali di dimensione infinita e confronta le due versioni; il Capitolo 4 tratta di esperimenti sulla riduzione dimensionale con *quasi* conservazione delle distanze a partire dallo spazio di Hilbert infinito-dimensionale $L^2([0,1],\mathbb{R})$. Le appendici includono definizioni e risultati ausiliari, e i codici MATLAB utilizzati negli esperimenti.

Introduction

The Johnson-Lindenstrauss Lemma, first formulated and proved in 1984 by the mathematicians William B. Johnson (born 1944) and Joram Lindenstrauss (1936-2012) in their article *Extensions of Lipschitz mappings into a Hilbert space* [1], is a theoretical result that has gained significant practical importance in recent years [2].

Today, this Lemma – named as such for historical reasons – is widely used in problems involving large-dimensional matrices and vectors (data), such as in machine learning, compressed sensing (see Section 3.1), computing *SVD* or other matrix decompositions [3, 4, 5], or approximating solutions to linear regression problems [6]. This is because it addresses dimensional reduction, a fundamental tool in these fields.

The Lemma asserts that a set of n points in a high-dimensional space can be embedded into a space of dimension $O\left(\frac{\log n}{\delta^2}\right)$ in such a way that the distance between any pair of points is preserved but for a relative error of $1 \pm \delta$, with $\delta \in (0, 1)$, i.e., it is *almost* preserved.

This result caught our attention not only because of its importance and usefulness but also due to the probabilistic techniques used in its proof [6, 7], and its strong connection with compressed sensing and the property known as *restricted isometry property (RIP)* [8] (see Chapter 3 later on).

Compressed sensing (CS) is a technique for efficiently acquiring and reconstructing signals by solving underdetermined linear systems [9, 10]. It relies on the fact that a *sparse* signal – a vector with mostly zero components – can be recovered from a relatively small number of, typically random, measurements [11]. This can be proven by demonstrating that the RIP [8] holds for certain sets [11].

The RIP concerns the *almost* preservation of norms (or, equivalently, distances) and has a close relationship with the Johnson-Lindenstrauss Lemma [2]. While in some cases verifying this property can be challenging, in others it can be shown to hold *with high probability* using the Lemma. Furthermore, the RIP is crucial, in its infinite-dimensional version, in the context of infinite-dimensional compressed sensing [12, 13], which is why it is addressed in [8], a key reference for the infinite-dimensional results of this thesis.

Rather than viewing the RIP as a tool for proving key results in (infinite) compressed sensing theory, we approach it as a property related to dimensional reduction with *almost* preservation of distances. We believe that investigating the use of this property in the field, e.g., of retarded functional differential equations [14, 15, 16] may lead to consider alternative and novel directions in analysing stability problems for this class of infinite-dimensional dynamical systems (see the conclusions for further discussion).

In this thesis, motivated also by this objective, we explore theoretical aspects of both the Johnson-Lindenstrauss Lemma and the infinite-dimensional RIP and present experiments on (finite and infinite) dimensional reduction with *almost* preservation of distances.

Specifically, our contributions related to the Johnson-Lindenstrauss Lemma include the results from the first experiments (Section 2.3) that we present. These experiments, carried out using the MATLAB code *JL* (Appendix B.1), concern the *embedding* of vectors in a space with a dimension smaller than the lower bound prescribed by the Lemma. The results show that the Johnson-Lindenstrauss *projection* behaves very effectively, even in low-dimensional spaces.

Additionally, we derive from the results of [8] an infinite-dimensional and probabilistic version of the Johnson-Lindenstrauss Lemma and explore a comparison between the parameters of this infinite-dimensional result and those of the finite-dimensional Lemma.

We also conduct experiments, using the MATLAB codes *JL_inf_Leg* and *Comp_JL_F_Leg* (Appendix B.2 and B.3), on dimensional reduction from infinite to finite-dimensional spaces, with *almost* preservation of distances. In particular, we consider the first n Legendre polynomials, *project* them in $\mathbb{R}^m$ using a random map that satisfies the conditions of the infinite-dimensional version of the Johnson-Lindenstrauss Lemma, and compare the results in terms of *almost* preservation of distances with those obtained by *projecting* each polynomial onto the first m coefficients of its Fourier-Legendre series. In these experiments as well, we work with the dimension m of the target space being significantly lower than the theoretical lower bound and obtain pleasingly positive results, along with insights that could further expand the theory, which we will discuss in the conclusions.

The thesis is structured as follows: the first chapter covers preliminary concepts of probability theory, which are essential for understanding the proofs and concepts presented later. In Chapter 2, we focus on the Johnson-Lindenstrauss Lemma, providing a probabilistic proof and presenting the experiments mentioned earlier. Chapter 3 is the most substantial of this work. In this chapter, we provide constructions and infinite-dimensional results, which, as previously mentioned, we compare with the finite-dimensional case, and which are fundamental for proceeding with the experiments on infinite-dimensional reduction with *almost* preservation of distances, detailed in Chapter 4. Finally, the appendices include auxiliary definitions and results (Appendix A), as well as the MATLAB codes (Appendix B), which we developed to conduct all the experiments, and are also freely available at `https://cdlab.uniud.it/software`. Some concluding remarks are collected in a final section that closes this work.

Chapter 1

Notations and preliminary notions of probability theory

In this chapter we give a list of the notations we use in this work, introduce some preliminary concepts of probability theory and state some fundamental results that are useful in the following. Most of the theorems in this chapter are basic results and have simple and easily available demonstrations [17], so we do not give their proofs.

1.1 Notations

Throughout this work we use the following notations:

- $(H, \langle \cdot, \cdot \rangle)$ is a Hilbert space with induced norm $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$
- $\mathcal{S} := \{x \in H : \|x\| = 1\}$ is the unit sphere in H
- if $H = \mathbb{R}^n$, the scalar product $\langle \cdot, \cdot \rangle$ is the canonical one – which induces the *Euclidean norm* $\| \cdot \|_2$ – defined for all $x, y \in \mathbb{R}^n$ by $\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i$
- the norm $\| \cdot \|_p$, for $p \in [1, +\infty)$, is the *p-norm* in $\mathbb{R}^n$, defined for all $x \in \mathbb{R}^n$ by $\|x\|_p := \left(\sum_{i=1}^{n} x_i^p \right)^{\frac{1}{p}}$
- $| \cdot |$ indicates the cardinality of a set
- given a set $\mathcal{E}$, $\mathcal{E}^c$ is its complementary set
- log is the natural logarithm.

1.2 Preliminary notions of probability theory

In this thesis we are dealing with different elements of probability theory, so in this section we start with the basics and we aim to introduce the concepts of probability theory that are useful for understanding the following chapters. Most of this section is taken from [18].

1.2.1 Basic notions

Definition 1.1. A *random experiment* is a process that can be repeated countless times under the same conditions, and with a well-defined set of results.

Definition 1.2. The set Ω containing the possible outcomes ω of a random experiment is called *sample space*, and its subsets $\mathcal{E}$ are called *events*.

The following definitions are needed to obtain a probability space from the set Ω.

Definition 1.3. A *σ-algebra* $\mathcal{F}$ on a non-empty set Ω is a collection of subset of Ω such that

- $\Omega \in \mathcal{F}$
- if $\mathcal{E} \in \mathcal{F}$, than $\mathcal{E}^c \in \mathcal{F}$
- if I is any subset of $\mathbb{N}$ and $\mathcal{E}_i \in \mathcal{F}$ for all $i \in I$, then $\bigcup\limits_{i \in I} \mathcal{E}_i \in \mathcal{F}$.

Example 1.4. An example of a σ-algebra on $\mathbb{R}$ is the *Borel σ-algebra*, which is the smallest σ-algebra containing all open (or, equivalently, closed) intervals of $\mathbb{R}$.

Definition 1.5. Given a couple $(\Omega, \mathcal{F})$, where Ω is a non-empty set and $\mathcal{F}$ is a σ-algebra on Ω, a *probability measure* on this space is a real-valued set function $P : \mathcal{F} \to [0, 1]$ such that

- $P(\mathcal{E}) \geq 0$ for all $\mathcal{E} \in \mathcal{F}$
- $P(\Omega) = 1$
- if I is any subset of $\mathbb{N}$ and $\{\mathcal{E}_i\}_{i \in I} \subseteq \mathcal{F}$ is a set of disjoint events in $\mathcal{F}$, then $P\big(\bigcup\limits_{i \in I} \mathcal{E}_i\big) = \sum\limits_{i \in I} P(\mathcal{E}_i)$.

Proposition 1.6 below contains results that follow easily from the definition of probability measure.

Proposition 1.6. *Let Ω be a non-empty set, $\mathcal{F}$ be a σ-algebra on Ω and P be a probability measure on $(\Omega, \mathcal{F})$. The following statements apply:*

1. $P(\emptyset) = 0$;
2. *for all* $\mathcal{E} \in \mathcal{F}$, $P(\mathcal{E}^c) = 1 - P(\mathcal{E})$;
3. *for all* $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{F}$, *if* $\mathcal{E}_1 \subseteq \mathcal{E}_2$ *then* $P(\mathcal{E}_1) \leq P(\mathcal{E}_2)$;
4. *for all* $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{F}$, $P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2) - P(\mathcal{E}_1 \cap \mathcal{E}_2)$;
5. *for all* $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{F}$, $P(\mathcal{E}_1) = P(\mathcal{E}_1 \cap \mathcal{E}_2) + P(\mathcal{E}_1 \cap \mathcal{E}_2^c)$;
6. *for all* $\mathcal{E}_1, \ldots, \mathcal{E}_n \in \mathcal{F}$, $P\big(\bigcup_{i=1}^{n} \mathcal{E}_i\big) \leq \sum_{i=1}^{n} P(\mathcal{E}_i)$.

We can now give the definition of a *probability space.*

Definition 1.7. A *probability space* is $(\Omega, \mathcal{F}, P)$ where Ω is a non-empty set, $\mathcal{F}$ is a σ-algebra on Ω and P is a probability measure on $(\Omega, \mathcal{F})$.

Other basic concepts in probability theory are those of *random variable* and *random vector.*

Definition 1.8. Given $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, a map $X : \Omega_1 \to \Omega_2$ is *measurable* with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$ if for all $\mathcal{E} \in \mathcal{F}_2$, $X^{-1}(\mathcal{E}) \in \mathcal{F}_1$, where $X^{-1}(\mathcal{E}) \coloneqq \{\omega \in \Omega_1 : X(\omega) \in \mathcal{E}\}$.

Definition 1.9. Given $(\Omega, \mathcal{F})$, a *(real) random variable* is a map $X : \Omega \to \mathbb{R}$ measurable with respect to $\mathcal{F}$ and the Borel σ-algebra on $\mathbb{R}$.

Definition 1.10. A *(real) random vector* is a vector whose components are random variables on the same probability space.

A random variable can be identified with the probability measure it induces on $\mathbb{R}$, and two (or more) random variables inducing the same probability measure on $\mathbb{R}$ can be thought of as the same random variable.

Definition 1.11. The *probability distribution* of a random variable X from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B})$ is the probability measure P_X induced by X on $\mathbb{R}$, defined for all $B \in \mathcal{B}$ by $P_X(B) := P(X^{-1}(B))$.

Definition 1.12. Two (or more) random variables defined on the same probability space are said to be *identically distributed* if they have the same probability distribution.

In this work we are interested in *continuous* random variables, namely those that have a continuous probability distribution, so in the following we simply refer to random variables, but we mean continuous random variables.

To conclude this paragraph, we give the definition of *independence* of two or more random variables and present a result that is useful in Chapter 3.

Definition 1.13. Two (or more) random variables defined on the same probability space are said to be *independent* if the values assumed by one do not influence in any way those assumed by the other.

In the lemma below and in the rest of this thesis we use the notation $P(X \in S)$ to denote $P(\{\omega \in \Omega : X(\omega) \in S\})$, where X is a random variable on $(\Omega, \mathcal{F}, P)$ and S is a set in $\mathcal{B}$, and by $P(X \geq a)$ and $P(X \leq a)$ we mean $P(X \in [a, +\infty))$ and $P(X \in (-\infty, a])$ respectively, of course using strict inequalities in the case of the open intervals.

Lemma 1.14. *Let X and Y be two random variables on $(\Omega, \mathcal{F}, P)$ and let $a, b \in \mathbb{R}$. Then $P(X + Y \geq a + b) \leq P(X \geq a) + P(Y \geq b)$.*

Proof. For 5. of Proposition 1.6 we have that

$$\begin{aligned} &P(X + Y \geq a + b) = \\ &\quad = P(X + Y \geq a + b \cap X \geq a) + P(X + Y \geq a + b \cap X < a) \\ &\quad = P(X + Y \geq a + b \cap X \geq a) + P(Y \geq b + a - X \cap a - X > 0). \end{aligned}$$

Then, thanks to 3. of Proposition 1.6, we have that

$$P(X + Y \geq a + b \cap X \geq a) \leq P(X \geq a),$$

and that

$$P(Y \geq b + a - X \cap a - X > 0) \leq P(Y \geq b),$$

which gives the thesis. □

1.2.2 Expected value and variance

The *expected value* and the *variance* of a random variable are deterministic values that express, respectively, the average value that the random variable assumes and the square of the deviation of the values assumed by the random variable from that value. Let us give the formal definitions.

Definition 1.15. The *probability density function* p_X of X is the real, non-negative and Lebesgue integrable function $p_X : \mathbb{R} \to \mathbb{R}$ which gives the probability distribution of X as

$$P_X(B) = \int_B p_X(x)dx, \text{ for all } B \in \mathcal{B}.$$

Definition 1.16. The *expected value*, also called *expectation*, of a random variable X is defined as

$$\mathbb{E}(X) := \int_{\mathbb{R}} x \cdot p_X(x)dx.$$

If $\mathbb{E}(X) = 0$, we say that X is a *centred* random variable.

Definition 1.17. The *variance* of a random variable X is defined as

$$Var(X) := \mathbb{E}([X - \mathbb{E}(X)]^2).$$

These two indices have many important properties. Let us state some of them without demonstration.

Proposition 1.18. *The expected value is linear, i.e., given n random variables $X_1, \dots, X_n$ and $a_1, \dots, a_n, b \in \mathbb{R}$, it holds that*

$$\begin{aligned}\mathbb{E}(a_1X_1 + \cdots + a_nX_n + b) &= \mathbb{E}(a_1X_1) + \cdots + \mathbb{E}(a_nX_n) + b \\ &= a_1\mathbb{E}(X_1) + \cdots + a_n\mathbb{E}(X_n) + b.\end{aligned}$$

Proposition 1.19. *The variance of a random variable X satisfies*

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

Also, given $a_1, \dots, a_n, b \in \mathbb{R}$, if the random variables $X_1, \dots, X_n$ are independent it holds that

$$\begin{aligned}Var(a_1X_1 + \cdots + a_nX_n + b) &= Var(a_1X_1) + \cdots + Var(a_nX_n) \\ &= a_1^2Var(X_1) + \cdots + a_n^2Var(X_n).\end{aligned}$$

1.2.3 Gaussian random variables

An important class of random variables with many valuable properties is that of *Gaussian random variables*, also called *normal random variables.*

Definition 1.20. A random variable Z has *standard normal distribution* if its probability density function is p_Z defined for all $z \in \mathbb{R}$ by

$$p_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}.$$

Briefly we write $Z \sim N(0,1)$ and we say that Z is a *standard normal random variable.*

Definition 1.21. A random variable X has *normal distribution* or *Gaussian distribution* if $X = \mu + \sigma Z$, with $\mu \in \mathbb{R}$, $\sigma > 0$ and $Z \sim N(0, 1)$. Briefly we write $X \sim N(\mu, \sigma^2)$ and say that X is a *normal* or *Gaussian random variable.*

We state without proof two fundamental results about these random variables that are useful in Chapter 2 and Chapter 3.

Proposition 1.22 ([18]). *If $X \sim N(\mu, \sigma^2)$, then*

- $\mathbb{E}(X) = \mu$;
- $Var(X) = \sigma^2$;
- *for all* $x \in \mathbb{R}$, $p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Proposition 1.23 ([18]). *Linear combinations of independent Gaussian random variables are Gaussian random variables.*
In particular, if $X_1 \sim N(\mu_1, \sigma_1^2), \dots, X_n \sim N(\mu_n, \sigma_n^2)$ are independent, and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, then

$$Y = \lambda_1 X_1 + \cdots + \lambda_n X_n \sim N(\mu, \sigma^2),$$

where $\mu = \lambda_1\mu_1 + \cdots + \lambda_n\mu_n$ and $\sigma^2 = \lambda_1^2\sigma_1^2 + \cdots + \lambda_n^2\sigma_n^2$.

1.2.4 Chi-squared random variables

Another class of random variables that are relevant to this work is that of *chi-squared random variables.* Let us therefore give their definition and state some of their properties.

Definition 1.24. A random variable W has *chi-squared distribution with n degrees of freedom* if $W = \sum_{i=1}^{n} Z_i^2$, where $Z_1, \dots, Z_n$ are independent standard normal random variables. Briefly we write $W \sim \chi_n^2$ and we say that W is a *chi-squared random variable with n degrees of freedom.*

Proposition 1.25 ([18]). *If $W \sim \chi_n^2$, then $\mathbb{E}(W) = n$.*

Proposition 1.26 ([19]). *If $W \sim \chi_n^2$, then for all $t \in (0, 1)$,*

$$P\left(\left|\frac{1}{n}W - 1\right| \geq t\right) \leq 2e^{-nt^2/8}.$$

Chapter 2

The Johnson-Lindenstrauss Lemma

In this chapter, after a brief introduction aimed at showing the importance of the Johnson-Lindenstrauss Lemma, we give the proof of the Lemma and present some experiments to highlight the power of this result in practice.

Most of the chapter presents results from the literature [6, 7], while the experiments in Section 2.3 and the MATLAB codes used for them – reported in Appendix B and at `https://cdlab.uniud.it/software` – are part of our own work.

For ease of reading, in the remaining of the thesis we refer to the Johnson-Lindenstrauss Lemma or related results by the acronym *JL*.

2.1 Dimensional reduction *almost* preserving distances

In many fields, such as, e.g., genetics, engineering and statistics, we deal with large amounts of data that we need to compare, use in regression analysis, or feed into machine learning algorithms [6].

In general, in order to represent them, it is useful to use vectors, often belonging to large spaces, because they have to contain a huge amount of data relating to the same object or case study. However, most algorithms that work with these vectors tend to become prohibitively expensive very quickly as the size of the vectors increases. To ensure that these algorithms remain viable, it is therefore desirable to reduce the dimension of the data in a way that preserves its relevant structure [20]. An intuitive way to score this goal is to represent the data with smaller dimensional vectors that retain some properties of the original vectors.

The JL Lemma is about exactly this: it states that a set of n points in a Euclidean space of large dimension can be embedded in a space of dimension $O\left(\frac{\log n}{\delta^2}\right)$ such that the distance between each pair of points varies by at most a factor of $1 \pm \delta$, with $\delta \in (0, 1)$, i.e., it is *almost* conserved.

This particularly powerful and influential theoretical result was first formulated and demonstrated in 1984 by the mathematicians William B. Johnson (born 1944) and Joram Lindenstrauss (1936-2012) for a very different purpose. They called it *Lemma*, name that has remained unchanged for historical reasons, because it was useful in proving the first important theorem in the article *Extensions of Lipschitz mappings into a Hilbert space* [1].

Today there is renewed interest in this result because it has gained great relevance in the context of *modern* problems such as (non-linear) dimension reduction in databases, compressed sensing (see Section 3.1), computing *SVD* or other decompositions of large matrices [3, 4, 5], or approximating solutions to linear regression problems [6]. Of course, not only has this Lemma been revisited in applications, but in recent years many theoretical results related to the JL Lemma have been proved. As an example, we point to [21], a negative result that in 2017 extended the work of N. Alon [22] by proving that there exist sets of n points that can *not* be embedded in a space of dimension less than $O\left(\frac{\log n}{\delta^2}\right)$, such that the distances between the points vary by at most a factor of $1 \pm \delta$.

2.2 The proof

As mentioned in [6], the original proof by Johnson and Lindenstrauss is probabilistic, showing that the *projection* of the n-point subset onto a (random) subspace of $O\left(\frac{\log n}{\delta^2}\right)$ dimension only changes the interpoint distances by $1 \pm \delta$ with positive probability.

The proof given in this work is simpler, but also relies on elementary probabilistic techniques. The body of the proof we present in this section is taken from [6], while its conclusion, which is not that clear in [6], is based on that of [7].

Theorem 2.1 (JL Lemma [6]). *Given $z_1, \dots, z_n \in \mathbb{R}^d$, for any $\delta \in (0, 1)$ there exists a linear map $f : \mathbb{R}^d \to \mathbb{R}^m$ such that*

$$(1-\delta)\|z_i - z_j\|_2 \leq \|f(z_i) - f(z_j)\|_2 \leq (1+\delta)\|z_i - z_j\|_2$$

holds for all $i, j = 1, \dots, n$, provided that

$$m \geq \frac{16}{\delta^2} \log n.$$

Remark 2.2. The JL Lemma makes no assumptions about the relationship between the dimension d of the initial space and m of the arrival space, and in fact, the result holds regardless of their relationship. However, it is clear that if $m \geq d$, the Lemma has no meaningful application.

The idea of the proof is to choose f randomly: let A be an $m \times d$ random matrix with independent entries $a_{i,j} \sim N(0,1)$, and let $\Pi := \frac{1}{\sqrt{m}}A$.
We define f as

$$\begin{aligned} f : \mathbb{R}^d &\to \mathbb{R}^m \\ x \mapsto f(x) &:= \Pi x = \frac{1}{\sqrt{m}}Ax. \end{aligned} \tag{2.1}$$

With this choice we have that for all $x \in \mathbb{R}^d$, $f(x)$ is a rescaling of Ax, which is a random vector with entries that are independent Gaussian random variables, being the linear combination of independent standard normal random variables with coefficients the entries of the vector x.
More formally, we have that for all $x \in \mathbb{R}^d$ the independent entries of Ax are

$$(Ax)_i \sim N\left(0, \|x\|_2^2\right),$$

and so that the independent entries of $f(x)$ are

$$(f(x))_i \sim N\left(0, \frac{\|x\|_2^2}{m}\right).$$

In fact, thanks to Propositions 1.18 and 1.19, for all $i = 1, \ldots, m$ we have that

- $\mathbb{E}((Ax)_i) = \mathbb{E}\left(\sum_{j=1}^d a_{i,j}x_j\right) = \sum_{j=1}^d \mathbb{E}(a_{i,j}x_j) = \sum_{j=1}^d x_j\mathbb{E}(a_{i,j}) = 0$
- $Var((Ax)_i) = Var\left(\sum_{j=1}^d a_{i,j}x_j\right) = \sum_{j=1}^d Var(a_{i,j}x_j) = \sum_{j=1}^d x_j^2 Var(a_{i,j}) = \|x\|_2^2.$

To prove that this choice of f satisfies the thesis of the Lemma, we proceed in three steps: first we state and prove the (δ, ξ)-*JL property* (Lemma 2.3 below), then we derive from it the *Probabilistic JL Lemma* (Theorem 2.4 below), and finally we show that the latter implies the JL Lemma.

Lemma 2.3 ((δ, ξ)-JL property [6])**.** *Given $\Pi \in \mathbb{R}^{m \times d}$ with independent entries $\frac{1}{\sqrt{m}}N(0,1)$, then for any $\xi \in (0,1)$ and $\delta \in (0,1)$, for all $x \in \mathbb{R}^d$*

$$(1-\delta)\|x\|_2^2 \leq \|\Pi x\|_2^2 \leq (1+\delta)\|x\|_2^2 \tag{2.2}$$

holds, provided that

$$m \geq \frac{8}{\delta^2} \log \frac{2}{\xi},$$

with probability at least $1 - \xi$.

Proof. Let us fix $x \in \mathbb{R}^d$ and consider the random vector $w = Ax$ with independent entries $w_i \sim N(0, \|x\|_2^2)$.

Since for all $i = 1, \dots, m$, $w_i = \|x\|_2 Z_i$ with $Z_i \sim N(0, 1)$, we have

$$\begin{aligned} \|\Pi x\|_2^2 &= \left\| \frac{1}{\sqrt{m}} Ax \right\|_2^2 = \left\| \frac{1}{\sqrt{m}} w \right\|_2^2 = \frac{1}{m} \sum_{i=1}^m w_i^2 \\ &= \frac{1}{m} \sum_{i=1}^m \|x\|_2^2 Z_i^2 = \frac{\|x\|_2^2}{m} \sum_{i=1}^m Z_i^2 \sim \frac{\|x\|_2^2}{m} \chi_m^2. \end{aligned}$$

Now, thanks to Proposition 1.26, we have

$$\begin{aligned} &P\left(\left| \frac{1}{m} \sum_{i=1}^m Z_i^2 - 1 \right| \geq \delta \right) \leq 2e^{-m\delta^2/8} \\ \Rightarrow\ &P\left(\left| \frac{\|x\|_2^2}{m} \sum_{i=1}^m Z_i^2 - \|x\|_2^2 \right| \geq \delta \|x\|_2^2 \right) \leq 2e^{-m\delta^2/8} \\ \Rightarrow\ &P\big(\big| \|\Pi x\|_2^2 - \|x\|_2^2 \big| \geq \delta \|x\|_2^2 \big) \leq 2e^{-m\delta^2/8} \\ \Rightarrow\ &P\big(\big| \|\Pi x\|_2^2 - \|x\|_2^2 \big| \leq \delta \|x\|_2^2 \big) \geq 1 - 2e^{-m\delta^2/8} \\ \Rightarrow\ &P\big((1-\delta)\|x\|_2^2 \leq \|\Pi x\|_2^2 \leq (1+\delta)\|x\|_2^2 \big) \geq 1 - 2e^{-m\delta^2/8}, \end{aligned}$$

and finally, taking

$$m \geq \frac{8}{\delta^2} \log \frac{2}{\xi},$$

we have

$$1 - 2e^{-m\delta^2/8} \geq 1 - 2e^{-\frac{8}{\delta^2} \log \frac{2}{\xi} \delta^2/8} = 1 - \xi,$$

which gives the thesis. □

It is clear that this lemma does not directly imply the JL Lemma: first, Theorem 2.1 is about preserving *all* distances between n vectors, whereas this result is about the probability of preserving *one* of them, and second, Lemma 2.3 states that an inequality holds with a certain probability, whereas the JL Lemma asserts that there *exists* a map f that *almost* preserves the distances between the n vectors. The idea is that such f can be found algorithmically by repeating the procedure of generating Π, which defines f, a sufficient number of times to obtain it sooner or later – *by throwing the dice, sooner or later the desired number will be obtained.*

In the following we prove the *Probabilistic JL Lemma*, which deals with the probability of *nearly* preserving all distances between n vectors, and we use a little more formal reasoning than this to obtain the thesis of the JL Lemma.

Theorem 2.4 (Probabilistic JL Lemma). *Given* $z_1, \dots, z_n \in \mathbb{R}^d$ *and* $\Pi \in \mathbb{R}^{m\times d}$ *with independent entries* $\frac{1}{\sqrt{m}}N(0,1)$*, then for any* $\xi_U \in (0,1)$ *and* $\delta \in (0,1)$*,*

$$(1-\delta)\|z_i - z_j\|_2 \le \|\Pi z_i - \Pi z_j\|_2 \le (1+\delta)\|z_i - z_j\|_2$$

holds for all $i, j = 1\dots, n$*, provided that*

$$m \ge \frac{8}{\delta^2}\log\frac{n(n-1)}{\xi_U},$$

with probability at least $1-\xi_U$*.*

Proof. Since $\delta \in (0,1)$, taking the square root of all members of (2.2) of Lemma 2.3 we have that for all $x \in \mathbb{R}^d$

$$(1-\delta)\|x\|_2 < \sqrt{1-\delta}\|x\|_2 \le \|\Pi x\|_2 \le \sqrt{1+\delta}\|x\|_2 < (1+\delta)\|x\|_2$$

holds, provided that $m \ge \frac{8}{\delta^2}\log\frac{2}{\xi}$, with probability at least $1-\xi$.

Now considering $x = z_i - z_j$, we have that for all $i, j = 1, \dots, n$ – the case $i = j$ is trivial – the distance between any z_i and any z_j is preserved by Π but for a relative error of $1 \pm \delta$, provided that $m \ge \frac{8}{\delta^2}\log\frac{2}{\xi}$, with probability at least $1-\xi$.

To have the thesis, we now fix

$$\xi := \frac{\xi_U}{\binom{n}{2}} = \frac{2\cdot\xi_U}{n(n-1)},$$

and consider for all $i = 1, \dots, n$

- the events $\mathcal{E}_i$, i.e., *the i-th distance out of* $\binom{n}{2}$ *is almost preserved by* Π
- the complementary events $\mathcal{E}_i^c$.

With this choice and notation, we have that the assumption of Lemma 2.3 on m becomes

$$m \ge \frac{8}{\delta^2}\log\frac{2}{\xi} = \frac{8}{\delta^2}\log\frac{n(n-1)}{\xi_U},$$

and that for all $i = 1, \ldots, n$ the probability that $\mathcal{T}$ *almost* preserves the i-th distance is

$$P(\mathcal{E}_i) \geq 1 - \frac{\xi_U}{\binom{n}{2}}.$$

So we have that for all $i = 1, \ldots, n$ the probability that $\mathcal{T}$ does not *almost* preserve the i-th distance is

$$P(\mathcal{E}_i^c) \leq \frac{\xi_U}{\binom{n}{2}},$$

and so that the probability that at least one of the $\binom{n}{2}$ distances is not *almost* preserved by $\mathcal{T}$ is

$$P(\mathcal{E}_1^c \cup \mathcal{E}_2^c \cup \ldots \cup \mathcal{E}_{\binom{n}{2}}^c) \leq \sum_{i=1}^{\binom{n}{2}} P(\mathcal{E}_i^c) \leq \binom{n}{2} \cdot \frac{\xi_U}{\binom{n}{2}} = \xi_U,$$

where in the first inequality we use 6. of Proposition 1.6.

This immediately gives that the probability that all the distances between the n vectors are *almost* preserved by $\mathcal{T}$ is at least $1 - \xi_U$, i.e., the thesis.

□

Remark 2.5. On the notation: in most of this work, we use the letters δ and ξ to denote the precision and the upper bound of the probability of failure, respectively. In the last theorem (Probabilistic JL Lemma), we use the subscript U on ξ to underline the fact that this result on the probability that $\binom{n}{2}$ inequalities hold follows, thanks to the union bound in 6. of Proposition 1.6, from Lemma 2.3, which concerns only one inequality, and from an appropriate choice of ξ.

We can now prove the JL Lemma.

Proof. (*of the JL Lemma 2.1*). Let us consider

$$\xi_U := \frac{n-1}{n},$$

namely

$$1 - \xi_U = \frac{1}{n}.$$

From Theorem 2.4 we have that $\mathcal{T}$ *nearly* preserves all the distances between the couples z_i and z_j for $i, j = 1, \ldots, n$, provided that

$$m \geq \frac{8}{\delta^2} \log \frac{n(n-1)}{\xi_U} = \frac{8}{\delta^2} \log \frac{n^2(n-1)}{n-1} = \frac{16}{\delta^2} \log n,$$

with probability at least $1 - \xi_U = 1/n$.

Repeating therefore $O(n)$ times the generation of Π, we have that *sooner or later* we get a Π that *almost* preserves all of the distances between the n vectors of the Lemma, provided that $m \geq \frac{16}{\delta^2} \log n$, so we have that there exists an f – defined as in (2.1) – that satisfies the Lemma. □

2.3 Experiments

An interesting result to experiment with is Theorem 2.4 (Probabilistic JL Lemma), which states that if the dimension m of the arrival space is not less than $\frac{8}{\delta^2} \log \frac{n(n-1)}{\xi_U}$, we can map n vectors $z_1, \ldots, z_n \in \mathbb{R}^d$ to $\mathbb{R}^m$ preserving all of their distances, but for a relative error of $1 \pm \delta$, with probability at least $1 - \xi_U$.

The aim of this section is to see if we can get better results in practice than in theory, and in particular if we can take m less than its lower bound while still retaining the thesis.

In the following experiments we consider n vectors with $d = 15000$ uniformly distributed random components, generated by the MATLAB function `rand(d,n)`[1], we imagine that we have an upper bound $\xi_U = 1/10$ for the *failure probability* and we compute how many times in 100 generations of f defined as in (2.1), as n, m and δ vary, all of the distances between the n vectors considered are preserved but for a relative error of $1 \pm \delta$, when f *projects* them into $\mathbb{R}^m$.

The MATLAB code *JL*, reported in Appendix B.1, takes as input the number n of vectors to consider, the dimensions d and m respectively of the initial and arrival space, and the precision parameter δ, generates 100 times the matrix Π defining f, and computes how often all of the distances between the n vectors are *nearly* preserved by the generated *projections*.

The two Tables 2.1 and 2.2 summarise the results which we have obtained – we remark that all the results are random – by experimenting with this code. With simple calculations we can see that, assuming $\xi_U = 1/10$, only the top left cells of both the tables are related to parameters that satisfy the lower bound for m of Theorem 2.4, while the others do not. As we expected, the tables clearly show that in practice, to have a *projection* that *almost* preserves the distances between the n vectors considered, when the dimension of the arrival space is not too small – but does not necessarily respect the theoretical bound – we can generate f once and be *reasonably certain* that we have obtained a map of the desired form.

[1] `https://www.mathworks.com/help/matlab/ref/rand.html`.

$n \setminus m$	1500	1000	500	250	125
10	100	100	100	100	96
50	100	100	100	99	29
100	100	100	100	98	1
200	100	100	100	90	0

Table 2.1: Results obtained by running *JL* (in Appendix B.1) with $d = 15000$, $\delta = 0.2$, and some values of n and m. Each entry of the table contains the number of times, out of 100 generations of f, that all of the distances between the n vectors considered are preserved by f, but for a relative error of $1 \pm \delta$.

$\delta \setminus m$	1500	1000	500	250	125
0.20	100	100	100	100	91
0.15	100	100	100	94	30
0.10	100	100	88	22	0
0.05	68	18	0	0	0

Table 2.2: Results obtained by running *JL* (in Appendix B.1) with $n = 13$, $d = 15000$, and some values of m and δ. Each entry of the table contains the number of times, out of 100 generations of f, that all of the distances between the n vectors considered are preserved by f, but for a relative error of $1 \pm \delta$.

By doing more targeted experiments, it can also be verified that in many cases where not all distances are *nearly* preserved by f, the map still *nearly* preserves most of them.

Chapter 3

Extending the Johnson-Lindenstrauss Lemma

In this chapter we present an extension of the JL Lemma to cases of potentially infinite-dimensional initial space and potentially infinite cardinality (but finite *intrinsic dimension*) of the set of elements that we want to reduce in dimension *nearly* preserving their distances.

We begin by explaining the motivation behind the literature to study these issues, continue with the central theorems of this thesis and with an example of their application, and conclude with Section 3.6, in which we derive an infinite-dimensional theoretical result comparable to the JL Lemma and compare it with the Lemma.

Most of the following is taken from [8], while Section 3.6 is our own work.

3.1 *CS*, *RIP* and more

Compressed sensing (CS) is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems [9], which are systems with a coefficient matrix that has fewer rows than columns and, if consistent, has infinitely many solutions.

The basic idea behind CS is that a *sparse* signal, namely a vector $x \in \mathbb{R}^n$ with most of its components equal to zero, can be recovered from measurements made by computing $m \ll n$ scalar products $b_i = \langle a_i^T, x \rangle$, where each a_i^T is a row vector that is typically generated randomly [11], i.e., more formally, that a linear system $Ax = b$, where $A \in \mathbb{R}^{m \times n}$ with $m \ll n$, has a unique sparse solution.

The *restricted isometry property (RIP)* is a fundamental tool in CS that allows one to show the uniqueness of the k-sparse solution – with at most

$k \ll n$ non-zero components – of such systems. In a finite-dimensional setting, a matrix $A \in \mathbb{R}^{m\times n}$ satisfies the RIP on a set $S \subset \mathbb{R}^n$ [8], if there exists a constant $\delta \in (0,1)$ such that for all $x \in S$,

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2,$$

and a sufficient condition to prove the uniqueness of the k-sparse solution of the system $Ax = b$, is that A satisfies the RIP on the set of the $2k$-sparse unit vectors $x \in \mathbb{R}^n$ (Lemma 2.5 in [11]). Unfortunately, this condition is not really any easier to verify for a given matrix A, but thanks to the JL Lemma (see, e.g., [2]) it can be shown to hold with high probability for large classes of matrices generated by certain random procedures – for example, for matrices with entries that are centred normal random variables with variance m^{-1} – which gives us confidence that these matrices work in a CS application [11].

Since the RIP plays such a key role in such an important area as that of CS, the authors of [8] have studied how to extend the construction of linear maps satisfying the RIP in a finite-dimensional ambient space to linear maps satisfying the RIP on subsets of a possibly infinite-dimensional space. The work in this direction is closely related to the theory of CS in infinite dimension [13], which was introduced to overcome "*the mismatch between computational and physical models, that can lead to critical errors when CS techniques are applied to real data arising from continuous models*" [12].

Our motivation lies in a completely different area: we would like to exploit the RIP in the context of retarded functional differential equations [14, 15, 16], and in particular we are interested in the *projection* without *too much* change in distances of finite cardinality subsets of the Hilbert space $L^2([0,1],\mathbb{R})$[1]. We provide more details in the conclusion, although it must be clear that we do not pursue directly this objective in the present work.

3.2 Preliminaries

In this section we introduce some of the notations and definitions (taken from [8]) that we use in the rest of this work.

Consider $(H, \langle\cdot,\cdot\rangle)$ a possibly infinite-dimensional Hilbert space with norm $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$ induced by the scalar product, and let $\Sigma \subset H$ be a subset of H of arbitrary cardinality $|\Sigma|$.

Definition 3.1. The *secant set* of Σ is defined as

$$\Sigma - \Sigma := \{z_1 - z_2 : z_1, z_2 \in \Sigma\}.$$

[1]See Section A.2 in Appendix A for the definition.

Definition 3.2. The *normalised secant set* of Σ is

$$S(\Sigma) := \left\{ \frac{y}{\|y\|} : y \in \Sigma - \Sigma \setminus \{0\} \right\}.$$

It follows immediately from the definition that $S(\Sigma) \subset \mathcal{S}$, where, recalling the notations introduced in Section 1.1, $\mathcal{S}$ is the unit sphere in H, and that S is *symmetric*, i.e., it holds that $x \in S$ if and only if $-x \in S$.

Since we are interested in *almost* preserving the distances between the elements of Σ, it is clear that, as long as the map that *projects* the elements is linear, it is equivalent to working on the set Σ or on the set $S(\Sigma)$, in the latter case, of course, by referring to the *almost* preservation of the norms.

Therefore, from now on, we work on an arbitrary set $S \subset \mathcal{S}$ lying on the unit sphere $\mathcal{S}$ in the ambient space H, keeping in mind that if we want to obtain results about the *nearly* preservation of the distances between the elements of a set Σ, we only need to substitute the normalised secant set $S(\Sigma)$ of Σ for S in the following.

In order to be able to state and prove most of the following results, we must first formally define the concept of the *intrinsic dimension* of a set S, since the following theory of *projecting* an infinite cardinality set without *too much* change in norms is based on the crucial assumption that the set has a small *intrinsic dimension*.

There are several definitions of dimension in the literature (see, for example, [23]). In this work we consider, as the authors of [8] do, the *upper box-counting dimension* which, if finite, ensures the possibility of *projecting nearly* preserving the distances. However, we must bear in mind that there exist other definitions of the *intrinsic dimension* of a set, not suitable for our context.

In order to give the definition of this dimension, we first need to define the *covering number* of a set and the concept of *minimal ϵ-net* for a set.

Definition 3.3. Let $\epsilon > 0$. The *covering number* of S is the minimum number $N_S(\epsilon)$ of closed balls of radius ϵ with centres in S needed to cover S.

Definition 3.4. A *minimal ϵ-net* for S is the set of centres of these balls.

We can now define the *intrinsic dimension* of a set.

Definition 3.5. The *upper box-counting dimension* of S is

$$\dim_B(S) := \limsup_{\epsilon \to 0} \frac{\log N_S(\epsilon)}{\log(1/\epsilon)}.$$

Remark 3.6. It is easy to see from the definition that if S has a finite number of elements, then $\dim_B(S) = 0$.

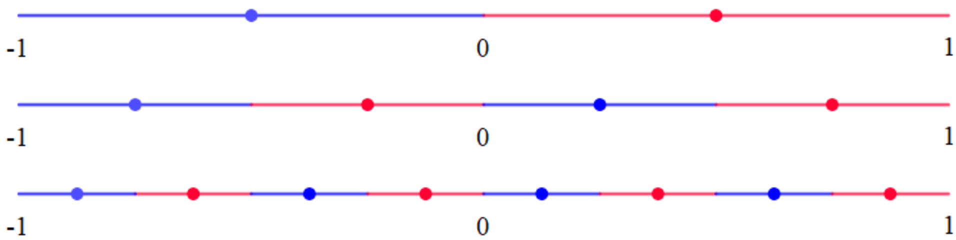


Figure 3.1: Covering with closed balls of radius $\epsilon = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ of the set $[-1, 1]$.

Before proceeding with the formulation of the *Assumption A*, which imposes a finite upper box-counting dimension on the set S, in order to better understand the latter definition, we give an example of calculating the upper box-counting dimension of a simple set, and we present some facts about this notion of dimension.

Example 3.7. Let us compute the upper box-counting dimension of the closed unit ball in $(\mathbb{R}, \|\cdot\|_2)$ centred at the origin, i.e., $I = [-1, 1]$.
If we consider $\epsilon = \frac{1}{2}$, then of course a minimal ϵ-net for I is $\{-\frac{1}{2}, \frac{1}{2}\}$ and so $N_I(\epsilon) = 2$. Similarly, if for simplicity we consider $\epsilon = \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n}, \ldots$ (see Figure 3.1) we have that $N_I(\epsilon) = \epsilon^{-1}$ for all ϵ. Applying the definition, it is now clear that the upper box-counting dimension of this set is $\dim_B(I) = 1$.

As we guess from the example above, the upper box-counting dimension of a set S (which, if the limit exists, can simply be called box-counting dimension) can be seen as the exponent d such that the covering number of S is $N_S(\epsilon) \approx C \cdot \epsilon^{-d}$, with $C > 0$, which is what one would expect in the trivial case where S is a *smooth space* of integer dimension d [24]. However there exists objects whose upper box-counting dimension is not integer, e.g., one can show that the dimension of the Cantor set is $\log(2)/\log(3)$ [25].

In general the upper box-counting dimension of a set is not so easy to calculate, so there are algorithms for doing so (see, for example, [26]). The reader who wishes to learn more about the properties of this dimension, which is of great importance for the study of fractals, should refer to [25].

As already mentioned, in this thesis the assumption to have a finite upper box-counting dimension is fundamental, so for convenience we summarise that in the following assumption.

Assumption A. *The set $S \subset \mathcal{S}$ has a finite upper box-counting dimension* $\dim_B(S)$ *which is strictly bounded by $s \geq 1$.*

It is easy to see that if *Assumption A* holds, there exists a set-dependent constant $\epsilon_S \in \left(0, \frac{1}{2}\right)$ such that $N_S(\epsilon) \leq \epsilon^{-s}$ for all $\epsilon \leq \epsilon_S$. This constant is important because it is part of the lower bound for the dimension m of the arrival space in Theorems 3.17 and 3.28 below, and in the following, whenever we assume that *Assumption A* holds, we consider this $\epsilon_S \in (0, \frac{1}{2})$ to be fixed.

Remark 3.8. Clearly, for any finite cardinality set, *Assumption A* holds.

3.3 Infinite-dimensional and probabilistic RIP

In this section we introduce the RIP in a potentially infinite-dimensional setting and provide sufficient conditions for a random linear map L of a certain general form to satisfy this property on sets S for which *Assumption A* holds, with high probability.

The following content is rather technical and is included for completeness, as it provides the foundation for Section 3.4, where we present a specific construction of a random linear map that satisfies the RIP.

Let L be a random linear map defined as

$$\begin{aligned} L : H &\to \mathbb{R}^m \\ x &\mapsto L(x) \coloneqq (\langle l_1, x\rangle, \ldots, \langle l_m, x\rangle)^T, \end{aligned} \tag{3.1}$$

where $(l_1, \ldots, l_m) \in H^m$ is an m-upla taken at random with a probability measure P on H^m.

In the following, we sometimes refer to a generic random linear map L without specifying that we are dealing with this one in particular, because some definitions and facts that we present are general, but in this section our interest is in this particular random linear and continuous map L.

Let us define a semi-norm related to a random linear map L, which is useful in the definition of the RIP.

Definition 3.9. Given a random linear map $L : H \to \mathbb{R}^m$, let $\|\cdot\|_L$ be the semi-norm defined for all $x \in H$ by

$$\|x\|_L \coloneqq \left[\mathbb{E}\left(\|L(x)\|_2^2\right)\right]^{\frac{1}{2}}.$$

Remark 3.10. On the notation: the authors of [8] define this semi-norm as $\|\cdot\|_P$ to emphasise the fact that the expected value is made respect to the probability measure P on H^m, but we prefer to use the subscript L, which we hope make it easier for the reader to remember how this semi-norm is defined.

To introduce the RIP in an infinite-dimensional context, we recall the following definition, which holds in a finite-dimensional one, and we extend it in an intuitive way.

Definition 3.11. A matrix $A \in \mathbb{R}^{m \times n}$ satisfies the *restricted isometry property (RIP)* on a generic set $S \subset \mathbb{R}^n$ if there exists a constant $\delta \in (0,1)$ such that for all $x \in S$

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2. \tag{3.2}$$

First, we note that the condition (3.2) is equivalent to

$$\left|\|Ax\|_2^2 - \|x\|_2^2\right| \leq \delta\|x\|_2^2.$$

Then, remarking that for some random matrices – for example with independent entries $N(0, m^{-1})$ – it holds that

$$\mathbb{E}\left(\|Ax\|_2^2\right) = \|x\|_2^2,$$

we can extend the definition of the RIP to the case of random matrices saying that a random matrix $A \in \mathbb{R}^{m \times n}$ satisfies the RIP on a generic set $S \subset \mathbb{R}^n$ if there exists a constant $\delta \in (0,1)$ such that for all $x \in S$

$$\left|\|Ax\|_2^2 - \mathbb{E}\left(\|Ax\|_2^2\right)\right| \leq \delta\|x\|_2^2.$$

If we now imagine moving from the finite-dimensional to the infinite-dimensional and to the unit sphere, we can give the following definition.

Definition 3.12 (RIP)**.** A random linear map $L : H \to \mathbb{R}^m$ satisfies the *RIP* on a set $S \subset \mathcal{S} \subset H$ if there exists a constant $\delta \in \left(0, \inf_{x \in S} \|x\|_L^2\right)$ such that for all $x \in S$

$$\left|\|L(x)\|_2^2 - \mathbb{E}\left(\|L(x)\|_2^2\right)\right| \leq \delta,$$

i.e.,

$$\left|\|L(x)\|_2^2 - \|x\|_L^2\right| \leq \delta. \tag{3.3}$$

Remark 3.13. In the definition above, δ is between 0 and $\inf_{x \in S} \|x\|_L^2$ for a very good reason: removing the absolute value in (3.3) we obtain the inequalities

$$\|x\|_L^2 - \delta \leq \|L(x)\|_2^2 \leq \|x\|_L^2 + \delta,$$

which have sense in terms of norm preserving if and only if the left term is greater than zero for all $x \in S$. The condition on δ in the definition, ensures that it is.

Since Definition 3.12 is a bit heavy, we now introduce some notations to make it more streamlined.

Definition 3.14. Given a random linear map $L : H \to \mathbb{R}^m$, let

- $\delta_{S,L} \coloneqq \sup_{x\in S}\left| \|L(x)\|_2^2 - \|x\|_L^2 \right|$, renamed to δ_S for ease of notation;
- $\underline{\delta}_{S,L} \coloneqq \inf_{x\in S} \|x\|_L^2$, renamed to $\underline{\delta}_S$ for ease of notation;
- $\overline{\delta}_{S,L} \coloneqq \sup_{x\in S} \|x\|_L^2$, renamed to $\overline{\delta}_S$ for ease of notation.

Using the above notations, we can easily redefine the RIP on $S \subset \mathcal{S} \subset H$.

Definition 3.15 (RIP, Definition II.1 of [8])**.** A random linear map $L : H \to \mathbb{R}^m$ satisfies the *RIP* on a set $S \subset \mathcal{S} \subset H$ if there exists a constant $\delta > 0$ such that

$$\delta_S \leq \delta < \underline{\delta}_S.$$

A condition which, if the *Assumption A* holds, ensures that a random linear map L constructed as in (3.1) satisfies the RIP on S with high probability, is that the following assumption, related to the probability measure P on H^m chosen to define it, holds. This is a very technical assumption, which will be replaced by another more specific condition when we construct a particular L in the next section.

Considering a random linear map L defined as in (3.1), let us define the function

$$\begin{aligned} h_L : H &\to \mathbb{R} \\ x &\mapsto h_L(x) \coloneqq \|L(x)\|_2^2 - \|x\|_L^2. \end{aligned} \tag{3.4}$$

Assumption B. *There exist two constants* $c_1, c_2 \in (0, +\infty]$ *such that for any fixed* $y, z \in S \cup \{0\}$*,*

$$P\{|h_L(y) - h_L(z)| \geq \lambda \|y - z\|\} \leq 2e^{-c_1 m \lambda^2} \text{ for every } 0 \leq \lambda \leq \frac{c_2}{c_1},$$

and

$$P\{|h_L(y) - h_L(z)| \geq \lambda \|y - z\|\} \leq 2e^{-c_2 m \lambda} \text{ for every } \lambda \geq \frac{c_2}{c_1}.$$

Remark 3.16. *Assumption B* includes the case of $c_1 = +\infty$ or $c_2 = +\infty$ to handle situations where one of the two bounds holds for all $\lambda \geq 0$.

We can now state the main theorem of this section, which concerns the satisfaction of the RIP on $S \subset \mathcal{S}$ by the random linear map L defined in (3.1).

Theorem 3.17 (Theorem II.2 of [8])**.** *Let* $S \subset \mathcal{S} \subset H$ *and* L *be a random linear map defined, as in* (3.1)*, as*

$$\begin{aligned} L : H &\to \mathbb{R}^m \\ x &\mapsto L(x) \coloneqq (\langle l_1, x\rangle, \ldots, \langle l_m, x\rangle)^T, \end{aligned}$$

where $(l_1, \dots, l_m) \in H^m$ *is an* m*-upla taken at random with a probability measure* P *on* H^m*. There exists an absolute constant* $K > 0$ *such that if Assumption A and Assumption B hold, then for any* $\xi \in (0, 1)$ *and* $\delta \in (0, 1)$*,*

$$\delta_S \leq \delta,$$

provided that

$$m \geq K(s, \delta, \xi) := \frac{K}{\min\{c_1, c_2\}\delta^2} \max\left\{ s \log \frac{1}{\epsilon_S}, \log \frac{6}{\xi} \right\},$$

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

Remark 3.18. The theorem ensures that, if *Assumptions A* and *B* hold and $m \geq K(s, \delta, \xi)$, a random linear map L defined as in (3.1) satisfies the RIP on the set $S \subset \mathcal{S}$ with constant $\delta \in (0, \underline{\delta}_S)$ with probability at least $1 - \xi$.

As the authors of [8] do, we divide the proof of this theorem into two steps: first we give some notations and prove Lemma 3.19, which gives a bound for δ_S that holds with probability at least $1 - 3\xi$, and then we use the estimates of certain sums (Lemma A.1) that are involved in the bound of the lemma to prove the theorem.

For the hypothesis of the theorem, we have, as already noted in Section 3.2, that there exists a set-dependent constant $\epsilon_S \in \left(0, \frac{1}{2}\right)$ such that the (finite) minimum number of closed balls with centres in S and radius ϵ needed to cover S is $N_S(\epsilon) \leq \epsilon^{-s}$ for all $\epsilon \leq \epsilon_S$, where s is an upper bound for $\dim_B(S)$, ensured exist by the validity of the *Assumption A*.

Considering fixed this ϵ_S, we can give the following notations.
For all $j \geq 0$, let

- $C_j \subset S$ be a minimal $(2^{-j}\epsilon_S)$-net for S, with cardinality

$$|C_j| = N_S(2^{-j}\epsilon_S);$$

- $\eta_j : H \to H$ be a map that at each $x \in H$ associates

$$\eta_j(x) \in \operatorname*{argmin}_{z \in C_j} \|x - z\|;$$

- D_j be the finite set $\{(\eta_{j+1}(x), \eta_j(x)) : x \in S\} \subset C_{j+1} \times C_j$, with cardinality

$$|D_j| \leq N_S^2(2^{-j-1}\epsilon_S). \tag{3.5}$$

It is easy to see that with this notations, for all $j \geq 0$ it holds the following:

$$\sup_{(y,z)\in D_j} \|y - z\| \leq 2^{-j+1}\epsilon_S, \tag{3.6}$$

in fact for all $(y, z) \in D_j$ there exists $x \in S$ such that $(\eta_{j+1}(x), \eta_j(x)) = (y, z)$, and

$$\begin{aligned}\|y - z\| &= \|\eta_{j+1}(x) - \eta_j(x)\| = \|\eta_{j+1}(x) - x + x - \eta_j(x)\| \\ &\leq \|\eta_{j+1}(x) - x\| + \|x - \eta_j(x)\| \leq 2^{-j-1}\epsilon_S + 2^{-j}\epsilon_S \\ &\leq 2 \cdot 2^{-j}\epsilon_S = 2^{-j+1}\epsilon_S.\end{aligned}$$

Furthermore, using $y = x \in S \subset \mathcal{S}$ and $z = 0$ in *Assumption B*, we have that there exist two constants $c_1, c_2 \in (0, +\infty]$ such that for any fixed $x \in S \subset \mathcal{S}$

$$P\{|h_L(x)| \geq \lambda\} \leq \begin{cases} 2e^{-c_1 m\lambda^2}, & \text{if } 0 \leq \lambda \leq \frac{c_2}{c_1} \\ 2e^{-c_2 m\lambda}, & \text{if } \lambda \geq \frac{c_2}{c_1}. \end{cases} \tag{3.7}$$

where the term $h_L(0)$ is clearly zero for the definition of h_L (3.4).

Let us now state and prove the following lemma, from which Theorem 3.17 follows.

Lemma 3.19 (Lemma A.1 of [8]). *Let $S \subset \mathcal{S} \subset H$ be a subset for which Assumption A holds, and L be a random linear map defined, as in* (3.1), *as*

$$\begin{aligned}L : H &\to \mathbb{R}^m \\ x &\mapsto L(x) \coloneqq (\langle l_1, x\rangle, \ldots, \langle l_m, x\rangle)^T,\end{aligned}$$

where $(l_1, \ldots, l_m) \in H^m$ is an m-upla taken at random with a probability measure P on H^m.

Let $\xi \in (0, 1)$ and Q_1, Q_2, Q_3 be defined as

$$\begin{aligned}Q_1(\epsilon_S, \xi) &\coloneqq \sqrt{\log\left[\frac{2}{\xi} N_S(\epsilon_S)\right]}, \\ Q_2(\epsilon_S, \xi) &\coloneqq \sum_{j=0}^{+\infty} 2^{-j+1}\sqrt{\log\left[\frac{2^{j+1}}{\xi} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right]}, \\ Q_3(\epsilon_S, \xi) &\coloneqq \sum_{j=0}^{+\infty} 2^{-j+1}\log\left[\frac{2^{j+1}}{\xi} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right].\end{aligned}$$

If Assumption B holds, then

$$\delta_S \leq \frac{Q_1 + \epsilon_S \cdot Q_2}{\sqrt{c_1 m}} + \frac{Q_1^2 + \epsilon_S \cdot Q_3}{c_2 m},$$

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

Proof. Consider the telescopic sum

$$h_L(x) = h_L(\eta_0(x)) + \sum_{j=0}^{+\infty}[h_L(\eta_{j+1}(x)) - h_L(\eta_j(x))].$$

The above equality holds for the following reasons:

- L is a continuous map with respect to the norm $\|\cdot\|$ induced by the scalar product $\langle\cdot,\cdot\rangle$;
- the maps $\|L(\cdot)\|_2^2$, $\mathbb{E}\left(\|L(\cdot)\|_2^2\right)$ and thus $h_L(\cdot)$ are continuous with respect to the norm $\|\cdot\|$;
- $h_L(x) - h_L(\eta_0(x)) - \sum_{j=0}^{N}[h_L(\eta_{j+1}(x)) - h_L(\eta_j(x))] = h_L(x) - h_L(\eta_{N+1}(x))$;
- $\lim_{N\to+\infty} \|\eta_{N+1}(x) - x\| = 0$;
- $\lim_{N\to+\infty} |h_L(x) - h_L(\eta_{N+1}(x))| = 0$;

where the last two points follow from the definitions of C_{N+1} and $\eta_{N+1}(\cdot)$, and from the continuity of $h_L(\cdot)$.

Due to the triangle inequality and the definitions of the map $h_L(\cdot)$ and the sets C_0 and D_j, we have

$$\begin{aligned}
\delta_S &= \sup_{x\in S}\left|\|L(x)\|_2^2 - \|x\|_L^2\right| \\
&= \sup_{x\in S}|h_L(x)| \\
&\leq \sup_{x\in S}|h_L(\eta_0(x))| + \sum_{j=0}^{+\infty}\sup_{x\in S}|h_L(\eta_{j+1}(x)) - h_L(\eta_j(x))| \\
&= \max_{x_0\in C_0}|h_L(x_0)| + \sum_{j=0}^{+\infty}\max_{(y,z)\in D_j}|h_L(y) - h_L(z)|.
\end{aligned}$$

Let now $a_j, b > 0$ be positive parameters for all $j \in \mathbb{N}$. We have

$$\begin{aligned}
&P\left(\delta_S \geq b + \sum_{j=0}^{+\infty} a_j\right) \leq \\
&\quad \leq P\left(\max_{x_0\in C_0}|h_L(x_0)| + \sum_{j=0}^{+\infty}\max_{(y,z)\in D_j}|h_L(y) - h_L(z)| \geq b + \sum_{j=0}^{+\infty} a_j\right) \qquad (3.8)\\
&\quad \leq P\left(\max_{x_0\in C_0}|h_L(x_0)| \geq b\right) + \sum_{j=0}^{+\infty} P\left(\max_{(y,z)\in D_j}|h_L(y) - h_L(z)| \geq a_j\right),
\end{aligned}$$

where the first inequality follows from 3. of Proposition 1.6 and in the second we use Lemma 1.14.

Thanks to 6. of Proposition 1.6 and the fact that $|C_0| = N_S(\epsilon_S)$, we have

$$\begin{aligned}
P\left(\max_{x_0\in C_0} |h_L(x_0)| \geq b\right) &= 1 - P\left(\max_{x_0\in C_0} |h_L(x_0)| < b\right) \\
&= 1 - P\big(|h_L(x_0)| < b \text{ for all } x_0 \in C_0\big) \\
&= 1 - \left[1 - P\left(\bigcup_{x_0\in C_0} \big(|h_L(x_0)| \geq b\big)\right)\right] \\
&= P\left(\bigcup_{x_0\in C_0} \big(|h_L(x_0)| \geq b\big)\right) \\
&\leq \sum_{x_0\in C_0} P\big(|h_L(x_0)| \geq b\big) \\
&\leq N_S(\epsilon_S) \cdot \max_{x_0\in C_0} P\big(|h_L(x_0)| \geq b\big).
\end{aligned}$$

Therefore, using the above to bound the first term in (3.8) and similar reasoning and (3.5) to bound the others, we have

$$\begin{aligned}
&P\left(\delta_S \geq b + \sum_{j=0}^{+\infty} a_j\right) \leq \\
&\qquad \leq N_S(\epsilon_S) \cdot \max_{x_0\in C_0} P\left(|h_L(x_0)| \geq b\right) + \\
&\qquad\qquad + \sum_{j=0}^{+\infty} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right) \cdot \max_{(y,z)\in D_j} P\left(|h_L(y) - h_L(z)| \geq a_j\right).
\end{aligned}$$

Now look at the first term on the right-hand side of the inequality above. Since $C_0 \subset S$, thanks to (3.7) we have

$$\max_{x_0\in C_0} P\left(|h_L(x_0)| \geq b\right) \leq \begin{cases} 2e^{-c_1 m b^2}, & \text{if } b \leq \frac{c_2}{c_1} \\ 2e^{-c_2 m b}, & \text{if } b \geq \frac{c_2}{c_1}. \end{cases}$$

Choosing

$$b := \begin{cases} \frac{Q_1}{\sqrt{c_1 m}} = \sqrt{\frac{\log(2N_S(\epsilon_S)/\xi)}{c_1 m}}, & \text{if } \frac{Q_1^2}{m} \leq \frac{c_2^2}{c_1} \\ \frac{Q_1^2}{c_2 m} = \frac{\log(2N_S(\epsilon_S)/\xi}{c_2 m}, & \text{if } \frac{Q_1^2}{m} \geq \frac{c_2^2}{c_1}, \end{cases}$$

with simple calculations we obtain in both cases

$$N_S(\epsilon_S) \cdot \max_{x_0\in C_0} P\left(|h_L(x_0)| \geq b\right) \leq \xi$$

and

$$b \leq \frac{Q_1}{\sqrt{c_1 m}} + \frac{Q_1^2}{c_2 m}. \tag{3.9}$$

Consider now any term $P\left(|h_L(y) - h_L(z)| \geq a_j\right)$ with $(y,z) \in D_j$, for any $j \in \mathbb{N}$. Due to (3.6), the *Assumption B* implies that

$$\begin{aligned} P\left(|h_L(y) - h_L(z)| \geq a_j\right) &= P\left(|h_L(y) - h_L(z)| \geq \frac{a_j \cdot 2^{j-1}}{\epsilon_S} \cdot \frac{\epsilon_S}{2^{j-1}}\right) \\ &\leq P\left(|h_L(y) - h_L(z)| \geq \frac{a_j \cdot 2^{j-1}}{\epsilon_S} \cdot \|y - z\|\right) \\ &\leq \begin{cases} 2e^{-c_1 m \left(a_j \cdot 2^{j-1}/\epsilon_S\right)^2} & \text{if } a_j \cdot 2^{j-1}/\epsilon_S \leq \frac{c_2}{c_1} \\ 2e^{-c_2 m a_j \cdot 2^{j-1}/\epsilon_S} & \text{if } a_j \cdot 2^{j-1}/\epsilon_S \geq \frac{c_2}{c_1}. \end{cases} \end{aligned} \tag{3.10}$$

Let $\mathcal{I} := \left\{ j \in \mathbb{N} : \frac{1}{m} \log\left(\frac{2^{j+1}}{\xi} \cdot N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right) \leq \frac{c_2^2}{c_1} \right\}$ and let we choose

$$a_j := \begin{cases} \frac{2^{-j+1}\epsilon_S}{\sqrt{c_1 m}} \sqrt{\log\left(\frac{2^{j+1}}{\xi} \cdot N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right)} & \text{if } j \in \mathcal{I} \\ \frac{2^{-j+1}\epsilon_S}{c_2 m} \log\left(\frac{2^{j+1}}{\xi} \cdot N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right) & \text{if } j \in \mathbb{N} \setminus \mathcal{I}. \end{cases}$$

With these choices we have

$$\frac{a_j \cdot 2^{j-1}}{\epsilon_S} \begin{cases} \leq \frac{c_2}{c_1} & \text{if } j \in \mathcal{I} \\ \geq \frac{c_2}{c_1} & \text{if } j \in \mathbb{N} \setminus \mathcal{I}, \end{cases}$$

and so, for (3.10) we have

$$P\left(|h_L(y) - h_L(z)| \geq a_j\right) \leq \begin{cases} 2e^{-c_1 m \left(a_j \cdot 2^{j-1}/\epsilon_S\right)^2} & \text{if } j \in \mathcal{I} \\ 2e^{-c_2 m a_j \cdot 2^{j-1}/\epsilon_S} & \text{if } j \in \mathbb{N} \setminus \mathcal{I}. \end{cases}$$

Therefore, replacing a_j and executing simple calculations, for all $j \in \mathbb{N}$ we obtain

$$N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right) \cdot \max_{(y,z) \in D_j} P\left(|h_L(y) - h_L(z)| \geq a_j\right) \leq 2^{-j}\xi,$$

that implies

$$\sum_{j=0}^{+\infty} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right) \cdot \max_{(y,z) \in D_j} P\left(|h_L(y) - h_L(z)| \geq a_j\right) \leq \sum_{j=0}^{+\infty} \frac{1}{2^j}\xi = 2\xi.$$

Note that, for these choices of the a_j, we have

$$\sum_{j=0}^{+\infty} a_j = \frac{\epsilon_S}{\sqrt{c_1 m}} \sum_{j\in\mathcal{I}} 2^{-j+1} \sqrt{\log\left(\frac{2^{j+1}}{\xi}\cdot N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right)} + \\ + \frac{\epsilon_S}{c_2 m} \sum_{j\in\mathbb{N}\setminus\mathcal{I}} 2^{-j+1} \log\left(\frac{2^{j+1}}{\xi}\cdot N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right)$$

and so, by bounding the two above sums on $\mathcal{I}$ and $\mathbb{N}\setminus\mathcal{I}$ with the sums on $\mathbb{N}$, we have

$$\sum_{j=0}^{+\infty} a_j \le \frac{\epsilon_S}{\sqrt{c_1 m}} Q_2 + \frac{\epsilon_S}{c_2 m} Q_3. \tag{3.11}$$

Therefore, it holds that

$$P\left(\delta_S \ge b + \sum_{j=0}^{+\infty} a_j\right) \le 3\xi,$$

and so

$$\begin{aligned} P\left(\delta_S \le b + \sum_{j=0}^{+\infty} a_j\right) &= 1 - P\left(\delta_S > b + \sum_{j=0}^{+\infty} a_j\right) \\ &\ge 1 - P\left(\delta_S \ge b + \sum_{j=0}^{+\infty} a_j\right) \\ &\ge 1 - 3\xi. \end{aligned}$$

Finally, using 3. of Proposition 1.6 and the bounds (3.9) and (3.11), we have

$$P\left(\delta_S \le \frac{Q_1 + \epsilon_S \cdot Q_2}{\sqrt{c_1 m}} + \frac{Q_1^2 + \epsilon_S \cdot Q_3}{c_2 m}\right) \ge P\left(\delta_S \le b + \sum_{j=0}^{+\infty} a_j\right) \ge 1 - 3\xi.$$

□

Using now the estimates for Q_1, Q_2 and Q_3 reported in Lemma A.1, we can prove Theorem 3.17.

Proof. (*of Theorem 3.17*). From Lemma 3.19 and Lemma A.1 we have

$$\begin{aligned}\delta_S &\leq \frac{Q_1 + \epsilon_S \cdot Q_2}{\sqrt{c_1 m}} + \frac{Q_1^2 + \epsilon_S \cdot Q_3}{c_2 m} \\ &\leq \frac{\sqrt{\log(2/\xi)}}{\sqrt{c_1 m}} + \frac{\sqrt{s\log(1/\epsilon_S)}}{\sqrt{c_1 m}} + \frac{\epsilon_S \cdot 8\sqrt{\log(2/\xi)}}{\sqrt{c_1 m}} + \frac{\epsilon_S \cdot 8\sqrt{2s\log(2)}}{\sqrt{c_1 m}} + \\ &+ \frac{\epsilon_S \cdot 4\sqrt{2s\log(1/\epsilon_S)}}{\sqrt{c_1 m}} + \frac{\log(2/\xi)}{c_2 m} + \frac{s\log(1/\epsilon_S)}{c_2 m} + \frac{\epsilon_S \cdot 8\log(2/\xi)}{c_2 m} + \\ &+ \frac{\epsilon_S \cdot 16 s\log(2)}{c_2 m} + \frac{\epsilon_S \cdot 8s\log(1/\epsilon_S)}{c_2 m},\end{aligned}$$

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

Let now $\delta \in (0,1)$. Remembering that $\epsilon_S \in \left(0, \frac{1}{2}\right)$, it is easy to see that if

$$m \geq \frac{16}{\min\{c_1, c_2\}\delta^2} \log\frac{2}{\xi},$$

all the terms in the above inequality that depend on ξ are upper bounded by δ. Similarly, if

$$m \geq \frac{8}{\min\{c_1, c_2\}\delta^2} s\log\frac{1}{\epsilon_S},$$

the terms that depend on $\log(1/\epsilon_S)$ are also bounded by δ, and if

$$m \geq \frac{32}{\min\{c_1, c_2\}\delta^2} s\log 2,$$

the remaining terms are bounded by δ as well.

Therefore, if

$$m \geq \frac{32}{\min\{c_1, c_2\}\delta^2} \max\left\{s\log\frac{1}{\epsilon_S}, \log\frac{2}{\xi}\right\},$$

we have that

$$\delta_S \leq 10\delta,$$

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

Finally, by the two changes of variables $\hat{\delta} = 10\delta$ and $\hat{\xi} = 3\xi$ and renaming the new parameters as the originals, we have the thesis. □

Remark 3.20. From the proof we can see that the absolute constant of Theorem 3.17 is $K = 3200$.

3.4 A random linear map that satisfies the RIP

In this section we provide a method (from [8]) for constructing a random linear map that satisfies the RIP on infinite-dimensional sets for which *Assumption A* holds, with high probability. Specifically, we present a two-step procedure for constructing such a random linear map and discuss the conditions on the probability measure chosen for the construction, which are sufficient to ensure the validity of *Assumption B* introduced in the previous section.

The first step involves reducing the dimension of the elements we want to *project* from infinite to finite (though potentially much large) dimension, while *essentially* preserving their norms (or equivalently, distances). The second step further reduces the dimension using random techniques, specifically through the use of a random matrix.

Note that, although this section deals with the infinite-dimensional case, it is clear that everything that follows can also be applied in a finite-dimensional context, such as in cases involving sets of infinite cardinality (but finite upper box-counting dimension), where the JL Lemma does not apply (see the example, also from [8], in Section 3.5).

3.4.1 First step: from infinite to finite dimension

We begin by constructing a linear and continuous map $b : H \to \mathbb{R}^d$, where d is potentially large but finite, that, in a certain sense, *almost* preserve the norms, i.e., it satisfies the RIP in an infinite-dimensional, non-random sense.

Let $S \subset \mathcal{S} \subset H$ be a subset of the infinite-dimensional Hilbert space $(H, \langle\cdot,\cdot\rangle)$ with induced norm $\|\cdot\|$, for which *Assumption A* holds.

Fix $\epsilon_* \in (0, 1)$. Since *Assumption A* ensures that S can be covered by a finite number of closed balls, we can consider a minimal ϵ_*-net $C(\epsilon_*)$ for S, i.e., a set of centres of closed balls of radius ϵ_* that cover S in a *minimal* way, with finite cardinality

$$|C(\epsilon_*)| = N_S(\epsilon_*) < +\infty.$$

Additionally, let us consider $V_{\epsilon_*} \subset H$ the finite-dimensional subspace of H generated by $C(\epsilon_*)$ (see Figure 3.2 – from [8] – that help to visualise the construction), and an orthonormal basis $\{b_1, \ldots, b_d\}$ for this space of dimension $\dim V_{\epsilon_*} = d < +\infty$.

Remark 3.21. It is clear that if the set $\Sigma \subset H$ that we want to *project*, whose normalised secant set is S, has finite cardinality, then the dimension d of V_{ϵ_*} is bounded by a constant that depends on the number of elements of Σ (see Section 3.6 for more on this). On the other hand, if Σ is an infinite-cardinality

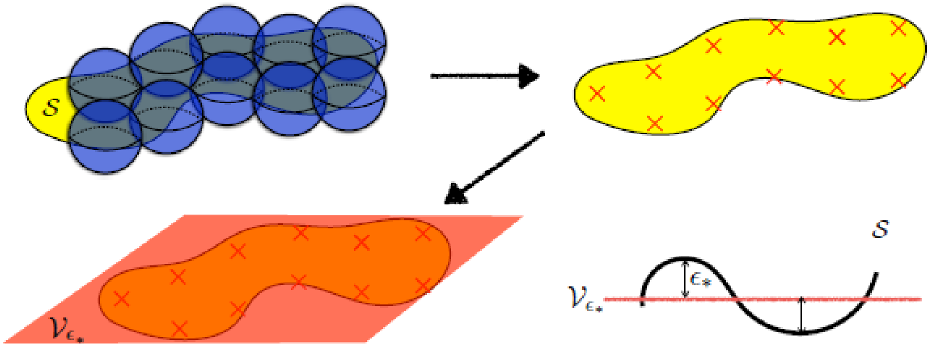


Figure 3.2: (Fig. 1. from [8], © [2017] IEEE) Figure illustrating the construction of the finite-dimensional subspace V_{ϵ_*}. In the top left, we see the set S covered by $N_S(\epsilon_*)$ closed balls of radius ϵ_*. The top right shows the corresponding minimal ϵ_*-net $C(\epsilon_*)$ for S, i.e., the set of the elements of S that are the centres of the balls covering S in the first image. The bottom images represent, respectively, the subspace V_{ϵ_*} spanned by $C(\epsilon_*)$ and the fact that V_{ϵ_*} *approximates* S with precision ϵ_* (Proposition 3.22).

ϵ_*	radius of the closed balls with centres in S needed cover S
$C(\epsilon_*)$	minimal ϵ_*-net for S
$N_S(\epsilon_*)$	cardinality of $C(\epsilon_*)$
V_{ϵ_*}	finite-dimensional subspace of H generated by $C(\epsilon_*)$
b_i	i-th element of an orthonormal basis for V_{ϵ_*}

Table 3.1: Summary of the main notations used from now on.

set, although such that its normalised secant set S satisfies *Assumption A*, the dimension d of V_{ϵ_*} may grow as ϵ_* decreases, since $N_S(\epsilon_*)$ increases.

With these ingredients and notations (summarised in Table 3.1) we can define the map b. Let b be the linear map defined as

$$\begin{aligned} b : H &\to \mathbb{R}^d \\ x &\mapsto b(x) := (\langle b_1, x\rangle, \dots, \langle b_d, x\rangle)^T. \end{aligned} \tag{3.12}$$

Note that for its construction, b is continuous with respect to the norm induced by the scalar product.

The following proposition is about the properties of b concerning the *almost* preservation of norms.

Proposition 3.22. *Let $\epsilon_* \in (0,1)$ and b be the linear map defined, as in (3.12), as*

$$\begin{aligned} b : H &\to \mathbb{R}^d \\ x &\mapsto b(x) := (\langle b_1, x\rangle, \dots, \langle b_d, x\rangle)^T, \end{aligned}$$

where $\{b_1, \dots, b_d\}$ is an orthonormal basis for the space V_{ϵ_*}, spanned by a minimal ϵ_*-net $C(\epsilon_*)$ for S.

The following apply:

$$\begin{aligned} \|b(x)\|_2 &\leq \|x\| \text{ for all } x \in H, \\ \|b(x)\|_2 &\geq 1 - \epsilon_* \text{ for all } x \in S, \end{aligned}$$

where $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^d$.

Proof. Consider the orthogonal projection $P_{V_{\epsilon_*}}$ on V_{ϵ_*}, defined as

$$\begin{aligned} P_{V_{\epsilon_*}} : H &\to V_{\epsilon_*} \\ x &\mapsto P_{V_{\epsilon_*}}(x) := \langle b_1, x\rangle b_1 + \cdots + \langle b_d, x\rangle b_d, \end{aligned}$$

and remember that for every $x \in H$ (see, e.g., [27]),

$$\langle x - P_{V_{\epsilon_*}}(x), y\rangle = 0 \text{ for all } y \in V_{\epsilon_*}, \tag{3.13}$$

and

$$\|x - P_{V_{\epsilon_*}}(x)\| \leq \|x - y\| \text{ for all } y \in V_{\epsilon_*}. \tag{3.14}$$

First we note that, for the orthonormality of the basis $\{b_1, \dots, b_d\}$,

$$\|b(x)\|_2 = \|P_{V_{\epsilon_*}}(x)\| \text{ for all } x \in H,$$

in fact, we have for all $x \in H$

$$\begin{aligned} \|P_{V_{\epsilon_*}}(x)\|^2 &= \langle P_{V_{\epsilon_*}}(x), P_{V_{\epsilon_*}}(x)\rangle \\ &= \langle b_1, x\rangle^2 \|b_1\|^2 + \cdots + \langle b_d, x\rangle^2 \|b_d\|^2 \\ &= \langle b_1, x\rangle^2 + \cdots + \langle b_d, x\rangle^2 \\ &= \|b(x)\|_2^2. \end{aligned}$$

So we have, thanks to (3.13), that for all $x \in H$

$$\begin{aligned} \|x\|^2 &= \|x - P_{V_{\epsilon_*}}(x) + P_{V_{\epsilon_*}}(x)\|^2 \\ &= \|x - P_{V_{\epsilon_*}}(x)\|^2 + \|P_{V_{\epsilon_*}}(x)\|^2 \\ &\geq \|P_{V_{\epsilon_*}}(x)\|^2 \\ &= \|b(x)\|_2^2, \end{aligned}$$

which gives the first part of the thesis.

Furthermore, from (3.14) we have that for every $x \in S \subset H$,

$$\|x - P_{V_{\epsilon_*}}(x)\| \leq \|x - y_0\| \text{ for all } y_0 \in C(\epsilon_*),$$

and so, we have that for every $x \in S$,

$$\|x - P_{V_{\epsilon_*}}(x)\| \leq \min_{y_0 \in C(\epsilon_*)} \|x - y_0\| \leq \epsilon_*,$$

where the last inequality comes from the definition of $C(\epsilon_*)$. This gives, together with the triangle inequality, that for all $x \in S$

$$\begin{aligned} \|b(x)\|_2 &= \|P_{V_{\epsilon_*}}(x) - x + x\| \\ &\geq \|x\| - \|x - P_{V_{\epsilon_*}}(x)\| \\ &\geq 1 - \epsilon_*, \end{aligned}$$

which completes the proof.

□

Proposition 3.22 above, as already mentioned, shows that b satisfies the RIP on $S \subset \mathcal{S}$, where, of course, we refer to an infinite-dimensional, non-random version of the property. In fact, the statement ensures that b satisfies

$$1 - \epsilon_* \leq \|b(x)\|_2 \leq \|x\| = 1 < 1 + \epsilon_* \text{ for all } x \in S \subset \mathcal{S},$$

i.e., it implies that b preserves the unitary norms of the elements of $S \subset \mathcal{S}$, but for a relative error of $1 \pm \epsilon_*$, when *projecting* them into $\mathbb{R}^d$, with d potentially large but finite.

Remark 3.23. It is possible to define b using a basis $\{b_1, \dots, b_d\}$ for the space V_{ϵ_*} that is not necessarily orthonormal. The linear map b defined in this way still satisfies the inequalities of Proposition 3.22, with $\mathbb{R}^d$ equipped with a norm defined *ad hoc* (which, in the orthonormal case, corresponds to the Euclidean norm). Readers interested in learning more about this should refer to [8].

3.4.2 Second step: into a lower dimension

Now that we are able to map elements from the infinite-dimensional space H to $\mathbb{R}^d$, without *too much* change in their norms, we can use random techniques to further reduce the dimension of these elements, say *projecting* them into $\mathbb{R}^m$, *almost* preserving their norms.

To do this, we randomly and independently take m vectors $a_1, \ldots, a_m \in \mathbb{R}^d$, according to a probability measure P on $\mathbb{R}^d$, and define, given the linear map $b : H \to \mathbb{R}^d$ defined in (3.12), the random linear map L as

$$\begin{aligned} L : H &\to \mathbb{R}^m \\ x &\mapsto L(x) := \frac{1}{\sqrt{m}} \left(a_1^T b(x), \ldots, a_m^T b(x) \right)^T . \end{aligned} \tag{3.15}$$

Remark 3.24. We note that this two-step construction first reduces the dimension to d, and then to m. It is clear that if we are not interested in further reducing the dimension of the elements of the set Σ, i.e., if the dimension m of the target space is equal to or greater than d, there is no need to continue with the second step. In this case, it suffices to stop at dimension d, to obtain the desired *projection*, i.e., the map b, which, as already noted in the previous section, *almost* preserves norms.

By appropriately choosing the probability measure P on $\mathbb{R}^d$, it can be shown that the map L defined in (3.15) satisfies the hypotheses of Theorem 3.17, and so it is a random linear map that, under *Assumptions* A and B, and under certain conditions on the dimension m of the arrival space, satisfies the RIP on the set $S \subset \mathcal{S}$ with constant $\delta \in (0, \underline{\delta}_S)$, with probability at least $1 - \xi$.

To prove this, we first need to introduce the following definitions, which are necessary to state a condition on the probability measure P that ensures the validity of the hypotheses of Theorem 3.17.

Definition 3.25. Let X be a random variable. The *sub-gaussian norm* of X is

$$|||X||| := \sup_{q \geq 1} \left\{ \frac{[\mathbb{E}(|X|^q)]^{\frac{1}{q}}}{\sqrt{q}} \right\} .$$

Definition 3.26. Let $a_1, \ldots, a_m$ be independent and identically distributed random vectors with d components, and b a linear map defined as in (3.12). We define Λ as the constant

$$\Lambda := \sup_{\substack{x \in \{S - S\} \cup S \\ x \neq 0}} \left\{ \frac{|||a_i^T b(x)|||}{\|b(x)\|_2} \right\} .$$

Remark 3.27. In Definition 3.26, Λ is well-defined because, for any fixed x, the sub-gaussian norm of the random variables $a_1^T b(x), \ldots, a_m^T b(x)$ is the same, as they are independent and identically distributed. This follows from the fact that these random variables are linear combinations of the components of the independent and identically distributed random vectors $a_1, \ldots, a_m$, with the same coefficients $b_1(x), \ldots, b_d(x)$.

A condition on P that ensures that the random linear map L defined in (3.15) satisfies the hypotheses of Theorem 3.17 is $\Lambda < +\infty$, as the proof of the following theorem shows.

Theorem 3.28 (Theorem III.4 of [8]). *Let $S \subset \mathcal{S} \subset H$ and L be a random linear map defined, as in* (3.15)*, as*

$$\begin{aligned} L : H &\to \mathbb{R}^m \\ x &\mapsto L(x) := \frac{1}{\sqrt{m}} \left(a_1^T b(x), \ldots, a_m^T b(x) \right)^T, \end{aligned}$$

where $a_1, \ldots, a_m$ are vectors taken independently and randomly with a probability measure P on $\mathbb{R}^d$. There exists an absolute constant $K' > 0$ such that if Assumption A holds, and

$$\Lambda = \sup_{\substack{x \in \{S - S\} \cup S \\ x \neq 0}} \left\{ \frac{\left\| \left| a_i^T b(x) \right| \right\|}{\|b(x)\|_2} \right\} < +\infty,$$

then for any $\xi \in (0,1)$ and $\delta \in (0,1)$, it holds that $\delta_S \leq \delta$, i.e.,

$$\|x\|_L^2 - \delta \leq \|L(x)\|_2^2 \leq \|x\|_L^2 + \delta$$

holds for all $x \in S$, provided that

$$m \geq K'(s, \delta, \xi) := \frac{K'}{\delta^2} \max\{8\Lambda^4, \Lambda^2\} \max \left\{ s \log \frac{1}{\epsilon_S}, \log \frac{6}{\xi} \right\},$$

with probability at least $1 - \xi$.

Remark 3.29. The theorem ensures that, if *Assumption A* holds, $\Lambda < +\infty$, and $m \geq K'(s, \delta, \xi)$, a random linear map L defined as in (3.15) satisfies the RIP on the set $S \subset \mathcal{S}$ with constant $\delta \in (0, \underline{\delta}_S)$ with probability at least $1 - \xi$.

To prove this theorem, it is necessary to provide the definitions of *sub-exponential random variable* and *sub-exponential norm.* These concepts are only required for this proof and are not central to our work. Therefore, before proceeding, we briefly introduce them, along with two related results needed in the proof, without dwelling on them. More details can be found in Section 5.2.4 of [10].

Definition 3.30 (Definition 5.13 of [10]). A *sub-exponential random variable* X is a random variable that satisfies

$$[\mathbb{E}(|X|^q)]^{\frac{1}{q}} \leq k \cdot q$$

for all $q \geq 1$, with $k > 0$.

Definition 3.31. Let X be a random variable. The *sub-exponential norm* of X is

$$\|X\|_{exp} := \sup_{q\geq 1} \left\{ \frac{[\mathbb{E}(|X|^q)]^{\frac{1}{q}}}{q} \right\}.$$

Proposition 3.32 (Remark 5.18 of [10])**.** *If X is a sub-exponential random variable, then so is $X - \mathbb{E}(X)$, and we have $\|X - \mathbb{E}(X)\|_{exp} \leq 2\|X\|_{exp}$.*

Lemma 3.33 (Corollary 5.17 of [10])**.** *There exists an absolute constant $c > 0$ such that for independent centred sub-exponential random variables $X_1, \dots, X_m$ with sub-exponential norm upper bounded by $\beta > 0$, we have*

$$P\left\{ \frac{1}{m}\left|\sum_{i=1}^m X_i\right| \geq t \right\} \leq 2 \cdot e^{-cmt^2/\beta^2}, \text{ for every } 0 \leq t \leq \beta,$$

and

$$P\left\{ \frac{1}{m}\left|\sum_{i=1}^m X_i\right| \geq t \right\} \leq 2 \cdot e^{-cmt/\beta}, \text{ for every } t \geq \beta.$$

Remark 3.34. From the proof (in [10]) it can be seen that the absolute constant c of the above lemma is $c = \frac{1}{2e}$.

We can now give the proof of Theorem 3.28.

Proof. (*of Theorem 3.28*). From

$$\Lambda = \sup_{\substack{x \in \{S-S\}\cup S \\ x \neq 0}} \left\{ \frac{|||a_i^T b(x)|||}{\|b(x)\|_2} \right\} < +\infty,$$

it follows that for all $i = 1, \dots, m$ and for all $x \in \{S - S\} \cup S, x \neq 0$,

$$\frac{|||a_i^T b(x)|||}{\|b(x)\|_2} \leq \Lambda < +\infty.$$

This gives, for Definition 3.25 of the sub-gaussian norm $|||\cdot|||$, that for all $i = 1, \dots, m$ and for all $x \in \{S - S\} \cup S, x \neq 0$,

$$\frac{\left[\mathbb{E}(|a_i^T b(x)|^q)\right]^{\frac{1}{q}}}{\sqrt{q}\|b(x)\|_2} \leq \Lambda < +\infty, \text{ for all } q \geq 1,$$

i.e., for all $i = 1, \dots, m$ and for all $x \in \{S - S\} \cup S$,

$$\left[\mathbb{E}(|a_i^T b(x)|^q)\right]^{\frac{1}{q}} \leq \Lambda\sqrt{q}\|b(x)\|_2, \text{ for all } q \geq 1. \tag{3.16}$$

Now fix $y, z \in S \cup \{0\}$, and let X be the random variable defined as

$$X \coloneqq \sum_{i=1}^{m} X_i,$$

where, for all $i = 1, \dots, m$,

$$X_i \coloneqq |a_i^T b(y)|^2 - |a_i^T b(z)|^2 - \mathbb{E}\left(|a_i^T b(y)|^2 - |a_i^T b(z)|^2\right).$$

Clearly, from its definition and the hypotheses on $a_1, \dots, a_m$, we immediately see that X is the sum of m independent centred random variables. Let us also prove that $X_1, \dots, X_m$ satisfy the others hypotheses of Lemma 3.33, so that we can apply it.

To do this, thanks to Proposition 3.32, it is sufficient to show that for all $i = 1, \dots, m$, the random variables $|a_i^T b(y)|^2 - |a_i^T b(z)|^2$, which are independent and identically distributed, are sub-exponential, with sub-exponential norm bounded by a certain constant.

Consider, therefore, any random variable $|a_i^T b(y)|^2 - |a_i^T b(z)|^2$, with $i \in \{1, \dots, m\}$. We have

$$\begin{aligned}
\left[\mathbb{E}\left(\left||a_i^T b(y)|^2 - |a_i^T b(z)|^2\right|^q\right)\right]^{\frac{1}{q}} &= \\
&= \left[\mathbb{E}\left(\left|\left(a_i^T\left(b(y)+b(z)\right)\right)\cdot\left(a_i^T\left(b(y)-b(z)\right)\right)\right|^q\right)\right]^{\frac{1}{q}} \\
&= \left[\mathbb{E}\left(\left|a_i^T b(y+z)\right|^q \cdot \left|a_i^T b(y-z)\right|^q\right)\right]^{\frac{1}{q}} \\
&\le \left[\mathbb{E}\left(\left|a_i^T b(y+z)\right|^{2q}\right)\right]^{\frac{1}{2q}} \cdot \left[\mathbb{E}\left(\left|a_i^T b(y-z)\right|^{2q}\right)\right]^{\frac{1}{2q}},
\end{aligned} \tag{3.17}$$

where we use the linearity of the map b and, in the last, the Cauchy-Schwartz inequality.

Now, since S is a normalised secant set and so it is symmetric, we have $y \pm z \in \{S - S\} \cup S$, and so we have that (3.16) holds for $y - z$ and $y + z$. More specifically, thanks to the inequalities of Proposition 3.22 satisfied by b, we have

$$\begin{aligned}
\left[\mathbb{E}\left(\left|a_i^T b(y \pm z)\right|^{2q}\right)\right]^{\frac{1}{2q}} &\le \Lambda\sqrt{2q}\cdot\|b(y \pm z)\|_2 \\
&\le \Lambda\sqrt{2q}\cdot\|y \pm z\| \\
&\le \begin{cases} \Lambda\sqrt{2q}\cdot 2 & \text{for } y+z \\ \Lambda\sqrt{2q}\cdot\|y-z\| & \text{for } y-z, \end{cases}
\end{aligned}$$

where the first case of the last inequality comes from the triangle inequality and the fact that $y, z \in S \cup \{0\} \subset \mathcal{S} \cup \{0\}$.

Using the above into (3.17), we have

$$\begin{aligned}\left[\mathbb{E}\left(\left||a_i^T b(y)|^2 - |a_i^T b(z)|^2\right|^q\right)\right]^{\frac{1}{q}} &\leq \Lambda^2 \cdot 2q \cdot 2\|y-z\| \\ &= 4\Lambda^2 q \cdot \|y-z\|,\end{aligned}$$

that shows that the random variables $|a_i^T b(y)|^2 - |a_i^T b(z)|^2$ are sub-exponential with sub-exponential norm upper bounded by $4\Lambda^2 \cdot \|y-z\|$.

As already mentioned, from Proposition 3.32 we have immediately that also the random variables $X_1, \ldots, X_m$ are sub-exponential, with sub-exponential norm

$$\|X_i\|_{exp} \leq 2 \cdot 4\Lambda^2 \cdot \|y-z\| = 8\Lambda^2 \cdot \|y-z\|,$$

and so we have that they satisfy the hypotheses of Lemma 3.33, with

$$\beta = 8\Lambda^2 \cdot \|y-z\|.$$

Remembering now the definition of the map h_L, defined in (3.4) as

$$\begin{aligned} h_L : H &\to \mathbb{R} \\ x &\mapsto h_L(x) := \|L(x)\|_2^2 - \|x\|_L^2,\end{aligned}$$

note that in this case

$$\begin{aligned} h_L(x) &= \|L(x)\|_2^2 - \mathbb{E}\left(\|L(x)\|_2^2\right) \\ &= \frac{1}{m}\sum_{i=1}^m |a_i^T b(x)|^2 - \frac{1}{m}\mathbb{E}\left(\sum_{i=1}^m |a_i^T b(x)|^2\right) \\ &= \frac{1}{m}\sum_{i=1}^m |a_i^T b(x)|^2 - \frac{1}{m}\sum_{i=1}^m \mathbb{E}\left(|a_i^T b(x)|^2\right) \\ &= \frac{1}{m}\sum_{i=1}^m \left(|a_i^T b(x)|^2 - \mathbb{E}\left(|a_i^T b(x)|^2\right)\right),\end{aligned}$$

and so

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

Therefore, thanks to Lemma 3.33, we have for every $0 \leq \lambda \leq 8\Lambda^2$,

$$\begin{aligned} P\left\{|h_L(y) - h_L(z)| \geq \lambda\|y-z\|\right\} &= P\left\{\frac{1}{m}\left|\sum_{i=1}^m X_i\right| \geq \lambda\|y-z\|\right\} \\ &\leq 2 \cdot e^{-cm\lambda^2/(64\Lambda^4)},\end{aligned}$$

and for every $\lambda \geq 8\Lambda^2$,

$$P\left\{|h_L(y) - h_L(z)| \geq \lambda\|y - z\|\right\} = P\left\{\frac{1}{m}\left|\sum_{i=1}^{m} X_i\right| \geq \lambda\|y - z\|\right\}$$
$$\leq 2 \cdot e^{-cm\lambda/(8\Lambda^2)},$$

where $c = \frac{1}{2e}$, and it is easy to see that this implies that *Assumption B* holds with

$$c_1 = \frac{c}{64\Lambda^4} \text{ and } c_2 = \frac{c}{8\Lambda^2}.$$

This prove the thesis, in fact, thanks to the fact that

$$\min\{c_1, c_2\} = \frac{c}{8} \cdot \min\left\{\frac{1}{8\Lambda^4}, \frac{1}{\Lambda^2}\right\}$$
$$= \frac{c}{8} \cdot \frac{1}{\max\{8\Lambda^4, \Lambda^2\}},$$

and since *Assumption B* holds, we can apply Theorem 3.17 and obtain that there exists an absolute constant $K > 0$ such that for any $\xi \in (0,1)$ and $\delta \in (0,1)$,

$$\delta_S \leq \delta,$$

provided that

$$m \geq \frac{K}{\min\{c_1, c_2\}\delta^2} \max\left\{s \log \frac{1}{\epsilon_S}, \log \frac{6}{\xi}\right\}$$
$$= \frac{K'}{\delta^2} \max\{8\Lambda^4, \Lambda^2\} \max\left\{s \log \frac{1}{\epsilon_S}, \log \frac{6}{\xi}\right\},$$

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

Remark 3.35. From the proof we can see that the absolute constant K' of Theorem 3.28 is

$$K' = \frac{K \cdot 8}{c},$$

where $K = 3200$ is the absolute constant of Theorem 3.17 and $c = \frac{1}{2e}$ is the absolute constant of Lemma 3.33, so

$$K' = 3200 \cdot 8 \cdot 2e.$$

As it can be understood from the discussion so far, we are interested in *projecting* elements while *almost* preserving their distances (or norms) when the arrival space is equipped with the Euclidean norm $\|\cdot\|_2$. However, it is interesting to note that the definitions and some of the theoretical results in this section and in Section 3.3 can also be restated in the case where the target space is equipped with the p-norm $\|\cdot\|_p$. In particular, Theorem 3.17 holds, with some adaptations, when $\mathbb{R}^m$ is equipped with the p-norm for every $p \in [1, +\infty)$, while a version of Theorem 3.28 can be formulated and proved – in a slightly different way – only when $\mathbb{R}^m$ is equipped with the 1-norm, by considering a random map $L : H \to \mathbb{R}^m$ defined for all $x \in H$ by $L(x) := \frac{1}{m}\left(a_1^T b(x) + \cdots + a_m^T b(x)\right)$. See [8] for more on this.

3.5 A finite-dimensional example: the RIP on *sparse* vectors

Before proceeding to study the case of interest – namely, involving a set with finite cardinality and infinite dimension that we want to *project* – let us present an example showing the application of the results from Section 3.4 in the rather simple case of a set with infinite cardinality but finite dimension.

This example, from [8], concerns the fact that matrices with Gaussian components satisfy the RIP on the set of unit $2k$-*sparse* vectors. As already mentioned in Section 3.1, this is a well-known result in the field of compressed sensing, which guarantees the uniqueness of a k-*sparse* solution to certain linear systems, i.e., the possibility of recovering a k-*sparse* signal from certain measurements made by computing *a few* scalar products.

Let us consider $k \ll n \in \mathbb{N}$ and the vector space $\mathbb{R}^n$ equipped with the Euclidean norm $\|\cdot\|_2$ induced by the canonical scalar product $\langle \cdot, \cdot \rangle$. Let $\Sigma \subset \mathbb{R}^n$ be the set of k-*sparse* vectors, i.e., the set of vectors with at most k non-zero components.

To *project* the elements of Σ without *too much* change in distances, we first compute the upper box-counting dimension of the normalised secant set $S := S(\Sigma)$ of Σ – which must be finite to apply the theory developed in Section 3.4 – and then consider a random linear map L as in (3.15), which satisfies the hypotheses of Theorem 3.28.

Clearly, the secant set of Σ is

$$\Sigma - \Sigma := \{y \in \mathbb{R}^n : y \text{ has at most } 2k \text{ non-zero components}\}$$

and so its normalised secant set is the set of the unit $2k$-sparse vectors, i.e.,

$$S := \{x \in \mathbb{R}^n : \|x\|_2 = 1 \text{ and } x \text{ has at most } 2k \text{ non-zero components}\} \subset \mathcal{S}.$$

Since it can be shown that an upper bound for the number of balls of radius ϵ

needed to cover S is $(3en/(2k\epsilon))^{2k}$ [28][2], we have that its upper box-counting dimension is $\dim_B(S) \leq 2k$.

Let us now consider the strict upper bound for the upper box-counting dimension of S that appears in *Assumption A* equal to $s = 4k$ and fix ϵ_S, which exists thanks to *Assumption A*, to $\epsilon_S = 2k/(3en)$ (note that $\epsilon_S < 1/2$). It is clear that these two constant works, in fact we have that

$$\frac{3en}{2k} \leq \frac{1}{\epsilon}, \text{ for all } \epsilon \leq \epsilon_S = \frac{2k}{3en},$$

and so

$$\begin{aligned} N_S(\epsilon) &\leq \left(\frac{3en}{2k\epsilon}\right)^{2k} \\ &= \left(\frac{3en}{2k}\right)^{2k} \cdot \epsilon^{-2k} \\ &\leq \epsilon^{-2k} \cdot \epsilon^{-2k} \\ &= \epsilon^{-s}, \end{aligned}$$

for all $\epsilon \leq \epsilon_S$.

In the case of finite dimension, the construction of the map L as in (3.15) is simpler. Using the notations from Table 3.1, we recall that in the general case, the first step in constructing L involves fixing a radius ϵ_*, finding a minimal ϵ_*-net $C(\epsilon_*)$ for S and then considering a basis $\{b_1, \dots, b_d\}$ of the finite-dimensional subspace V_{ϵ_*} generated by $C(\epsilon_*)$. However, since we are already working in finite dimension, we can skip this step by choosing an ϵ_* sufficiently small so that $\dim V_{\epsilon_*} = n$ and we can directly choose $\{b_1, \dots, b_n\}$ to be the canonical basis $\{e_1, \dots, e_n\}$, thus the map b is the identity map.

We can therefore move on to the second step in the construction of L. Let us consider a random matrix $A \in \mathbb{R}^{m \times n}$ with rows a_i^T and independent standard normal entries a_{ij}, and define L as

$$\begin{aligned} L : \mathbb{R}^n &\to \mathbb{R}^m \\ x &\mapsto L(x) := \frac{1}{\sqrt{m}} Ax. \end{aligned} \tag{3.18}$$

[2]The idea is the following: the set of unit $2k$-sparse vectors with a certain configuration, e.g., those with the first $2k$ non-zero components, can be viewed as a subset of the unit sphere in $\mathbb{R}^{2k}$. This sphere, which has integer dimension $2k-1$, can clearly be covered by $C \cdot 1/\epsilon^{2k-1}$ balls of radius ϵ, where $C > 0$. Therefore, considering all the possible configurations, in number $\binom{n}{2k}$, we have that a sufficient number of balls of radius ϵ to cover all unit $2k$-sparse vectors is $\binom{n}{2k} \cdot C \cdot 1/\epsilon^{2k-1}$, which, by working on the constant C and bounding the binomial coefficient with $(en/(2k))^{2k}$, gives the desired bound.

Since we have already proved that *Assumption A* holds, in order to apply Theorem 3.28 to this map, we only have to prove that $\Lambda < +\infty$, where in this case

$$\Lambda := \sup_{\substack{x \in \{S-S\} \cup S \\ x \neq 0}} \left\{ \frac{|||a_i^T x|||}{\|x\|_2} \right\},$$

where $|||a_i^T x|||$ is the sub-gaussian norm of $a_i^T x$ (Definition 3.25).
To do so, we use the fact that a centred normal random variable X with variance σ^2 has sub-gaussian norm $|||X||| \leq \alpha \cdot \sigma$, where $\alpha > 0$ is an absolute constant (Example 5.8 of [10]). Thanks to the results in Section 1.2.3, we have that for all $i = 1, \dots, m$ and for all $x \in \mathbb{R}^n$,

$$a_i^T x = a_{i1}x_1 + \cdots + a_{in}x_n \sim N(0, \|x\|_2^2),$$

and so for all $x \in \mathbb{R}^n$,

$$|||a_i^T x||| \leq \alpha \cdot \|x\|_2,$$

which implies

$$\Lambda \leq \alpha < +\infty.$$

Now, thanks to Theorem 3.28, we have that there exists a constant $K' > 0$ (see Remark 3.35 for the exact value of K') such that for any $\xi \in (0, 1)$ and $\delta \in (0, 1)$, the random linear map L defined in (3.18) satisfies

$$\|x\|_L^2 - \delta \leq \|L(x)\|_2^2 \leq \|x\|_L^2 + \delta \tag{3.19}$$

for all $x \in S$, provided that

$$m \geq K'(k, n, \delta, \xi) := \frac{K'}{\delta^2} \max\{8\Lambda^4, \Lambda^2\} \max\left\{4k \log \frac{3en}{2k}, \log \frac{6}{\xi}\right\},$$

with probability at least $1 - \xi$.

From inequality (3.19) it is easy to see that, if $m \geq K'(k, n, \delta, \xi)$, this map L – namely, the random matrix $A/\sqrt{m}$ – is a *projection* that *almost* preserves the distances between the elements of the set Σ, with probability $1 - \xi$. In fact, again thanks to the results of Section 1.2, we have that for all $x \in S$,

$$\begin{aligned}
\|x\|_L^2 &= \mathbb{E}\left(\|L(x)\|_2^2\right) \\
&= \mathbb{E}\left(\left\|\frac{Ax}{\sqrt{m}}\right\|_2^2\right) \\
&= \frac{1}{m} \cdot \mathbb{E}\left[(a_{11}x_1 + \cdots + a_{1n}x_n)^2 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}x_n)^2\right] \\
&= \frac{1}{m} \cdot m\|x\|_2^2 \\
&= \|x\|_2^2 \\
&= 1,
\end{aligned}$$

and so, under the above conditions, we have that

$$1-\delta \le \frac{\|Ax\|_2^2}{m} \le 1+\delta, \text{ for all } x \in S,$$

that, remembering that S is the normalised secant set of Σ, is equivalent to

$$(1-\delta)\|z_1-z_2\|_2^2 \le \frac{\|Az_1-Az_2\|_2^2}{m} \le (1+\delta)\|z_1-z_2\|_2^2, \text{ for all } z_1, z_2 \in \Sigma,$$

which is exactly the desired condition of *almost* preservation of distances.

To sum up, with this example of applying the results from Section 3.4, we have shown that if $m \ge K'(k,n,\delta,\xi)$, the random matrix $A/\sqrt{m} \in \mathbb{R}^{m\times n}$ with independent entries $N(0,m^{-1})$ satisfies the RIP (3.19) on the set of unit $2k$-sparse vectors in $\mathbb{R}^n$ (a well-known result in compressed sensing) with constant δ such that

$$0 < \delta < \underline{\delta}_S = \inf_{x\in S} \|x\|_L^2 = 1,$$

and *almost* preserves the distances between the k-sparse vectors in $\mathbb{R}^n$, with probability at least $1-\xi$.

3.6 Finite cardinality case

Since, as anticipated in Section 3.1, we are interested in the *projection* with *nearly* preserving distances of finite cardinality subsets of a certain infinite-dimensional Hilbert space, we devote this section to the case of $\Sigma \subset H$ with $|\Sigma| = n > 2$, where $(H, \langle\cdot,\cdot\rangle)$ is an infinite-dimensional Hilbert space with induced norm $\|\cdot\|$.

We derive from Theorem 3.28 a dimensionality reduction theorem which holds on this finite cardinality set Σ and which is of the same form as the JL Lemma, and compare it with the latter.

The following discussions are based on our own ideas and work and, as far as we know, have not appeared in the literature.

Let $\Sigma := \{z_1,\dots,z_n\} \subset H$ and

$$S := \left\{ \frac{z_i - z_j}{\|z_i - z_j\|} : z_i, z_j \in \Sigma, z_i \ne z_j \right\} \subset \mathcal{S}$$

be the normalised secant set of Σ, with cardinality $|S| \le n(n-1)$.

Remark 3.36. If there are pairs of vectors in Σ with the same normalised difference, then $|S| < n(n-1)$.

ϵ_*	radius of the closed balls with centres in S needed cover S
$C(\epsilon_*)$	minimal ϵ_*-net for S
$N_S(\epsilon_*)$	cardinality of $C(\epsilon_*)$
V_{ϵ_*}	finite-dimensional subspace of H generated by $C(\epsilon_*)$
b_i	i-th element of an orthonormal basis for V_{ϵ_*}

Table 3.2: Summary of the main notations used in the construction of the map b defined in (3.20) (same table as in Table 3.1).

In this context *Assumption A* holds, in fact if we consider

$$\bar{\epsilon} := \min_{x \neq y \in S} \|x - y\|,$$

we have $N_S(\epsilon) = |S| \leq n(n-1)$ for all $\epsilon < \bar{\epsilon}$, which implies $\dim_B(S) = 0 < s$, with $s := 1$.

For simplicity, let us fix the radius ϵ_* (see Table 3.2 for the notations) of the closed balls with centres in S needed to cover S, in $(0, \min\{\bar{\epsilon}, 1\})$. With this choice, of course, we have S itself as a minimal ϵ_*-net for S, so $C(\epsilon_*) = S$, with

$$|C(\epsilon_*)| = N_S(\epsilon_*) \leq n(n-1),$$

and we have (Proposition 3.37 below) that the dimension $d := \dim V_{\epsilon_*}$ of the vector space $V_{\epsilon_*} = \mathrm{span}(C(\epsilon_*)) = \mathrm{span}(S)$ satisfies $d \leq n-1$.

Proposition 3.37. *If*

$$\epsilon_* \in \left(0, \min\left\{\min_{x \neq y \in S} \|x - y\|, 1\right\}\right),$$

then the vector space V_{ϵ_*}*, i.e., the span of the minimal* ϵ_**-net* $C(\epsilon_*) = S$ *for* S*, has dimension* $\dim V_{\epsilon_*} \leq n-1$.

Proof. The set S can be written as

$$S = \left\{\pm \frac{z_i - z_j}{\|z_i - z_j\|} : i < j \in \{1, \dots, n\}\right\},$$

which shows that

$$V_{\epsilon_*} = \mathrm{span}(S) = \mathrm{span}\left\{\frac{z_i - z_j}{\|z_i - z_j\|} : i < j \in \{1, \dots, n\}\right\}.$$

Since for all $1 < i < j \leq n$, $\frac{z_i - z_j}{\|z_i - z_j\|}$ can be written as a linear combination of the two vectors $z_1 - z_i$ and $z_1 - z_j$, we have that a basis for V_{ϵ_*} is

$$\{z_1 - z_i : i \in I\},$$

where I is a subset of $\{2, \dots, n\}$, which gives the thesis. □

Remark 3.38. It is not difficult to show that if $z_1, \dots, z_n$ are independent, the dimension of V_{ϵ_*} is exactly $n - 1$.

Let now $\{b_1, \dots, b_d\}$ be an orthonormal basis for V_{ϵ_*} and let b be the linear map defined, as in (3.12), as

$$\begin{aligned} b : H &\to \mathbb{R}^d \\ x &\mapsto b(x) := (\langle b_1, x\rangle, \dots, \langle b_d, x\rangle)^T. \end{aligned} \tag{3.20}$$

As shown in Proposition 3.22, we have

$$\begin{aligned} &\|b(x)\|_2 \leq \|x\| \text{ for all } x \in H, \\ &\|b(x)\|_2 \geq 1 - \epsilon_* \text{ for all } x \in S. \end{aligned} \tag{3.21}$$

Defining now A as an $m \times d$ matrix with rows $a_1^T, \dots, a_m^T$ and entries a_{ij} independent standard normal random variables, we can consider the linear map L defined as

$$\begin{aligned} L : H &\to \mathbb{R}^m \\ x &\mapsto L(x) := \frac{Ab(x)}{\sqrt{m}} \end{aligned} \tag{3.22}$$

and apply Theorem 3.28 to this L, after proving that $\Lambda < +\infty$ (Lemma 3.40 below).

Remark 3.39. As already remarked for the general case in Section 3.4.2, if $m \geq d$ there is no point in constructing the application L, and in order to have the RIP it is sufficient to stop at the construction of b, which satisfies the inequalities

$$1 - \epsilon_* \leq \|b(x)\|_2 \leq \|x\| = 1 < 1 + \epsilon_*$$

for all $x \in S$, and so

$$(1 - \epsilon_*)\|z_i - z_j\| \leq \|b(z_i) - b(z_j)\|_2 \leq (1 + \epsilon_*)\|z_i - z_j\|$$

for all $z_i, z_j \in \Sigma$.
In particular, since we have seen that $d \leq n - 1$, it makes no sense to construct the application L if we have $m \geq n - 1$.

Lemma 3.40. *Let $a_1, \dots, a_m$ be random vectors with independent standard normal entries a_{ij} for all $i = 1, \dots, m$ and $j = 1, \dots, d$, and let b be defined as in* (3.20)*. Then*

$$\Lambda = \sup_{\substack{x \in \{S-S\} \cup S \\ x \neq 0}} \left\{ \frac{|||a_i^T b(x)|||}{\|b(x)\|_2} \right\} = \sqrt{\frac{2}{\pi}}.$$

Proof. We remark that for all $i = 1, \dots, m$ and for all $x \in H$, thanks to Proposition 1.23, the random variables

$$a_i^T b(x) = a_{i1} \cdot \langle b_1, x \rangle + \dots + a_{id} \cdot \langle b_d, x \rangle$$

have a Gaussian distribution with zero expectation and variance $\|b(x)\|_2^2$, being the sums of d independent centred Gaussian random variables with variances $\langle b_j, x \rangle^2$, for all $j = 1, \dots, d$.

It is now easy to show that for all $i = 1, \dots, m$ and for all $x \in H$,

$$|||a_i^T b(x)||| = \|b(x)\|_2 \sqrt{\frac{2}{\pi}},$$

indeed for the linearity of the expectation and for the discussion in Section 5.2.3 of [10], if X is a centred normal random variable with variance $\sigma^2 > 0$, then for all $q \geq 1$,

$$[\mathbb{E}(|X|^q)]^{\frac{1}{q}} = \sigma\sqrt{2} \left[\frac{\Gamma\left(\frac{1+q}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \right]^{\frac{1}{q}},$$

where Γ is the Gamma function, so for all $q \geq 1$,

$$\left[\mathbb{E}(|a_i^T b(x)|^q)\right]^{\frac{1}{q}} = \|b(x)\|_2 \sqrt{2} \left[\frac{\Gamma\left(\frac{1+q}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \right]^{\frac{1}{q}},$$

and (see Lemma A.2 for the last equality)

$$\begin{aligned} |||a_i^T b(x)||| &= \sup_{q \geq 1} \left\{ \frac{\left[\mathbb{E}(|a_i^T b(x)|^q)\right]^{\frac{1}{q}}}{\sqrt{q}} \right\} \\ &= \|b(x)\|_2 \sup_{q \geq 1} \left\{ \frac{\sqrt{2}}{\sqrt{q}} \left[\frac{\Gamma\left(\frac{1+q}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \right]^{\frac{1}{q}} \right\} \\ &= \|b(x)\|_2 \sqrt{\frac{2}{\pi}}. \end{aligned}$$

Therefore we have

$$\begin{aligned}\Lambda &= \sup_{\substack{x\in\{S-S\}\cup S\\ x\neq 0}} \left\{ \frac{|||a_i^T b(x)|||}{\|b(x)\|_2} \right\} \\ &= \sup_{\substack{x\in\{S-S\}\cup S\\ x\neq 0}} \left\{ \frac{\|b(x)\|_2\sqrt{\frac{2}{\pi}}}{\|b(x)\|_2} \right\} \\ &= \sqrt{\frac{2}{\pi}}.\end{aligned}$$

□

We can now state the following theorem.

Theorem 3.41. *Let $S \subset \mathcal{S} \subset H$ be the normalised secant set of $\Sigma \subset H$ with $|\Sigma| = n$, and L be a random linear map defined, as in (3.22), as*

$$\begin{aligned} L &: H \to \mathbb{R}^m \\ x &\mapsto L(x) := \frac{Ab(x)}{\sqrt{m}}, \end{aligned}$$

where A is an $m \times d$ matrix with rows $a_1^T, \dots, a_m^T$ and entries a_{ij} independent standard normal random variables. There exists an absolute constant $\hat{K} > 0$ such that for any $\xi \in (0,1)$ and $\delta \in (0,1)$,

$$\|b(x)\|_2^2 - \delta \leq \|L(x)\|_2^2 \leq \|b(x)\|_2^2 + \delta$$

holds for all $x \in S$, provided that

$$m \geq \hat{K}(n,\delta,\xi) := \frac{\hat{K}}{\delta^2} \max\left\{ \log[n(n-1)], \log\frac{6}{\xi} \right\},$$

with probability at least $1 - \xi$.

Proof. It is clear that if we prove (a) that $\|x\|_L^2 = \|b(x)\|_2^2$ for all $x \in S$ and (b) that $\epsilon_S := \frac{1}{n(n-1)}$ is well chosen, then we obtain the thesis from Theorem 3.28, having shown that *Assumption A* holds and that $\Lambda < +\infty$.

(a) For the linearity of expectation, the fact that $a_{ij} \sim N(0,1)$ are independent for all i, j and thanks to Proposition 1.25, we have, for all $x \in S$,

$$\begin{aligned}
\|x\|_L^2 &= \mathbb{E}\left[\frac{(a_1^T b(x))^2}{m} + \cdots + \frac{(a_m^T b(x))^2}{m}\right] \\
&= \mathbb{E}\left[(a_i^T b(x))^2\right] \\
&= \mathbb{E}\left[(a_{i1} \cdot \langle b_1, x\rangle + \cdots + a_{id} \cdot \langle b_d, x\rangle)^2\right] \\
&= \mathbb{E}\left[\langle b_1, x\rangle^2 \cdot a_{i1}^2 + \cdots + \langle b_d, x\rangle^2 \cdot a_{id}^2\right] \\
&= \langle b_1, x\rangle^2 \cdot \mathbb{E}(a_{i1}^2) + \cdots + \langle b_d, x\rangle^2 \cdot \mathbb{E}(a_{id}^2) \\
&= \|b(x)\|_2^2.
\end{aligned}$$

(b) Recalling that in this case $s = 1$ and for all $\epsilon \in (0,1)$, $N_S(\epsilon) \le n(n-1)$, it is easy to see that $\epsilon_S := \frac{1}{n(n-1)}$ ensures that $N_S(\epsilon) \le \epsilon^{-s}$ for all $\epsilon \le \epsilon_S$, and is therefore well chosen.

□

Remark 3.42. Note that here ϵ_S depends on the cardinality of S, and since $n > 2$, choosing $\epsilon_S = \frac{1}{n(n-1)}$ ensures that $\epsilon_S \in (0, \frac{1}{2})$.

Remark 3.43. The absolute constant $\hat{K} > 0$ of Theorem 3.41 is the product of the absolute constant $K' > 0$ from the Theorem 3.28 and the quantity $\max\{8\Lambda^4, \Lambda^2\} = 8 \cdot \frac{4}{\pi^2}$.
More specifically, it is given by

$$\hat{K} = K' \cdot 8\Lambda^4 = \frac{K \cdot 8}{c} \cdot 8\Lambda^4,$$

where $K = 3200$ and $c = \frac{1}{2e}$ are the absolute constants of Theorem 3.17 and of Lemma 3.33, so it is

$$\hat{K} = 3200 \cdot 8 \cdot 2e \cdot 8 \cdot \frac{4}{\pi^2}.$$

Remark 3.44. Since

$$\underline{\delta}_S = \inf_{x \in S} \|x\|_L^2 = \inf_{x \in S} \|b(x)\|_2^2 \ge (1 - \epsilon_*)^2,$$

we have that for any $\xi \in (0,1)$ and $\delta \in (0, (1-\epsilon_*)^2)$, L satisfies the RIP on S with constant δ, provided that $m \ge \hat{K}(n, \delta, \xi)$, with probability at least $1 - \xi$.

To compare this situation with the JL Lemma, we need a result with inequalities of the same form. To obtain this, we proceed in three steps: first we use (3.21) to make the inequalities of Theorem 3.41 independent of b, second we consider any $\delta' \in (2\epsilon_* - \epsilon_*^2, 1)$ and fix $\delta := \delta' - 2\epsilon_* + \epsilon_*^2$ to get the inequalities $1 - \delta' \leq \|L(x)\|_2^2 \leq 1 + \delta'$, for all $x \in S$, and finally we do the same as in the first part of the proof of the Probabilistic JL Lemma (Theorem 2.4) to get the desired inequalities.

Recalling that

$$(1 - \epsilon_*)^2 \leq \|b(x)\|_2^2 \leq \|x\|^2 = 1$$

holds for all $x \in S$, thanks to Theorem 3.41 we have that for any $\xi \in (0, 1)$ and $\delta \in (0, (1 - \epsilon_*)^2)$,

$$(1 - \epsilon_*)^2 - \delta \leq \|L(x)\|_2^2 \leq 1 + \delta$$

holds for all $x \in S$, provided that $m \geq \hat{K}(n, \delta, \xi)$, with probability at least $1 - \xi$.

Remark 3.45. Actually, this last result holds for any $\delta \in (0, 1)$, but we are interested in the left member being non-negative.

Consider now any $\delta' \in (2\epsilon_* - \epsilon_*^2, 1)$ and fix $\delta = \delta' - 2\epsilon_* + \epsilon_*^2$.
Since $2\epsilon_* - \epsilon_*^2 \geq 0$ for all $\epsilon_* \in (0, 1)$, with this choices we have

$$\begin{cases} (1 - \epsilon_*)^2 - \delta = 1 - 2\epsilon_* + \epsilon_*^2 - \delta = 1 - \delta' \\ 1 + \delta \leq 1 + \delta + 2\epsilon_* - \epsilon_*^2 = 1 + \delta'. \end{cases}$$

Noting now that for all $\delta' \in (0, 1)$ we have the inequalities $\sqrt{1 - \delta'} > 1 - \delta'$ and $\sqrt{1 + \delta'} < 1 + \delta'$, and that for the definition of S and the linearity of L we have

$$\begin{aligned} & 1 - \delta' \leq \|L(x)\|_2 \leq 1 + \delta', \text{ for all } x \in S \\ \iff & 1 - \delta' \leq \left\| L\left(\frac{z_i - z_j}{\|z_i - z_j\|} \right) \right\|_2 \leq 1 + \delta', \text{ for all } z_i \neq z_j \in \Sigma \\ \iff & (1 - \delta')\|z_i - z_j\| \leq \|L(z_i) - L(z_j)\|_2 \leq (1 + \delta')\|z_i - z_j\|, \text{ for all } z_i, z_j \in \Sigma, \end{aligned}$$

where in the last step if $z_i = z_j$ the inequalities are trivially valid, we get the following infinite-dimensional and probabilistic version of the JL Lemma.

Theorem 3.46 (Infinite-dimensional and probabilistic JL Lemma)**.** *Let* Σ *be the subset* $\Sigma := \{z_1, \dots, z_n\} \subset H$ *and* L *be a random linear map defined, as in* (3.22)*, as*

$$\begin{aligned} L : H &\to \mathbb{R}^m \\ x &\mapsto L(x) := \frac{Ab(x)}{\sqrt{m}}, \end{aligned}$$

	Theorem 2.4	Theorem 3.46
cardinality of Σ	n	n
upper bound for failure probability	$\xi_U \in (0,1)$	$\xi \in (0,1)$
precision	$\delta \in (0,1)$	$\delta' \in (2\epsilon_* - \epsilon_*^2, 1)$
lower bound for m	$\frac{8}{\delta^2} \log \frac{n(n-1)}{\xi_U}$	$\frac{\hat{K}}{(\delta' - 2\epsilon_* + \epsilon_*^2)^2} \max \left\{ \log[n(n-1)], \log \frac{6}{\xi} \right\}$
absolute constant	8	$3200 \cdot 8 \cdot 2e \cdot 8 \cdot \frac{4}{\pi^2}$

Table 3.3: Summary of the parameters in Theorems 2.4 (Probabilistic JL Lemma) and 3.46 (Infinite-dimensional and probabilistic JL Lemma).

where A is an $m \times d$ matrix with rows $a_1^T, \dots, a_m^T$ and entries a_{ij} independent standard normal random variables. There exists an absolute constant $\hat{K} > 0$ such that for any $\xi \in (0,1)$ and $\delta' \in (2\epsilon_* - \epsilon_*^2, 1)$,

$$(1 - \delta')\|z_i - z_j\| \leq \|L(z_i) - L(z_j)\|_2 \leq (1 + \delta')\|z_i - z_j\|$$

holds for all $z_i, z_j \in \Sigma$, provided that

$$m \geq \hat{K}(n, \delta', \xi) := \frac{\hat{K}}{(\delta' - 2\epsilon_* + \epsilon_*^2)^2} \max \left\{ \log[n(n-1)], \log \frac{6}{\xi} \right\},$$

with probability at least $1 - \xi$.

Remark 3.47. If we choose ϵ_* very close to 0, we have $\delta' - 2\epsilon_* + \epsilon_*^2 \approx \delta'$ and an almost zero left bound for δ'.

Remark 3.48. On the notation: as in the case of Theorem 2.4, in the last theorem (Infinite-dimensional and probabilistic JL Lemma), we don't use the notation as in the rest of the work, in order to highlight some facts about its proof. In particular, we use δ' instead of δ to denote the precision, to emphasise the fact that this result is obtained from Theorem 3.41 using some asymmetric inequalities, which, by choosing δ appropriately, give a symmetric result.

Now that we have a result in the same form as the JL Lemma, let us compare it with the Lemma by looking at and commenting on Table 3.3, which compares the parameters of the two results:

- the cardinality of the initial set Σ and the probability of success affect the lower bound for m equally in both cases;

- as the precision tends to its lower bound (0 or $2\epsilon_* - \epsilon_*^2$), the lower bounds for m have the same behaviour. Moreover, as already noted, taking ϵ_* very close to 0, the precision can be the same in both cases;
- for all $n > 2$ and for all $\xi_U \in (0,1)$ it holds that $\log \frac{n(n-1)}{\xi_U} > \log[n(n-1)]$ and, if $\xi_U = \xi$, $\log \frac{n(n-1)}{\xi_U} \geq \log \frac{6}{\xi}$;
- the absolute constant $\hat{K}$ of Theorem 3.46 is much larger than that of Theorem 2.4, since it is the product of $8\Lambda^4$ and the constant K' of Theorem 3.28, which is large because that theorem applies to much more general cases than the JL Lemma (cardinality of S and dimension of the initial space H potentially infinite).

Chapter 4

A first test in infinite dimension

In this chapter, with the idea of going in the direction mentioned in Section 3.1 (*projection* of finite cardinality subsets of $L^2([0,1],\mathbb{R})$ with *almost* preserving distances), we consider the case where Σ is a collection of a finite number of functions, i.e., a set with finite cardinality and infinite dimension.

Our purpose is to reduce the dimension of the elements of the set Σ, first by using a MATLAB code that exploits the theory developed in Chapter 3, and then by mapping the functions in the coefficients of the truncations of their Fourier-Legendre series, in order to study which is the best way to follow when we want to *almost* preserve the distances between these elements. We reasonably expect that the *projection* made using the random techniques from Chapter 3 is the most effective in terms of *almost* preserving distances, as it is specifically designed for this purpose, unlike the *Fourier-Legendre projection*.

In particular, we consider a simple and structured set Σ which allows us to rely on the theory for the definition of the application b in (4.2), the calculation of the coefficients of the Fourier-Legendre series and the counting of the number of distances *almost* preserved by *projecting* the elements of Σ on the first m coefficients of their Fourier-Legendre series.

All of the results we report in this chapter, and all of the MATLAB codes that we used to obtain them – reported in Appendix B and at `https://cdlab.uniud.it/software` – are new contributions to the literature.

4.1 The *JL-projection*

Consider $H = L^2([0,1],\mathbb{R})$ with the usual scalar product $\langle \cdot, \cdot \rangle$ and with the induced norm $\|\cdot\|$, and let $\Sigma \subset H$ be the set of the first n normalised Legendre

ϵ_*	radius of the closed balls with centres in S needed cover S
$C(\epsilon_*)$	minimal ϵ_*-net for S
V_{ϵ_*}	finite-dimensional subspace of H generated by $C(\epsilon_*)$
ψ_i	i-th element of an orthonormal basis for V_{ϵ_*}

Table 4.1: Summary of the main notations used in the construction of the map b defined in (4.2) (similar to Tables 3.1 and 3.2).

polynomials[1], i.e., $\Sigma := \{\phi_0, \dots, \phi_k\}$ with $k = n-1$, where ϕ_i is the normalised Legendre polynomial in $L^2([0,1],\mathbb{R})$ of degree i, for $i = 0, \dots, k$.

Let us reduce the dimension of such polynomials by using the constructions of Section 3.6.

For ease of reading, we summarise the main notations used in the first step of the construction of the *JL-projection* in Table 4.1.

The normalised secant set of Σ is

$$S = \left\{ \frac{\phi_i - \phi_j}{\sqrt{2}} : i \neq j \in \{0, \dots, k\} \right\}$$

and its cardinality is $|S| = (k+1)k = n(n-1)$.

Let us fix

$$\epsilon_* \in \left(0, \min_{x \neq y \in S} \|x - y\|\right) = (0, 1),$$

which gives $C(\epsilon_*) = S$.
A basis for $V_{\epsilon_*} = \operatorname{span}(C(\epsilon_*)) = \operatorname{span}(S)$ is

$$\mathcal{B}_{V_{\epsilon_*}} = \{\phi_0 - \phi_i : i = 1, \dots, k\},$$

which shows that the dimension of V_{ϵ_*} is $d = k = n - 1$.
This basis is not orthonormal, but to get an orthonormal basis for V_{ϵ_*} to use for constructing the map b in (4.2), it is sufficient to use the Gram-Schmidt process on $\mathcal{B}_{V_{\epsilon_*}}$ and then normalise the basis obtained, as we see in the following proposition.

Proposition 4.1. *Applying the Gram-Schmidt process to the basis $\mathcal{B}_{V_{\epsilon_*}}$ for V_{ϵ_*} gives the basis $\{\hat{\psi}_1, \dots, \hat{\psi}_k\}$, where, for all $i = 1, \dots, k$,*

$$\hat{\psi}_i = \frac{1}{i}(\phi_0 + \dots + \phi_{i-1}) - \phi_i. \tag{4.1}$$

[1]See Section A.2 in Appendix A for a brief introduction to the space $L^2([0,1],\mathbb{R})$ and to the Legendre polynomials.

Then, normalising gives the orthonormal basis $\mathcal{B} = \{\psi_1, \dots, \psi_k\}$, where, for all $i = 1, \dots, k$,

$$\psi_i = \sqrt{\frac{i}{i+1}}\hat{\psi}_i.$$

Proof. We get the first part of the thesis by induction, and the second by performing a simple calculation.

Let $i = 1$. The first step of the Gram-Schmidt process gives $\hat{\psi}_1 = \phi_0 - \phi_1$, so we have the expected result.

Now let $i > 1$. Suppose that (4.1) holds for $\hat{\psi}_1, \dots, \hat{\psi}_{i-1}$, and prove it for $\hat{\psi}_i$.
Noting that for all $j = 0, \dots, i-2$ we have

$$\frac{1}{j+1} - \sum_{k=j+1}^{i-1} \frac{1}{k(k+1)} = \frac{1}{i},$$

which is easy to demonstrate by creating a common denominator, collecting and simplifying, we have:

$$\begin{aligned}
\hat{\psi}_i &= \phi_0 - \phi_i - \sum_{j=1}^{i-1} \frac{\langle \phi_0 - \phi_i, \hat{\psi}_j \rangle}{\langle \hat{\psi}_j, \hat{\psi}_j \rangle} \hat{\psi}_j \\
&= \phi_0 - \phi_i - \sum_{j=1}^{i-1} \frac{1/j}{1+1/j} \left[\frac{1}{j} \sum_{k=0}^{j-1} \phi_k - \phi_j \right] \\
&= \phi_0 - \phi_i + \sum_{j=1}^{i-1} \frac{1}{j+1} \phi_j - \sum_{j=1}^{i-1} \frac{1}{j(j+1)} \sum_{k=0}^{j-1} \phi_k \\
&= -\phi_i + \sum_{j=0}^{i-1} \frac{1}{j+1} \phi_j - \sum_{k=0}^{i-2} \sum_{j=k+1}^{i-1} \frac{1}{j(j+1)} \phi_k \\
&= -\phi_i + \sum_{j=0}^{i-2} \frac{1}{j+1} \phi_j - \sum_{j=0}^{i-2} \sum_{k=j+1}^{i-1} \frac{1}{k(k+1)} \phi_j + \frac{1}{i} \phi_{i-1} \\
&= -\phi_i + \sum_{j=0}^{i-2} \left(\frac{1}{j+1} - \sum_{k=j+1}^{i-1} \frac{1}{k(k+1)} \right) \phi_j + \frac{1}{i} \phi_{i-1} \\
&= \sum_{j=0}^{i-2} \frac{1}{i} \phi_j + \frac{1}{i} \phi_{i-1} - \phi_i \\
&= \frac{1}{i} (\phi_0 + \dots + \phi_{i-1}) - \phi_i,
\end{aligned}$$

where in the second equality we use the inductive hypothesis and thus the fact that for all $j \leq i-1$,

$$\langle \hat{\psi}_j, \hat{\psi}_j \rangle = \frac{1}{j^2} \cdot j + 1 = 1 + \frac{1}{j}.$$

Remarking now that for all $i = 1, \dots, k$ the square of the norm of $\hat{\psi}_i$, thanks to the orthonormality of $\phi_0, \dots, \phi_i$, is

$$\langle \hat{\psi}_i, \hat{\psi}_i \rangle = \frac{1}{i^2} \cdot i + 1 = \frac{i+1}{i},$$

we get the thesis. □

Using the orthonormal basis $\mathcal{B}$ for V_{ϵ_*} obtained in Proposition 4.1, we can define – as done in (3.20) in Section 3.6 – the map b that maps each element of H into the k-upla of its scalar products with the elements of the basis $\mathcal{B}$. Let b be defined as

$$\begin{aligned} b : H &\to \mathbb{R}^k \\ f &\mapsto b(f) \coloneqq (\langle f, \psi_1 \rangle, \dots, \langle f, \psi_k \rangle)^T. \end{aligned} \tag{4.2}$$

It easy to see that for all $\phi_i \in \Sigma$, the j-th component of $b(\phi_i)$ is

$$\langle \phi_i, \psi_j \rangle = \begin{cases} 0 & \text{if } j < i \\ -\sqrt{\frac{j}{j+1}} & \text{if } j = i \\ \frac{1}{\sqrt{j(j+1)}} & \text{if } j > i, \end{cases}$$

namely $b(\phi_i)$ is the vector of the i-th coordinates of the elements of the basis $\mathcal{B}$ in the canonical basis $\{\phi_0, \dots, \phi_k\}$.

To *project* now the n normalised Legendre polynomials in dimension m, once we have their images by the map b, we need to generate a matrix $A \in \mathbb{R}^{m \times k}$ with independent standard normal entries and to perform the matrix-vector products $Ab(\phi_i)/\sqrt{m}$ for all $i = 0, \dots, k$.
It is possible to do these operations by executing the code *JL_inf_Leg*, reported in Appendix B.2, which takes as input the number n of the first normalised Legendre polynomials to consider and the dimension m of the space on which we wish to *project* them and, generating only once the random matrix A, outputs their n *projections*.

Remark 4.2. Remember that for this construction to make sense, it must be $m < k = n - 1$.

4.2 The *F-projection*

As mentioned at the beginning of the chapter, one method of dimension reduction which is very easy to apply in this particular case, is to map each function into the first m coefficients of its Fourier-Legendre series[2].

Let F be the map that do this *projection*, so

$$\begin{aligned} F &: H \to \mathbb{R}^m \\ f &\mapsto F(f) := (\langle f, \phi_0 \rangle, \ldots, \langle f, \phi_{m-1} \rangle)^T. \end{aligned}$$

It is easy to see that we can explicitly compute the *projections* via F of all the n elements of Σ. In fact, thanks to the orthonormality of the polynomials, we trivially have that for all $i = 0, \ldots, n-1$ and for all $j = 1, \ldots, m$, the j-th component of $F(\phi_i)$ is

$$\langle \phi_i, \phi_{j-1} \rangle = \begin{cases} 1 & \text{if } i = j-1 \\ 0 & \text{if } i \neq j-1. \end{cases} \tag{4.3}$$

Remark 4.3. Bearing in mind that we are interested in the case where $m < k$, i.e., where $n \geq m + 2$, we have that

$$F(\phi_i) = \begin{cases} e_{i+1} & \text{if } i = 0, \ldots, m-1 \\ 0 & \text{if } i = m, \ldots, n, \end{cases}$$

where e_i is the i-th vector of the canonical basis of $\mathbb{R}^m$, and 0 is the zero vector of $\mathbb{R}^m$.

4.3 Numerical experiments

Now that we know how to compute the *projections* to $\mathbb{R}^m$ of the first n normalised Legendre polynomials by the two methods, all that remains is to compare them. The idea is to observe which type of *projection* causes more distances between the elements of Σ to be *nearly* preserved, i.e., maintained but for a relative error of $1 \pm \delta$.

For practical reasons it is not possible for us to consider a number n of polynomials large enough to choose m greater than or equal to the theoretical lower bounds imposed by the theorems of Chapter 3 and at the same time less than $n - 1$. In fact if we choose, e.g., $\delta' - 2\epsilon_* + \epsilon_*^2 = 0.1$ and $\xi = 0.1$, to satisfy the assumptions about m of the Theorem 3.46 and that $m < n - 1$, we would

[2]See Section A.2 in Appendix A for the definition.

$$
\begin{matrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & \underbrace{0 \quad 0}_{\sqrt{2}} & & 0 & \underbrace{1 \quad 0}_{1} & & 0 & \underbrace{0 \quad 0}_{0} &
\end{matrix}
$$

Table 4.2: Matrix with columns the *F-projections* in $\mathbb{R}^5$ of the first 9 normalised Legendre polynomials $\phi_0, \dots, \phi_8$.

have to consider an n of the order of 10^9, which is of course impossible to do with our tools. Therefore, in the following experiments we consider reasonable numbers of elements in Σ and evaluate, as the dimension m of the arrival space (chosen independently of n but always less than $n-1$) varies, what happens to the distances between the elements, by *projecting* in the two ways.[3]

Actually, this choice of parameters is not only due to the practical difficulties, but also to the fact that, according to the conclusions of the experiments in Section 2.3 on *classical JL-projection*, we expect the practical results to be better than the theoretical ones.

Remark 4.4. Note that both in Section 2.3 and here we consider values of the parameter m that do not respect the lower bounds imposed by the theory, but while in Section 2.3 the focus is on how many times in 100 generations of the *projection* all of the distances between the n vectors considered are preserved but for a relative error of $1 \pm \delta$, here we want to study how much is the percentage of distances *nearly* preserved generating only once the *JL-projection*, and compare it with the percentage of distances *nearly* preserved by the *F-projection*.

As already mentioned, in the particular case of this set Σ, it is not necessary to make any experiment regarding the number of distances between the polynomials *almost maintained* by the *F-projection*, in fact, as can be easily seen by looking at the Table 4.2, we have that:

- the distance in $\mathbb{R}^m$ between any pair of images of the m polynomials that are mapped by F into a vector of the canonical basis of $\mathbb{R}^m$ is $\sqrt{2}$;

- the distance in $\mathbb{R}^m$ between the image via F of a polynomial mapped

[3]On the notation: in this section we use δ to indicate the error, even if we do experiments on Theorem 3.46, precisely because the parameters that we use are not linked as in the theorem, and δ is an error parameter independent of the others.

m	2000		1000		500		200	
n	*JL*	*F*	*JL*	*F*	*JL*	*F*	*JL*	*F*
5000	99.86	16.00	99.55	4.00	88.56	1.00	68.45	0.16
2500	99.85	63.99	97.52	15.99	89.07	3.99	67.83	0.64
1500	×	×	97.62	44.43	88.60	11.10	68.30	1.77
750	×	×	×	×	88.39	44.41	67.80	7.09
300	×	×	×	×	×	×	68.96	44.37

Table 4.3: Results obtained by running *Comp_ JL_ F_ Leg* (in Appendix B.3) with $\delta = 0.05$ and some values of n and m. Each entry of the table contains the percentage of distances between the first n normalised Legendre polynomials that are preserved by the *JL-projection* or by the *F-projection* but for a relative error of $1 \pm \delta$. The × indicate the cases where the construction of the *JL-projection* does not make sense, i.e., where $m \geq n - 1$.

by F into a vector of the canonical basis of $\mathbb{R}^m$ and the image of a polynomial mapped into the zero vector of $\mathbb{R}^m$ is exactly 1;

- the distance in $\mathbb{R}^m$ between any pair of images of the $n-m$ polynomials mapped by F into the zero element of $\mathbb{R}^m$ is exactly 0;

and so, since the distance between any pair of Σ elements is $\|\phi_i - \phi_j\| = \sqrt{2}$, we have that if $\delta < \frac{\sqrt{2}-1}{\sqrt{2}}$, the number of the distances *nearly* preserved by F is $m(m-1)/2$, and that otherwise it is $m(m-1)/2 + m(n-m)$.

Remark 4.5. Note that, due to its construction, the *F-projection* generally reduces distances and can map two different elements onto the same vector, which the *JL-projection* does not do.

On the other hand, to calculate how many distances are *nearly* preserved by the *JL-projection*, we cannot use theoretical results. The MATLAB code *Comp_ JL_ F_ Leg*, reported in Appendix B.3, takes as input the number n of the normalised Legendre polynomials we are considering, the dimension m of the arrival space, and the parameter δ which indicates the precision in distance preservation we want to have when *projecting* the polynomials in the two ways. It calculates, by calling *JL_ inf_ Leg* (in Appendix B.2) – in the case of the *JL-projection* – and by relying on the value of δ – in the case of the *F-projection* – what is the number of distances that are *nearly* preserved by the two *projections*, and gives as output their percentages.

The three Tables 4.3, 4.4 and 4.5 summarise the results of the experiments carried out with the code *Comp_ JL_ F_ Leg* (in Appendix B.3), varying n and m for different values of the precision δ. Each entry gives the percentage

m	2000		1000		500		200	
n	*JL*	*F*	*JL*	*F*	*JL*	*F*	*JL*	*F*
5000	100	16.00	100	4.00	99.84	1.00	95.53	0.16
2500	100	63.99	100	15.99	99.85	3.99	95.64	0.64
1500	×	×	100	44.43	99.82	11.10	95.33	1.77
750	×	×	×	×	99.86	44.41	95.47	7.09
300	×	×	×	×	×	×	95.71	44.37

Table 4.4: Results obtained by running *Comp_ JL_ F_ Leg* (in Appendix B.3) with $\delta = 0.1$ and some values of n and m. Each entry of the table contains the percentage of distances between the first n normalised Legendre polynomials that are preserved by the *JL-projection* or by the *F-projection* but for a relative error of $1 \pm \delta$. The $\times$ indicate the cases where the construction of the *JL-projection* does not make sense, i.e., where $m \geq n - 1$.

m	2000		1000		500		200	
n	*JL*	*F*	*JL*	*F*	*JL*	*F*	*JL*	*F*
5000	100	64.00	100	36.00	100	19.00	100	7.84
2500	100	96.01	100	64.01	100	36.01	100	15.36
1500	×	×	100	88.90	100	55.57	100	24.90
750	×	×	×	×	100	88.92	100	46.25
300	×	×	×	×	×	×	100	88.96

Table 4.5: Results obtained by running *Comp_ JL_ F_ Leg* (in Appendix B.3) with $\delta = 0.3$ and some values of n and m. Each entry of the table contains the percentage of distances between the first n normalised Legendre polynomials that are preserved by the *JL-projection* or by the *F-projection* but for a relative error of $1 \pm \delta$. The $\times$ indicate the cases where the construction of the *JL-projection* does not make sense, i.e., where $m \geq n - 1$.

of distances between the first n normalised Legendre polynomials that are preserved but for a relative error of $1 \pm \delta$ by the *JL-projection* or by the *F-projection*, and some entries have an $\times$ because, as already remarked, the construction of the *JL-projection* does not make sense in cases where $m \geq n-1$.

Remembering that the results for the *JL-projection* are random, while those for the *F-projection* are obtained by simply calculating the value $m(m-1)/2$ or $m(m-1)/2 + m(n-m)$ (depending on the value of δ) and the percentage, we can now conclude with some observations on the results obtained:

- due to the fact that $\delta = 0.05$ and $\delta = 0.1$ are both less than $\frac{\sqrt{2}-1}{\sqrt{2}}$, the percentages of distances *nearly* preserved by the *F-projection* are the same in the first two cases, while they are greater in the case of $\delta = 0.3 > \frac{\sqrt{2}-1}{\sqrt{2}}$;

- it is clear that the *F-projection* is more suitable when we do not require too much precision and when the number n of polynomials to be *projected* is not too large compared to the dimension of the arrival space m, since the percentage of distances *nearly* preserved by this *projection*, as is obvious from the theory, increases as n decreases, with m being equal;

- as we expected, for each value of the parameter δ considered, the *JL-projection* is always better than the *F-projection* in terms of *almost* preservation of the distances. Moreover, the results obtained for the *JL-projection*, in particular those for $\delta = 0.1$ and $\delta = 0.3$, show that in this case we have a better behaviour of the *projection* in practice than what is predicted by the theory, which imposes that m respects a lower bound much larger than the values of m fixed in these experiments;

- finally, we note that the tables suggest that there may be a relationship between the dimension of the arrival space and the percentage of distances that are *nearly* preserved by the *JL-projection*. In particular, these experiments show that the percentage of success may depend only on the dimension of the arrival space m and be independent of the number of elements considered n. We do not address this issue in this work, but we reserve to do so in the future.

Conclusions and future developments

The main message this thesis aims to convey is the following: the theoretical lower bound for the dimension m of the target space in the JL Lemma is much higher than the m that can be used in practice to obtain good results in terms of *almost* preserving distances, at least in the finite case, with promising prospects even in the infinite one.

In particular, the experiments in Section 2.3 (conducted using the MATLAB code *JL* in Appendix B.1) show that even with an m considerably smaller than the minimum imposed by the theory – even 10 times smaller – most of the time when we *project* n vectors of dimension d with a random matrix in dimension m, all of the distances between vectors are maintained, except for a relative error of $1 \pm \delta$, for some values of δ in the range $[0.1, 0.2]$. This allows us to *trust* that we can generate the random *projection* only once when we want to *almost* preserve distances, even if we aim to *project* into a much lower dimensional space.

In addition, by comparing the infinite and finite-dimensional versions of the probabilistic JL Lemma (Section 3.6), we observe that the version of the Lemma that leads us to *project* elements of infinite dimension into $\mathbb{R}^m$, apart from the multiplicative constant in the lower bound for m, does not differ significantly from the classical version. This clearly makes sense, as the constant depends on the fact that the infinite-dimensional and probabilistic JL Lemma applies to much more general cases – and partly also on the construction of the *projection* – and suggests that even in this case we could achieve very good practical results as well.

The numerical experiments in Chapter 4 (MATLAB codes in Sections B.2 and B.3) show an excellent behaviour of the *JL-projection* from the infinite dimensional space $L^2([0,1], \mathbb{R})$ to $\mathbb{R}^m$, even with an m much lower than its theoretical lower bound, in the specific case of the normalised Legendre polynomials. More specifically, our experiments show that *projecting* the polynomials using the random techniques of the infinite-dimensional JL Lemma is always better in terms of *almost* distance preservation than *projecting* them onto the first m coefficients of their Fourier-Legendre series, considering *particularly low* values of m and relative errors of $1 \pm \delta$ with $\delta = 0.1, 0.2, 0.3$. Furthermore, these experiments lead to the hypothesis that there may be no correlation between the percentage of distances *nearly* preserved by the random *projection* and the number of polynomials considered, but as already mentioned, we not address this issue, which provides a good starting point for future work.

However, the experiments presented in Chapter 4 are not intended to suggest that performing *projections* that *almost* preserve distances from an infinite-dimensional space into a finite one is easy. In fact, in the general case of non-structured finite subsets Σ of an infinite-dimensional Hilbert space, it is not straightforward to construct the *projection*, and in particular, it is difficult to find an (orthonormal) basis for the space V_{ϵ_*}, i.e., the space generated by the normalised secant set of Σ.

We aim to overcome this problem in the future, as our work is driven by the interest in *projecting* generic functions from the Hilbert space $\mathbb{R}^d \times L^2([-\tau, 0], \mathbb{R}^d)$. This space plays an important role in the study of retarded functional differential equations[1] [14, 15, 16], as, in certain cases, it serves as the domain for certain infinite-dimensional operators (see, for instance, Section 2.4 of [14] or [29]), which are strictly connected to the infinite-dimensional dynamical system generated by the equation, and whose spectra (eigenvalues, singular values, etc.) are closely related to the stability of the solutions to the equation.

Our long term objective is to *project*, by leveraging the developed theory, elements of the space $\mathbb{R}^d \times L^2([-\tau, 0], \mathbb{R}^d)$, in order to *reduce* these infinite-dimensional operators and use the spectra of the *reduced* operators for the stability analysis of the dynamical system. In this thesis we present in fact infinite-dimensional experiments on the space $L^2([0, 1], \mathbb{R})$, with the intention of extending them in the future to the more general case of the space $\mathbb{R}^d \times L^2([-\tau, 0], \mathbb{R}^d)$.

In conclusion, this thesis highlights the discrepancy between the theoretical constraints and the practical results in applying the JL Lemma for dimensionality reduction with *almost* preservation of distances. The experimental results show that it is possible to achieve excellent practical outcomes, even when the dimension of the arrival space is much smaller than the theoretical lower bound, thus suggesting promising applications in both practical and theoretical contexts. Despite the difficulty in constructing *projections* from infinite-dimensional spaces, our experiments pave the way for significant future developments, particularly in the field of retarded functional differential

[1]The constant $\tau > 0$ is the *delay*.

equations, where the stability of solutions is closely related to the spectra of infinite-dimensional operators.

In the future, it will be crucial to overcome the difficulties associated with constructing *projections* in arbitrary spaces, so that dimensionality reduction techniques can be more widely applied in real-world contexts. The theory developed in this thesis, combined with numerical experiments, represents the beginning of a body of work aimed at extending the dimensional reduction with *almost* preservation of distances to other spaces and applications, with the ultimate goal of leveraging it in the stability analysis of infinite-dimensional dynamical systems.

Appendix A

Useful definitions and results

In this appendix we present some definitions and results that are useful for understanding some of the proofs or discussions in this thesis.

A.1 Auxiliary results

In this section we give some useful estimations of the sums Q_1, Q_2 and Q_3 defined in Lemma 3.19 and involved in the proof of Theorem 3.17, and state and trace the proof of Lemma A.2, which is useful in the proof of Lemma 3.40.

Lemma A.1. *Let $S \subset \mathcal{S} \subset H$ be a subset for which Assumption A holds, $\xi \in (0,1)$ and Q_1, Q_2 and Q_3 be defined as*

$$Q_1(\epsilon_S, \xi) \coloneqq \sqrt{\log\left[\frac{2}{\xi} N_S(\epsilon_S)\right]},$$

$$Q_2(\epsilon_S, \xi) \coloneqq \sum_{j=0}^{+\infty} 2^{-j+1} \sqrt{\log\left[\frac{2^{j+1}}{\xi} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right]},$$

$$Q_3(\epsilon_S, \xi) \coloneqq \sum_{j=0}^{+\infty} 2^{-j+1} \log\left[\frac{2^{j+1}}{\xi} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right].$$

The following apply:

- $Q_1 \leq \sqrt{\log(2/\xi)} + \sqrt{s \log(1/\epsilon_S)}$;
- $Q_1^2 \leq \log(2/\xi) + s \log(1/\epsilon_S)$;
- $Q_2 \leq 8\sqrt{\log(2/\xi)} + 8\sqrt{2s\log(2)} + 4\sqrt{2s\log(1/\epsilon_S)}$;
- $Q_3 \leq 8\log(2/\xi) + 16s\log(2) + 8s\log(1/\epsilon_S)$.

Proof. We first note (see, e.g., Section 4.2.3 in [30]) that

$$\sum_{j=0}^{+\infty} j \cdot 2^{-j} = 2.$$

We have

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

and similarly,

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

We also have,

$$\begin{aligned}
\sum_{j=0}^{+\infty} 2^{-j}\sqrt{\log\left(2^{2(j+1)s} \cdot \epsilon_S^{-2s}\right)} &= \sum_{j=0}^{+\infty} 2^{-j}\sqrt{2(j+1)s\log(2) + 2s\log(1/\epsilon_S)} \\
&\leq \sqrt{2s\log(2)}\sum_{j=0}^{+\infty} 2^{-j}\sqrt{j+1} + 2\sqrt{2s\log(1/\epsilon_S)} \\
&\leq \sqrt{2s\log(2)}\sum_{j=0}^{+\infty} 2^{-j}(j+1) + 2\sqrt{2s\log(1/\epsilon_S)} \\
&= 4\sqrt{2s\log(2)} + 2\sqrt{2s\log(1/\epsilon_S)},
\end{aligned}$$

and similarly,

$$\sum_{j=0}^{+\infty} 2^{-j}\log\left(2^{2(j+1)s} \cdot \epsilon_S^{-2s}\right) \leq 8s\log(2) + 4s\log(1/\epsilon_S).$$

Now for the definition of ϵ_S, which comes from *Assumption A*, we have

$$\begin{aligned} Q_1^2 &= \log\left[\frac{2}{\xi} N_S(\epsilon_S)\right] \\ &\leq \log(2/\xi) + \log(\epsilon_S^{-s}) \\ &= \log(2/\xi) + s\log(1/\epsilon_S), \end{aligned}$$

which implies

$$Q_1 \leq \sqrt{\log(2/\xi)} + \sqrt{s\log(1/\epsilon_S)},$$

and we have, for all $j \in \mathbb{N}$,

$$N_S\left(\frac{\epsilon_S}{2^{j+1}}\right) \leq \left(\frac{\epsilon_S}{2^{j+1}}\right)^{-s} = 2^{(j+1)s} \cdot \epsilon_S^{-s}.$$

Thanks to the last inequalities and the estimations above, we have so

$$\begin{aligned} Q_2 &= \sum_{j=0}^{+\infty} 2^{-j+1} \sqrt{\log\left[\frac{2^{j+1}}{\xi} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right]} \\ &\leq 2\sum_{j=0}^{+\infty} 2^{-j} \sqrt{\log\left(2^{j+1}/\xi\right) + \log\left(2^{2(j+1)s} \cdot \epsilon_S^{-2s}\right)} \\ &\leq 2\sum_{j=0}^{+\infty} 2^{-j} \left[\sqrt{\log\left(2^{j+1}/\xi\right)} + \sqrt{\log\left(2^{2(j+1)s} \cdot \epsilon_S^{-2s}\right)}\right] \\ &\leq 8\sqrt{\log(2/\xi)} + 8\sqrt{2s\log(2)} + 4\sqrt{2s\log(1/\epsilon_S)}, \end{aligned}$$

and similarly,

$$\begin{aligned} Q_3 &= \sum_{j=0}^{+\infty} 2^{-j+1} \log\left[\frac{2^{j+1}}{\xi} N_S^2\left(\frac{\epsilon_S}{2^{j+1}}\right)\right] \\ &\leq 8\log(2/\xi) + 16s\log(2) + 8s\log(1/\epsilon_S). \end{aligned}$$

□

Lemma A.2. *Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined for all $x \in \mathbb{R}$ by*

$$f(x) = \sqrt{\frac{2}{x}} \left[\frac{\Gamma\left(\frac{1+x}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\right]^{\frac{1}{x}},$$

where Γ is the Gamma function. The maximum of f in $[1, +\infty)$ is

$$\max_{x\geq 1} f(x) = f(1) = \sqrt{\frac{2}{\pi}}.$$

Proof. (Trace) If we prove that the continuous function f is monotonically decreasing in $[1,+\infty)$, we have that it assumes its maximum at $x=1$, which is

$$f(1)=\sqrt{2}\frac{\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)}=\sqrt{\frac{2}{\pi}}.$$

For doing this, since the logarithm maintains monotony, we show that the function $g(x) \coloneqq \log f(x)$ has negative derivative for all $x\in[1,+\infty)$.

It is easy to see that the derivative of the function

$$g(x)=\log\sqrt{\frac{2}{x}}+\frac{1}{x}\log\frac{\Gamma\left(\frac{1+x}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}=-\frac{1}{2}\log\frac{x}{2}+\frac{1}{x}\log\frac{\Gamma\left(\frac{1+x}{2}\right)}{\sqrt{\pi}}$$

is

$$\begin{aligned}g'(x)=&-\frac{1}{2}\cdot\frac{2}{x}\cdot\frac{1}{2}-\frac{1}{x^2}\cdot\log\frac{\Gamma\left(\frac{1+x}{2}\right)}{\sqrt{\pi}}+\frac{1}{x}\cdot\frac{\sqrt{\pi}}{\Gamma\left(\frac{1+x}{2}\right)}\cdot\frac{1}{\sqrt{\pi}}\cdot\Gamma'\left(\frac{1+x}{2}\right)\cdot\frac{1}{2}\\=&-\frac{1}{2x}-\frac{1}{x^2}\cdot\log\frac{\Gamma\left(\frac{1+x}{2}\right)}{\sqrt{\pi}}+\frac{1}{2x}\cdot\frac{\Gamma'\left(\frac{1+x}{2}\right)}{\Gamma\left(\frac{1+x}{2}\right)}\\=&-\frac{1}{2x}-\frac{1}{x^2}\cdot\log\frac{\Gamma\left(\frac{1+x}{2}\right)}{\sqrt{\pi}}+\frac{1}{2x}\cdot\psi\left(\frac{1+x}{2}\right),\end{aligned}$$

where

$$\psi(x)=\frac{\Gamma'(x)}{\Gamma(x)}$$

is the digamma function [31].

To show that $g'(x)<0$ for all $x\in[1,+\infty)$, it is sufficient to show that:

(a) $g'(1)<0$

(b) g' is monotonically increasing in $[1,+\infty)$

(c) $\lim_{x\to\infty} g'(x)=0$.

Remembering that $\psi(1)=-\gamma$ [31], where γ is the Euler-Mascheroni constant, it is easy to see that

$$g'(1)=-\frac{1}{2}-\log\frac{\Gamma(1)}{\sqrt{\pi}}+\frac{1}{2}\psi(1)=-\frac{1}{2}-\log\frac{1}{\sqrt{\pi}}-\frac{1}{2}\gamma<0.$$

Also (b) is not hard to show: recalling that the Gamma function in $(0,+\infty)$ is a *logarithmically convex* function [32], i.e., is such that the composition of the logarithm with it is a convex function [33], and using some properties of logarithmically convex functions [33], we have that $f(x)$ is also a logarithmically convex function in this interval, so that $g''(x) < 0$ for all $x \in (0, +\infty)$ and that g' is monotonically increasing in the desired interval.

Finally, to prove (c) and thus the thesis, it is sufficient to use the Stirling's approximation [34] and note that Γ and ψ are both definitively monotonically increasing functions in order to have that the second and the third terms of g' tend to zero, as x increases. □

A.2 Legendre polynomials and Fourier-Legendre series

In this section we give some definitions and results that are useful for understanding the discussions in Chapter 4 on the Legendre polynomials.

Definition A.3. The space $L^2([a,b],\mathbb{R})$ is the space of measurable functions from the interval $[a,b] \subset \mathbb{R}$ to $\mathbb{R}$ with second power having finite integral, i.e.,

$$L^2([a,b],\mathbb{R}) \coloneqq \left\{ f : [a,b] \to \mathbb{R} : f \text{ measurable}, \int_{[a,b]} f(x)^2 dx < +\infty \right\}.$$

Definition A.4. The *usual scalar product* $\langle \cdot, \cdot \rangle$ in $L^2([a,b],\mathbb{R})$ is defined for all $f, g \in L^2([a,b],\mathbb{R})$ by

$$\langle f, g \rangle \coloneqq \int_{[a,b]} f(x) \cdot g(x) dx.$$

Definition A.5. The *2-norm* $\|\cdot\|_2$, or simply $\|\cdot\|$, in $L^2([a,b],\mathbb{R})$ is the induced norm from the usual scalar product, and it is defined for all $f \in L^2([a,b],\mathbb{R})$ by

$$\|f\| \coloneqq \sqrt{\langle f, f \rangle} = \left(\int_{[a,b]} f(x)^2 dx \right)^{\frac{1}{2}}.$$

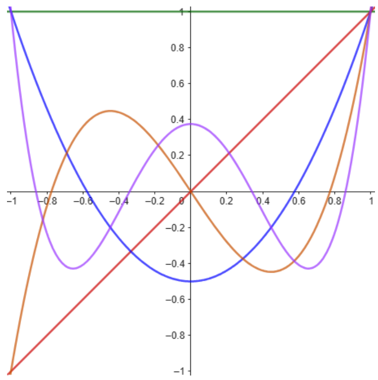
Proposition A.6. *$(L^2([a,b],\mathbb{R}), \langle \cdot, \cdot \rangle)$ is a Hilbert space.*

Definition A.7 ([35])**.** The *Legendre polynomial of degree n*, for $n = 0, 1, \ldots,$ is $\Phi_n \in L^2([-1,1],\mathbb{R})$ defined for all $x \in [-1,1]$ by

$$\Phi_n(x) \coloneqq \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n.$$

Proposition A.8. *The Legendre polynomials are orthogonal and form a basis for the space $L^2([-1,1],\mathbb{R})$. In particular , for all $n, m = 0, 1, \ldots,$ we have [35]*

$$\langle \Phi_n, \Phi_m \rangle = \int_{[-1,1]} \Phi_n(x) \cdot \Phi_m(x) dx = \begin{cases} 0 & \text{if } n \neq m \\ \frac{2}{2n+1} & \text{if } n = m. \end{cases}$$



$$\Phi_0(x) = 1$$

$$\Phi_1(x) = x$$

$$\Phi_2(x) = \tfrac{1}{2}(3x^2 - 1)$$

$$\Phi_3(x) = \tfrac{1}{2}(5x^3 - 3x)$$

$$\Phi_4(x) = \tfrac{1}{8}(35x^4 - 30x^2 + 3).$$

Figure A.1: Graphs and expressions of the first five Legendre polynomials.

Definition A.9. The *normalised Legendre polynomial in* $L^2([0,1],\mathbb{R})$ *of degree* n, for $n = 0, 1, \ldots,$ is $\phi_n \in L^2([0,1],\mathbb{R})$ defined for all $x \in [0,1]$ by

$$\phi_n(x) = \sqrt{2n+1} \cdot \Phi_n(2x-1).$$

Proposition A.10. *The normalised Legendre polynomials in* $L^2([0,1],\mathbb{R})$ *are orthonormal and form a basis for the space* $L^2([0,1],\mathbb{R})$.

Definition A.11. The *Fourier-Legendre series* of a function $f \in L^2([0,1],\mathbb{R})$ is defined as

$$\sum_{n=0}^{+\infty} \langle f, \phi_n \rangle \phi_n,$$

where for $n = 0, 1, \ldots$ the coefficient $\langle f, \phi_n \rangle$ is called the $n+1$*-th coefficient of the Fourier-Legendre series of* f.

Appendix B

MATLAB codes

This appendix contains the MATLAB codes used in this work, which are also available at `https://cdlab.uniud.it/software`.

B.1 JL

```
function pres=JL(n,d,m,delta)
%JL number of JL-projections that almost preserve all distances.
%   pres=JL(n,d,m,delta) computes how many random projections preserve all the
%   n(n-1)/2 distances between n random vectors of dimension d, but for a relative
%   error between 1-delta and 1+delta, when project in dimension m.
%   INPUT:
%    n: number of random vectors (1x1)
%    d: dimension of the initial space (1x1)
%    m: dimension of the arrival space (1x1)
%    delta: precision (1x1)
%   OUTPUT:
%    pres: number of JL-projections that almost preserve all the distances (1x1)

pres=0;
B=rand(d,n);
for k=1:100
    P=randn(m,d)/sqrt(m);
    J=P*B;
    num=0;
    for i=1:n-1
        for j=i+1:n
            num=num+(abs(norm(J(:,i)-J(:,j))/norm(B(:,i)-B(:,j))-1)<=delta);
        end
    end
    if num==n*(n-1)/2
        pres=pres+1;
    end
end
```

B.2 JL_inf_Leg

```
function J=JL_inf_Leg(n,m)
%JL_INF_LEG infinite JL for normalised Legendre polynomials.
%   J=JL_INF_LEG(n,m) maps in R^m via the JL-projection the first n normalised
%   Legendre polynomials from [0,1] to R.
%   INPUT:
%    n: number of normalised Legendre polynomials (1x1)
%    m: dimension of the arrival space (1x1)
%   OUTPUT:
%    J: matrix with projections as columns (mxn)

B=zeros(n-1,n);
for j=1:n-1
    B(j,j+1)=-sqrt(j/(j+1));
    B(j,1:j)=1/sqrt(j*(j+1));
end
J=randn(m,n-1)*B/sqrt(m);
end
```

B.3 Comp_JL_F_Leg

```
function [percentJL,percentF]=Comp_JL_F_Leg(n,m,delta)
%COMP_JL_F_LEG comparison of JL-projection and F-projection.
%   [percentJL,percentF]=COMP_JL_F_LEG(n,m,delta) computes the percentages of
%   distances preserved but for a relative error between 1-delta and 1+delta by the
%   JL-projection and by the F-projection when project the first n normalised
%   Legendre polynomials from [0,1] to R. It is based on the theory to calculate
%   the percentage relative to the F-projection.
%   INPUT:
%    n: number of normalised Legendre polynomials (1x1)
%    m: dimension of the arrival space (1x1)
%    delta: precision (1x1)
%   OUTPUT:
%    percentJL: percentage of distances preserved by JL (1x1)
%    percentF: percentage of distances preserved by F (1x1)

numJL=0;
numF=0;
J=JL_inf_Leg(n,m);
for i=1:n-1
    for j=i+1:n
        numJL=numJL+(abs(norm(J(:,i)-J(:,j))/sqrt(2)-1)<=delta);
    end
end
if delta>=(sqrt(2)-1)/sqrt(2)
    numF=m*(m-1)/2+m*(n-m);
else
    numF=m*(m-1)/2;
end
percentJL=200*numJL/(n*(n-1));
percentF=200*numF/(n*(n-1));
end
```

Bibliography

[1] W. B. Johnson, J. Lindenstrauss. *Extensions of Lipschitz mappings into a Hilbert space*. Contemporary Mathematics, https://doi.org/10.1090/conm/026/737400, pp. 189-206. 1984.

[2] D. Lin. *Compressed sensing*. Northeastern University, Khoury College of Computer Sciences, Seminar in Theoretical Computer Science, pp. 6-7. 2023.

[3] L. N. Trefethen, D. Bau III. *Numerical linear algebra*. Philadelphia, Society for Industrial and Applied Mathematics. 1997.

[4] N. Kutz. *Applied Linear Algebra: Randomized Linear Algebra*. https://www.youtube.com/watch?v=mQlzdDTRJpU. 2021.

[5] N. Kutz. *Applied Linear Algebra: Implementing randomized LA*. https://www.youtube.com/watch?v=dyeSHiwntxY. 2021.

[6] M. Weinberg. *Lecture 9: Dimensionality Reduction and the Johnson-Lindenstrauss Lemma*. Princeton University, Course: Advanced Algorithm Design, par. 1-3. 2018.

[7] S. Dasgupta, A. Gupta. *An Elementary Proof of a Theorem of Johnson and Lindenstrauss*. Random Structures & Algorithms, vol. 22(1), pp. 60–65 (2003), https://doi.org/10.1002/rsa.10073, https://cseweb.ucsd.edu/~dasgupta/papers/jl.pdf, par. 1-2. 2002.

[8] G. Puy, M. E. Davies, R. Gribonval. *Recipes for Stable Linear Embeddings From Hilbert Spaces to* $\mathbb{R}^m$. IEEE Transactions on Information Theory, vol. 63, no. 4, pp. 2171-2187, doi: 10.1109/TIT.2017.2664858, https://arxiv.org/abs/1509.06947. 2017.

[9] *Compressed sensing*. https://en.wikipedia.org/wiki/Compressed_sensing. Visited in Jan 2025.

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

[11] K. Bryan, T. Leise. *Making Do with Less: An Introduction to Compressed Sensing*. SIAM Review, vol. 55, no. 3, JSTOR, `http://www.jstor.org/stable/24248565`, pp. 547-66. 2013.

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

[13] B. Adcock, A. C. Hansen. *Generalized Sampling and Infinite-Dimensional Compressed Sensing*. Found Comput Math 16, 1263-1323 (2016), `https://doi.org/10.1007/s10208-015-9276-6`. 2015.

[14] R. F. Curtain, H. J. Zwart. *An Introduction to Infinite-Dimensional Linear System Theory*. Springer New York, NY, `https://link.springer.com/book/10.1007/978-1-4612-4224-6`. 1996.

[15] J. K. Hale, S. M. Verduyn Lunel. *Introduction to functional differential equations*. Applied Mathematical Sciences, no. 99, Springer Verlag, New York, second edition. 1993.

[16] O. Diekmann et al. *Delay Equations - Functional, Complex and Nonlinear Analysis*. New York etc.: Springer-Verlag. 1995.

[17] M. Lefebvre. *Basic Probability Theory with Applications*. Springer-Verlag New York Inc. 2012.

[18] L. Pace. *Lezioni di Statistica I*. University of Udine, Course: Statistics I, lecture notes n. 2-16. 2021.

[19] M. J. Wainwright. *Basic tail and concentration bounds* in *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, `https://www.stat.berkeley.edu/%7Emjwain/stat210b/Chap2_TailBounds_Jan22_2015.pdf`, pp. 21. 2019.

[20] *Johnson–Lindenstrauss lemma*. `https://en.wikipedia.org/wiki/Johnson-Lindenstrauss_lemma`. Visited in Jan 2025.

[21] K. G. Larsen, J. Nelson. *Optimality of the Johnson - Lindenstrauss lemma*. `FOCS`. 2017.

[22] N. Alon. *Problems and results in extremal combinatorics - I.* Discrete Mathematics. 2003.

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

[24] *Minkowski–Bouligand dimension.* https://en.wikipedia.org/wiki/Minkowski%E2%80%93Bouligand_dimension. Visited in Jan 2025.

[25] V. Arifi. *Una introduzione alla dimensione frattale.* University of Padova, degree course in Mathematics, pp. 13-20. 2024.

[26] J. Theiler. *Estimating fractal dimension.* J. Opt. Soc. Am. A 7, pp. 1055-1073. 1990.

[27] D. G. Luenberger. *Optimization by Vector Space Methods.* New York, Wiley, https://sites.science.oregonstate.edu/~show/old/142_Luenberger.pdf. 1997.

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

[29] D. Breda. *Nonautonomous delay differential equations in Hilbert spaces and Lyapunov exponents.* Differential and Integral Equations, vol. 23 (9-10), 935-956. 2010.

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

[31] *Polygamma function.* https://en.wikipedia.org/wiki/Polygamma_function. Visited in Jan 2025.

[32] M. Merkle. *Logarithmic Convexity and Inequalities for the Gamma Function.* Journal of Mathematical Analysis and Applications 203, article n. 0385, pp. 369-380 (1996). 1995.

[33] *Logarithmically convex function.* https://en.wikipedia.org/wiki/Logarithmically_convex_function. Visited in Jan 2025.

[34] *Stirling's approximation.* https://en.wikipedia.org/wiki/Stirling%27s_approximation. Visited in Jan 2025.

[35] G. C. Barozzi. *Ortogonalità.* University of Bologna, lecture notes, www.ciram.unibo.it/âĹijbarozzi/Orto, pp. 56-57. 2004.