Asymmetric-Kernel Diffusion Maps, Inverse Problems, and Applications

Alvaro Almeida Gomez

Instituto Nacional de Matemática Pura e Aplicada

A thesis submitted for the degree of

Doctor of Philosophy

Rio de Janeiro 2021

In memory of our loved ones who are no longer with us

Acknowledgments

First of all, I would like to thank God, the force that moves this universe and that allows me to enjoy this path called life. I also want to dedicate this research to my family for their support in difficult moments. I wish to express my deepest gratitude to the loved ones who were with me in this earthly world, but that I will soon find them in another dimension.

I sincerely thank Professor Jorge Zubelli for his patience, support, valuable guidance, reflections, and especially for his good energies. I do not have enough words to describe all the good things that he has done for me. I hope God gratifies his soul for all that he has done.

I would like to thank Professor Antônio José da Silva Neto for his wise words, good advice, for the contributory comments, and also for his excellent charisma every time. I hope that God bless him for his help.

I also want to thank all the people with whom I have had the opportunity to share part of this earthly life with them.

I want to thank IMPA for all the support and help, they have given me over the years, it will be impossible to forget the moments I lived in this great institution.

Finally, I want to thank CNPq for the financial support, which was essential to developing this project.

Abstract

The purpose of this thesis is to study and extend the theory of diffusion-maps, and the regularization methods associated with the inverse problem of the growth fragmentation equation for models involving asymmetric kernels.

We extend the diffusion-map formalism to datasets that are induced by asymmetric kernels. Analytical convergence results of the resulting expansion are proved, and an algorithm is proposed to perform the dimensional reduction of datasets lying in a high dimensional space. We compare our results with those obtained in [6, 18, 19, 52], and illustrate the efficiency of the algorithms with synthetic data, as well as with real data from applications including climate change studies.

Based on the asymmetric kernel model, we recover the Riemannian gradient of a given function defined on interior points of a Riemannian submanifold in the Euclidean space based on a sample of function evaluations at points in the submanifold. This approach is based on the estimates of the Laplace-Beltrami operator proposed in the diffusion-maps theory. The Riemannian gradient estimates do not involve differential terms. Analytical convergence results of the Riemannian gradient expansion are proved. We apply the Riemannian gradient estimate in a gradient-based algorithm providing a derivative-free optimization method. We test and validate several applications, including tomographic reconstruction from an unknown random angle distribution, and the sphere packing problem in dimensions 2 and 3.

Furthermore, we consider the inverse problem of determining the fragmentation rate from noisy measurements in the growth-fragmentation equation. We use the Fourier transform theory on locally compact groups to treat this problem for more general fragmentation kernels. We develop a regularization method based on spectral filtering, which allows us to deal with the inverse problem in weighted $\mathbb{L}^2$ spaces. Our approach regularizes the signal generated by differential operators in the frequency domain.

Contents

Chapter 1

Introduction

This thesis focuses on extending the diffusion-map theory for asymmetric kernels and on some applications of our approach to the tomographic reconstruction problem from an unknown random angle distribution, and optimization problems in Euclidean submanifolds. In addition, we also deal with the inverse problem associated with the growth-fragmentation equation for a wide class of kernels that includes the self-similar kernels. Part of our research was published in the following paper:

- Diffusion representation for asymmetric kernels (with Jorge P. Zubelli, and Antônio J. Silva Neto), Applied Numerical Mathematics.

Chapters 2 and 3 of this thesis are based on our contribution to the theory of diffusion-maps, and Chapter 4 is based on our research on the inverse problem. In Chapter 2, we extend the theory of diffusion-maps to asymmetric kernels, and we also find analytical representations in a similar way to classical theory. Chapter 3 is intended to show how to use our asymmetric kernel approach to reconstruct the Riemannian gradient of a smooth function defined on a Euclidean submanifold, as well as some applications of our approach to optimization problems on submanifolds. Furthermore, we apply our theory to the tomographic reconstruction problem from an unknown random angle distribution. In

Chapter 4, we explain in more detail our approach to the inverse problem of determining the fragmentation rate from noisy measurements in the growth-fragmentation equation.

This introduction is intended to briefly highlight our contribution to the topics discussed. Technical details are in the following chapters.

Diffusion Representation for Asymmetric Kernels

Data compression has been studied extensively in many applications. See [70, 74]. Several dimensionality reduction algorithms are based on the spectral decomposition of symmetric linear operators which induce geometric structures in the dataset. A classical example of such operators is the Laplacian matrix associated with an undirected graph [22]. In Ref. [6], the eigenvalue decomposition of the Laplacian matrix is used to reduce the dimensionality of datasets in such a way that the local information is preserved. The diffusion-map approach is based on using the symmetric normalized Laplacian matrix [19]. Compared with other dimensionality reduction methods diffusion-maps assume that the dataset resides in a lower-dimensional manifold, and uses approximations of the Laplace-Beltrami operator to reveal relevant parameters in the dataset.

The spectral decomposition theorem does not hold for integral operators with asymmetric kernels. Therefore, we cannot use the diffusion-map framework to represent diffusion distances induced by asymmetric kernels. Moreover, if we use the spectral decomposition in some symmetric normalization of an asymmetric kernel, its performance requires a computational complexity $O(n^3)$, for $n \times n$ matrices.

To reduce the above $O(n^3)$ complexity and to deal with more general kernels, we present a new framework to represent the diffusion geometry induced by asymmetric kernels. We use the 2-D FFT to compute this representation. The main advantage of using this representation is that compared to the eigenvector representation, the computation time de-

creases. In fact, for a matrix of dimension $n \times n$, the complexity of the 2-D FFT is $O(n^2 \times log(n))$. The choice of the 2-D FFT to represent the dataset structure is based on the fact that the Fourier basis diagonalizes the Laplacian defined on the Euclidean Torus.

We deal with datasets whose structure is induced by an asymmetric kernel. Based on Refs. [37, 68, 67], we use alternative orthonormal bases to reduce the dimensionality of the datasets in such a way that the geometric structure is preserved. Here, we work with the representation theory of diffusion distances in the context of changing data, proposed in Ref. [18]. First, we find a representation form for the classical diffusion geometry, and then we find a representation form for changing data. To do that, we start with the case $t = 1$ and then extend it for any time t.

The methodology is employed in meteorological applications to identify regions of largest temperature variation, as well as geometric features of complex datasets, which are revealed in key examples such as the Möbius strip and the sphere.

A diffusion-map-based algorithm for gradient computation on manifolds and applications

Several iterative minimization algorithms in Euclidean spaces rely on the fact that the negative gradient determines the steepest descent direction. The applications in inverse problems and imaging abound [12, 23]. Some examples of these algorithms are the Gradient Descent and Newton's method [8]. These methods can be generalized to Riemannian submanifolds of $\mathbb{R}^n$ using a retraction function [1]. An important task in these methods is to compute the Riemannian gradient. In many cases it is not straight computable due to the complexity of the function's local behavior, as well as whenever the available information consists of high-dimensional unsorted sample points, lying in an unknown nonlinear lower-dimensional submanifold. The latter issue does not allow the tangent

space to be efficiently and economically computed from noisy sample points. The purpose of this thesis is to address the difficulties mentioned above. We emphasize that we focus on giving Riemannian gradient estimates instead of proposing an optimization method. We compute approximations of the Riemannian gradient of a function using sample points. An important feature of our approximations is that do not depend on differential conditions of the function. The main tool to compute these estimates is the diffusion-map theory. The latter is a dimensional reduction methodology that is based on the diffusion process in a manifold. See Refs. [19, 18, 20] for more details. An important feature of this theory is that it recovers the Laplace-Beltrami operator when the dataset approximates a Riemannian submanifold of $\mathbb{R}^n$. The diffusion-map theory is based on a symmetric kernel defined on the dataset. The symmetric kernel measures the connectivity between two points. Our approach is based on implementing this theory in the recently developed case of asymmetric kernels [2]. Compared to symmetric kernels, asymmetric kernels provide more details on how the information is distributed along each direction. This characteristic allows us to know which direction has the greatest variations. Another sample point-based approximations for the gradient is proposed in the learning gradient theory [56, 55]. However, these approximations are based on solving optimization problems which are expensive in high-dimensional numerical implementations. As an application of our methodology, we use the operator $\overline{P}_t f(x)$ defined in Eq. (3.2) as the main direction in a gradient-based algorithm. See Ref. [1]. The main advantage of using this operator is that it does not depend on some *a priori* knowledge of the Riemannian gradient of the function. Furthermore, since the operator is defined as an integral, then it is robust to noise in the data. We test the proposed gradient-based algorithm for the sphere packing problem in dimensions 2 and 3. Regarding the previous literature, in Chapter 2 of [13], an optimization algorithm is proposed to address the sphere packing problem on a Grassmannian manifold. However, in the present experiment, we consider the sphere packing in the Euclidean space,

which is more difficult due to the symmetries of the lattices in the ambient space. Compared with the approach in [13], we solve optimization problems with more general cost functions. We remark that the MANOPT TOOLBOX, described in [10, 9] and especially the derivative-free solvers PSO and NELDER-MEAD, are not suitable to deal with this problem, since it presents instabilities due to the dependence of the logarithm of certain matrices. Indeed, the logarithm cannot be defined for matrices for which the Jordan block decomposition has non-positive eigenvalues occurring an odd number of times. Furthermore, we apply the proposed methodology to the tomographic reconstruction problem from samples of unknown angles. This post-processing algorithm is parallelizable. It also has a similar flavor to the algorithm developed in [49, 50] since we are trying to solve a high dimensional optimization problem with a swarm of computed auxiliary data. In the latter case, this is done with the approximation to the roots of a high degree polynomial. Our reconstruction method is based on using the diffusion-maps for a partition of the dataset, instead of considering the complete database as proposed in [21]. We remark that we reconstruct the image except for a possible rotation and reflection. Compared to traditional reconstruction methods [21, 5], our method does not assume the hypothesis that the distribution of the angles is previously known, which makes it a more general and practical method for numerical implementations. Furthermore, our method runs faster and more efficiently than the method proposed in [21]. In fact, if the number of sample points is $qs+r$ with $r < s < q$, then the complexity of the algorithm proposed in [21] is $O(q^3\, s^3)$, while our algorithm runs with complexity $O(q\, s^3)$. On the other hand, the numerical implementation described in [51] of the methodology proposed in [5], uses brute force which is not suitable when the number of sample points is large.

Recovering the Fragmentation Rate in the Growth-Fragmentation Equation for Transport Kernels.

The growth-fragmentation equation describes, in a quantitative way, the evolution process of the density of an ensemble of particles. We assume that each particle grows over time, and splits into smaller particles, in such a way that the total mass is conserved. This model is used in biological phenomena for instance, in cell division processes [62, 53, 4, 47], protein polymerization [35], and in telecommunications [3]. The growth-fragmentation equation can be expressed using the following integro-differential equation,

$$\begin{cases} \frac{\partial}{\partial t}n(t,x) + \frac{\partial}{\partial x}(g(x)n(t,x)) = -B(x)n(t,x) + k\int_0^\infty \mathbb{K}(x,y)\,B(y)\,n(t,y)dy, \\ \\ n(0,x) = n_0(x), \\ \\ g(0)n(t,0) = 0, \end{cases}$$

where $n(t,x)$ represents the density of particles of mass x at time t, with initial condition $n_0(x)$. The function $g(x)$ is the growth rate for particles of mass x. The fragmentation kernel $\mathbb{K}(x,y)$ represents the probability that a particle of mass y splits into k smaller particles of mass x. The function $B(x)$ is the fragmentation rate of particles of mass x.

A natural question concerns to the asymptotic behaviour of the population density $n(t,x)$, when $t \to \infty$. In [64, 54] is proved that under certain conditions on the coefficients

g, B, and $\mathbb{K}$, there exist an unique eigenpair (N, λ) solution of

$$\begin{cases} \frac{d}{dx}(g(x)N(x)) + (B(x) + \lambda)N(x) = k \int_0^\infty \mathbb{K}(x,y)\, B(y)\, N(y) dy, \\ \\ g(0)N(0) = 0, \\ \\ N(x > 0) > 0, \\ \\ \int_0^\infty N(x) dx = 1. \end{cases}$$

Moreover, in a certain weighted norm we have that $n(t,x)e^{-\lambda t} \to N(x)$, when the time t goes to infinity. Note that it is possible to estimate $n(t,x)$ from $N(x)$ when the time t is large. The importance of studying $N(x)$ instead of $n(t,x)$ is that the variable t is removed, which reduces the study of a two-dimensional problem to a one-dimensional.

We are interested on how to recover in a stable way the fragmentation rate $B(x)$ from noisy measurements of $N(x)$ and λ. Our strategy is to study $H = BN$ rather than B, as proposed in [29, 11], and then use truncated division by N to recover B. Thus, the inverse problem under consideration is how to recover in a stable way H from approximate knowledge of N and λ, in which H is the solution of

$$k\mathcal{K}(H) - H = \frac{d}{dx} gN + \lambda N, \tag{1.1}$$

where

$$\mathcal{K}(H) = \int_0^\infty \mathbb{K}(x,y)\, H(y)\, dy. \tag{1.2}$$

In [29, 63, 28] the inverse problem was treated for the equal mitosis case, that is, when a

particle splits into two small particles with half of mass. In this case, the parameters are

$$k = 2 \quad \text{and} \quad \mathbb{K}(x, y) = \delta_{x=\frac{y}{2}}.$$

A more general kind of kernels are the self-similar kernels, that is, when

$$\mathbb{K}(x, y) = \frac{1}{y} P(\frac{x}{y}),$$

where P is a probability measure in $[0, 1]$. Self-similar kernels arise when the fragmentation depends on the ratio between the size of the mother particle and the size of next generation particle. The inverse problem for self-similar kernels was treated in papers [11, 30].

The aim of this thesis is to address this inverse problem for more general kernels. Namely, we consider kernels of the form

$$\mathbb{K}(x, y) = P(x \odot_\rho y^{-1}) \ \frac{d}{dz} \rho^{-1}(y),$$

where ρ is an increasing diffeomorphism from $\mathbb{R}$ to $\mathbb{R}^+$, and P is a probability on $(0, \rho(0))$. The transport operation $\odot_\rho$ is defined by

$$a \odot_\rho b := \rho(\, \rho^{-1}(a) + \rho^{-1}(b)\,).$$

We shall denoted such kernels as transport kernels. Observe that if we take the exponential function $\rho(x) = e^x$, then the transport operation for this function is the usual product on the positive real line. Hence, transport kernels generalize self-similar kernels.

We treat the inverse problem described in (1.1), for this kind of kernels. Here, we do not require that the mass must be conserved in the division process. This, in order not to restrict this method only to biological models, and also to use it in real-world applications.

treating this problem from a more general point of view. We do that in two steps. Firstly, we guarantee that under some assumptions in N, there exists a unique H solution of (1.1). To do that, we establish that on proper spaces the operator $k\mathcal{K} - Id$ is an isomorphism. The second step is to guarantee the stability of the inverse problem. For that, we propose a new regularization method. This method is based on the implementation of the Fourier transform theory, and spectral filtering techniques.

We show that the Quasi-Reversibility approach [63] is a particular case of our method. Compared to regularization methods based on convolution [11, 30], our method improves the error order. Namely, we obtain an error of order $O(\varepsilon^{\frac{2m}{2m+1}})$.

This thesis is organized as follows: in Section 4.1, we study the transport operation and some relations with the Fourier transform on locally compact abelian groups. In Section 4.2, we discuss the invertibility of the operator $k\mathcal{K} - Id$ in proper spaces. In Section 4.3, we present a new regularization method to treat the stability of the inverse problem. In Section 4.4, we give examples for which some of the Hypotheses (4.2.1) or (4.3.1) are satisfied. Finally, in Section 4.5, we present the numerical implementation of our method.

Chapter 2

Diffusion Representation for Asymmetric Kernels

Highlights

- A novel methodology for asymmetric-kernel data dimension reduction is developed.
- The dimension reduction is based on the highly efficient FFT algorithm in high dimensions.
- Numerical evidence indicates that the tensor product of the FFT basis performance is faster than eigenvalue-based methods.
- Geometric features of complex datasets are revealed in key examples such as the Möbius strip and the sphere.
- The methodology is employed in meteorological applications to identify regions of the largest temperature variation.

This chapter is a compilation of paper [2]. The chapter is organized as follows, in Section 2.1 we give a brief exposition of the classical representation theory for diffusion

distances proposed in Refs. [19, 18, 20, 52]. In Section 2.2 we present our framework to represent diffusion distances when the structure in the dataset is induced by an asymmetric kernel. Finally, in Section 2.3, we show some experiments with applications and results.

2.1 Classical Diffusion-Map Theory

Nowadays, science, engineering, and many tech companies process a large amount of high-dimensional information. Refs. [45, 15]. Due to the relevance of data processing in real-world applications, it is necessary to implement algorithms that optimize the computational resources in the processing of datasets. In many cases, the high-dimensional information lies in a lower-dimensional submanifold. Then, it is possible to extract the relevant information using a few coordinates, such that the principal features of the dataset are not affected. This procedure is the main purpose of dimensionality reduction algorithms.

If the dataset lies in a hyperplane, it is possible to use linear reduction techniques such as principal component analysis (PCA). Ref. [42]. On the other hand, when the data lies in a different submanifold of a hyperplane, linear reduction methods do not perform well. In that case, it is necessary to apply a reduction method that extracts the most relevant nonlinear parameters. Diffusion-map is a nonlinear dimensionality reduction method based on the heat equation. Compared to other methods, this algorithm offers robustness to noise perturbation. Ref. [19]. We illustrate this consideration in the following example. We consider the nonlinear dataset Z, which consists of 100 sample points distributed on the curve γ, defined as

$$\gamma(t) = (\cos t, \sin t, t), \tag{2.1}$$

for $t \in [0, 6\pi]$. We plot this dataset in Figure 2.1. In addition, in Figure 2.2, we plot the two-dimensional embedding using the PCA and diffusion-map. Observe that the diffusion-map algorithm is the one that best preserves the geometric features of the dataset, this

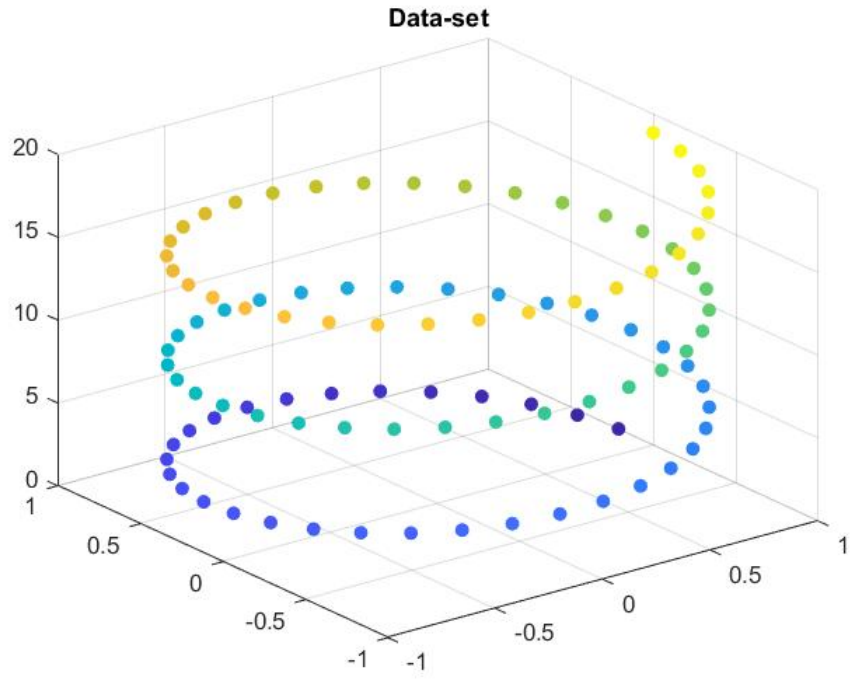


Figure 2.1: dataset Z with 100 sample points uniformly distributed over the curve given by Eq. 2.1.

is since the information lies in a nonlinear submanifold. Furthermore, we illustrate the robustness to noise perturbations. For that, we consider the additive noise perturbation Z_ϵ, defined by

$$Z_\epsilon = Z + \epsilon W, \tag{2.2}$$

where ϵ is the error order, and W is the white noise. Figure 2.3 shows that even in the case of noise perturbations Z_ϵ, the two-dimensional embedding preserves the geometric features of the dataset. Observe that in this example, the noise perturbations do not significantly affect the embedding using the diffusion-map algorithm.

In many cases, asymmetric kernels provide better information than symmetric kernels, for instance, the space of probability distributions endowed with the Kullback-Leiber divergence. Ref [48]. Thus, reducing its dimensionality, we optimize the processing in many

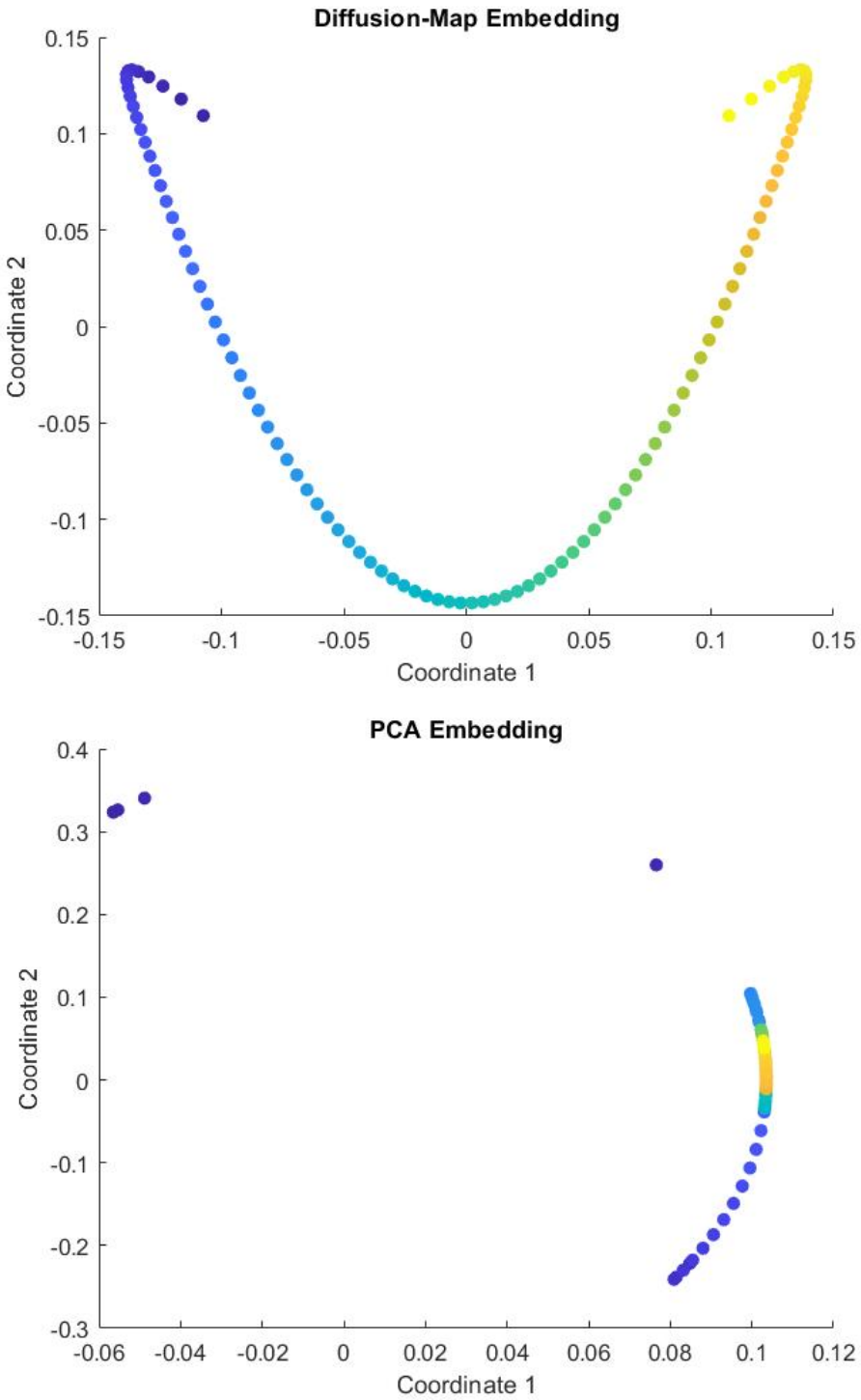
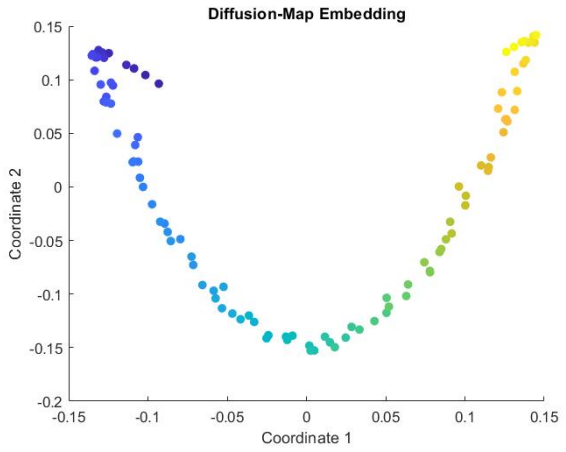
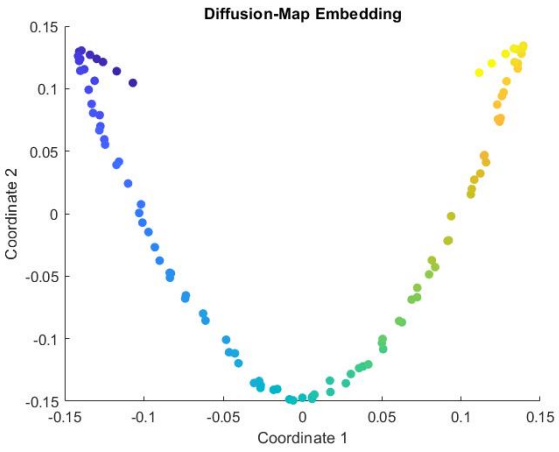


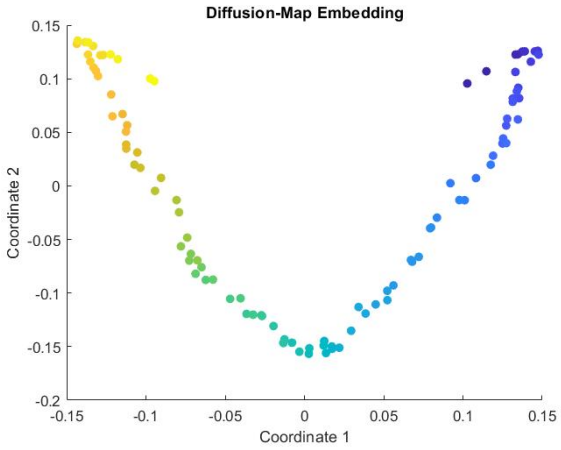
Figure 2.2: Two-dimensional dimensionality reduction using diffusion-map and PCA.



(a) $\epsilon = 0.1$



(b) $\epsilon = 0.5$



(c) $\epsilon = 1$

Figure 2.3: Two-dimensional dimensionality reduction using diffusion-map to noise perturbation Z_ϵ defined in Eq. (2.2) for several values of ϵ.

required tasks. However, the classical approach does not allow to reduce the dimensionality in the case of asymmetric kernels. The purpose of this chapter is to extend the diffusion-map theory to the asymmetric case and also to improve the execution time using the Fourier basis.

In this section, we review some classical results on diffusion-map theory. We define diffusion distances in a measure space and we recall some results related to the representation of such diffusion distances.

2.1.1 Diffusion distance

Assume that our dataset (X, μ) is a measure space, and let $k : X \times X \to \mathbb{R}_{\geq 0}$ be a non-negative symmetric kernel, which is used to measure the local connectivity between two points x and y. If X is also a metric space, the most classical example of these kernels is the Gaussian kernel given by

$$e^{-d^2(x,y)/2\sigma^2},$$

where d is the distance function, and σ^2 is the scaling parameter. We define the associated Markov kernel $\rho(x, y)$ by

$$\rho(x, y) := \frac{k(x, y)}{\sqrt{v(x)}\sqrt{v(y)}} \ , \tag{2.3}$$

where $v(x)$ is the volume form defined as $v(x) = \int_X k(x, y)\, d\mu(y)$. Assuming that the volume form never vanishes and that $k \in L^2(X \times X)$, then the operator $A : L^2(X) \to L^2(X)$ given by

$$A(f)(x) := \int_X \rho(x, y)\, f(y)\, d\mu(y), \tag{2.4}$$

is compact and self-adjoint. By the spectral theorem, we can write

$$\rho(x, y) = \sum_{n \in \mathbb{N}} \lambda n \phi_n(x)\ \phi_n(y) \ ,$$

where $\{\phi_n, \lambda n\}$ is the spectral decomposition of the operator A. For any natural number t, we define the diffusion distance at time t between two points x and y by

$$D^t(x,y) := \| \rho^t(x,\cdot) - \rho^t(y,\cdot)\|_{L^2(X)} \ , \tag{2.5}$$

where ρ^t is the kernel of the integral operator A^t. Here, A^t is the composition of the operator A, a total of t times. In this case, the kernel ρ^t has the representation

$$\rho^t(x,y) = \int_X \int_X \cdots \int_X \int_X \rho(x,z_1)\rho(z_1,z_2)\cdots\rho(z_{t-2},z_{t-1})\rho(z_{t-1},y)dz_1dz_2\cdots dz_{t-2}dz_{t-1}.$$

The kernel $\rho^t(x,y)$ measures the probability that the points x and y are connected by a random walk of length t. Observe that the distance D^t is an average over all of the paths in time t connecting x to y. Therefore, the diffusion distance is robust to noisy data. We see that the quantity D^t is small when there are many random walks of length t connecting x and y. Using the spectral decomposition of the operator A, we can write the diffusion distance as

$$D^t(x,y) = \sqrt{\sum_{n\in\mathbb{N}} \lambda n^t(\phi_n(x) - \phi_n(y))^2} \ . \tag{2.6}$$

The above expression allows us to reduce the dimensionality of the diffusion geometry, namely, we embed our dataset in a lower dimensional space using the diffusion-map $\psi_k^t : X \to \mathbb{R}^k$, where

$$\psi_k^t(x) = (\lambda i^{t/2}\, \phi_i(x))_{i=1}^k \ .$$

2.1.2 Diffusion-map for changing data

The diffusion distance for changing data proposed in Ref. [18] compares data points between parametric datasets. We define X_α as the set X endowed with the kernel structure k_α. As above, for each kernel k_α, we consider the associated Markov kernel ρ_α defined in

Eq. (2.3), and the operator A_α as in Eq. (2.4). To compare the data structure X_α with X_β, we define the dynamic diffusion distance $D^t : X_\alpha \times X_\beta \to \mathbb{R}_{\geq 0}$ by

$$D^t(x_\alpha, y_\beta) = \| \rho_\alpha^t(x, \cdot) - \rho_\beta^t(y, \cdot) \|_{L^2(X)} .$$

Furthermore, in Ref. [18], the global diffusion distance at time t is defined by

$$\mathbb{D}^t(X_\alpha, X_\beta) = \sqrt{\int_X (D^t(x_\alpha, x_\beta))^2 \, d\mu(x)} .$$

The global diffusion distance measures the change from the data structure X_α to X_β. Under mild assumptions on the dataset X and the family of kernels $\{k_\alpha\}$, the dynamic diffusion distance and the global diffusion distance can be computed using the spectral decomposition of the operators A_α, as in Eq. (2.6). See Ref. [18].

2.2 Diffusion Representation for Asymmetric Kernels

Assume that (X, μ) is a measure space, and consider an asymmetric kernel k, which is any square integrable measurable function $k : X \times X \to \mathbb{R}_{\geq 0}$. As example of these kernels, we assume that (X, d) is a metric space, and consider the weighted Gaussian kernel defined by

$$w(x, y) \, e^{-d^2(x,y)/2\sigma^2} ,$$

where w is the weight function. Weighted Gaussian kernels measure how information is distributed locally from x to y. Note that the distribution of the information may not be uniform. To deal with more general models, we do not use the Markov normalization given by Eq. (2.3) to define the diffusion distance, instead, we use the diffusion kernel k to define

it. More specifically, we work with the diffusion distance at time t given by

$$D^t(x,y) := \| k^t(x,\cdot) - k^t(y,\cdot)\|_{L^2(X)} \ , \tag{2.7}$$

where k^t is the kernel of the operator A^t, and A is the integral operator defined as

$$A(f)(x) := \int_X k(x,y)\, f(y)\, d\mu(y). \tag{2.8}$$

We remark that the Markov normalization is avoided to optimize computational performance.

2.2.1 Diffusion representation for t=1

We now design a representation for the diffusion distance given by Eq. (2.7), where k is an asymmetric kernel. Suppose that $\{W_{m_1}\}_{m_1\in\mathbb{Z}}$ and $\{W_{m_2}\}_{m_2\in\mathbb{Z}}$ are two orthonormal bases of $L^2(X)$, and that there exists a positive constant M such that for any $m_1 \in \mathbb{Z}$, and all $x \in X$,

$$|W_{m_1}(x)| \leq M.$$

We recall that in such case the tensor product $\{W_{m_1} \otimes W_{m_2}\}_{(m_1,m_2)\in\mathbb{Z}\times\mathbb{Z}}$ defined by

$$W_{m_1} \otimes W_{m_2}(x,y) = W_{m_1}(x)\, W_{m_2}(y), \tag{2.9}$$

is an orthonormal basis of $L^2(X \times X)$, (for more details, see [61, 34, 71]). We note, in passing, that our approach also works if the $L^2(X)$ space is finite-dimensional, in which case $\mathbb{Z}$ should be substituted by a finite index set. To develop our theory we assume the following hypothesis on the kernel k.

Hypothesis 2.2.1. *Suppose that $k \in L^2(X \times X)$, and for a.e $x \in X$, the kernel function*

$k(x,\cdot)$ *belongs to the space* $L^2(X)$.

If we assume the above hypothesis, we can use the basis given by Eq. (2.9) to write the kernel $k(x,y)$ as

$$k(x,y) = \sum_{(m_1,m_2)\in\mathbb{Z}\times\mathbb{Z}} a(m_1,m_2)\, W_{m_1} \otimes W_{m_2}(x,y), \tag{2.10}$$

where $a(m_1,m_2)$ are the coefficients. Using this decomposition we obtain a representation form for the diffusion distance at time $t=1$.

Theorem 2.2.1 (Diffusion representation for $t=1$)**.** *Assume that the kernel k satisfies the Hypothesis 2.2.1, and that the representation of k in the coordinate system (2.9) is given by Eq. (2.10). If the coefficients satisfy the summability condition*

$$\sum_{(m_1,m_2)\in\mathbb{Z}\times\mathbb{Z}} |a(m_1,m_2)| < \infty, \tag{2.11}$$

then the diffusion distance at time $t=1$ has the representation form

$$(D^1(x,y))^2 = \sum_{m_2\in\mathbb{Z}} |\sum_{m_1\in\mathbb{Z}} a(m_1,m_2)\,(W_{m_1}(x) - W_{m_1}(y))|^2 \ . \tag{2.12}$$

Proof. Using the summability condition of Eq. (2.11), we can write the function $k(x,\cdot) - k(y,\cdot)$ as

$$\sum_{m_2\in\mathbb{Z}} \left(\sum_{m_1\in\mathbb{Z}} a(m_1,m_2)\,(W_{m_1}(x) - W_{m_1}(y)) \right) W_{m_2}(\cdot).$$

Since the set $\{W_{m_2}\}$ is an orthonormal basis, we conclude that

$$(D^1(x,y))^2 = \sum_{m_2\in\mathbb{Z}} |\sum_{m_1\in\mathbb{Z}} a(m_1,m_2)\,(W_{m_1}(x) - W_{m_1}(y))|^2.$$

□

For practical purposes, we do not use the representation formula given in Theorem 2.2.1 to approximate the diffusion distance, because this representation includes two sums with many terms. Instead, we use an approximation involving sums a of few terms, this is established in Theorem 2.2.2. To prove this theorem, we first prove an auxiliary lemma.

Lemma 2.2.1 (Approximation lemma)**.** *Consider the function*

$$f_{k_1,k_2} : L^1(\mathbb{Z}\times\mathbb{Z}) \times L^1(\mathbb{Z}\times\mathbb{Z}) \times X \times X \to \mathbb{R}_{\geq 0},$$

defined by

$$f_{k_1,k_2}(a_1,a_2,x,y) := \sum_{|m_2|\leq k_2} \Big| \sum_{|m_1|\leq k_1} a_1(m_1,m_2)\, W_{m_1}(x) - a_2(m_1,m_2)\, W_{m_1}(y)\Big|^2 \ .$$

Suppose that a_1 and a_2 are sequences in $L^1(\mathbb{Z}\times\mathbb{Z})$, then for each $\delta > 0$, there exist positive integers $\overline{k_1}$ and $\overline{k_2}$, such that if $\overline{k_1} \leq k_1$, $\overline{k_2} \leq k_2$, and if $x, y \in X$, the following inequality holds

$$|f_{k_1,k_2}(a_1,a_2,x,y) - f(a_1,a_2,x,y)| < \delta \ ,$$

uniformly in the variables x, y, where $f(a_1,a_2,x,y) = \lim_{k\to\infty} f_{k,k}(a_1,a_2,x,y)$.

Proof. Let $b(m_1,m_2)$ be defined by

$$b(m_1,m_2) = a_1(m_1,m_2)\, W_{m_1}(x) - a_2(m_1,m_2) W_{m_1}(y).$$

Since the $L^2(\mathbb{Z}\times\mathbb{Z})$ norm is smaller than, or equal to, the $L^1(\mathbb{Z}\times\mathbb{Z})$ norm,then, by triangular inequality, we obtain that the expression $|f_{k_1,k_2}(a_1,a_2,x,y) - f(a_1,a_2,x,y)|$ is

bounded from above by

$$\left(\sum_{m_2 \in \mathbb{Z}} \sum_{|m_1|>k_1} |b(m_1,m_2)| + \sum_{|m_2|>k_2} \sum_{m_1 \in \mathbb{Z}} |b(m_1,m_2)| \right)^2 .$$

Therefore, by Assumption (2.11), we conclude that for a given $\delta > 0$, we can take k_1 and k_2 large, such that the above term is less than or equal to δ. □

As a consequence of the above lemma, we prove the following theorem, which states that we can approximate the diffusion distance with finite sums.

Theorem 2.2.2. *Assume that the kernel k satisfies the same hypotheses of Theorem 2.2.1. Then, for each $\delta > 0$, there exist positive integers $\overline{k_1}$ and $\overline{k_2}$ such that for any $\overline{k_1} \leq k_1$, and $\overline{k_2} \leq k_2$, and all $x, y \in X$, the following inequality holds*

$$|f_{k_1,k_2}(a_1,a_1,x,y) - (D^1(x,y))^2| \leq \delta \ .$$

Proof. The proof is a consequence of Theorem 2.2.1 together with Lemma 2.2.1. □

2.2.2 Diffusion representation for arbitrary time

Suppose that t is a positive integer denoting an arbitrary time of the Markov chain. We now use the coordinate system of Eq. (2.9) to find the representation form for the kernel k^{t+1} in terms of the coefficients $a(m_1,m_2)$. Let k be an asymmetric kernel, and suppose that for any $1 \leq j \leq t$, the kernel k^j satisfies all the hypotheses of Theorem 2.2.1. Under this assumption, we can use the Fubini's theorem to write recursively the kernel k^{t+1} of the operator A^{t+1} as

$$k^{t+1}(x,y) = \langle k(x,\cdot), k^t(\cdot,y) \rangle_{L^2(X)} \ . \tag{2.13}$$

Assuming that the kernel k^t has the series representation

$$k^t(x,y) = \sum_{(m_1,m_2)\in\mathbb{Z}\times\mathbb{Z}} a(m_1,m_2)^t\, W_{m_1}\otimes W_{m_2}(x,y).$$

then by Eq. (2.13) we obtain that

$$a(m_1,m_2)^{t+1} = \sum_{k\in\mathbb{Z}} a(m_1,k)\, a(k,m_2)^t\ . \tag{2.14}$$

Recursively, we obtain that the expression for the coefficients of the kernel k^{t+1}

$$a(m_1,m_2)^{t+1} = \sum_{n_1\in\mathbb{Z}}\sum_{n_2\in\mathbb{Z}}\sum_{n_3\in\mathbb{Z}}\cdots\sum_{n_t\in\mathbb{Z}} a(m_1,n_1)\, a(n_1,n_2)a(n_2,n_3)\ldots a(n_t,m_2)\ . \tag{2.15}$$

Again, the above expression contains infinitely many sums. We now prove that we can approximate the coefficients of k^t using finite sums. The following lemma establishes this result.

Lemma 2.2.2. *For any $\delta>0$, there exist k_0, such that for any $k_0\le k$ we have*

$$|a^{t+1} - h_k^{t+1}(a)|_{L^1(\mathbb{Z}\times\mathbb{Z})} \le \delta\ ,$$

where $h_k^{t+1}(a)(m_1,m_2)$ is the finite sum

$$h_k^{t+1}(a)(m_1,m_2) = \sum_{|n_1|\le k}\sum_{|n_2|\le k}\sum_{|n_3|\le k}\cdots\sum_{|n_t|\le k} a(m_1,n_1)\, a(n_1,n_2)a(n_2,n_3)\ldots a(n_t,m_2).$$

Proof. We prove by induction over t, for $t = 1$ is clear since

$$\sum_{(m_1,m_2)\in\mathbb{Z}\times\mathbb{Z}} |a(m_1,m_2)^2 - h_k^2(m_1,m_2)| \leq \sum_{|n_1|\geq k}\sum_{m_1\in\mathbb{Z}}\sum_{m_2\in\mathbb{Z}} |a(m_1,n_1)a(n_1,m_2)|$$
$$\leq \|a\|_{L^1} \sum_{|n_1|\geq k}\sum_{m_1\in\mathbb{Z}} |a(m_1,n_1)| \ .$$

Then, by Assumption 2.11, we have that for k large the above inequality is less than or equal to δ. We now assume that the claim holds for t, and we prove that also holds for $t+1$. For k large we have that

$$\|a^t - h_k^t(a)\|_{L^1(\mathbb{Z}\times\mathbb{Z})} \leq \delta.$$

The above inequality implies

$$\|h_k^t(a)\|_{L^1(\mathbb{Z}\times\mathbb{Z})} \leq (\delta + \|a\|_{L^1(\mathbb{Z}\times\mathbb{Z})}) \ .$$

Furthermore, by Eq. (2.14), we have that

$$|a(m_1,m_2)^{t+1} - h_k^{t+1}(a)(m_1,m_2)|,$$

is less or equal to

$$\sum_{i\in I} |a(m_1,i)|\,|a(i,m_2)^t - h_k^t(a)(i,m_2)| + \sum_{|i|\geq k} |a(m_1,i)||h_k^t(a)(i,m_2)| \ .$$

Using the above inequalities, we obtain the following estimate

$$\|a^{t+1} - h_k^{t+1}(a)\|_{L^1} \leq \delta\|a\|_{L^1} + (\delta + \|a\|_{L^1}) \sum_{|i|\geq k}\sum_{m_1\in\mathbb{Z}} |a(m_1,i)| \ .$$

Therefore, by Assumption (2.11), we conclude that the claim holds for $t+1$. □

We now use the above result to design a representation for the diffusion distance at time $t+1$. This representation is based on finite sums of the coefficients $a(m_1, m_2)$. We establish this result in Theorem 2.2.3, the proof involves the following technical lemma.

Lemma 2.2.3 (Continuity)**.** *Consider the function f_{k_1,k_2} as in Lemma 2.2.1, and suppose that a and b are two sequences in $L^1(\mathbb{Z}\times\mathbb{Z})$. Then, for any positive number ϵ, there exists a positive number δ, such that for any pair of sequences c and d satisfying*

$$\|a-c\|_{L^1(\mathbb{Z}\times\mathbb{Z})} \leq \delta \qquad \text{and} \qquad \|b-d\|_{L^1(\mathbb{Z}\times\mathbb{Z})} \leq \delta,$$

then, the following inequality holds for all $x, y \in X$, and for all positive integers k_1, k_2,

$$|f_{k_1,k_2}(a,b,x,y) - f_{k_1,k_2}(c,d,x,y)| \leq \epsilon,$$

where the inequality is uniform in the variables x, y.

Proof. Let k_1, k_2 be positive integers, and define the function R_{m_2} as

$$R_{m_2}(a_1,a_2,x,y) := \sum_{|m_1|\leq k_1} a_1(m_1,m_2)\, W_{m_1}(x) - a_2(m_1,m_2)\, W_{m_1}(y).$$

Observe that

$$|R_{m_2}(a,b,x,y) - R_{m_2}(c,d,x,y)| \leq M \sum_{|m_1|\leq k_1} \left(|(a-c)(m_1,m_2)| + |(b-d)(m_1,m_2)|\right), \quad (2.16)$$

where M is a constant independent of x and y. For any real numbers B and C, the following inequality holds

$$|B^2 - C^2| = |(B-C)^2 + 2BC - 2C^2| \leq (B-C)^2 + 2|C||B-C|. \quad (2.17)$$

Applying the above inequality to $B = |R_{m_2}(c,d,x,y)|$, and $C = |R_{m_2}(a,b,x,y)|$, together with the fact that the L^2 norm is smaller than or equal to the L^1 norm, we obtain that $|f_{k_1,k_2}(a,b,x,y) - f_{k_1,k_2}(c,d,x,y)|$ is bounded from above by

$$M(E^2 + 2(\|a\|_{L^1} + \|b\|_{L^1})E),$$

where M is a constant which does not depend on k_1, k_2, x, y, and

$$E := \|a - c\|_{L^1} + \|b - d\|_{L^1} \ .$$

□

Theorem 2.2.3. *Suppose that for any $1 \leq j \leq t+1$, the kernel k^j satisfies the same hypotheses of Theorem 2.2.1. Then, for any positive number ϵ, there exist positive integers $\overline{k_1}$, $\overline{k_2}$, and $\overline{k_3}$, such that for any natural numbers k_1, k_2, k_3, satisfying $\overline{k_1} \leq k_1$, $\overline{k_2} \leq k_2$ and $\overline{k_3} \leq k_3$, the following inequality holds*

$$|f_{k_1,k_2}(h_{k_3}^{t+1}(a), h_{k_3}^{t+1}(a), x, y) - (D^{t+1}(x,y))^2| \leq \epsilon \ ,$$

uniformly in the variables x, y, where f_{k_1,k_2} is defined as in Lemma 2.2.1, and h_A^{t+1} as in Lemma 2.2.2.

Proof. Applying Theorem 2.2.2 to the kernel k^{t+1}, we have that there exist $\overline{k_1}$, $\overline{k_2}$, such that for any $\overline{k_1} \leq k_1$, and $\overline{k_2} \leq k_2$, the following inequality holds

$$|f_{k_1,k_2}(a^{t+1}, a^{t+1}, x, y) - (D^t(x,y))^2| \leq \epsilon/2 \ ,$$

for any $x, y \in X$. Moreover, using Lemmas 2.2.2 and 2.2.3, there exists a positive integer

$\overline{k_3}$ with the property that if $\overline{k_3} \leq k_3$, then for all $x, y \in X$,

$$|f_{k_1,k_2}(a^{t+1}, a^{t+1}, x, y) - f_{k_1,k_2}(h_{k_3}^{t+1}(a), h_{k_3}^{t+1}(a), x, y)| \leq \epsilon/2 \ .$$

Thus, by the triangular inequality we obtain the desired result. □

2.2.3 Diffusion representation for changing data

In this section we use the coordinate system of Eq. (2.9) to represent the dynamic diffusion distance induced by asymmetric kernels. Suppose that $\{k_\gamma\}$ is a family of asymmetric kernels defined in the dataset X. Again, we consider the diffusion distance without the Markov normalization, that is, we work with the dynamic diffusion distance given by

$$D^t(x_\gamma, y_\beta) = \| k_\gamma^t(x, \cdot) - k_\beta^t(y, \cdot)\|_{L^2(X)}.$$

We assume that for each parameter γ, the kernel k_γ satisfies all the hypotheses of Theorem 2.2.1. In this case, we can write the kernel k_γ as

$$k_\gamma(x, y) = \sum_{(m_1,m_2)\in\mathbb{Z}\times\mathbb{Z}} a^\gamma_{(m_1,m_2)}\, W_{m_1} \otimes W_{m_2}(x, y).$$

The following theorem gives a representation for the dynamic diffusion distance. The proof of the theorem is similar to that of in Theorem 2.2.1.

Theorem 2.2.4. *Assume that for all γ, the kernel k_γ satisfies all the hypotheses of Theorem 2.2.1, then the dynamic diffusion distance $(D^t(x_\gamma, y_\beta))^2$ can be written as*

$$\sum_{m_2\in\mathbb{Z}} | \sum_{m_1\in\mathbb{Z}} (a^\gamma)^t(n, m)\, W_{m_1}(x) - (a^\beta)^t(n, m) W_{m_1}(y)|^2.$$

Remark 2.2.1. *Using the same ideas of the proof of Theorem 2.2.3, we can prove that it is possible to approximate the dynamic diffusion distance by sums involving few terms of a^γ and a^β . To be more specific, the same statement of Theorem 2.2.3 holds if we replace the classical diffusion distance by the dynamic diffusion distance $D^t(x_\gamma, y_\beta)$, and the function $f_{k_1,k_2}(h_C^{t+1}(a), h_C^{t+1}(a), x, y)$ by the function $f_{k_1,k_2}(h_C^{t+1}(a^\gamma), h_C^{t+1}(a^\beta), x, y)$.*

We use the previous result to compute the global diffusion distance in terms of the coefficients a^γ. We recall that the global diffusion distance between X_α, and X_β, is defined by

$$(\mathbb{D}^t(X_\gamma, X_\beta))^2 = \int_X (D^t(x_\gamma, x_\beta))^2 d\mu(x).$$

Theorem 2.2.5. *Under the same assumptions of Theorem 2.2.4, the global diffusion distance at time t can be written as*

$$(\mathbb{D}^t(X_\gamma, X_\beta))^2 = \sum_{(n,m)\in\mathbb{Z}\times\mathbb{Z}} |((a^\gamma)^t - (a^\beta)^t)(n,m)|^2 \ .$$

Proof. Theorem 2.2.4 implies that

$$(\mathbb{D}^t(X_\gamma, X_\beta))^2 = \int_X \sum_{m\in\mathbb{Z}} |\sum_{n\in\mathbb{Z}} ((a^\gamma)^t - (a^\beta)^t)(n,m) W_n(x)|^2 d\mu(x) \ .$$

Expanding the quadratic form, we obtain that

$$\mathbb{D}^t(X_\gamma, X_\beta)^2 = \int_X \sum_{m\in\mathbb{Z}} \sum_{(i,j)\in\mathbb{Z}\times\mathbb{Z}} ((a^\gamma)^t - (a^\beta)^t)(i,m)\, ((a^\gamma)^t - (a^\beta)^t)(j,m)\, W_i(x)\, W_j(x) d\mu(x).$$

Using Hölder's inequality and the fact that W_n is orthonormal, we conclude that the expression

$$S = \sum_{m\in\mathbb{Z}} \sum_{(i,j)\in\mathbb{Z}\times\mathbb{Z}} \int_X |((a^\gamma)^t - (a^\beta)^t)(i,m)\, ((a^\gamma)^t - (a^\beta)^t)(j,m)\, W_i(x)\, W_j(x)| d\mu(x).$$

is bounded from above by

$$\sum_{m\in\mathbb{Z}}\left(\sum_{n\in\mathbb{Z}}|((a^\gamma)^t-(a^\beta)^t)(n,m)|\right)^2\leq\left(\sum_{m\in\mathbb{Z}}\sum_{n\in\mathbb{Z}}|((a^\gamma)^t-(a^\beta)^t)(n,m)|\right)^2<\infty,$$

where we used the fact that the L^2 norm is smaller than or equal to the L^1 norm. By the dominated convergence theorem, we can change the order between the integral and the sums. Moreover, using the fact that W_n is an orthonormal basis of $L^2(X)$, we conclude that

$$(\mathbb{D}^t(X_\gamma,X_\beta))^2=\sum_{m\,\in\,\mathbb{Z}}\sum_{n\,\in\,\mathbb{Z}}|((a^\gamma)^t-(a^\beta)^t)(n,m)|^2\ .$$

□

2.2.4 Weak representation

The framework developed up to this point uses the absolute summability condition (2.11). We now relax this assumption, and under an *a priori* assumption on the diffusion distance, we design a representation for the diffusion distance of points lying in a set of large measure. Without loss of generality, we work with the changing data framework.

Theorem 2.2.6. *Let t be a nonnegative integer, assume that the kernel k_γ^{t+1} satisfies the Hypothesis 2.2.1, and also that for any integer j, with $1\leq j\leq t$, the kernel k_γ^j satisfies all the hypotheses of Theorem 2.2.1. If we suppose that the dynamic diffusion distance at time $t+1$ is bounded from above, that is, there exists a positive constant C such that for any $x,y\in X$, and all indices γ,β,*

$$D^{t+1}(x_\gamma,y_\beta)\leq C.$$

Then, for any positive real number ϵ, there exists a positive integer k_0, such that for any integer $k\geq k_0$, there exists a measurable set E_k, with the property that the measure of

$X \setminus E_k$ satisfies $\mathcal{L}(X \setminus E_k) < \epsilon$, and such that for any $x, y \in E_k$ the following inequality holds

$$|f_{k,k}(h_k^{t+1}(a_k^\gamma), h_k^{t+1}(a_k^\beta), x, y) - (D^{t+1}(x_\gamma, y_\beta))^2| \leq \overline{C}(\sqrt{\epsilon} + \epsilon),$$

where $\overline{C}$ is a constant and the functions $f_{k,k}$, h_k^{t+1} are defined in Lemmas 2.2.1, and 2.2.2, and a_k^γ is the truncated sequence defined by

$$a_k^\gamma(n,m) = \begin{cases} a^\gamma(n,m), & \text{if} \quad \|(n,m)\| \leq k \\ 0, & \text{otherwise} \end{cases} .$$

Proof. Define the truncated kernel $k_{\gamma,k}$ by

$$k_{\gamma,k}(x,y) = \sum_{(m_1,m_2)\in\mathbb{Z}\times\mathbb{Z}} a_k^\gamma(m_1,m_2)\, W_{m_1} \otimes W_{m_2}(x,y) \ .$$

Using the recursive formula in Eq. (2.13), and the fact that $\lim_{k\to\infty} k_{\gamma,k} = k_\gamma$, we see inductively that $\lim_{k\to\infty} k_{\gamma,k}^{t+1} = k_\gamma^{t+1}$ for any non negative integer t, where the convergence is in the $L^2(X \times X)$ norm. Therefore, for any positive number ϵ, there exists k_0 such that for any integer k satisfying $k_0 \leq k$, we have $\|k_\gamma^{t+1} - k_{\gamma,k}^{t+1}\|_{L^2(X\times X)} \leq \epsilon$. Define the set

$$E_k = \{x \in X| \ \ \|k_{\gamma,k}^{t+1}(x,\cdot) - k_\gamma^{t+1}(x,\cdot), \|_{L^2(X)} \leq \sqrt{\epsilon}\} \ .$$

Then, by Chebyshev's inequality and Fubini's theorem we obtain that

$$\mathcal{L}(X \setminus E_k) \leq \frac{1}{\epsilon} \int_X \|k_{\gamma,k}^{t+1}(x,\cdot) - k_\gamma^{t+1}(x,\cdot), \|_{L^2(X)}^2 dx \leq \epsilon \ .$$

On the other hand, if $x, y \in E_k$, then by Minkowski inequality

$$|D^{t+1}(x_{\gamma,k}, y_{\beta,k}) - D^{t+1}(x_\gamma, y_\beta)| \leq 2\sqrt{\epsilon} \ ,$$

where $x_{\gamma,k}$ is the data point x endowed with the kernel structure $k_{\gamma,k}^{t+1}$. We apply Inequality (2.17) of Lemma 2.2.3 with $B = D^{t+1}(x_{\gamma,k}, y_{\beta,k})$, and $C = D^{t+1}(x_\gamma, y_\beta)$, to obtain

$$|D^{t+1}(x_{\gamma,k}, y_{\beta,k})^2 - D^{t+1}(x_\gamma, y_\beta)^2| \leq (1+2M)(\epsilon + \sqrt{\epsilon}) \ .$$

Since the set of non-zero coefficients of the kernel $k_{\gamma,k}$ is finite, then, by Eq. (2.15), we conclude that $a_{\gamma,k}^{t+1}(n,m) = 0$ whenever $\|(n,m)\| > k$, and also that for any integers n, m we have $a_{\gamma,k}^{t+1}(n,m) = h_k^{t+1}(a)(n,m)$. Using these facts we conclude that

$$D^{t+1}(x_{\gamma,k}, y_{\beta,k})^2 = f_{k,k}(h_k^{t+1}(a_k^\gamma), h_k^{t+1}(a_k^\beta), x, y) \ .$$

□

2.3 Applications, Experiments, and Results

2.3.1 Dimensionality reduction

In this section, we use our representation framework to reduce the dimension of datasets, in such a way that the Euclidean norm of the reduced dataset approximates the diffusion distance. Assume that the dataset X is endowed with the kernel structure k_α. Then, by Remark 2.2.1 the map $\phi_{k_1,k_2} : X \to \mathbb{C}^{2k_2+1}$ defined by

$$\phi_{k_1,k_2}(x)(i) = \sum_{|m_1| \leq k_1} (a^\gamma)^t(m_1, i-k_2-1)\, W_{m_1}(x), \tag{2.18}$$

approximates the diffusion distance at time t. We summarize the above in Algorithm 1.

In order to work with a small parameter k_2, and thus embed our dataset in a low dimensional Euclidean space, we need to use a proper orthonormal basis of $L^2(X)$. Proper

bases are those for which the information of the kernel k is concentrated on the coefficients $a(m_1, m_2)$, whose pair of integers (m_1, m_2) are near to the origin.

Algorithm 1 Dimensionality reduction algorithm.

1. Take the data in the form of an $M \times M$ (possibly asymmetric) matrix L representing the kernel structure on the dataset
2. Compute the first components of the matrix L^t in a proper orthonormal tensor basis.
3. Embed the dataset using the function ϕ_{k_1,k_2} of Eq. (2.18)

We note that depending on the data and the mathematical model, we may want to use a proper normalization for the matrix L in Algorithm 1.

In practical situations, our measure space is a finite dataset $X = (x_i)_{i=1}^n$, endowed with the counting measure. If the kernel matrix $\mathbf{A}$ is symmetric, then one can use the eigenvector basis. More generally, if the matrix $\mathbf{A}$ is asymmetric, one can use the singular value decomposition (SVD) to write

$$\mathbf{A} = \sum_{i=1}^{n} c_i \, L_i \otimes R_i \ ,$$

where L_i and R_i $(i = 1, \cdots, n)$ are the left and right singular vectors of $\mathbf{A}$, respectively, and where c_i $(i = 1, \cdots, n)$ are the singular values of $\mathbf{A}$. The computational complexity of the SVD of an $n \times n$ matrix is $O(n^3)$. In the following experiments, we use the Fourier basis to improve the computational complexity. In fact, the computational complexity of the 2-D FFT is $O(n^2 \log n)$. We recall that the Fourier basis $(E_k)_{k=0}^{n-1}$ in the complex vector

space $\mathbb{C}^n$ is given by

$$E_k = \frac{1}{\sqrt{n}}(1, e^{2\pi ik/n}, e^{2\pi i2k/n}, ..., e^{2\pi i(n-1)k/n}) \ . \tag{2.19}$$

In all experiments we compute the coefficients using the 2D-FFT algorithm. All experiments were run in MATLAB software, using a desktop computer with the following configuration: Intel core i7-2600 3.4 GHz processor, and 16 GB RAM.

2.3.2 Synthetic data using a symmetric kernel

In this experiment, our dataset X consists of n random points in the sphere S^2 (Figure 2.4). Here, we use the parametrization

$$\begin{aligned} x(u,v) &= \cos u \sin v, \\ y(u,v) &= \sin u \sin v, \\ z(u,v) &= \cos v, \end{aligned}$$

for $0 \leq u \leq \pi$ and $0 \leq v \leq 2\pi$. We endow the dataset X with the Markov normalization ρ defined in Eq. (2.3), of the Gaussian kernel

$$k(x,y) = e^{-||x-y||^2} \ .$$

Our goal is to compare the efficiency between the representation given in Eq. (2.6) using the eigenvector basis and the representation of Theorem 2.2.1, using the Fourier basis of Eq. (2.19). In Figure 2.5, we show the first two coordinates for each representation with a dataset of 512 points. Observe that the first two coordinates of both representations are similar. In Figure 2.6, we plot the error and computational time (in seconds) of the first

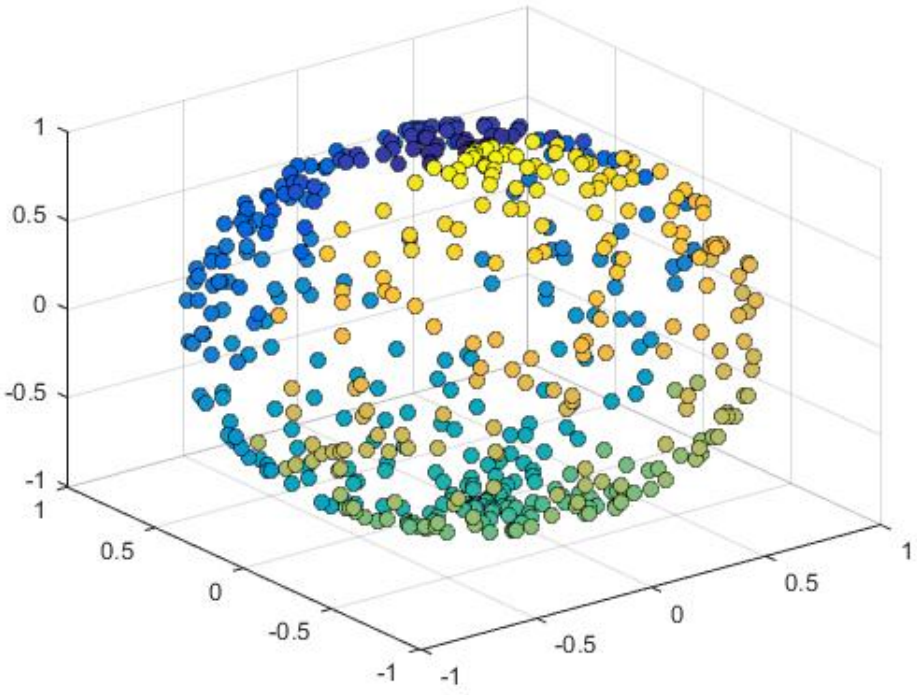
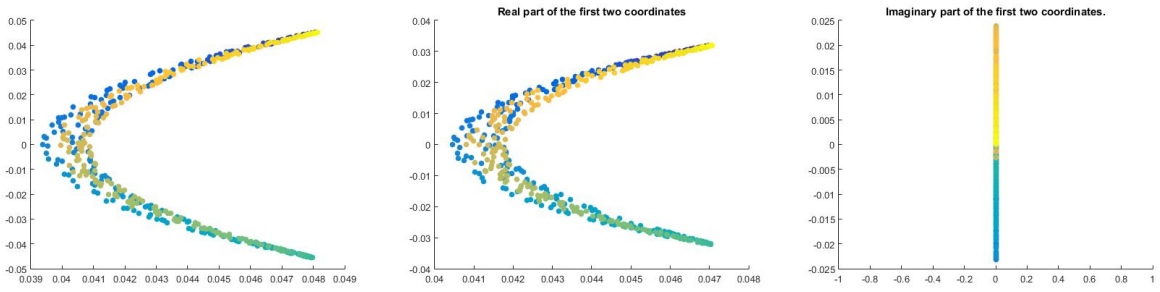


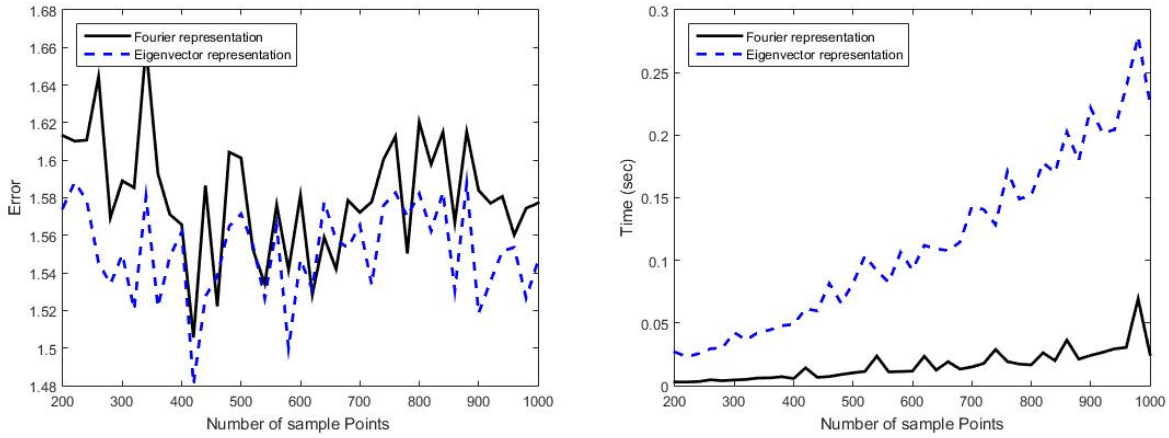
Figure 2.4: dataset X with 512 random points in the sphere S^2.



(a) Eigenvector basis

(b) Real and imaginary part of the Fourier basis

Figure 2.5: Plot of the two-dimensional embedding for the dataset X using the eigenvector basis coefficients (a), and the Fourier basis coefficients (b). Note the scale.



(a) Error of the approximation

(b) Computational time

Figure 2.6: Plot of the L^2 error and computational time of the first two coordinates for different $n \times n$ kernel-sizes, for the dataset of random points in the sphere.

two coordinates for several values of n. We also remark that by using the Fourier basis, the performance of the representation is faster than using the eigenvector basis, and also that the Fourier basis gives an acceptable error when compared to the eigenvector method.

2.3.3 Synthetic data on the Möbius strip

Here, we assume that our dataset X is a set of 300 data points distributed along of the Möbius strip (Figure 2.7). We use the parametrization

$$\begin{aligned} x(u,v) &= (1 + \frac{v}{2}\cos\frac{u}{2})\cos u, \\ y(u,v) &= (1 + \frac{v}{2}\cos\frac{u}{2})\sin u, \\ z(u,v) &= \frac{v}{2}\sin\frac{u}{2}, \end{aligned}$$

for $0 \leq u \leq 2\pi$ and $-\frac{1}{2} \leq v \leq \frac{1}{2}$. We endow the dataset X with the Markov normalization ρ defined in Eq. (2.3), of the weight Gaussian kernel

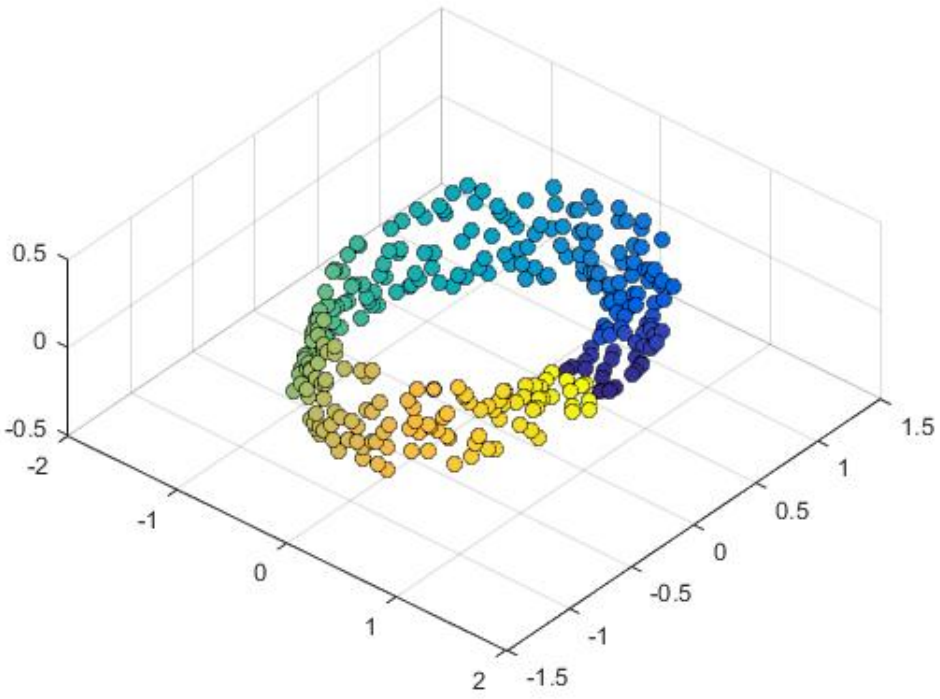
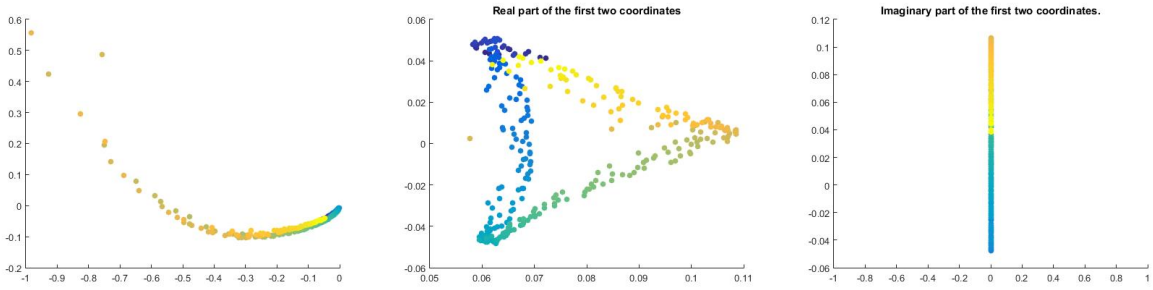


Figure 2.7: dataset X with 300 random points in the Möbius strip M^2.

$$k(x, y) = (S(x - y) + 1)e^{-||x-y||^2},$$

where $S(z)$ is the sign function of the angle (in cylindrical coordinates) of the vector z. This kernel measures local information taking into account if the first two components are rotating clockwise. In this experiment, we compared the performance of the representation using the SVD, and the representation using the Fourier basis. In Figure 2.8, we plot the first two coordinates for each representation. Note that the real part of the representation given by the Fourier basis allows us to see in more detail the distribution of the dataset. The representation that uses the Fourier basis recognizes the rotation of the dataset. However, their presentation that uses SVD does not allow recognizing this feature of X. This is since the representation using the SVD approximates the kernel $k^T k$, instead of the kernel k (Figure 2.9). Therefore, the representation using the SVD does not distinguish some

geometric properties of the dataset X.



(a) svd (b) Real and imaginary part of the Fourier basis

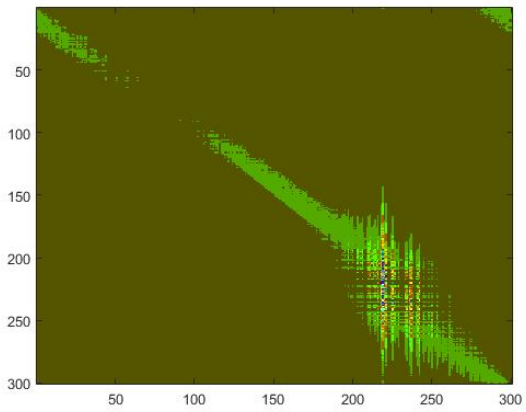
Figure 2.8: Dimensionality reduction using the Eigenvector basis, and Fourier basis

2.3.4 Synthetic data using an asymmetric kernel

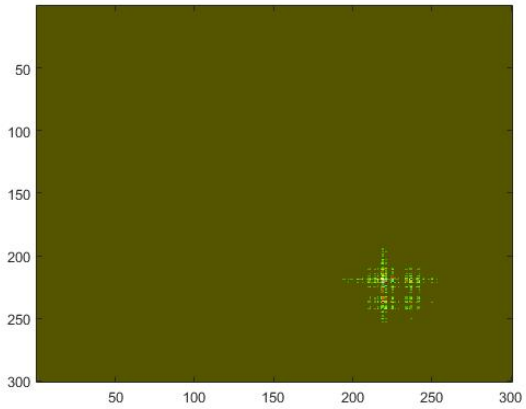
Here, we assume that our dataset X is a random set of 256 data points. We endow this dataset with the kernel structure k given by the Tom Jobim picture of Figure 2.10 whose dimensions are 256×256 pixels. That is, $k(x, y)$ is defined to be the grayscale value of the pixel coordinates (x, y). As in the previous experiment, we use the Markov normalization of the kernel k to compare the performance of the approximations. In this experiment, we compared the performance of the representation using the SVD and the representation using the Fourier basis.

We stress that our objective with this example is not to try to do image processing, but rather to use a picture so that we can visually assess the quality of the approximation. This point will be furthered in the sequel.

In Figure 2.11, we plot the approximation of the kernel using the SVD and the Fourier basis, both using the parameters $k_1 = 256$ and $k_2 = 64$. One may notice that we see a horizontal modulation both under the SVD and the Fourier basis methods. However, this modulation is stronger in the Fourier method. This is since we have only used high frequencies to approximate this image.



(a) Kernel matrix k



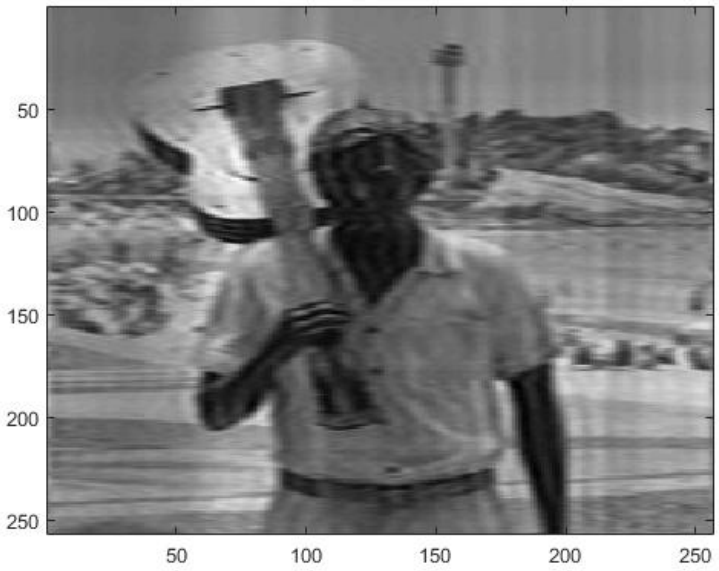
(b) Kernel matrix $k^T k$

Figure 2.9: Kernel matrix k (a), and normalization $k^T k$ (b)

Figure 2.10: Synthetic asymmetric kernel structure chosen from a picture of Tom Jobim in Ipanema Beach (Rio de Janeiro, Brazil). Source [24].

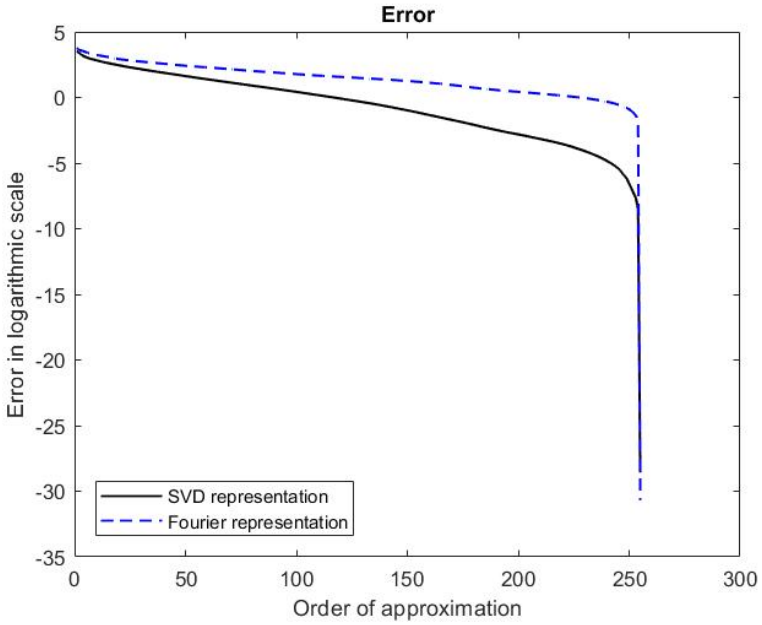
In this case we observe that despite using a smaller number of parameters, it is possible to obtain a good approximation of the original image. In Figure 2.12, we plot the L^2 error and computational time (in seconds), in a logarithmic scale, of the embedding dataset using several approximation orders. As in the previous experiment, we see that using the Fourier basis, the performance is faster, and provides an acceptable error when compared to the SVD method. Furthermore, we point out that in the two previous experiments we did not obtain a better performance concerning computational time when we used the truncated SVD instead of the SVD.

(a) Approximation using the SVD

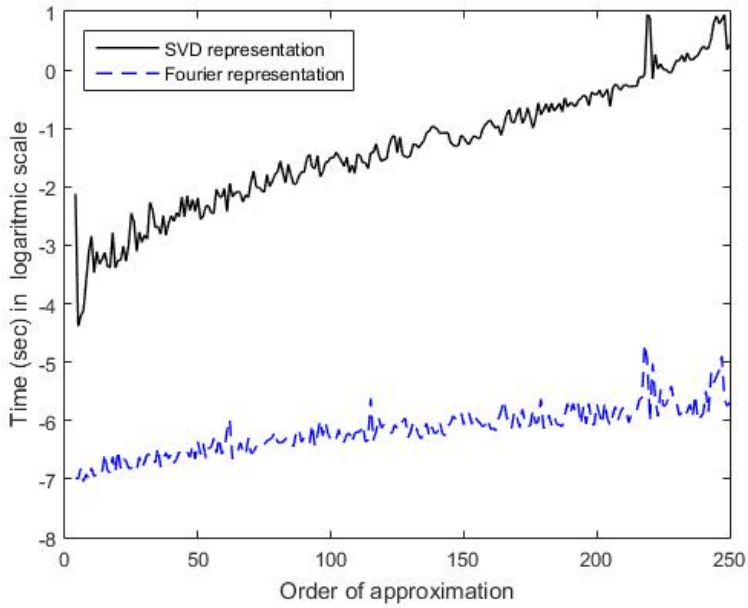


(b) Approximation using the Fourier basis

Figure 2.11: Plot of the approximation using the SVD and the Fourier basis, both using the parameters $k_1 = 256$ and $k_2 = 64$



(a) Approximation error



(b) Computational time

Figure 2.12: Plot of the L^2 error and computational time for the dataset representing the picture of Tom Jobim.

2.3.5 Temperature changes in Brazil

In the last decades, the world temperature distribution has presented drastic changes, in part, this is likely due to human activities [39, 58, 60]. Here, we use the diffusion distance for changing data to detect the regions of Brazil in which the local average annual temperature variation was the highest in the years 2010 and 2018 (Figure 2.13), compared with the year 2000 (Figure 2.13). If a certain region has a great diffusion distance, it means that this particular region has presented significant changes in its temperature. This experiment is based on the change detection on hyperspectral imagery data proposed in Ref. [18]. Our dataset $\{x_{i,j}\}$ consists of $N = 13,974$ points, and each point $x_{i,j}$ represents a pixel coordinate of the Brazilian map. These points are a subset of a picture of size 170×170 pixels. Here we do not take into account the blank pixels, which correspond to places outside the Brazilian territory. For each year we endow the dataset with the un-normalized kernel K which is defined on $X \times X$ by

$$K(x_{i,j}, x_{k,l}) = \frac{1}{\sqrt{N}} T_\alpha(x_{k,l})\, e^{-\|(i,j)-(k,l)\|^2/2\sigma^2} \ ,$$

where $T_\alpha(x_{k,l})$ is the temperature in the rectangular pixel $x_{k,l}$ in the year $\alpha \in \{2000,$ 2010, 2018$\}$, and where $\|\cdot\|$ is the Euclidean distance. In this experiment we use the scaling parameter $2\sigma^2 = 650$. We obtained similar results with a parameter in a range of $600 \leq 2\sigma^2 \leq 700$. We use this kernel without normalization to avoid its high computational cost. This dataset was taken from the Brazilian National Institute of Meteorology website [41]. This asymmetric kernel represents the distribution of the local temperature around the rectangular pixel $x_{k,l}$.

We use Theorem 2.2.4 to approximate the dynamic diffusion distance $(D^1(x_{2000}, x_\gamma))^2$, for all data points, where $\gamma \in \{2010, 2018\}$. Due to the high dimensionality of the kernel matrix, the SVD algorithm did not run in the computer whose configuration is given in

Year	Global diffusion distance (°C)
2010	25.5506
2018	37.5000

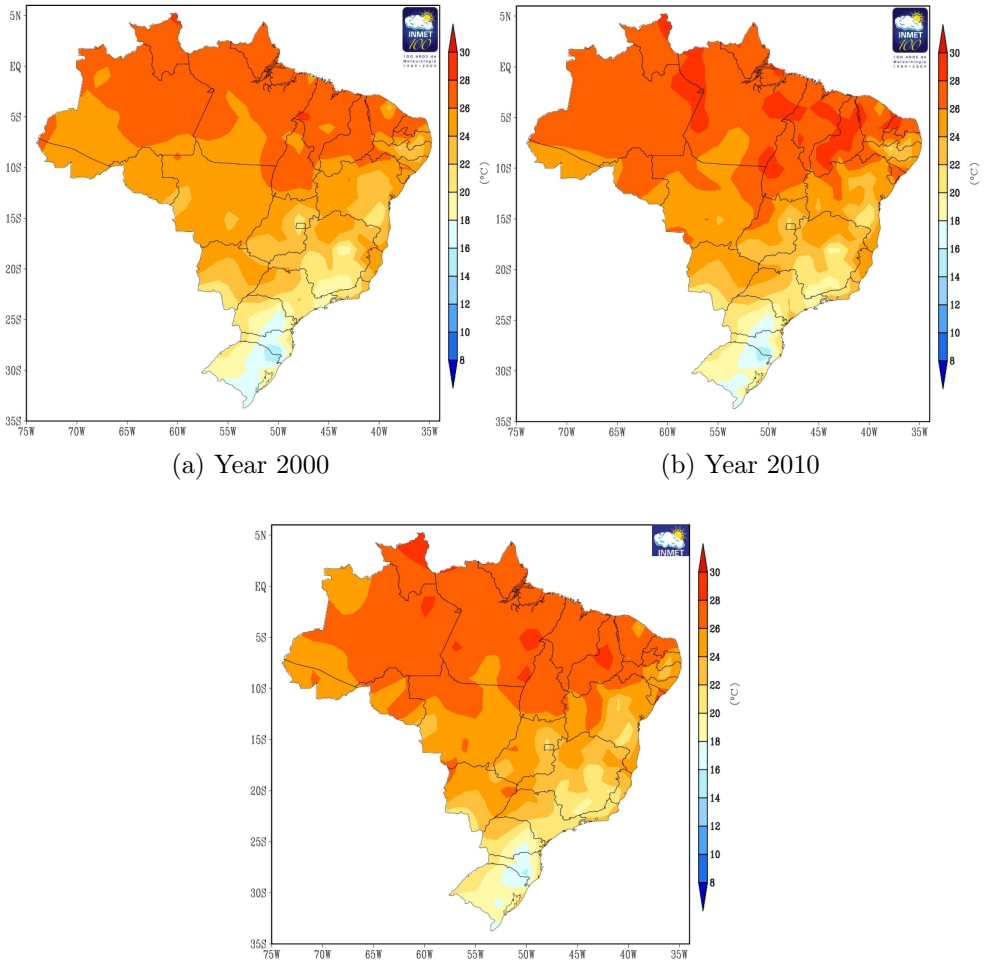
Table 2.1: Global diffusion distance in the years 2010 and 2018, with respect to year 2000.

Section 2.3.1. Therefore, we cannot use the singular vector basis to represent the diffusion distance. Here, we use the Fourier basis defined in Eq. (2.19) to represent this diffusion distance. We approximate the dynamic diffusion distance using the parameters $k_1 = 13,974$, and $k_2 \in \{5, 100\}$. See Figures 2.14 and 2.16 for 2010 and 2018, respectively. The green-yellow scale represents the intensity of the dynamic diffusion distance, in which the yellow regions have a greater diffusion distance as compared to the green regions. To detect which regions have the greatest positive variation, that is, regions in which the temperature has increased, we use contour plots of the diffusion distance, taking into account regions where the temperature increased. See Figures 2.15 and 2.17 for 2010 and 2018, respectively.

In Table 2.1, we show the global diffusion distance for each year. Observe that the distance is greater in 2018 than in 2010. This suggests that during 2018, there were more temperature changes when compared to 2010. In Figures 2.18 and 2.19, we plot the error and computational time of the performance for several approximation orders k_2, for 2010 and 2018, respectively. We evaluate the performance of the orders using the metric

$$M_B = \frac{1}{N} E \times t_{cpu} \ , \tag{2.20}$$

where E is the absolute error between $(D^1(x_{2000}, x_\gamma))^2$ and its approximation, and t_{cpu} is the computational time (in seconds) to compute the approximation. We see that even using small orders, it is possible to obtain a good performance compared to larger orders.

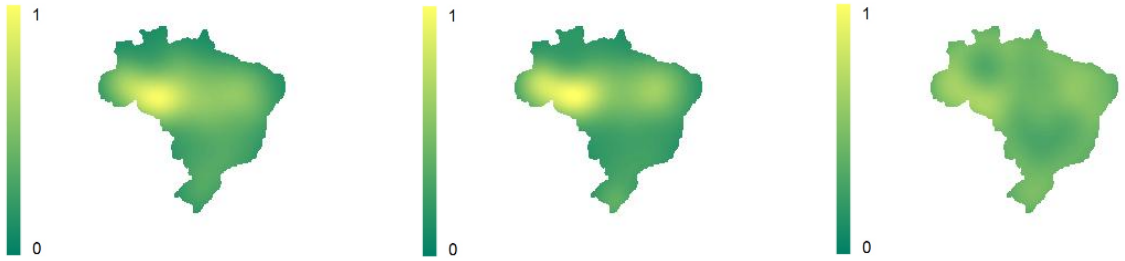


(a) Year 2000

(b) Year 2010

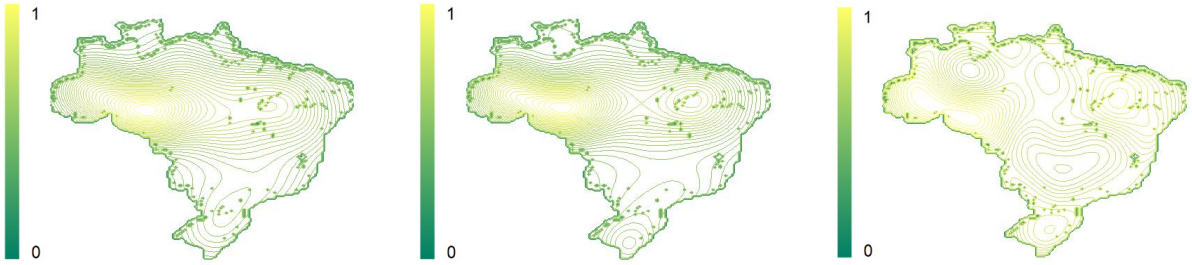
(c) Year 2018

Figure 2.13: Plot of the average Brazilian temperature distribution in years 2000, 2010, 2018.



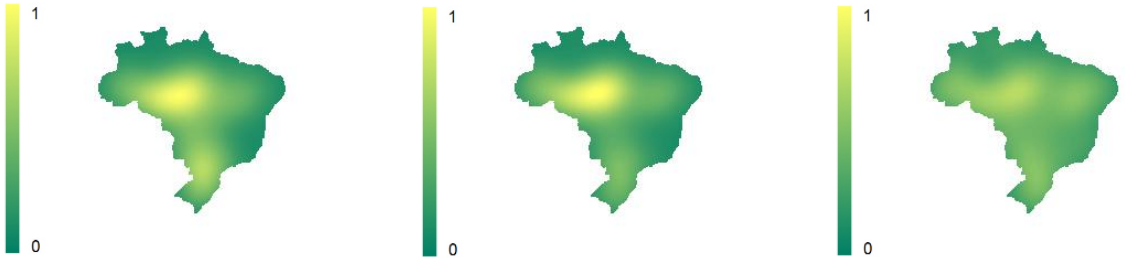
(a) $k_2 = 5$ (b) $k_2 = 100$ (c) $D^1(x_{2000}, x_{2010})^2$

Figure 2.14: Plot of the approximations for the normalized dynamic diffusion distance $(D^1(x_{2000}, x_{2010}))^2$ with different values of k_2 (a), (b), and plot of the dynamic diffusion distance (c).



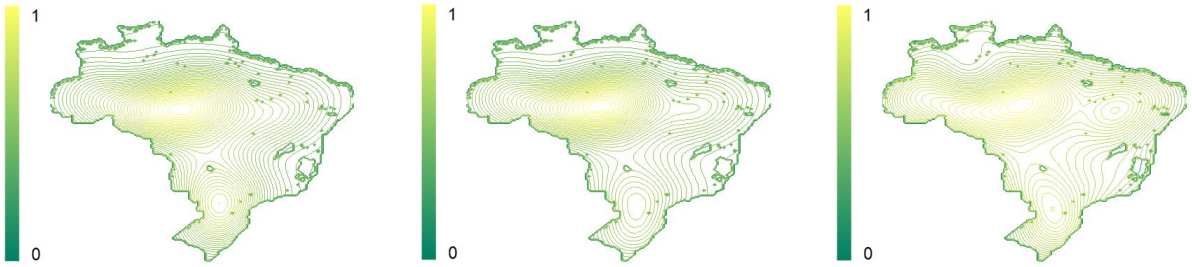
(a) $k_2 = 5$ (b) $k_2 = 100$ (c) $D^1(x_{2000}, x_{2010})^2$

Figure 2.15: Contour plot of the representation approximation for the normalized diffusion distance $(D^1(x_{2000}, x_{2010}))^2$ taking into account the temperature increase.



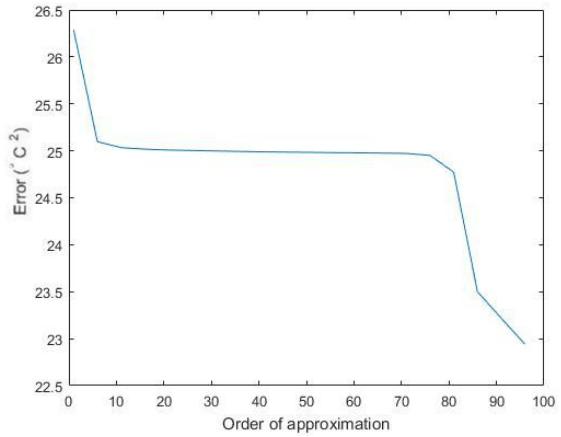
(a) $k_2 = 5$ (b) $k_2 = 100$ (c) $D^1(x_{2000}, x_{2010})^2$

Figure 2.16: Plot of the approximations for the normalized dynamic diffusion distance $(D^1(x_{2000}, x_{2018}))^2$ with different values of k_2 (a), (b), and plot of the dynamic diffusion distance (c).

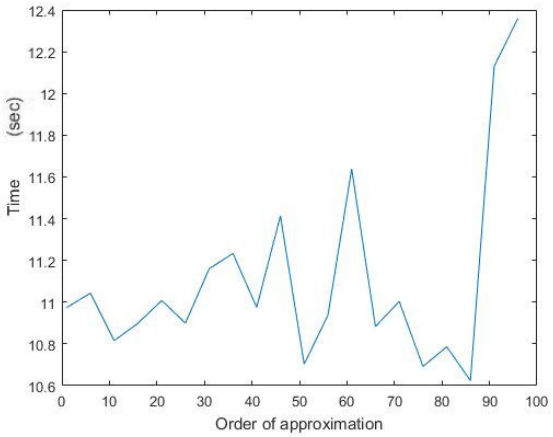


(a) $k_2 = 5$ (b) $k_2 = 100$ (c) $D^1(x_{2000}, x_{2010})^2$

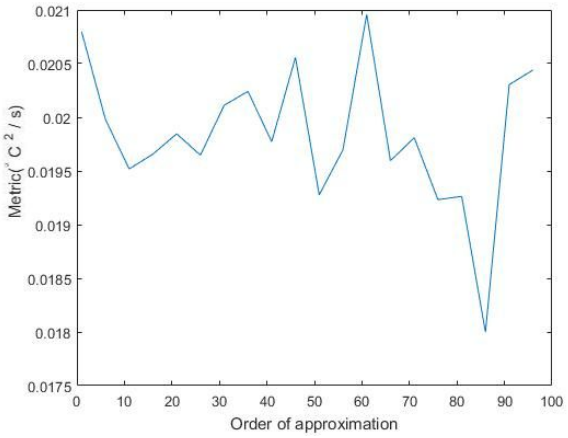
Figure 2.17: Contour plot of the representation approximation for the normalized diffusion distance $(D^1(x_{2000}, x_{2018}))^2$ taking into account the temperature increase.



(a) Error of the approximation

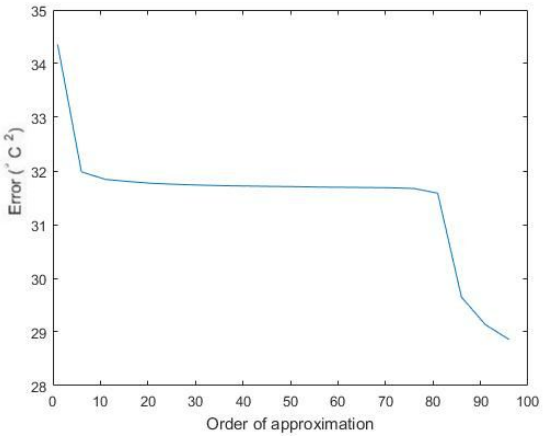


(b) Computational time

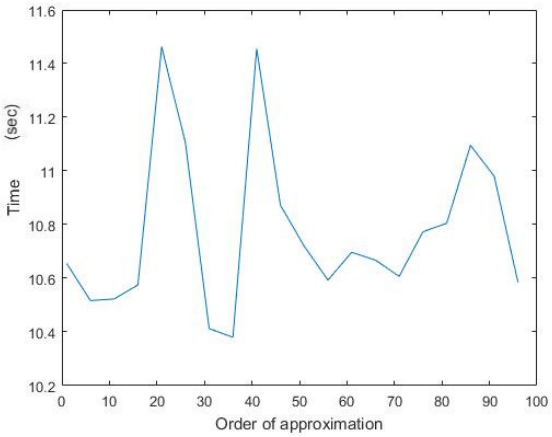


(c) Metric M_B

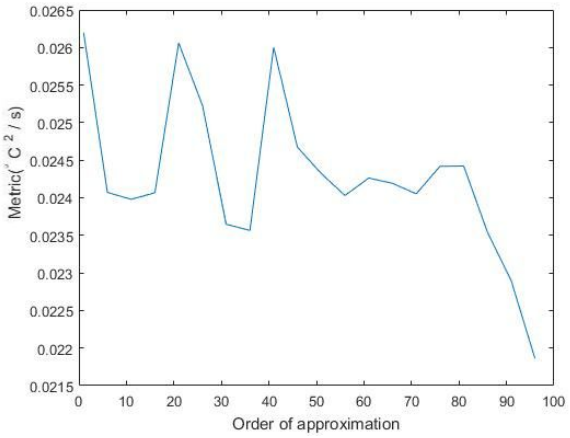
Figure 2.18: Plot of the absolute error, computational time, and the metric given by Eq. (2.20) with different orders to approximate $(D^1(x_{2000}, x_{2010}))^2$, for the dataset with the temperature distribution in Brazil.



(a) Error of the approximation



(b) Computational time



(c) Metric M_B

Figure 2.19: Plot of the absolute error, computational time, and the metric given by Eq. (2.20) with different orders to approximate $(D^1(x_{2000}, x_{2018}))^2$, for the dataset with the temperature distribution in Brazil.

Chapter 3

A diffusion-map-based algorithm for gradient computation on manifolds and applications

Highlights

- A novel methodology is developed to efficiently retrieve the Riemannian gradient defined on a Riemannian submanifold of Euclidean space.
- The proposed approach is based on the estimates of Diffusion-Maps for the Laplace-Beltrami operator.
- The gradient-estimate is used to analyze gradient flows on lower-dimensional submanifolds embedded in high-dimensional spaces.
- The methodology is employed for minimizing non-differentiable real functions defined on Riemannian submanifolds.

- The method's effectiveness in the two and three-dimensional sphere packing problem is verified.

- The theory is applied to efficiently and effectively provide a tomographic reconstruction from an unknown random angle distribution.

This chapter is organized as follows, in Section 3.1, we give a brief exposition of the classical representation theory for diffusion distances proposed in Refs. [19, 18, 20], and we state our main result in Theorem 3.1.1. In Section 3.2, we review facts about flows defined over manifolds, and we show how to use the flow generated by the approximations to find minimizers. In Section 3.3, we show some experiments related to the sphere packing problem, and we also show the effectiveness of our tomographic reconstruction method when the angles are unknown. Finally, in Appendices A and B, we cover the technical details of the proof of the main result.

3.1 Recalling diffusion-maps

We assume that our dataset $X = \{x_i\}_{i=1}^k$ satisfies $X \subset \mathcal{M} \subset \mathbb{R}^n$, where $\mathcal{M}$ is a $d-$*dimensional* Riemannian submanifold of the ambient space $\mathbb{R}^n$. In this case the dimension d of $\mathcal{M}$ is assumed to be much smaller than n. In our approach, we use asymmetric vector-valued kernels as in Chapter 2. The main advantage of using these kernels is that we have a more specific description of the distribution of the dataset in certain directions. Based on the expansion for the Laplace-Beltrami operator proposed in Ref. [19] we recover the Riemannian gradient. Firstly, we consider the vector-valued kernel

$$\overline{K}_t : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^n,$$

defined as

$$\overline{K}_t(x,y) = (y-x)e^{\frac{-\|y-x\|^2}{2t^2}}.$$

We fix the exponent $\delta \in (1/2,1)$, and let $d_t(x)$ be defined by

$$d_t(x) = \int_{U(x,t^\delta)} e^{\frac{-\|y-x\|^2}{2t^2}} dy,$$

where

$$U(x,t) = \{y \in \mathcal{M} | \|y-x\| \leq t\}. \tag{3.1}$$

Here, the parameter δ has to be in $(1/2,1)$ to guarantee convergence of the estimates as shown in Lemma B.0.1. We consider the Markov normalized kernel given by

$$\rho_t(x,y) = \frac{\overline{K}_t(x,y)}{d_t(x)}.$$

For a function f, we define the operator

$$\overline{P}_t f(x) = \int_{U(x,t^\delta)} \rho_t(x,y)(f(y)-f(x))dy. \tag{3.2}$$

We now show that this operator approximates the Riemannian gradient of a given function on some Riemannian submanifold. The technical details of the proof are given in Appendices A and B.

Theorem 3.1.1. *Let $\mathcal{M}$ be a Riemannian submanifold of $\mathbb{R}^n$ and assume that the function f is smooth, and x is an interior point of $\mathcal{M}$. Then, the following estimate holds*

$$\overline{P}_t f(x) = \nabla f(x)\, t^2 + O(t^{4\delta}), \tag{3.3}$$

where ∇f is the Riemannian gradient of f. In particular, we have that

$$\lim_{t\to 0} \frac{\overline{P}_t f(x)}{t^2} = \nabla f(x). \tag{3.4}$$

Note that the operator $\overline{P}_t$ does not depend on differentiability condition. Furthermore, since the operator is defined as an integral one, then it is robust to noise perturbation. Considering these characteristics, we use this operator as a substitute for the Riemannian gradient as the main direction of a gradient-based algorithm on manifolds detailed in Ref. [1].

3.2 Flows and optimization methods on submanifolds

In this section, we review some facts of flows defined on submanifolds and we show how the flow generated by the vector field $\overline{P}_t f(\cdot)$ can be used in optimization methods.

Assume that $h : \mathcal{M} \to \mathbb{R}^n$ is a continuous function defined on the submanifold $\mathcal{M} \subset \mathbb{R}^n$. We say that a curve b starts at x_0, if $b(0) = x_0$. The Peano existence theorem guarantees that for all $x_0 \in \mathcal{M}$, there exists a smooth curve $c_{h,x_0} : (-\varepsilon, \varepsilon) \to \mathcal{M}$ starting at x_0, which is solution of

$$c'_{h,x_0}(s) = -h(c_{h,x_0}(s)). \tag{3.5}$$

We refer the reader to Ref. [72] for a complete background about ordinary differential equations. We observe that assuming only the continuity condition, the uniqueness of the curve is not guaranteed. Since the solution of Eq. (3.5) may not be unique, we can concatenate solutions as follows. Let c_{h,x_0} be a solution of Eq. (3.5) starting at the point x_0. For a fix s_1 in the domain of c_{h,x_0}, we define $x_1 = c_{h,x_0}(s_1)$. If c_{h,x_1} is a solution of

Eq. (3.5) starting in x_1, we define a new curve c_{h,x_0,x_1} as

$$c_{h,x_0,x_1}(s) = \begin{cases} c_{h,x_0}(s), & \text{for } s \leq s_1 \\ c_{h,x_1}(s - s_1), & \text{for } s_1 < s \end{cases}.$$

Proceeding recursively, we obtain a piecewise differentiable curve $c_{h,x_0,x_1,x_2\cdots}(s)$ starting at x_0, and satisfying Eq. (3.5) (except in a discrete set). See Figure 3.1 for a graphic description. In this case, we say that the curve $c_{h,x_0,x_1,x_2\cdots}(s)$ is a piecewise solution of Eq. (3.5). We focus on curves which are solutions (except in a discrete set) of Eq. (3.5), because these curves allow updating the direction in which we look for stationary points.

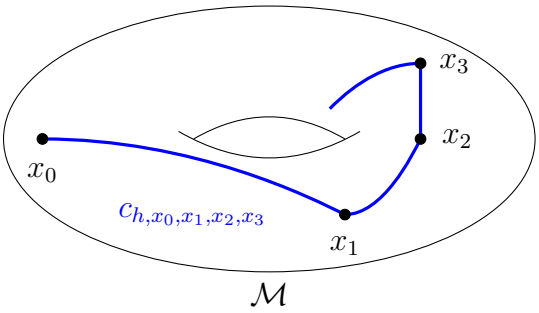


Figure 3.1: Piecewise curve obtained by concatenating four curves.

Suppose that $f : \mathcal{M} \to \mathbb{R}$ defines a smooth function. In this case we consider the vector field $h = \nabla f$. If $c_{h,x_0,x_1,x_2\cdots}$ is a piecewise solution of Eq. (3.5) starting at x_0. Then, for all t (except in a discrete set), we have that

$$\|c'_{h,x_0}(s)\|^2 = -\frac{d}{ds} f(c_{h,x_0}(s)). \tag{3.6}$$

Therefore, the function $f(c_{h,x_0}(\cdot))$ is decreasing. Thus, we can use use the flow c to find local minimum for the function f .

3.2.1 Lipschitz functions

We recall that f is a locally Lipschitz function if for all $x \in \mathcal{M}$ there exists a neighborhood $x \in U \subset \mathcal{M}$ and a positive constant C, such that for all $y \in U$ it holds that

$$|f(x) - f(y)| \leq C\|x - y\|_{L^2}.$$

We also recall that the Sobolev space $H^1(0, T, \mathcal{M})$ is defined as the set of all square integrable functions from $[0, T]$ to $\mathcal{M}$ whose weak derivative has also finite L^2 norm.

Our goal is to use the gradient approximation in Theorem 3.1.1 to find minimal points of locally Lipschitz functions. Recall that Rademacher's theorem states that for a locally Lipschitz function f, the gradient operator ∇f exists almost everywhere. See Ref. [32] for more details. However, for a locally Lipschitz function f, the gradient ∇f may not exist for all points. In this case, it is not possible to define the gradient flow.

To address this problem, we propose to use the flow generated by the approximation $\frac{\overline{P}_t f(x)}{t^2}$ defined in Eq. (3.2) instead of the gradient. The operator $\overline{P}_t f$ is defined as an integral, and thus it is continuous. This fact guarantees the existence of a flow associated with $\frac{\overline{P}_t f(x)}{t^2}$ for arbitrarily small positive t.

Now we show that at the points where the function is smooth, this flow approximates a curve for which the function decreases with time. To do that, we first prove a technical result.

Proposition 3.2.1. *Suppose that f is continuously differentiable in an open neighborhood of x_0. We define the function $J : [0, T] \times \overline{B(x_0, R) \cap \mathcal{M}} \to \mathbb{R}^n$ as*

$$J(t, x) = \begin{cases} \frac{P_t f(x)}{t^2}, & \text{for } t > 0 \\ \nabla f(x), & \text{for } t = 0 \end{cases},$$

where $B(x_0, R)$ is the ball in $\mathbb{R}^n$ with center x_0 and radius R. Then, for small enough numbers T, R, the function J is uniformly continuous. In particular, there exists a positive constant M such that for all $(t, x) \in (0, T] \times \overline{B(x_0, R) \cap \mathcal{M}}$ the following estimate holds.

$$\frac{P_t f(x)}{t^2} \leq M. \tag{3.7}$$

Proof. Since the set $[0, T] \times \overline{B(x_0, R) \cap \mathcal{M}}$ is compact, it is enough to show that J is continuous. Firstly, we show that J is continuous on $(0, T] \times \overline{B(x_0, R) \cap \mathcal{M}}$. For that, we claim that for a continuous vector-valued function $\omega : (0, T] \times \overline{B(x_0, R) \cap \mathcal{M}} \times \overline{B(x_0, R) \cap \mathcal{M}} \to \mathbb{R}^m$, the operator

$$\Omega(t, x) = \int_{U(x, t^\delta)} \omega(t, x, y)\, dy,$$

is continuous. In fact, we observe that

$$\Omega(t, x) - \Omega(t_1, x_1) = \int_{U(x, t^\delta)} \omega(t, x, y) - \omega(t_1, x_2, y)\, dy + \int_{G(t, t_1, x, x_1)} \omega(t_1, x_2, y) dy, \tag{3.8}$$

where

$$G(t, t_1, x, x_1) = U(x_1, t_1^\delta) \setminus U(x, t^\delta) \cup U(x, t^\delta) \setminus U(x_1, t_1^\delta).$$

On the other hand, a straightforward computation shows that

$$\lim_{(t_1, x_1) \to (t, x)} 1_{G(t, t_1, x, x_1)} = 0,$$

where the convergence is pointwise almost everywhere, therefore

$$\lim_{(t_1, x_1) \to (t, x)} \int_{G(t, t_1, x, x_1)} \omega(t_1, x_2, y) dy = 0 \tag{3.9}$$

In addition, since the function ω is continuous, then

$$\lim_{(t_1,x_1)\to(t,x)} \int_{U(x,t^\delta)} \omega(t,x,y) - \omega(t_1,x_2,y)\, dy = 0. \tag{3.10}$$

Using Eqs. (3.9) (3.10) in Eq. (3.8), we conclude that Ω is a continuous function. We apply the previous result to the function $w_1(t,x,y) = e^{\frac{-\|y-x\|^2}{2t^2}}$ to obtain that $\Omega_1(t,x) = d_t(x)$ is a continuous function. This implies that the function

$$w_2(t,x,y) = \frac{\rho_t(x,y)(f(y)-f(x))}{t^2},$$

is continuous for on $(0,T]\times\overline{B(x_0,R)\cap\mathcal{M}}$. Again, we apply the same result to the function

$$w_2(t,x,y),$$

to conclude that $J(t,x)$ is a continuous function on $(0,T]\times\overline{B(x_0,R)\cap\mathcal{M}}$.

Moreover, using Estimate (B.2) of the proof of Theorem 3.4 and Lemma B.0.2, we conclude that the function J is continuous for all points of the form $(0,x)$. This proves our result. □

The estimate of Proposition 3.2.1 states that for a fixed x_0, and small T, the family of curves $\{c_{h(t_n),x_0}\}_{t_n}$ is uniformly bounded on the Sobolev space $H^1(0,T,\mathcal{M})$. Thus, the Rellich-Kondrachov theorem states that for any sequence $t_n\to 0$, there exists a subsequence $t_{n_k}\to 0$ such that $c_{h(t_{n_k}),x_0}$ converges to some curve c in the L^2-norm. Observe that by Arzela-Ascoli theorem, we can also suppose that the sequence $c_{h(t_n),x_0}$ converges uniformly to c. Finally we prove the main result in this section.

Proposition 3.2.2. *Assume the same assumptions and notations of Proposition 3.2.1.*

Then, for $t_1 < t_2$ we have that

$$f(c(t_1)) \geq f(c(t_2)).$$

Proof. We claim that $\frac{\overline{P}_t f(c_{h(t_n),x_0}(\cdot))}{t^2}$ converges pointwise to $\nabla f(c(\cdot))$, where c is the curve previously described. In fact, for all s, we have by Proposition 3.2.1 that

$$\lim_{n\to\infty} \frac{\overline{P}_t f(c_{h(t_n),x_0}(s))}{t^2} - \nabla f(c_{h(t_n),x_0}(s)) = 0.$$

The continuity of the gradient guarantees that

$$\lim_{n\to\infty} \nabla f(c_{h(t_n),x_0}(s)) = \nabla f(c(s)).$$

The above estimates prove our claim. Using inequality (3.7) together with the dominated convergence theorem, we obtain that

$$\lim_{n\to\infty} \int_0^T \left\| \frac{\overline{P}_t f(c_{h(t_n),x_0}(s))}{t^2} - \nabla f(c(s)) \right\|^2 ds = 0. \tag{3.11}$$

On the other hand, since $c_{h(t_n),x_0}(I)$ is solution of Eq. (3.5), then

$$\begin{aligned} 0 &\geq \langle \tfrac{\overline{P}_t f(c_{h(t_n),x_0}(s))}{t^2}, c'_{h(t_n),x_0}(s) \rangle \\ &\geq \langle \tfrac{\overline{P}_t f(c_{h(t_n),x_0}(s))}{t^2} - \nabla f(c(s)), c'_{h(t_n),x_0}(s) \rangle + \langle \nabla f(c(s)), c'_{h(t_n),x_0}(s) - c'(s) \rangle \\ &+ \langle \nabla f(c(s)), c'(s) \rangle. \end{aligned}$$

Using the weak convergence assumption, together with Eq. (3.11), we conclude that for all points $t_1 < t_2$, the following inequality holds

$$0 \geq \int_{t_1}^{t_2} \langle \nabla f(c(s)), c'(s) \rangle ds = f(c(t_2)) - f(c(t_1)).$$

□

The previous result establishes that the flow generated by $\frac{\overline{P}_t f(x)}{t^2}$ approximates a curve c for which the function f is decreasing.

3.3 Algorithm Development

In this section we propose a computational algorithm to approximate the Riemannian gradient of a function defined on a Riemannian submanifold of the Euclidean space using a set of sample points. We use these approximations as principal directions in gradient-based algorithms as described in [1]. If the function is not differentiable at a point x, we say that x is a singularity. Here, we assume that the singularity points form a discrete set.

Theorem 3.1.1 states that the operator $\overline{P}_t f(x)$ can be used to approximate the Riemannian gradient. An important task is to compute the integrals involving the operator $\overline{P}_t$, defined in Eq. (3.2). In practical applications, we only have access to a finite sample points $x_1, x_2, x_3, \cdots, x_m$ on $U(x,t)$, which are the realizations of i.i.d random variables with probability density function *(PDF)* q. However, the integral in Eq. (3.2) does not depend on the *(PDF)* q. To address this issue, for a fixed x, we consider the normalized points

$$(x_i - x)(f(x_i) - f(x))\ e^{\frac{-\|x_i - x\|^2}{2t^2}}\ /q(x_i),$$

$(i = 1, \cdots, m)$ which are realizations of i.i.d random variables regarding the *PDF* $q(x)$. In that case, the Law of Large Numbers *LLN* guarantees that

$$\overline{P}_t f(x) = \lim_{m\to\infty} \frac{1}{m\ d_t(x)} \sum_{i=1}^{m} (x_i - x)\ (f(x_i) - f(x))\ e^{\frac{-\|x_i - x\|^2}{2t^2}}\ /q(x_i),$$

where $d_t(x)$ can be computed similarly using the LLN

$$d_t(x) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} e^{\frac{-\|x_i - x\|^2}{2t^2}} \ /q(x_i).$$

In particular, when the PDF is the function

$$q(y) = e^{\frac{-\|y-x\|^2}{2t^2}} \ /d_t(x), \tag{3.12}$$

we can approximate $\overline{P}_t f(x)$ using $\mathcal{V}$, where

$$\mathcal{V} = \frac{1}{m} \sum_{i=1}^{m} (x_i - x) \ (f(x_i) - f(x)). \tag{3.13}$$

This vector is analogous to the weighted gradient operator defined for graphs. See [7] for more details. The parameter t controls how much we approximate the true gradient. Needless to say, a choice of an extremely small t would lead to numerical instabilities, and thus t in a certain sense would work as a regularization parameter. We shall call t the *gradient approximation parameter* and it will be provided as an input to the Algorithm 2.

Algorithm 2 Approximate Gradient Sampling Algorithm

input Sample points $x_1, x_2, x_3 \cdots x_m$ on $U(x,t)$ with *PDF* q, and gradient approximation parameter t.

1. **for** $i = 1$ to m **do**
   - $c_i \leftarrow e^{\frac{-\|x_i - x\|^2}{2t^2}} / q(x_i)$
2. **end for**
3. $d_t \leftarrow \sum_{i=1}^{m} c_i$
4. $\mathcal{V} \leftarrow \frac{1}{d_t} \sum_{i=1}^{m} (x_i - x)\ (f(x_i) - f(x))\ c_i$

return $\mathcal{V}$ which is an approximation for $\overline{P}_t f(x)$

We apply Algorithm 2 in a gradient-based optimization method. Intuitively, Proposition 3.2.2 says that the energy associated with the gradient decreases along the curve c. Therefore, we can use this curve to find a better approximation for local minimizers, ultimately leading to a derivative-free optimization method. The proposed algorithm is useful in situations where it is not straightforward to compute the gradient of a function.

Using Proposition 3.2.2, we have that the flow generated by

$$Dir = \frac{\overline{P}_t f(x)}{t^2}, \tag{3.14}$$

approximates a curve along which the function f decreases. Thus, suggesting that if we use the direction Dir defined in Eq. (3.14) as the main direction in a gradient-based algorithm, then in a certain way we are approximating the gradient descent method. The

gradient-based optimization method generated by the direction *Dir* is described by

$$x_{k+1} = \beta_{x_k}(x_k - \lambda \frac{\overline{P}_t f(x_k)}{t^2}),$$

where λ is some relaxation parameter which defines the step size and β_x is a local retraction of $\mathcal{M}$ around the point x.

We recall that a local retraction β consists of a locally defined continuous map from the ambient space onto the manifold $\mathcal{M}$ such that it coincides with the identity when restricted to $\mathcal{M}$. In other words, $\beta \circ \iota = I_A$, where A is a open set in $\mathcal{M}$, and ι is the inclusion map from $\mathcal{M}$ into the ambient space.

The parameter λ must be regularly reduced to avoid instabilities in our iteration. We propose to reduce the *relaxation parameter* λ by a step-scale factor s_f after l consecutive numerical iterations. This procedure is similar to Armijo point rule described in [1] . We shall call l the *sub-iteration control number*.

We update the size λ of the step such that after a certain number of iterations, it decreases to a pre-conditioned proportion. We do this since the interval for which the curve is defined can be limited, and iterating with a fixed size would generate instabilities in the algorithm. Therefore, if we take smaller step sizes as the number of iterations increases, we obtain better estimates for the minimizer. As the iteration numbers increases, we get closer to a local minimum. For this reason, our stopping criteria is achieved when

$$|f(x_k) - f(x_{k+1})| \leq \epsilon,$$

for a certain accuracy ϵ. The latter will be called the *termination tolerance on the function value* and will be provided as an input parameter. Results on the convergence of this algorithm, as well as stopping criteria are described in Ref. [1].

We summarize the above discussion in Algorithm 3.

Algorithm 3 Diffusion-map-based optimization

input Initial guess x_0, gradient approximation parameter t, relaxation parameter λ, sub-iteration control number l, termination tolerance ϵ, and step-scale factor s_f.

initialization

$k \leftarrow 0$

$counter \leftarrow 0$

$x_{min} \leftarrow x_0$

$x_{-1} \leftarrow x_0$

while $|f(x_{k-1}) - f(x_k)| \geq \epsilon$ **or** $k = 0$

1. $x_{k+1} \leftarrow \beta_{x_k}(x_k - \lambda \frac{\overline{P}_t f(x_k)}{t^2})$
2. **if** $f(x_{k+1}) < f(x_{min})$ **do**
   - $x_{min} \leftarrow x_{k+1}$
3. **end if**
4. $k \leftarrow k + 1$
5. **if** $l < counter$ **do**
   - $counter \leftarrow 0$
   - $x_k \leftarrow x_{min}$
   - $\lambda \leftarrow \lambda / s_f$
6. **end if**
7. $counter \leftarrow counter + 1$

end while

return x_{min}

3.3.1 High-dimensional datasets

In many optimization problems, the dataset consists of sample points lying in an unknown lower-dimensional submanifold embedded in a high-dimensional space. We propose to use the dimensional reduction method and then, Algorithm 3 to solve the optimization problem in the embedded space. This will be done without directly involving the *a priori* knowledge of the manifold.

To be more specific, we assume that the optimization problem under consideration consists on minimizing the cost function f over the dataset $X = \{x_i\}_{i=1}^k$. Regarding the dataset, we suppose that $X \subset \mathcal{M} \subset \mathbb{R}^n$, where n is a large number, and $\mathcal{M}$ is a lower-dimensional Riemannian submanifold. Since the information contains a large number of irrelevant data that make the computing process inefficient, we use the diffusion-maps approach to embed our dataset in a lower-dimensional space. This embedding process allows us to work only with the most important features, and thus, we obtain a better computational performance of the optimization algorithm. We denote the embedded points by

$$y_i = \psi_m^t(x_i), \tag{3.15}$$

where ψ_m^t is the diffusion-map. We apply Algorithm 3 to the dataset $Y = \{y_i\}_{i=1}^k$, and the function $\tilde{f}$. Here, the function $\tilde{f}$ is defined as $\tilde{f}(y_i) = f(x_i)$, for all $x_i \in X$, and y_i the associated point (3.15). In this case, we use the retraction β, defined as the projection on Y, that is,

$$\beta(x) = \underset{y_i \in Y}{\arg\min} \ \|x - y_i\|.$$

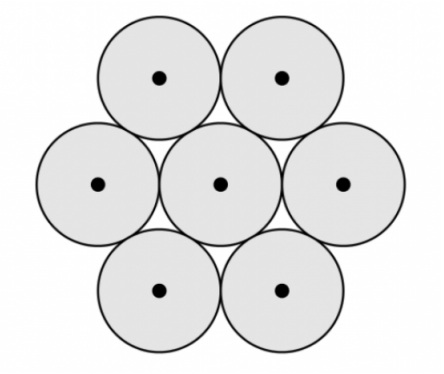


Figure 3.2: The hexagonal lattice and the corresponding sphere packing.

3.4 Numerical Experiments and Applications

The following experiments were implemented in MATLAB software, using a desktop computer with the following configuration: Intel i5 9400 4.1 GHz processor, and 16 GB RAM.

3.4.1 Sphere packing problem in dimensions 2 and 3

The sphere packing problem in the Euclidean space asks the following question: How to arrange congruent balls as densely as possible, such that they touch each other at most on the boundary. This problem has exact solution in dimensions $1, 2, 3, 8$, and 24. See [73, 17]. The one-dimensional sphere packing problem is the interval packing problem on the line, which is trivial. The two and three-dimensional are far from trivial. The two-dimensional case of packing circles in the plane was first solved in 1892 by the Norwegian mathematician Thue. He showed that the hexagonal packing gives the largest density; see Figure 3.2. The three-dimensional case of packing spheres in $\mathbb{R}^3$ was first posed by Johannes Kepler in 1611, so it is often referred to as the Kepler Conjecture. In 1831 Gauss proved that the pyramid arrangement of equally sized spheres filling space is optimal among lattice packings, but this still leaves out a lot of possible packings; see Figure 3.3. In 1900 Hilbert included this problem in his famous list of unsolved problems (as part of the 18th).

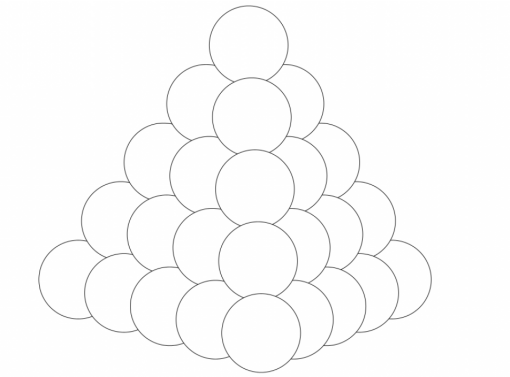


Figure 3.3: The pyramid sphere packing in $\mathbb{R}^3$.

The three-dimensional problem was solved by Thomas Hales in 2005, he gave a complex proof, which makes intensive use of computers [38]. In 2017, Maryna Viazovska solved the problem in dimensions eight and twenty-four. See Refs. [73, 17].

A remarkable feature of this problem is that each dimension has its peculiarities. It does not seem likely that a single and simple construction will give the best packing in every dimension.

We now discuss the problem in more detail. We denote Vol the volume form associated with the Lebesgue measure, and for $x \in \mathbb{R}^n$ and r a positive real number, we denote by $B(x, r)$ the ball in $\mathbb{R}^n$ with center x and radius r.

How do we define a sphere packing in the n dimensional space? To this end, we assume that $C \subset \mathbb{R}^n$ be a discrete set of points such that $2r \leq \|x - y\|$, for any two distinct $x, y \in C$, where r is a positive real number. Then, the union

$$S = \bigcup_{x \in C} B(x, r),$$

is a sphere packing, and its density Δ_S is defined as

$$\Delta_S = \limsup_{r \to \infty} \frac{Vol(S \cap B(0, r))}{Vol\left(B(0, r)\right)}.$$

Intuitively, the density of a sphere packing is the fraction of space covered by the spheres of the packing. The sphere packing problem consists in knowing what is the supremum Δ_n over all possible packing densities in $\mathbb{R}^n$. The number Δ_n is called the n dimensional sphere packing constant.

One important way to create a sphere packing is to start with a lattice $\Lambda \subset \mathbb{R}^n$, and center the spheres at the points of Λ, with radius half the length of the shortest non-zero vectors in Λ. Such packing is called lattice packing. A more general notion than lattice packing is periodic packing. In periodic packings, the spheres are centered on the points in the union of finitely many translates of a lattice Λ. Not every sphere packing is a lattice packing, and, in all sufficiently large dimensions, there are packings denser than every lattice packing. In contrast, it is proved in Ref. [36] that periodic packings get arbitrarily close to the greatest packing density. Moreover, in Ref. [36] it is shown that for every periodic packing P of the form

$$P = \bigcup_{i=1}^{k} \bigcup_{x \in \Lambda} (x_i + B(x, r)),$$

where Λ is a lattice, its density is given by

$$\Delta_P = k \frac{Vol\,(B(0,r))}{Vol\,(\Lambda)},$$

where $r = \min_{x,y \in P} \|x - y\|$.

Observe that the density packing is invariant under scaling, that is, for a lattice Λ and a positive constant α we have $\Delta_{\alpha\Lambda} = \Delta_\Lambda$. Thus, without loss of generality and normalizing if necessary, we can assume that the volume of the lattice is $Vol\,(\Lambda) = 1$. If $b_1, \cdots b_n$ is a

basis for Λ, then our problem can be reformulated as

$$\begin{array}{ll} \max\limits_{b_1,\cdots b_n} & Vol\,(B(0,1))\,(\dfrac{g(b_1,\cdots b_n)}{2})^n \\ \text{subject to} & det\,(b_1,\cdots,b_n)=1. \end{array} \tag{3.16}$$

where $det(\cdot)$ is the determinant function, and the function $g(b_1,\cdots b_n)$ is defined as the minimum value of $\|z_1\,b_1+\cdots+z_nb_n\|_2$ over all possible $(z_1,\cdots z_n)\in\mathbb{Z}^n\setminus 0$.

Since the function g is defined as a minimum, then this function is non-differentiable at least in the set of orthonormal matrices. In fact, if we consider an orthonormal set $b_1,\cdots,b_n$, then $g(b_1,\cdots,b_n)=1$. In that case, the smooth curve defined as

$$c(t)=(tb_1,\frac{1}{t}b_2,b_3,\cdots,b_n),$$

for $t>0$, satisfies

$$g(c(t))=\begin{cases} \frac{1}{t}, & \text{for } t\geq 1 \\ t & \text{for } t<1 \end{cases}.$$

Since $g(c(t))$ is non-differentiable, then g is not differentiable in $(b_1,\cdots,b_n)$

To apply our approach, we first prove that the function g is locally Lipschitz. We write the matrices A and B as the column form $A=[a_1,\cdots a_n]$ and $B=[b_1,\cdots b_n]$, and the special linear group as $SL(n)=\{A\,|\,det(A)=1\}$. Since the inverse of a matrix is a continuous function on $SL(n)$, then for $A\in SL(n)$, there exists an open set $U\ni A$ and a positive constant D such that for all $B\in U$

$$\|B^{-1}\|_2\leq D.$$

Assume that $g(a_1, \cdots, a_n) = \|A\,\vec{z}\|_2$ for some $\vec{z} \in \mathbb{Z}^n \setminus 0$. Then, we have that

$$\begin{aligned} g(b_1, \cdots, b_n) - g(a_1, \cdots, a_n) &\leq \|(A-B)\|_2\,\|\vec{z}\|_2 \\ &\leq \|A^{-1}\|_2 \|A-B\|_2 \|A\,\vec{z}\|_2. \end{aligned}$$

Minkowski's theorem for convex sets, Ref. [59], guarantees that for any matrix A with $det(A) = 1$, the estimate $g(A) \leq \sqrt{n}$ is satisfied. Thus, we obtain that

$$g(b_1, \cdots, b_n) - g(a_1, \cdots, a_n) \leq \sqrt{n} D \|A-B\|_2.$$

By symmetry, the above inequality is still valid if we change the order of A and B. This proves that g is locally Lipschitz.

In dimensions 2 and 3 the solutions of the problem in Eq. (3.16) are $\Delta_2 = \frac{\pi}{2\sqrt{3}}$ and $\Delta_3 = \frac{\pi}{3\sqrt{2}}$,, respectively. In these dimensions the maximizers are the hexagonal lattice, Figure 3.2, and the pyramid lattice packing, Figure 3.3.

Observe that the problem in Eq. (3.16) can be considered as an optimization problem on the manifold $SL(n)$. We use our approach to find the maximizers in dimensions 2 and 3. Since maximizing the function g is equivalent to minimizing $-g$, then we apply our approach to the function $-g$. We use Algorithm 3 to minimize the function $-g$, and thus Algorithm 2 to compute $\overline{P}_t f(x)$. In this experiment, we use the *PDF* function q defined as in Eq (3.12) to compute the gradient. In this case, the Approximation is given in Eq. (3.13). We generate a total of $m = 20$ sample points from the normal distribution using the MATLAB function *normrnd*, and then projected to the manifold $SL(n)$ using the retraction given in Eq. (3.17). Since $\Delta_n \leq 1$, then, we take a small initial step size to get a better performance of our methodology. Our initial guess x_0, is the identity matrix and initial parameters $t = 10^{-5},\ \lambda = 0.1,\ l = 20,\ \epsilon = 10^{-10},\ s_f = 1.1$. We use the local

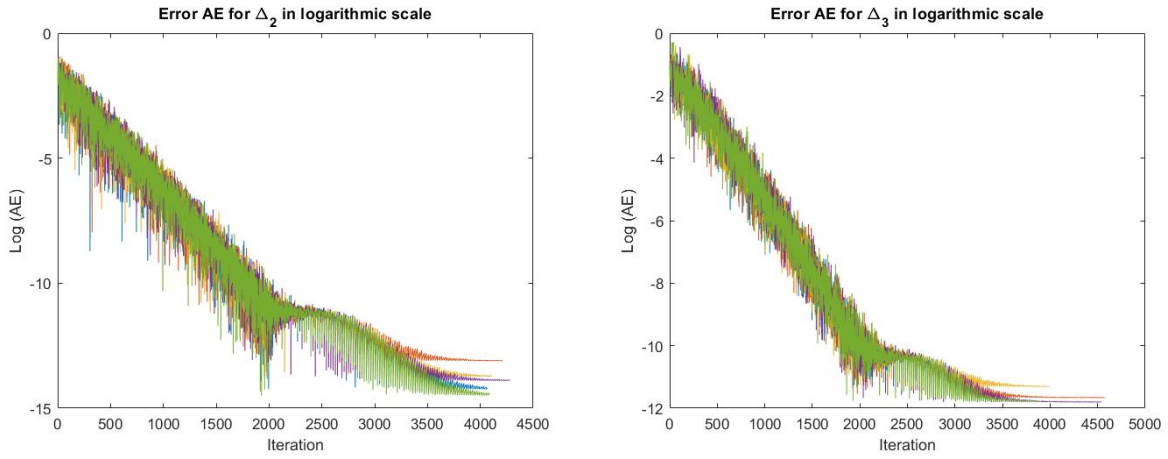
retraction

$$\beta_A(b_1, \cdots, b_n) = \frac{(\text{sign}(\det(B))\, b_1, b_2, \cdots, b_n)}{|\det(B)|^{\frac{1}{n}}}. \tag{3.17}$$

We note that the parameters for which we obtain better results were those described above.

We use the Exhaustive Enumeration Algorithm proposed in Ref. [69] to compute the function g. The implementation of this algorithm is provided in the GitHub repository [14] using MATLAB.

We test our algorithm through five executions. In Figure 3.4, we plot the absolute error (AE) of approximating Δ_2 and Δ_3 for the iteration value x_n. Each color represents a different execution. Observe that the error absolute error is small, which shows the effectiveness of our algorithm to solve the optimization problem in both dimensions.



(a) Approximation error for Δ_2 (b) Approximation error for Δ_3

Figure 3.4: Plot of the absolute error (AE) generated by five executions using Algorithm 3. Here we use the logarithmic scale.

In Figures 3.5 and 3.6, we plot the final step of each execution of our algorithm in dimensions 2 and 3. Observe that in all executions, the final step approximates the optimal sphere packing illustrated in Figures 3.2 and 3.3 in each dimension (to rotations). This fact was verified by calculating the error in the Figure 3.4.

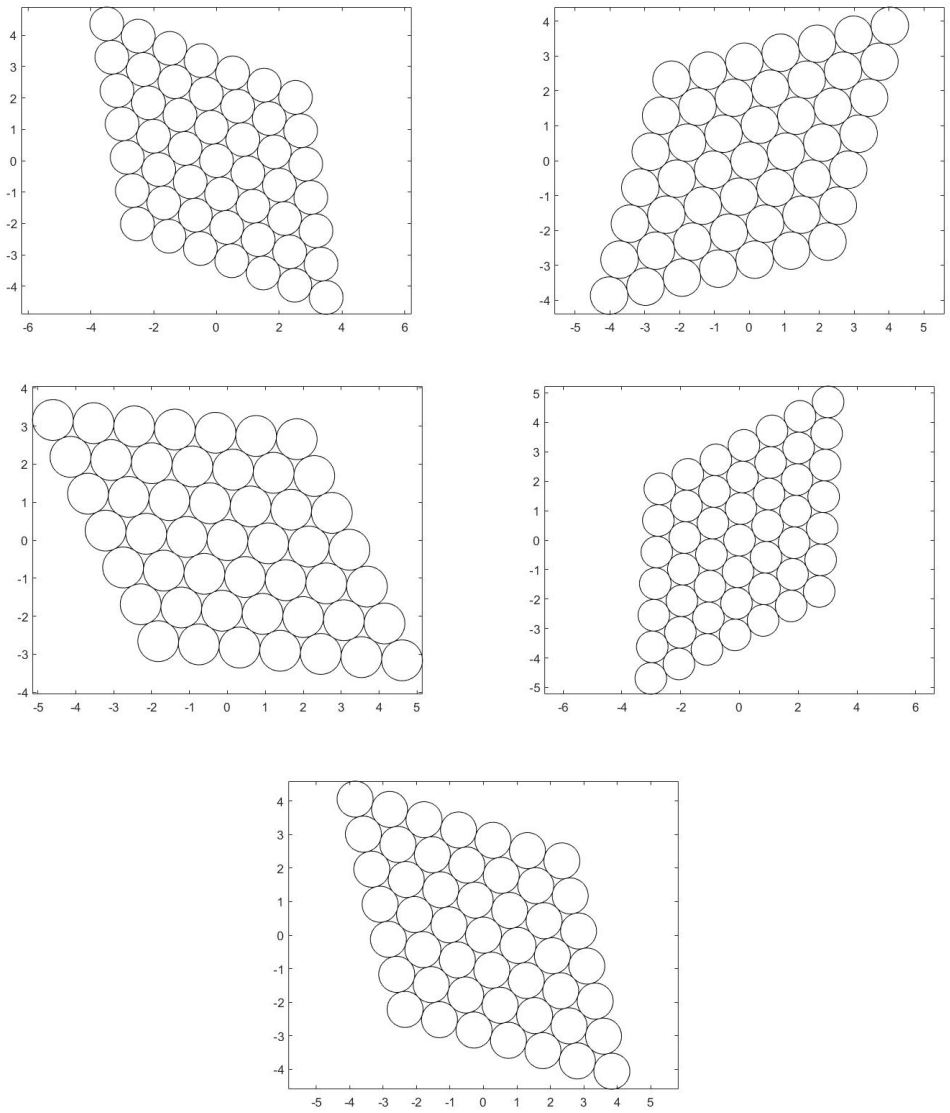


Figure 3.5: Plot of final lattice packing step of five executions to approximate the density Δ_2.

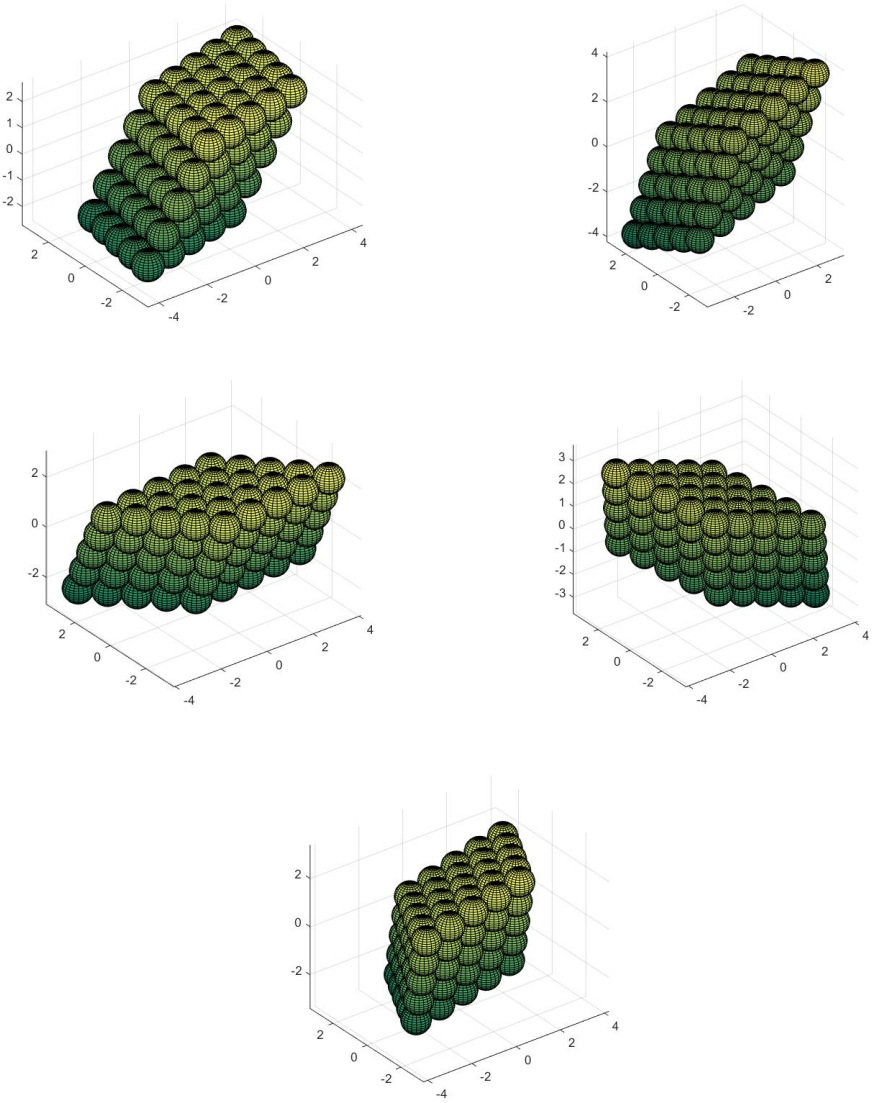


Figure 3.6: Plot of final lattice packing step of five executions to approximate the density Δ_3.

3.4.2 Tomographic reconstruction from unknown random angles

Tomographic reconstruction is a widely studied problem in the field of inverse problems. Its goal is to reconstruct an object from its angular projections. This problem has many applications in medicine, optics and other areas. We refer the reader to Ref. [40, 25, 43, 57] for more details .

Classical reconstruction methods are based on the fact that the angular position is known. See [40]. In contrast, there are many cases for which the angles of the projections are not available, for instance, when the object is moving. The latter is a nonlinear inverse problem, which can be more difficult compared to the classical linear inverse problem.

Now, we explain the problem in more details. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ describes the density of an object, and let θ be an angle. We define the one-dimensional tomographic projection over the angle θ as

$$\mathbb{P}_\theta f(x) = \int f(R_\theta(x, y))\, dy,$$

where $R_\theta(x, y)$ is the counterclockwise rotation of the two-dimensional vector (x, y) with respect to the angle θ. Since

$$\int |\mathbb{P}_{\theta_i} f(x)|\, dx = \int f(x, y)\, dydx,$$

thus, normalizing if necessary, we also assume that $\|\mathbb{P}_{\theta_i} f\|_{L^1} = 1$. The problem under consideration consists in reconstructing the density f with the knowledge of projections $\mathbb{P}_{\theta_1} f, \mathbb{P}_{\theta_2} f, \cdots \mathbb{P}_{\theta_k} f$, where the angles $\theta_1, \theta_2, \cdots \theta_k$ are unknown. If through some method the rotations are known, then we can obtain the density function f using classical reconstruction methods.

In Ref. [21] an approach using the graph Laplacian is proposed to deal with this

problem. However, the difficulty in using the previous approach is that it assumes *a priori* the knowledge of the distribution of the angles $\{\theta_i\}_{i=1}^k$. That is, it is necessary to assume the Euclidean distance between two consecutive angles.

We use our methodology to tackle the latter problem, the road-map of our approach is established in Algorithm 4. Let DS be the dataset defined as the set of all tomographic projections

$$DS = \{\mathbb{P}_{\theta_i} f\}_{i=1}^k. \tag{3.18}$$

If we assume that the density function f has compact support, then a straightforward computation gives

$$\begin{aligned}\int \mathbb{P}_{\theta_i} f(x)\, x\, dx &= \int\int \langle (x,y), (f(R_\theta(x,y),0))\rangle dxdy \\ &= \int\int \langle (x,y), R_\theta(f(x,y),0))\rangle dxdy \\ &= \langle \tilde{V}, R_{\theta_i}(1,0)\rangle, \end{aligned} \tag{3.19}$$

where $\tilde{V}$ is the two-dimensional vector

$$\tilde{V} = (\int\int x f(x,y)\, dxdy, \int\int y f(x,y)\, dxdy).$$

For practical purposes, we consider the discretization of the projection $\mathbb{P}_{\theta_i} f$ as the multi-dimensional vector given by

$$\overline{\mathbb{P}_{\theta_i} f} = (\mathbb{P}_{\theta_i} f(x_1), \mathbb{P}_{\theta_i} f(x_2), \cdots, \mathbb{P}_{\theta_i} f(x_l)),$$

where $x_1 < x_2 < \cdots < x_l$ are equally spaced fixed points on the x axis that describe the projection onto the angle θ_i. See Figure 3.7.

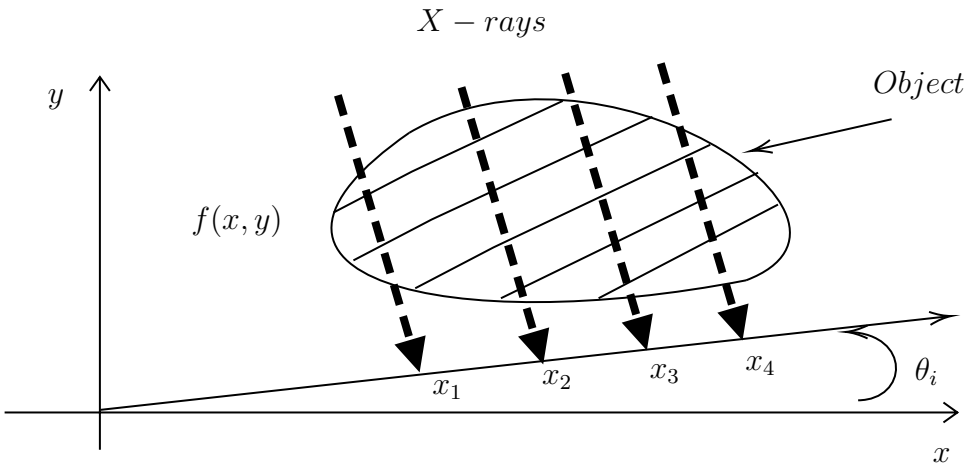


Figure 3.7: Tomography of an object.

Let X be the multidimensional vector

$$X = (x_1, x_2, \cdots, x_l).$$

The discretization of the integrals in Eq. (3.19) gives

$$\frac{1}{h}\langle \overline{\mathbb{P}_{\theta_i}} f, X\rangle \approx \langle \tilde{V}, R_{\theta_i}(1,0)\rangle, \tag{3.20}$$

where h is the distance between two consecutive points. The Eq. (3.20) allows to estimate, except for a possible sign and translation, the angle θ_i. Namely, if the two-dimensional vector $\tilde{V}$ has angle $\tilde{\theta}$, then, we recover θ_i using the expression

$$\cos(\theta_i - \tilde{\theta}) \approx \frac{1}{h\ \|\tilde{V}\|_2}\langle \overline{\mathbb{P}_{\theta_i}} f, X\rangle. \tag{3.21}$$

In this case, we use Eq. (3.20) to compute the value $\|\tilde{V}\|$ as

$$\|\tilde{V}\|_2 \approx \max_{\theta_i}\left|\frac{1}{h}\langle \overline{\mathbb{P}_{\theta_i}} f, X\rangle\right|. \tag{3.22}$$

We remark that in this approach we do not compute the two-dimensional vector $\tilde{V}$, instead, we compute the norm $\|\tilde{V}\|$ using Eq. (3.22). Observe that to solve the optimization problem in Eq. (3.22) is sufficient to assume that $\theta_i \in [0, \pi]$.

Once we solve the previous optimization problem, we use Eq. (3.20) to calculate the angle $\theta_i - \tilde{\theta}$. Observe that if we do not determine the sign of the $\theta_i - \tilde{\theta}$, then a flipping effect appears on the reconstructed object, resulting in an image with many artifacts. We apply our gradient estimates to determine the sign of the angle. For that, we assume that the angles are distributed on the interval $I = [0, \pi]$, and consider the numbers

$$m_1 = \min_i |\theta_i - \tilde{\theta}| \qquad and \qquad M_1 = \max_i |\theta_i - \tilde{\theta}|. \tag{3.23}$$

Since the maximum of the optimization problem in Eq. (3.22) is reached for some θ_i, then $m_1 = 0$ or $M_1 = \pi$. Without loss of generality, it is enough to consider the case $m_1 = 0$. In fact, if $M_1 = \pi$, then we reflect the angles over the y-axis. Furthermore, changing the order if necessary we assume that

$$0 = |\theta_1 - \tilde{\theta}| < |\theta_2 - \tilde{\theta}| < \cdots < |\theta_k - \tilde{\theta}|. \tag{3.24}$$

We observe that our dataset (DS) defined as in Eq. (3.18) lies in the curve $c(I)$, which is parameterized by

$$c(\theta) = \mathbb{P}_\theta f,$$

in our case this parametrization is unknown. The main idea in our algorithm is to use the gradient flow of the function g on the manifold $c(I)$, where $g : c(I) \to \mathbb{R}$ is defined as

$$g(Y) = \frac{1}{h}\langle Y, X\rangle. \tag{3.25}$$

The importance of the gradient flow in our method lies in the fact that in a local neighborhood of the vector associated with the angle 0, the gradient flow divides the dataset into two different clusters that determine the sign of the associated angles.

Before initializing our algorithm we divide the indices $\tilde{A} = \{i\}_{i=1}^{k}$ as follows. We select a fixed number s, which represents the size of the partition, and we consider the decomposition $k = qs + r$, where q and r are non-negative integers with $r < s$. Then, we define the sets

$$\tilde{A}_i = \{is + 1, is + 2, \cdots, (i + 1)s\}, \tag{3.26}$$

for $i \in \{0, 1, 2, \cdots, q - 1\}$, and

$$\tilde{A}_q = \tilde{A} \setminus \bigcup_{i=0}^{q-1} \tilde{A}_i. \tag{3.27}$$

We use the partition $\{\tilde{A}_i\}_{i=1}^{q}$ to represent the local geometry of the dataset. For that, we consider the subset DS_i of DS, defined as

$$DS_i = \{\mathbb{P}_{\theta_j} f \,|j \in A_i\}. \tag{3.28}$$

The first step in our algorithm is to determine the sign of angles in a local neighborhood of 0, for that, we use the diffusion-map algorithm to embed the dataset $\overline{DS}_1 = DS_1 \cup DS_2 \cup DS_3$ into the two-dimensional space $\mathbb{R}^2$. We endow this embedded dataset with the counting measure. Once the dataset is embedded, we proceed to compute the approximation for $\overline{P}_1 \tilde{g}$ as described in Algorithm 2. Here, we select the points $x_1, x_2, x_3 \cdots x_m$ as the m closest points to x. Since we only are interested in the direction induced by the gradient, then we propose to reduce the computational cost of the execution using the approximation

$$\overline{\mathcal{V}} = \sum_{i=1}^{m} (x_i - x)\ (\tilde{g}(x_i) - \tilde{g}(x))\ e^{\frac{-\|x_i - x\|^2}{2}}, \tag{3.29}$$

where, the function $\tilde{g}$ is such that for each two-dimensional embedded point $y \in \mathbb{R}^2$ associated to vector $Y \in DS$, the value of $\tilde{g}(x)$ is defined as

$$\tilde{g}(y) = g(Y). \tag{3.30}$$

The two-dimensional representation of the dataset allows determining the sign of the angles $\theta_i - \tilde{\theta}$ regarding the orientation of the flow generated by the function $\tilde{g}(y)$. This is done by observing that locally the set of gradient vectors associated with positive angles and the set of gradient vectors associated with negative angles are separated by a hyperplane. Since $\theta_2 - \tilde{\theta}$ is the smallest nonzero angle, then we use its gradient to define a hyperplane that separates the sets mentioned above. To be more specific, we separate the sets according to the sign of the inner product of its gradient with the gradient associated with $\theta_2 - \tilde{\theta}$. We remark that in the first step we only classify the sign of angles associated to points lying in $DS_1 \cup DS_2$, to avoid instabilities generated by computing the gradient of the boundary points lying in DS_3.

The second step is to proceed inductively to determine the sign of the remaining angles as follows. Assume that for $2 \leq i$ the sign of the angles associated to points lying in the set DS_i is determined, and consider the dataset $\overline{DS}_i = DS_i \cup DS_{i+1}$. As in the first step, we use diffusion-maps to embed this dataset into $\mathbb{R}^2$. Observe that the function g has not critical points on $\overline{DS}_i$. Then, the two-dimensional representation is divided at most into two clusters, for which each cluster represents the set of points with the same sign. We determine the sign of each cluster according to the sign of angles associated to points in DS_i lying in the corresponding cluster. For practical purposes, we define the sign of each angle $\theta_i - \tilde{\theta}$ as the sign of angle previously determined with the closest two-dimensional representation. We run this step until all the signs are determined. We summarize this reconstruction method in Algorithm 4. We remark that the choice of parameters s, and m

has to be modestly small to avoid instabilities in our algorithm.

Algorithm 4 Tomographic reconstruction from unknown random angles

input Tomographic projections $DS = \{Y_i\}_{i=1}^k$, where $Y_i = \mathbb{P}_{\theta_i} f$, size of the partition s, number of closest points m.

1. Normalize the dataset DS such that $\|\mathbb{P}_{\theta_i} f\|_{L^1} = 1$ for all i.
2. Compute $\|\tilde{V}\|_2$ solving the optimization problem 3.22.
3. Determine the angles $\theta_i - \tilde{\theta}$ using Eq. (3.21).
4. Compute M_1 as in Eq. (3.23).
5. If $M_1 = \pi$, then we proceed to reflect the angles $\tilde{\theta}_i$ over the y-axis.
6. Construct DS_i following (3.26), (3.27), and (3.28).
7. Use the diffusion-map approach to embed the dataset $DS_1 \cup DS_2 \cup DS_3$ into $\mathbb{R}^2$.
8. Compute $\overline{P}_1 \tilde{g}$ using Approximation (3.29), where $\tilde{g}$ is defined in (3.25) and (3.30).
9. Determine the sign of the angles $\tilde{\theta}_i$ associated to points in $DS_1 \cup DS_2$, according to the sign of the inner product of the its gradient with the gradient associated to θ_2.
10. **for** $j = 2$ to m **do**
    - Use the diffusion-map approach to embed the dataset $DS_j \cup DS_{j+1}$ into $\mathbb{R}^2$.
    - Determine the sign of each angle $\tilde{\theta}_i$ in DS_{j+1} as the sign of angle previously determined with the closest two-dimensional representation.
11. **end for**
12. Reconstruct the signed angles.

The computational complexity of all the embeddings is $O(qs^3)$, which corresponds to the complexity of the eigenvalue decomposition. On the other hand, the complexity of all gradient computations is $O(s)$, and the computational complexity of the other procedures described in Algorithm 4 is $O(s)$. Thus, Algorithm 4 runs with a $O(qs^3)$ complexity which improves the $O(q^3 s^3)$ complexity of the algorithm proposed in [21].

We test our algorithm on the tomographic reconstruction of two objects. The first is

the Shepp–Logan phantom, and the second is a computed tomography of a knee taken from Ref. [16]. See figure 3.8. In this experiment, we generate $k = 2 \times 10^3$ random points uniformly distributed in $[0, \pi]$. The parameters used in Algorithm 4 are $s = 20$, and $m = 10$. The tomographic projections $\mathbb{P}_{\theta_1} f, \mathbb{P}_{\theta_2} f, \cdots \mathbb{P}_{\theta_k} f$ are computed using MATLAB 's `radon` function. We add random noise to these projections, for that, we consider the dataset of the form

$$\mathbb{P}^{\varepsilon}_{R_i} f = \mathbb{P}_{R_i} f + \varepsilon W, \tag{3.31}$$

where W is a white noise. Our purpose is to recover the density f, using only the measurements $\mathbb{P}^{\varepsilon}_{R_i} f$, regardless of their respective angles.

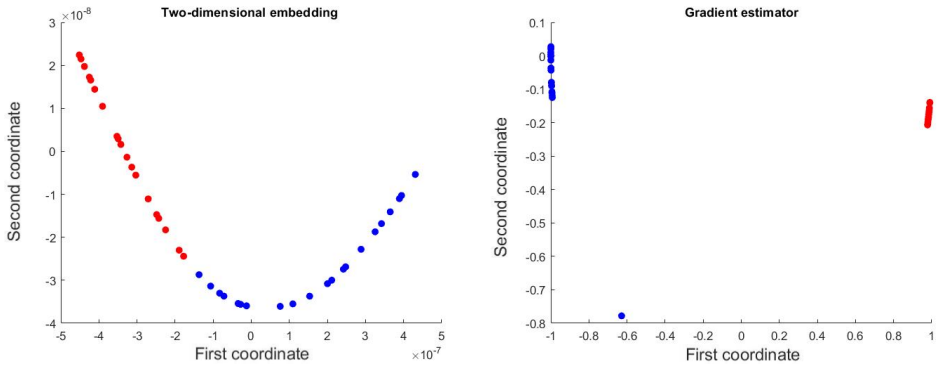
To illustrate how Algorithm 4 works, we plot the two essential steps in the method. In Figure 3.9, we plot the first two-dimensional embedding and their respective gradient Approximation defined in Eq. (3.29). Points with blue color are associated with positive angles and those with red color with negative angles. Furthermore, in Figure 3.10 , we plot the second two-dimensional embedding of our method. We observe that our method performs effectively in dividing the dataset into two different clusters according to the sign of the corresponding angle.

In Figures 3.11 and 3.12, we plot the reconstructed images of the Shepp–Logan phantom and the knee tomography. Here, the samples of the angles are uniformly distributed over $[0, \pi]$. We consider different levels of additive error ϵ as in Eq. (3.31). We remark that we obtained similar results to the previous ones using multiple executions of our method. To measure the effectiveness of our method, we compare the L^2 error generated when our algorithm is implemented. The L^2 error is computed in Tables 3.1 and 3.2. Observing the computational error and image quality, we conclude that our reconstruction algorithm works efficiently with relatively low computational cost.

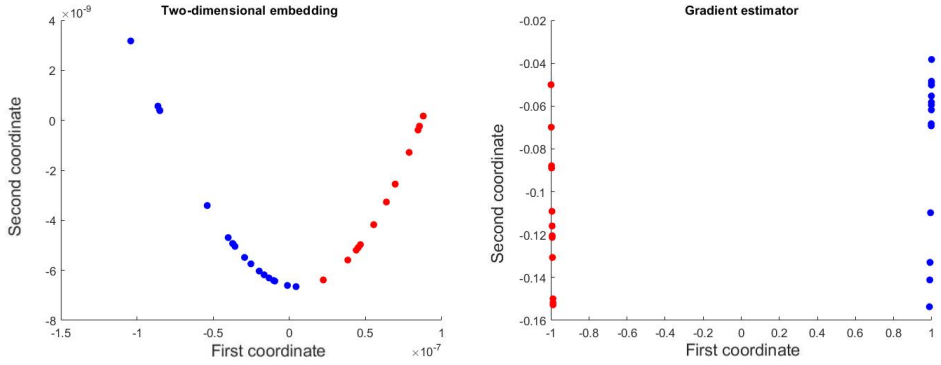
(a) Shepp–Logan phantom

(b) Sample image of a knee

Figure 3.8: Picture of the Shepp–Logan phantom, and a knee sample image. Source [16].

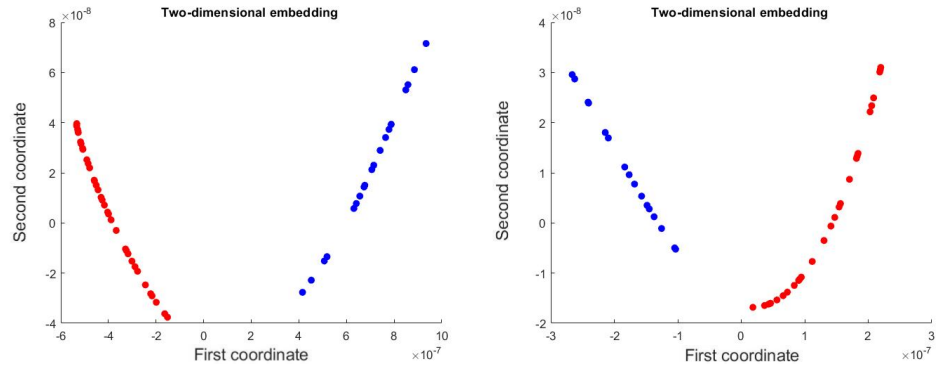


(a) Shepp–Logan phantom



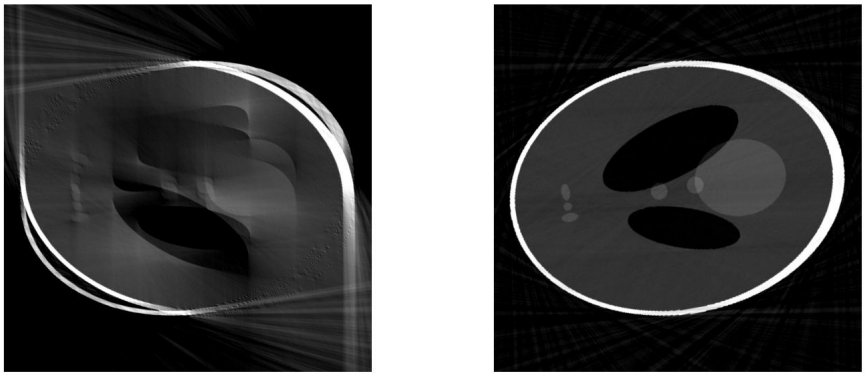
(b) Sample image of the knee

Figure 3.9: Plot of the first two-dimensional embedding (left), and their associated gradient Approximation (right). In this experiment, the angle sample is uniformly distributed on $[0, \pi]$. Each color represents a different sign. Figure (A) corresponds to the Shepp–Logan phantom, and Figure (B) to the image of the knee.

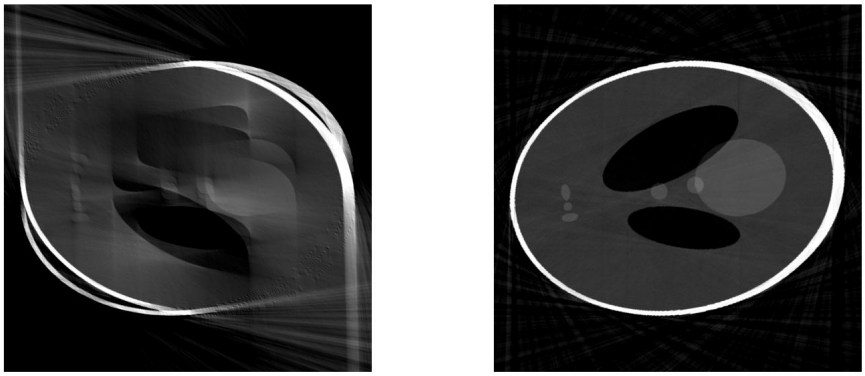


(a) Shepp–Logan phantom (b) Shepp–Logan phantom

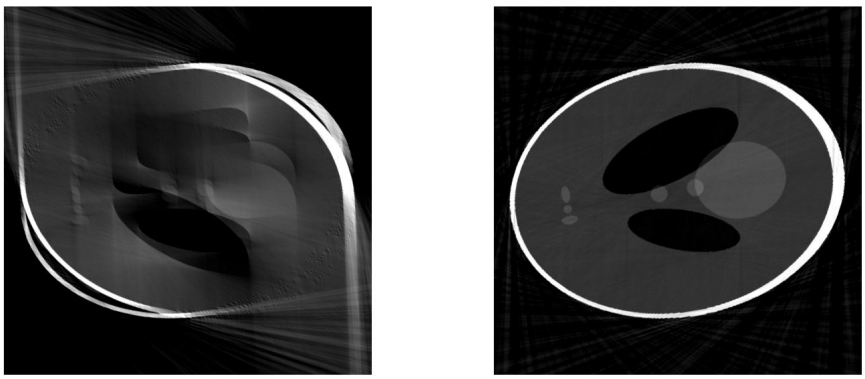
Figure 3.10: Plot of the second two-dimensional embedding (left). Figure (A) corresponds to the Shepp–Logan phantom, and Figure (B) to the image of the knee.



(a) $\epsilon = 0$

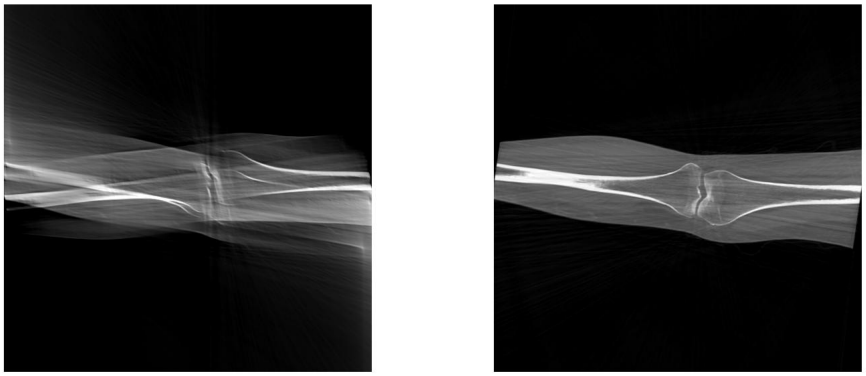


(b) $\epsilon = 0.05$

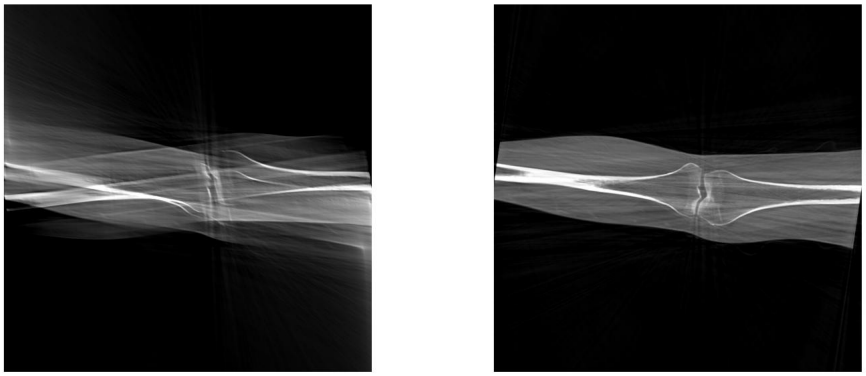


(c) $\epsilon = 0.1$

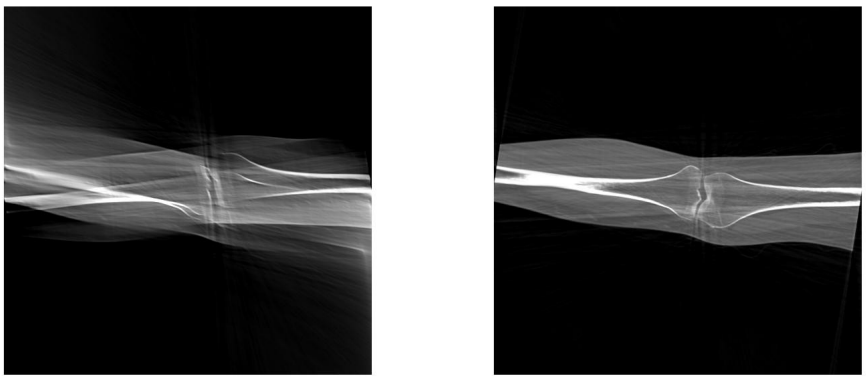
Figure 3.11: Reconstructed Shepp–Logan phantom for several additive errors ϵ as in Eq. (3.31). The images on the left are obtained without determining the sign of each angle, and the images on the right are obtained by implementing our algorithm.



(a) $\epsilon = 0$



(b) $\epsilon = 0.05$



(c) $\epsilon = 0.1$

Figure 3.12: Reconstructed knee tomography for several additive errors ϵ as in Eq. (3.31). The images on the left are obtained without determining the sign of each angle, and the images on the right are obtained by implementing our algorithm.

Value of ϵ	With determination of the sign	Without determination of the sign
0	0.0814	0.2087
0.05	0.0816	0.2101
0.1	0.0824	0.2129

Table 3.1: Error of the reconstructed Shepp–Logan phantom. We use the L^2 norm to compute the errors. Here, the sample angles are uniformly distributed over $[0, \pi]$.

Value of ϵ	With determination of the sign	Without determination of the sign
0	0.1001	0.1411
0.05	0.1053	0.1425
0.1	0.1114	0.1445

Table 3.2: Error of the reconstructed knee tomography . We use the L^2 norm to compute the errors. Here, the sample angles are uniformly distributed over $[0, \pi]$.

Appendix A

Review of differential geometry

We review some facts of differential geometry. We refer the reader to Ref. [26] for a more detailed description. Given an interior point $x \in \mathcal{M}$, there exists a positive real number ε such that the map $\psi = \exp_x \circ T : B(0,\varepsilon) \subset \mathbb{R}^d \to \mathcal{M}$ is a local chart. Here, $\exp_x$ is the exponential map at the point x, and $T : \mathbb{R}^d \to T_x\mathcal{M}$ is a rotation from $\mathbb{R}^d$ onto $T_x\mathcal{M}$, both sets considered subsets of $\mathbb{R}^n$. The chart ψ defines the normal coordinates at point x.

Given a smooth function $f \in C^\infty(\mathcal{M})$, the gradient operator $\nabla f(x) \in T_x\mathcal{M}$ is given in normal coordinates by

$$\nabla f(x) = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i} T(e_i).$$

Here, e_i is the standard basis in $\mathbb{R}^d$. Now, we recall some estimates that use normal coordinates that are useful when estimating approximations for differential operators. The Taylor series of ψ around the point 0 is given by

$$\psi(v) = x + T(v) + \frac{1}{2} D^2\psi_0(v,v) + O(\|v\|^3). \tag{A.1}$$

Let $v \in B(0,\varepsilon) \subset \mathbb{R}^d$, and consider the geodesic $\gamma_{T(v)}$, with initial tangent vector $T(v) \in$

$T_x\mathcal{M}$, then using Estimate (A.1) we obtain

$$\gamma_{T(v)}(t) = x + T(v)\, t + \frac{1}{2} D^2\psi_0\,(v,v)t^2 + O(\|v\|^3)t^3.$$

Since the covariant derivative of a geodesic vanishes, then $\gamma''_{T(v)}$ is orthogonal to $T_x\mathcal{M}$. Thus, we have the following estimates

$$\|\psi(v) - x\|^2 = \|T(v)\|^2 + O(\|v\|^4), \tag{A.2}$$

and

$$\mathcal{P}_x(\psi(v) - x) = T(v) + O(\|v\|^3), \tag{A.3}$$

where $\mathcal{P}_x$ is the orthogonal projection on $T_x\mathcal{M}$. Using the Estimates (A.2) and (A.3), we obtain that there exist positive constants M_1 and M_2 such that for $\|v\|$ small

$$\|v\| - M_2\|v\|^3 \leq \|\psi(v) - x\| \leq M_1\|v\|.$$

Thus, if $\|v\|^2 \leq \frac{1}{2M2}$ we have

$$\frac{1}{2}\|v\| \leq \|\psi(v) - x\| \leq M_1\|v\|.$$

This says that for t small

$$B(0, t/M_1) \subseteq \psi^{-1}(U(x, t^\delta)) \subseteq B(0, 2t). \tag{A.4}$$

Appendix B

Expansion of the gradient operator

Here, we show the technical details of the proof of Theorem 3.1.1. The main idea is to use the Taylor expansion of the function f around the point x.

Lemma B.0.1. *Assume that $\frac{1}{2} < \delta < 1$, and let $K : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^m$ be a vector value kernel. Define*

$$P_t(x) = \int_{U(x,t^\delta)} K(x,y)\, e^{\frac{-\|y-x\|^2}{2t^2}} dy,$$

where $U(x,t^\delta)$ is defined as in Eq. (3.1). Assume that for t small, the function $\psi : B(0,2t^\delta) \to \mathcal{M}$ defines normal coordinates in a neighborhood of x, and let S be a vector value function defined in $\mathbb{R}^d$ such that

$$K(x,\psi(v)) - S(v) = O(\|v\|^r),$$

and

$$K(x,y) = O(\|x-y\|^s).$$

Then, we have

$$P_t(x) = O((e^{C_2 t^{4\delta-2}} - 1)t^{s+d} + t^{r+d}) + \int_{\psi^{-1}(U(x,t^\delta))} S(v)\, e^{\frac{-\|T(v)\|^2}{2t^2}} dv.$$

Proof. Using Eq. (A.4), we assume that for t small, the set $U(x, t^\delta)$ lies in the image of a normal chart $\psi : B(0, 2t^\delta) \to \mathcal{M}$ centered in x. Thus,

$$\begin{aligned}
\int_{U(x,t^\delta)} K(x,y)\, e^{\frac{-\|y-x\|^2}{2t^2}} dy &= \int_{\psi^{-1}(U(x,t^\delta))} K(x,\psi(v)) e^{\frac{-\|\psi(v)-x\|^2}{2t^2}} dv \\
&= \int_{\psi^{-1}(U(x,t^\delta))} K(x,\psi(v))(e^{\frac{-\|\psi(v)-x\|^2}{2t^2}} - e^{\frac{-\|T(v)\|^2}{2t^2}})dv \\
&+ \int_{\psi^{-1}(U(x,t^\delta))} (K(x,\psi(v)) - S(v)) e^{\frac{-\|T(v)\|^2}{2t^2}} dv \\
&+ \int_{\psi^{-1}(U(x,t^\delta))} S(v)\, e^{\frac{-\|\mathbb{T}(v)\|^2}{2t^2}} dv.
\end{aligned}$$

We now estimate

$$A = \int_{\psi^{-1}(U(x,t^\delta))} K(x,\psi(v))(e^{\frac{-\|\psi(v)-x\|^2}{2t^2}} - e^{\frac{-\|T(v)\|^2}{2t^2}})dv.$$

Using Eq. (A.2), and the inequality $|e^x - 1| \leq e^{|x|} - 1$ we obtain

$$\begin{aligned}
|e^{\frac{-\|\psi(v)-x\|^2}{2t^2}} - e^{\frac{-\|T(v)\|^2}{2t^2}}| &= e^{\frac{-\|T(v)\|^2}{2t^2}} |e^{\frac{O(\|v\|^4)}{2t^2}} - 1| \\
&\leq e^{\frac{-\|T(v)\|^2}{2t^2}} (e^{\frac{C_1\|v\|^4}{2t^2}} - 1).
\end{aligned}$$

Therefore, by Equation (A.4) we obtain

$$\begin{aligned}
\|A\| &\leq C_3\ t^s (e^{C_2 t^{4\delta-2}} - 1)t^d \int_{\mathbb{R}^d} \|v\|^s e^{-\|v\|^2/2} dv \\
&= O((e^{C_2 t^{4\delta-2}} - 1)t^{s+d}).
\end{aligned}$$

On the other hand, by assumption we have

$$\int_{\psi^{-1}(U(x,t^\delta))} (K(x,\psi(v)) - S(v)) e^{\frac{-\|T(v)\|^2}{2t^2}} dv = O(t^{r+d}).$$

□

Lemma B.0.2. *Under the same assumptions of Lemma B.0.1, we define*

$$E = \int_{\psi^{-1}(U(x,t^\delta))} Q(v) e^{\frac{-\|T(v)\|^2}{2t^2}} g(v) dv,$$

where g is a smooth function and Q is a homogeneous polynomial of degree l. Then, we have

$$E = \int_{\mathbb{R}^d} Q(v) e^{\frac{-\|T(v)\|^2}{2t^2}} (g(0) + \sum \frac{\partial g}{\partial v_i}(0) v_i) dv \; + O(t^{d+l} e^{-M_2 t^{2(\delta-1)}} + t^{d+2+l}).$$

Proof. Using the Taylor expansion of g around 0 we have

$$E = \int_{\psi^{-1}(U(x,t^\delta))} Q(v) e^{\frac{-\|T(v)\|^2}{2t^2}} (g(0) + \sum \frac{\partial g}{\partial v_i}(0)\, v_i + O(\|v\|^2)) dv.$$

Let B be defined as

$$B = \| \int_{\mathbb{R}^d \setminus \psi^{-1}(U(x,t^\delta))} Q(v) e^{\frac{-\|T(v)\|^2}{2t^2}} (g(0) + \sum \frac{\partial g}{\partial v_i}(0)\, v_i) dv \|.$$

Using Eq. (A.4) and the fast decay of the exponential function, we obtain that

$$B \leq C_4 t^{d+l} e^{-M_2 t^{2(\delta-1)}} \int_{\mathbb{R}^d \setminus B(0, t^{\delta-1}/M_1)} P(\|v\|) e^{\frac{-\|T(v)\|^2}{4}} dv.$$

for a certain polynomial P. Therefore, we have

$$B = O(t^{d+l} e^{-M_2 t^{2(\delta-1)}}),$$

for a proper constant M_2. Finally, we observe that

$$\int_{\psi^{-1}(U(x,t^\delta))} Q(v) e^{\frac{-\|T(v)\|^2}{2t^2}} O(\|v\|^2) dv = O(t^{d+2+l}).$$

□

We recall the following computations related to the moments of the normal distribution that are useful in proving Theorem 3.1.1. For all index i

$$\int_{\mathbb{R}^d} v_i e^{\frac{-\|T(v)\|^2}{2t^2}} dv = 0,$$

and

$$\int_{\mathbb{R}^d} v_i^2 e^{\frac{-\|T(v)\|^2}{2t^2}} dv = (2\pi)^{\frac{d}{2}} t^{d+2},$$

moreover, if $i \neq j$ then

$$\int_{\mathbb{R}^d} v_i \, v_j e^{\frac{-\|T(v)\|^2}{2t^2}} dv = 0.$$

Lemma B.0.3. *Under the same assumptions of Lemmas B.0.1 and B.0.2 we have*

$$d_t(x) = (2\pi)^{\frac{d}{2}} t^d + O(t^{d+4\delta-2}). \tag{B.1}$$

Proof. We apply Lemmas B.0.1 and B.0.2 to the functions $K(x,y) = 1$, $S(v) = 1$, $Q(v) = 1$, and $g(x) = 1$. We use the parameters $r = 2$, $s = 0$ and $l = 0$. Using the exponential decay we obtain the following estimate

$$d_t(x) = (2\pi)^{\frac{d}{2}} t^d + O(t^{d+4\delta-2}).$$

□

Proof of Theorem 3.1.1. We apply Lemmas B.0.1 and B.0.2 to the functions $K(x,y) = (y-$

$x)(f(y)-f(x))$, $S(v) = T(v)(f(\psi(v))-f(x)) = \sum v_i(f(\psi(v))-f(x))T(e_i)$, $Q(v) = v_i$ and $g(v) = (f(\psi(v)) - f(x))$. Since $\psi(v) - x - T(v) = O(\|v\|^2)$ and $f(\psi(v)) - f(x) = O(\|v\|^1)$, then the parameters that we use are $r = 3$, $s = 2$ and $l = 1$. Again, using the exponential decay we have that

$$\int_{U(x,t^\delta)} \overline{K}(x,y)\, e^{\frac{-\|y-x\|^2}{2t^2}} dy = (2\pi)^{\frac{d}{2}}\, t^{d+2} \sum \frac{\partial f}{\partial v_i}(0)\, T(e_i) + O(t^{d+4\delta}). \tag{B.2}$$

Finally we use Eq. (B.1) of Lemma B.0.3 to conclude the result . □

Chapter 4

The Inverse Problem of Recovering the Fragmentation Rate in the Growth-Fragmentation Equation for Transport Kernels

Highlights

- We consider the problem of recovering the fragmentation rate in the growth-fragmentation equation for a wide class of fragmentation kernels.
- A new regularization method for the growth-fragmentation equation is developed.
- We show that the quasi-reversibility methodology is a particular case of ours.
- We provide estimates to the error order in the regularization process.
- Numerical evidence indicates that the Landweber filter performs better.

This Chapter is concerned with the inverse problem for structured population models which describe quantitatively, the evolution process of the density of an ensemble of particles. The present chapter is organized as follows: in Section 4.1, we study the transport operation and some relations with the Fourier transform on locally compact abelian groups. In Section 4.2, we discuss the invertibility of the operator $k\mathcal{K} - Id$ in proper spaces. In Section 4.3, we present a new regularization method to treat the stability of the inverse problem. In Section 4.4, we give examples for which some of the Hypotheses 4.2.1 or 4.3.1 are satisfied. Finally, in Section 4.5, we present the numerical implementation of our method.

4.1 Preliminaries

We briefly review some concepts on Haar measure and Fourier transform for locally compact groups, we refer the reader to [66] for more details. We recall that the Haar measure on an abelian group G is the unique non-negative and regular measure, up to a positive multiplicative constant, which is translation invariant.

We define the transport operation $\odot_\rho$ as

$$a \odot_\rho b := \rho(\, \rho^{-1}(a) + \rho^{-1}(b)\,),$$

where ρ is an increasing diffeomorphism from $\mathbb{R}$ to $\mathbb{R}^+$. The pair $(\mathbb{R}^+, \odot_\rho)$, equipped with the induced topology of $\mathbb{R}$, is a locally compact group. Moreover, the Haar measure μ_ρ is given by the measure

$$d\mu_\rho = \frac{d}{dx}\rho^{-1} dx,$$

where dx is the Lebesgue measure on the positive real line. In this context, we can develop the theory of the Fourier transform on the group $(\mathbb{R}^+, \odot_\rho, \mu_\rho)$.

4.1.1 Fourier transform on $(\mathbb{R}^+, \odot_\rho, \mu_\rho)$.

The Fourier transform $\mathcal{F}_\rho$ on the group $(\mathbb{R}^+, \odot_\rho)$ is defined for a function $f(x) \in L^1(\mathbb{R}^+, \mu_\rho)$ and a real number ξ as

$$\mathcal{F}_\rho f(\xi) = \int_0^\infty f(x)\, e^{-2\pi i \xi \rho^{-1}(x)}\, d\mu_\rho.$$

In this general context, the Fourier transform theory on $(\mathbb{R}^+, \odot_\rho)$ can be developed in a similar fashion as in the standard case $(\mathbb{R}, +)$. Specially, we obtain inversion and isometry theorems. See [66] for more details.

Theorem (Inversion theorem)**.** *Suppose that $f(x) \in L^1(\mathbb{R}^+, \mu_\rho)$ and $\mathcal{F}_\rho f(\xi) \in L^1(\mathbb{R}, dx)$, then for a.e positive number x we have*

$$f(x) = \int_{-\infty}^{+\infty} \mathcal{F}_\rho f(\xi)\, e^{2\pi x \rho^{-1}(\xi)}\, d\xi.$$

Theorem (Plancherel Theorem)**.** *The Fourier transform $\mathcal{F}_\rho$, restricted to $L^1(\mathbb{R}^+, \mu_\rho) \cap L^2(\mathbb{R}^+, \mu_\rho)$ is an isometry (with respect to the L^2 norms) onto a dense linear subspace of $L^2(\mathbb{R}, dx)$. Hence, it may be extended, in a unique manner, to an isometry from $L^2(\mathbb{R}^+, \mu_\rho)$ to $L^2(\mathbb{R}, dx)$.*

4.1.2 The $L^2_{\rho,s}$ spaces.

In order to extend the Plancherel theorem, we define the $L^2_{\rho,s}$ space. They are a natural generalization of the $L^2(\mathbb{R}^+, \mu_\rho)$ space. More precisely, the $L^2_{\rho,s}$ space is defined as $L^2(\mathbb{R}^+, e^{2\pi s \rho^{-1}(\cdot)} \mu_\rho)$. Observe that there exists a canonical isomorphism T_s between $L^2_{\rho,s}$ and $L^2_{\rho,0} = L^2(\mathbb{R}^+, \mu_\rho)$, which is given by

$$T_s f(x) = f(x)\, e^{\pi s \rho^{-1}(x)}.$$

We now define the Fourier transform $\mathcal{F}_{\rho,s}$ on $L^2_{\rho,s}$ as

$$\mathcal{F}_{\rho,s} = \mathcal{F}_\rho \circ T_s.$$

Observe that by the Plancherel theorem the Fourier transform $\mathcal{F}_{\rho,s}$ is an isometry.

4.1.3 The Fourier transform of probability measures

Let P be a probability measure on R^+ satisfying

$$\int_0^\infty e^{\pi s \rho^{-1}(z)} \frac{d}{dx} \rho^{-1}(z) dP(z) < \infty. \tag{4.1}$$

We defined the Fourier transform $\mathcal{F}_{\rho,s}$ of the probability P by

$$\mathcal{F}_{\rho,s} P(\xi) = \int_0^\infty e^{(-2\xi i + s)\pi \rho^{-1}(z)} \frac{d}{dz} \rho^{-1}(z) dP(z),$$

for any real number ξ.

4.1.4 The Fourier transform and convolution operator

We now consider the convolution operator with kernel P, given by

$$\tilde{\mathcal{K}}(f)(x) = \int_0^\infty P(x \odot_\rho y^{-1}) f(y)\, d\mu_\rho(y). \tag{4.2}$$

In the same manner as in the case of $\mathbb{R}$, and under some conditions of integrability, the Fourier transform $\mathcal{F}_{\rho,s}$ of a convolution operator becomes a multiplicative operator.

Property. *Let P be a probability measure satisfying Equation* (4.1)*, then the convolution*

operator defined in Eq. (4.2) *satisfies the multiplier property*

$$\mathcal{F}_{\rho,s}\tilde{\mathcal{K}}(f)(\xi) = \mathcal{F}_{\rho,s}P(\xi)\ \cdot\ \mathcal{F}_{\rho,s}f(\xi), \tag{4.3}$$

for all f *in* $L^2_{\rho,s}$.

4.1.5 The Fourier transform of $\frac{d}{dx}gN$.

The Fourier transform on the real line has the property of diagonalizing differential operators. We now state an analogous version of this fact for the Fourier transform $\mathcal{F}_{\rho,s}$.

Proposition 4.1.1. *Suppose that* S *is on the Sobolev space* $H^1_0(0,\infty)$*, and that the functions*

$$\frac{d}{dx}S\,,\ S\frac{d}{dx}\rho^{-1}\,,\ S(\frac{d^2}{dx^2}\rho^{-1})(\frac{d}{dx}\rho^{-1})^{-1}, \tag{4.4}$$

are in $L^2_{\rho,s}$*. Then, we have*

$$\begin{aligned}\mathcal{F}_{\rho,s}(\frac{d}{dx}S)(\xi) = (2\xi i - s)\pi\,\mathcal{F}_{\rho,s}(S\frac{d}{dx}\rho^{-1})(\xi)\\ -\mathcal{F}_{\rho,s}(S(\frac{d^2}{dx^2}\rho^{-1})(\frac{d}{dx}\rho^{-1})^{-1})(\xi).\end{aligned} \tag{4.5}$$

Proof. Using integration by parts, and the smoothness assumption of the function S we have that

$$\mathcal{F}_{\rho,s}(\frac{d}{dx}S)(\xi) = -\int_0^\infty S(z)\frac{d}{dz}\,(e^{(-2\xi i+s)\pi\rho^{-1}(z)}\ \frac{d}{dz}\rho^{-1}(z))\ dz.$$

Observe that

$$\frac{d}{dz}\,(e^{(-2\xi i+s)\pi\rho^{-1}(z)}\ \frac{d}{dz}\rho^{-1}(z)) = ((-2\xi i+s)\pi(\frac{d}{dz}\rho^{-1}(z))^2 + \frac{d^2}{dz^2}\rho^{-1})e^{(-2\xi i+s)\pi\rho^{-1}(z)},$$

and this proves our result. □

4.2 Invertibility of the operator $k\mathcal{K} - Id$.

Let $\mathcal{K} : L^2_{\rho,s} \to L^2_{\rho,s}$ be the operator defined as

$$\mathcal{K}(H) = \int_0^\infty \mathbb{K}(x,y)\, H(y)\, dy. \tag{4.6}$$

We use the ideas of [11] in order to guarantee the invertibility of $k\mathcal{K} - Id$. We prove that under additional assumptions on the kernel P, the operator $k\mathcal{K} - Id$, has a bounded inverse in some subspaces of $L^2_{\rho,s}$. If we apply the Fourier transform and the Equation (4.3), we get that $k\mathcal{K} - Id$ is in fact a multiplier operator

$$\mathcal{F}_{\rho,s}(k\mathcal{K} - Id)f(\xi) = (k\,\mathcal{F}_{\rho,s}P(\xi) - 1)\,\mathcal{F}_{\rho,s}f(\xi).$$

Thus, the operator $k\mathcal{K} - Id$ has a bounded inverse on $L^2_{\rho,s}$ if the function

$$|k\,\mathcal{F}_{\rho,s}P(\cdot) - 1|,$$

is bounded from below by positive constant on the real line. In some cases, the function $|k\,\mathcal{F}_{\rho,s}P(\cdot) - 1|$ never vanishes, but goes to zero when x converges to 0 or ∞, in this situation, the function $|k\,\mathcal{F}_{\rho,s}P(\cdot) - 1|$ is bounded from below on compact sets of the real line. Moreover, for computational purposes, we are interested in a reconstruction on compact intervals. In this case, we focus on the invertibility of the operator on compact sets. To find a bounded inverse on an open set U, we assume that the kernel P satisfies the following hypothesis.

Hypothesis 4.2.1. *There exists an open set U, such that the function $|k\,\mathcal{F}_{\rho,s}P(\cdot) - 1|$ is bounded from below by a positive constant on U.*

In Section 4.4 we give examples where the previous hypothesis is satisfied.

We consider the Paley-Wiener spaces $PW_{\rho,s}(U)$, as the subspace of $L^2_{\rho,s}$ of all the functions f whose Fourier transform $\mathcal{F}_{\rho,s}$ has support on U. Observe that if f is on $PW_{\rho,s}(U)$, then $\mathcal{K}f$ also lies on $PW_{\rho,s}(U)$. Thus, the operator $k\mathcal{K} - Id$ from $PW^{\alpha}_{\rho,s}$ to itself is well defined. Using the Fourier transform $\mathcal{F}_{\rho,s}$, and Eq. (4.3), we conclude the following proposition, which guarantees the existence of a bounded inverse.

Proposition 4.2.1. *Suppose that the probability P satisfies Hypothesis 4.2.1, then the operator $k\mathcal{K} - Id$ from $PW_{\rho,s}(U)$ to itself has a bounded inverse.*

4.3 Regularization of the inverse problem

In this section, we discuss some methods to regularize the inverse problem, i.e, recover $H = BN$ from noisy measurements of N in some $L^2_{\rho,s}$ norm, in such a way that we can control the approximation error.

The main difficulty arises from the fact that we cannot estimate the $L^2_{\rho,s}$ norm of $\frac{d}{dx}gN$ just with the information $\|N\|$. The principal idea in this section is to use the Fourier transform to turn the differential operator $\frac{d}{dx}$ onto a Fourier multiplier operator [75], and then, we regularize the inverse problem using spectral filters as described in [31].

4.3.1 Regularization by spectral filtering

We now present our strategy to regularize the inverse problem as follows. Using the Proposition 4.1.1, we see that under the Fourier transform, the differential operator is simply a quasi multiplier operator, whose multiplier part is given by $(2i\xi - s)\pi$. Unfortunately, this function is unbounded in non-compact sets. Then, noisy data with small error can approximate solutions with large error, that is, it has a de-regularizing effect. To regularize the inverse problem, we define a multiplier approximation for $\frac{d}{dx}gN$, for which the multiplier part is given by multiplication of $(2i\xi - s)\pi$ by a regularized filter $f(\alpha, \xi)$. To be more

precise, we define the approximation $H_\alpha(N_\varepsilon, \lambda_\varepsilon)$ by

$$\mathcal{F}_{\rho,s} H_\alpha(N,\lambda)(\xi) = f(\alpha,\xi)\, L(N)(\xi) + h(\alpha,\xi)\, \lambda\, \mathcal{F}_{\rho,s} N(\xi), \tag{4.7}$$

where

$$L(N)(\xi) = (2i\xi - s)\pi\, \mathcal{F}_{\rho,s}\left(gN \frac{d}{dx}\rho^{-1}\right)(\xi) - \mathcal{F}_{\rho,s}\left(gN\left(\frac{d^2}{dx^2}\rho^{-1}\right)\right)\left(\left(\frac{d}{dx}\rho^{-1}\right)^{-1}\right)(\xi).$$

Let $(N_\varepsilon, \lambda_\varepsilon)$ be a noisy measurement of (N,λ) in the product space $PW_{\rho,s}(U) \times L^\infty(0,\infty)$, and let $H = H(N,\lambda)$ be the unique solution of (1.1) in $PW_{\rho,s}(U)$. The following theorem gives error estimates for $\|H_\alpha(N_\varepsilon,\lambda_\varepsilon) - H(N,\lambda)\|$. It was inspired by [29], yet, the method is different because it focuses on regularizing the spectral signal of the differential operators.

the method is different because it focuses on regularizing the spectral signal of the differential operators.

Theorem 4.3.1 (Spectral regularization)**.** *Let K be a subspace of $PW_{\rho,s}(U)$. Suppose that there exists a positive constant C, such that for all $\alpha \in [0, C)$ the bilinear operator H_α defined from $K \times L^\infty(0,\infty)$ to $PW_{\rho,s}(U)$ is bounded. Assume that $N \in L^2_{\rho,s}$ satisfies*

$$\|H(N,\lambda) - H_\alpha(N,\lambda)\| \le \alpha M,$$

where the constant M does not depend on α. Then, we have the estimate

$$\|H_\alpha(N_\varepsilon,\lambda_\varepsilon) - H(N,\lambda)\| \le \|H_\alpha(N_\varepsilon - N, \lambda_\varepsilon - \lambda)\| + \alpha M.$$

Proof. By triangular inequality we obtain

$$\|H_\alpha(N_\varepsilon,\lambda_\varepsilon) - H(N,\lambda)\| \le \|H_\alpha(N_\varepsilon,\lambda_\varepsilon) - H_\alpha(N,\lambda)\| + \|H_\alpha(N,\lambda) - H(N,\lambda)\|,$$

which implies the result. □

To apply the above result we need to guarantee that for a fixed α, the operator H_α is well defined and bounded. For simplicity, we first consider the case when

$$g\frac{d}{dx}\rho^{-1}, \quad g\left(\frac{d^2}{dx^2}\rho^{-1}\right)\left(\frac{d}{dx}\rho^{-1}\right)^{-1}, \tag{4.8}$$

and

$$\xi \to \quad \xi f(\alpha,\xi), \quad f(\alpha,\xi), \quad h(\alpha,\xi), \tag{4.9}$$

are bounded functions on U. If we want to show that the operator H_α is well defined, it is necessary to find proper subspaces K. For that, we consider the subspace $\mathcal{D}(U)$ of $PW_{\rho,s}(U)$, which consist of all functions N such that

$$gN\frac{d}{dx}\rho^{-1}, \quad gN\left(\frac{d^2}{dx^2}\rho^{-1}\right)\left(\frac{d}{dx}\rho^{-1}\right)^{-1},$$

are in $PW_{\rho,s}(U)$. Then, the operator H_α from $\mathcal{D}(U) \times L^\infty(0,\infty)$ to $PW_{\rho,s}(U)$ is well defined, and bounded. We use modified versions of Tikhonov and Landweber filters, such that, all the functions in Equation (4.9) are bounded. These filters are commonly used in regularization theory for compact operators [31].

4.3.2 Tikhonov filtering

In the classical case of linear operators, the solution x to the problem $Ax = y$, where $A : \mathcal{H} \to \mathcal{H}$ is a non-negative linear and compact operator, can be regularized by using the filter

$$\frac{1}{\lambda + \alpha},$$

in the spectral variable, such that

$$x^{\alpha} = \sum_{n=1}^{\infty} \frac{< x, u_n >}{\lambda_n + \alpha} v_n,$$

and where

$$A(x) = \sum_{n=1}^{\infty} \lambda_n < x, u_n > v_n,$$

is the singular value decomposition of the operator A. See [31] for more details. For the problem under consideration, we modified the above function and consider the filter

$$f(\alpha, \xi) = \frac{1}{k\,\mathcal{F}_{\rho,s}P(\xi) - 1}\left(\frac{1}{1+\alpha|\xi|}\right) \quad \text{and} \quad h(\alpha, \xi) = \frac{1}{k\,\mathcal{F}_{\rho,s}P(\xi) - 1}. \tag{4.10}$$

We assume that the probability P satisfies Hypothesis 4.2.1. Using these filters, we observe that the functions in Equation (4.9) are bounded with respect to the variable ξ. Under these assumptions, the bilinear operator H_α is bounded. In fact, a straightforward computation shows that

$$\|H_\alpha(N, \lambda)\| \leq C(1 + \frac{1}{\alpha} + |\lambda|)\|N\|,$$

for some positive constant C, which only depends on s, P, g, ρ^{-1}.

We now state the following regularization method, based on the filter functions defined as above.

Theorem 4.3.2 (Tikhonov regularization)**.** *Assume that the probability P satisfies Condition 4.1, and Hypothesis 4.2.1. Moreover, we assume that $N \in \mathcal{D}(U)$ satisfies all the assumptions of Proposition 4.1.1, and*

$$\xi\,\mathcal{F}_{\rho,s}(\frac{d}{dx}gN) \in L^2(\mathbb{R})\,.$$

Then, for a noisy measurement $(N_\varepsilon, \lambda_\varepsilon)$ of (N, λ) in $\mathcal{D}(U) \times L^\infty(0, \infty)$, the approximation

$H_\alpha(N_\varepsilon, \lambda_\varepsilon)$ using the filters (4.10) *satisfies the estimate*

$$\|H - H_\alpha(N_\varepsilon, \lambda_\varepsilon)\| \leq M\,(\alpha + (1 + \frac{1}{\alpha} + |\lambda_\varepsilon - \lambda|)\,\|N_\varepsilon - N\|),$$

for some positive constant M which depends on s, P, g, ρ^{-1}, and N.
If $(N_\varepsilon, \lambda_\varepsilon)$ satisfy $\|N_\varepsilon - N\| \leq \varepsilon$, and $\|\lambda_\varepsilon - \lambda\| \leq \varepsilon$, we can choose the optimal parameter $\alpha = \sqrt{\varepsilon}$ to conclude

$$\|H - H_\alpha(N_\varepsilon, \lambda_\varepsilon)\| \leq M(\sqrt{\varepsilon} + \varepsilon + \varepsilon^2).$$

Proof. Using the Fourier transform in Eq. (1.1), together with Proposition 4.1.1, we have

$$\mathcal{F}_{\rho,s}(H(N,\lambda) - H_\alpha(N,\lambda)) = \frac{1}{k\,\mathcal{F}_{\rho,s}P(\xi) - 1}\,\frac{\alpha|\xi|}{1+\alpha|\xi|}\,\mathcal{F}_{\rho,s}(\frac{d}{dx}gN).$$

Therefore, we obtain

$$\|H - H_\alpha(N,\lambda)\| \leq \alpha B \left\| \xi\,\mathcal{F}_{\rho,s}\left(\frac{d}{dx}gN\right)\right\|.$$

Hence, by Theorem 4.3.1 we obtain the result. □

4.3.3 The quasi-reversibility Tikhonov filtering

The quasi-reversibility method proposed in [63] regularizes the inverse problem for the case of equal-mitosis. The idea of quasi-reversibility goes back to the work [46]. This method is based on adding a perturbation of a differential operator. In [30], an extension of this method was proposed for more kernels. We now present a different generalization using spectral filters. We show that there exists a relation between the quasi-reversibility method and our approach, especially in the case of self-similar kernels.

Let us first describe our method without going to technical details. For each real

number j, we defined the following modified Tikhonov filters

$$f^j(\alpha,\xi) = h^j(\alpha,\xi) = \frac{1}{k\,\mathcal{F}_{\rho,s}P(\xi) - 1 + (2\pi i\xi + j)\alpha}.$$

If the quotient factor blows up, then the last equation is not well defined. To avoid this kind of problem, we assume that the kernel P satisfies a stronger hypothesis than 4.2.1, namely, we assume that

Hypothesis 4.3.1. *There exist positive constants M and C such that $|k\,\mathcal{F}_{\rho,s}P(\xi) - 1 + (2\pi i\xi + j)\alpha| \geq M$ for all $\xi \in U$, and $\alpha \in [0, C]$.*

In Section 4.4, we give some examples for which the above hypothesis is satisfied. If we assume that N is a smooth function satisfying Condition 4.13, we see that under the Fourier transform $\mathcal{F}_{\rho,s}$, the approximation in Eq. (4.7) solves the following differential equation

$$S_\alpha(N,\lambda) + k\mathcal{K}H_\alpha(N,\lambda) - H_\alpha(N,\lambda) = \frac{d}{dx}gN + \lambda N, \tag{4.11}$$

where

$$S_\alpha(N,\lambda) = \alpha\frac{d}{dx}(H_\alpha(N,\lambda)(\frac{d}{dx}\rho^{-1})^{-1}) + \alpha H_\alpha(N,\lambda)(\frac{d^2}{dx^2}\rho^{-1})(\frac{d}{dx}\rho^{-1})^{-2} + j\alpha H_\alpha(N,\lambda).$$

Thus, our method can be seen as a perturbation method Ref. [44]. For the self-similar case, that is, when $\rho(x) = e^x$, we see that

$$S_\alpha(N,\lambda) = \alpha\frac{d}{dx}(xH_\alpha(N,\lambda)) + \alpha(j-1)H_\alpha(N,\lambda).$$

Taking $j = 1$, the perturbation method defined in Eq. (4.11) reduces to the quasi-reversibility method proposed in [30]. With the same ideas of the proof of Theorem 4.3.2, we prove the next result.

Theorem 4.3.3 (Quasi-reversibility method)**.** *Assume that the probability P satisfies Equation* (4.1)*, and Hypothesis 4.3.1. Moreover, we assume that $N \in \mathcal{D}(U)$, and also that satisfies all the assumptions of Proposition 4.1.1. If the functions*

$$\xi\, \mathcal{F}_{\rho,s}(\frac{d}{dx} gN) \quad \textit{and} \quad \xi\, \mathcal{F}_{\rho,s}(\frac{d}{dx} N),$$

are in $L^2(\mathbb{R})$, then, for all noisy measurement $(N_\varepsilon, \lambda_\varepsilon)$ of (N, λ) in $\mathcal{D}(U) \times L^\infty(0,\infty)$, the approximation $H_\alpha(N_\varepsilon, \lambda_\varepsilon)$ satisfies the error estimate

$$\|H - H_\alpha(N_\varepsilon, \lambda_\varepsilon)\| \leq\ M\,(\alpha + (1 + \frac{1}{\alpha} + \|\lambda_\varepsilon - \lambda\|)\,\|N_\varepsilon - N\|,$$

for some constant M which depends on P, g, ρ^{-1}, and N.
If $(N_\varepsilon, \lambda_\varepsilon)$ satisfy $\|(N_\varepsilon - N, \lambda_\varepsilon - \lambda)\| \leq \varepsilon$, we can choose the regularization parameter $\alpha = \sqrt{\varepsilon}$, to conclude

$$\|H - H_\alpha(N_\varepsilon, \lambda_\varepsilon)\| \leq M(\sqrt{\varepsilon} + \varepsilon + \varepsilon^2).$$

4.3.4 The Landweber‘s method

Using the Tikhonov and quasi-reversibility filters, we obtain an approximation error of order $O(\sqrt{\varepsilon})$. We improve this error order, and thus we obtain a better approximation. For that, we use a new filter, which is a modification of the classical Landweber filter

$$f(\alpha,\xi) = \frac{1}{k\,\mathcal{F}_{\rho,s}P(\xi) - 1}\left(1 - \left(1 - \frac{1}{1+\xi^2}\right)^{\alpha}\right) \quad \text{and} \quad h(\alpha,\xi) = \frac{1}{k\,\mathcal{F}_{\rho,s}P(\xi) - 1}. \tag{4.12}$$

To implement these filters in Theorem 4.3.1, we first establish the following inequalities.

Lemma 4.3.1. *There exists a positive constant A, such that for all $x \in \mathbb{R}_{>0}$ and for all*

positive α, the following estimate holds

$$|xf(x,\alpha)| \leq A\sqrt{\alpha},$$

where A depends on s, P, g, ρ^{-1}. Moreover, for all $m \in \mathbb{R}$, such that $m < 2\alpha$ the following inequality holds

$$|\frac{1}{x^m}(1-\frac{1}{1+x^2})^\alpha| \leq (\frac{m}{2\alpha})^m.$$

Proof. Let us prove the first inequality. Using the change of variables $z = 1 - \frac{1}{1+x^2}$, we see that

$$|xf(x,\alpha)|^2 \leq M\frac{z}{1-z}(1-z^\alpha)^2.$$

Since $z \in [0,1]$, then applying the mean value inequality to the function z^α, we obtain

$$(1-z^\alpha)^2 \leq (1-z^\alpha) \leq \alpha(1-z),$$

thus $|xf(x,\alpha)| \leq M\sqrt{\alpha}$ as desired. To prove the second inequality, we observe that the point $\omega = \frac{2\alpha-m}{2\alpha}$ attains the maximum for the function $(1-z)^m z^{2\alpha-m}$ in $[0,1]$. Thus, for all $x \in \mathbb{R}$ we have

$$|\frac{1}{x^m}(1-\frac{1}{1+x^2})^\alpha|^2 = (1-z)^m z^{2\alpha-m} \leq (1-\omega)^m = (\frac{m}{2\alpha})^m.$$

□

As consequence of the previous lemma, we have that the functions in (4.9) are bounded. Thus, the bilinear operator H_α is bounded. In fact, a straightforward computation shows that

$$\|H_\alpha(N,\lambda)\| \leq A(1+\sqrt{\alpha}+\|\lambda\|)\|N\|,$$

for some positive constant A, which only depends on s, P, g, ρ^{-1}. Now, we apply Theorem 4.3.1 to the Landweber filter. For that, we require a smoothness condition of order $m \in \mathbb{R}_{>0}$ for the function N

Theorem 4.3.4 (Landweber regularization)**.** *Assume that the probability P satisfies* (4.1)*, and Hypothesis 4.2.1. Moreover, assume that $N \in \mathcal{D}(U)$, satisfies all the assumptions of Proposition 4.1.1, and the smoothness condition with order $m \in \mathbb{R}_{>0}$*

$$\xi^{m+1}\, \mathcal{F}_{\rho,s}(\frac{d}{dx} gN) \in L^2(\mathbb{R})\,.$$

Then, for all noisy measurement $(N_\varepsilon, \lambda_\varepsilon)$ of (N, λ) in $\mathcal{D}(U) \times L^\infty(0, \infty)$, the approximation $H_\alpha(N_\varepsilon, \lambda_\varepsilon)$ defined in Eq. (4.7) *using the spectral filter of Eq.* (4.12) *satisfies the error estimate*

$$\|H - H_\alpha(N_\varepsilon, \lambda_\varepsilon)\| \leq\ M\,((\frac{m}{\alpha})^m + (1 + \sqrt{\alpha} + \|\lambda_\varepsilon - \lambda)\|)\,\|N_\varepsilon - N\|),$$

for some constant M which depends on s, m, P, g, ρ^{-1}, and N. Suppose that $(N_\varepsilon, \lambda_\varepsilon)$ satisfies $\|(N_\varepsilon - N, \lambda_\varepsilon - \lambda)\| \leq \varepsilon < 2^{\frac{2m+3}{2}}\sqrt{m}$. We can choose the optimal parameter given by

$$\alpha = (\frac{2m^{m+1}}{\varepsilon})^{\frac{2}{2m+1}},$$

which satisfies $m < 2\alpha$, to conclude

$$\|H - H_\alpha(N_\varepsilon, \lambda_\varepsilon)\| \leq M(\varepsilon + \varepsilon^{\frac{2m}{2m+1}}).$$

Proof. Using the Fourier transform in Equation (1.1), and Proposition 4.1.1, we have that

$$\mathcal{F}_{\rho,s}(H(N, \lambda) - H_\alpha(N, \lambda)) = \frac{1}{k\,\mathcal{F}_{\rho,s}P(\xi) - 1}(1 - \frac{1}{1 + x^2})^\alpha \mathcal{F}_{\rho,s}(\frac{d}{dx} gN).$$

Next, we apply the second estimate of Lemma 4.3.1 to obtain

$$\|H(N,\lambda) - H_\alpha(N,\lambda)\| \leq M(\frac{m}{\alpha})^m \|\xi^{m+1}\, \mathcal{F}_{\rho,s}(\frac{d}{dx} gN)\|,$$

hence by Theorem 4.3.1 we obtain the result. □

4.3.5 The unbounded case

In many cases, the functions in Equation (4.8) are not bounded. For instance, in the self-similar case. We now extend our results to the unbounded case. The idea is to regularize the functions in Equation (4.8) using a spectral filter. To do that, we write

$$T_1 = \left(\frac{d}{dx}\rho^{-1}\right)^2 \quad T_2 = \left(\left(\frac{d^2}{dx^2}\rho^{-1}\right)\left(\frac{d}{dx}\rho^{-1}\right)^{-1}\right)^2,$$

and

$$g_\alpha = \frac{g}{1 + \alpha\left(\exp(g^2) + \exp(g^2 T_1) + \exp(g^2 T_2)\right)},$$

where $\alpha \in \mathbb{R}_{>0}$, is the regularization parameter of the functions (4.8). Observe that the exponential decay guarantees that using g_α in the place of g, then the functions (4.8) are bounded. Now, we show that if the function N has fast decay, then the bounded function g_α can be used to regularize the inverse problem, even if g is not bounded.

Proposition 4.3.1. *Assume that the function N satisfies*

$$\frac{d}{dx}(g_\alpha(\exp(g^2) + \exp(g^2 T_1) + \exp(g^2 T_2))N) \ \in L^2_{\rho,s}. \tag{4.13}$$

We define H^α as the solution of (1.1)*, where the solution is associated with the function*

g_α instead of g. Then, we have that

$$\|H - H^\alpha\| \leq C\alpha \| \frac{d}{dx}(g(exp(g^2) + exp(g^2 T_1) + exp(g^2 T_2))N)\|,$$

where C is a constant, which depends on the kernel $\mathcal{K}$ and the number k.

Proof. Observe that

$$(k\mathcal{K} - Id)(H - H^\alpha) = \frac{d}{dx}(g - g_\alpha)N.$$

Since

$$g - g_\alpha = g\left(\frac{\alpha \exp(g^2) + \exp(g^2 T_1) + \exp(g^2 T_2)}{1 + \alpha\left(\exp(g^2) + \exp(g^2 T_1) + \exp(g^2 T_2)\right)}\right),$$

then, using the Fourier transform $\mathcal{F}_{\rho,s}$, we obtain

$$\|H - H^\alpha\| \leq \alpha \left\| \frac{1}{\mathcal{F}_{\rho,s}(k\mathcal{K} - Id)} \right\|_\infty \left\| \frac{d}{dx}(g(exp(g^2) + exp(g^2 T_1) + exp(g^2 T_2))N) \right\|$$

□

Thus, the function H^α is a controlled approximation for H. In this case, the function H^α is the solution of Eq. 1.1, associated with the bounded function g_α. Therefore, we can use some of the previous methods (Tikhonov or Landweber) to regularize H^α, and by Proposition 4.3.1 to regularize H.

4.4 Examples

To use the previous regularization methods, we need to verify that the probability P satisfies (4.1), and some of the Hypotheses 4.2.1 or 4.3.1. In this section, we study some examples which satisfy the previous conditions. First, we discuss examples arising from the self-similar fragmentation kernels, that is, when $\rho(x) = e^x$ and $k = 2$.

To check Hypothesis 4.3.1, it is sufficient to consider the case $j = 0$. In fact, if Hypothesis 4.3.1 holds for $j = 0$, then by triangular inequality we see that for all $j \neq 0$ and $\alpha \in [0, \min(C, M/2|j|)]$

$$\frac{M}{2} \leq M - |j|\alpha \leq |k\mathcal{F}_{\rho,s}(P)(\xi) - 1 + 2\pi i\alpha\xi + j\alpha|.$$

Thus, the Hypothesis 4.3.1 holds for all real number j. We now study for which values of s, and open sets U, the Hypothesis 4.3.1 holds. Without loss of generality, we assume that $j = 0$.

4.4.1 The equal-mitosis

An important example of a self-similar case is the equal-mitosis. In equal-mitosis the probability kernel is given by $P = \delta_{x=\frac{1}{2}}$. In this case, the Fourier transform of P is given by

$$\mathcal{F}_{\rho,s}(P)(\xi) = 2\left(\frac{1}{2}\right)^{\pi s - 2\pi i\xi}.$$

Thus, by triangular inequality

$$|2^{2-\pi s} - 1| \leq |\mathcal{F}_{\rho,s}(P)(\xi) - 1|. \tag{4.14}$$

Therefore, we conclude that this probability satisfies the Condition (4.1) and Hypothesis 4.2.1, for all $s \in \mathbb{R} \setminus \{\frac{2}{\pi}\}$. In order to verify Hypothesis 4.3.1, we consider the following cases.

- **First case:** $s > 2/\pi$.

  For such s, the Hypothesis 4.3.1 is satisfied for all $\xi \in \mathbb{R}$. In fact, by triangular

inequality

$$1 - 2^{2-\pi s} < \sqrt{1 + (2\pi\alpha\xi)^2} - 2^{2-\pi s} \leq |2\mathcal{F}_{\rho,s}(P)(\xi) - 1 + 2\pi i\alpha\xi|.$$

- **Second case:** $s < 2/\pi$.

  In this case, the Hypothesis 4.3.1 holds for each open bounded set, and α small enough. In fact, using triangular inequality we have

$$|\mathcal{F}_{\rho,s}(P)(\xi) - 1| - 2\pi i\alpha\xi \leq |2\mathcal{F}_{\rho,s}(P)(\xi) - 1 + 2\pi i\alpha\xi|.$$

  Then, using estimate 4.14, we conclude that for α small the expression $|\mathcal{F}_{\rho,s}(P)(\xi) - 1| - 2\pi i\alpha\xi$ is positive.

Thus, we can apply our methodology to deal with the inverse problem for these cases. In Section 4.5, we will develop some numerical simulations for the previous cases and see the effectiveness of our approach.

4.5 Numerical solution of the inverse problem

In this section, we recover numerically the solution of the inverse problem using the regularization methods proposed in Section 4.3. That is, we recover H from noisy measurements of N. If we assume that the noisy measurement N_ε is a smooth function satisfying Condition 4.13, then there exists a unique solution H_α^ε for the Equation (1.1) associated with N_ε. The purpose of this section is to explore numerically how close is H_α^ε to H, when the noisy data N_ε is close to N.

To do that, we first construct an approximation for N, using the numerical schemes proposed in [29, 30], and then, we add random noise to the data N. We construct the approx-

imation H_α^ε using Equation (4.7), where α is the optimal regularization parameter for each method.

4.5.1 Parametric specification.

We now present some numerical simulations for the equal-mitosis case. Here, the parameters are $s = 0$, $k = 2$ and $\rho(x) = e^x$, and the probability distribution is given by

$$P = \delta_{\frac{1}{2}}.$$

To guarantee the existence and uniqueness of the solution of the direct problem, and also the condition (4.8), we select fragmentation and growth rates with fast decay at 0 and ∞. See [27]. To be more specific, we use the growth rate $g(x) = xe^{-(x+1/x)}$ and the fragmentation rate

$$B(x) = x^2 e^{-(x+1/x)}.$$

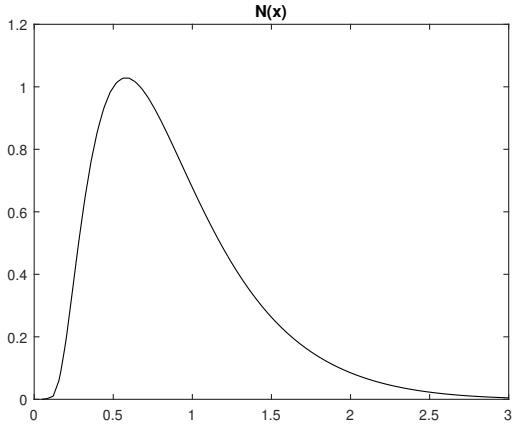
In this experiment, we use the space $L^2_{\rho,0}$ to recover the data H on $(0, 10]$. We also use the parameter $j = 1$ for the quasi-reversibility method, and $m = 10$ for the Landweber filter.

4.5.2 Construction of N.

We construct the function N on the interval $(0, 3]$ using the numerical scheme proposed in [29, 30]. Here, the initial condition is given by

$$n_0(x) = e^{-(x-8)^2/2}.$$

We use a regular grid on $(0, 3]$ with 500 points. For the evolution process, we use a regular grid on $(0, 200]$, with 10^4 points. We plot the function N in Figure 4.1.



(a) Equal-mitosis case

Figure 4.1: Construction of N, solution of the direct problem for the equal- mitosis case.

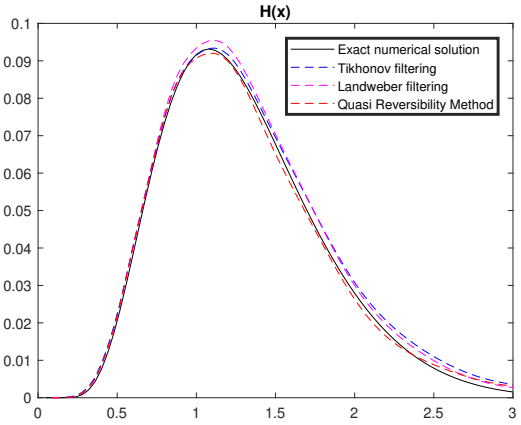
4.5.3 Reconstruction of H and B.

We now consider a noisy approximation $(N_\varepsilon, \lambda_\varepsilon)$ for the eigenpair (N, λ), obtained by adding a random noise to the data. That is, we assume that

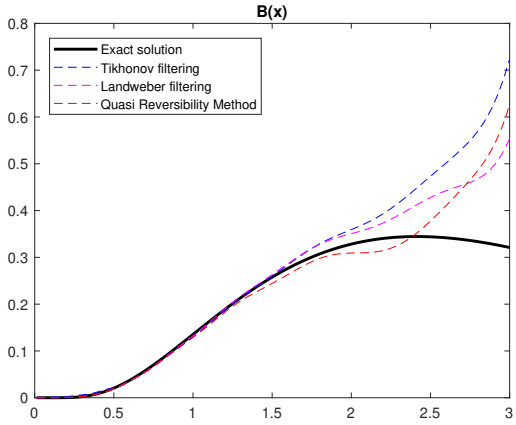
$$N_\varepsilon = \max(N + R_\varepsilon, 0), \quad \lambda_\varepsilon = \max(\lambda + S_\varepsilon, 0),$$

where $R_\varepsilon, S_\varepsilon$ are random noises uniformly distributed in $[-\varepsilon, \varepsilon]$. We recover the approximation H_α^ε using the noisy measurement $(N_\varepsilon, \lambda_\varepsilon)$. We plot the approximation H_α^ε for different noise levels ε. Here, we use the values $\varepsilon = 10^{-2}$ (Figure 4.2), and $\varepsilon = 10^{-3}$ (Figure 4.3). The parameters used are $m = 10$ for the Tikhonov method and $j = 1$ for the Landweber method.

To recover the fragmentation rate B from $H = BN$ we use the truncate division by N. That is, we define $B(x) = H(x)/N(x)$ if $N(x) \neq 0$, and zero otherwise. The following figures show the recovered function B for the parameters $\varepsilon = 10^{-2}$, and $\varepsilon = 10^{-3}$. Observe that the instabilities near $x = 3$ of the reconstructed function B are due to the fact that the fast decay of N near $x = 3$ affects the truncated division.

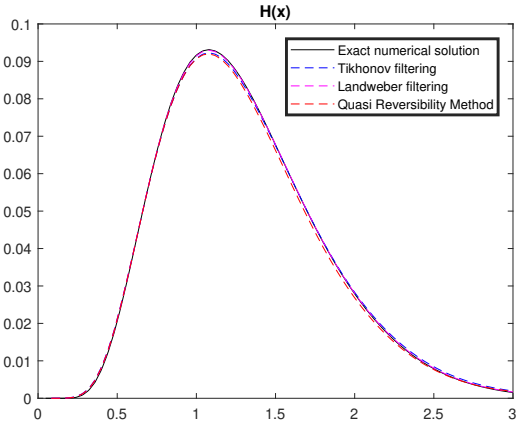


(a) Reconstruction of H

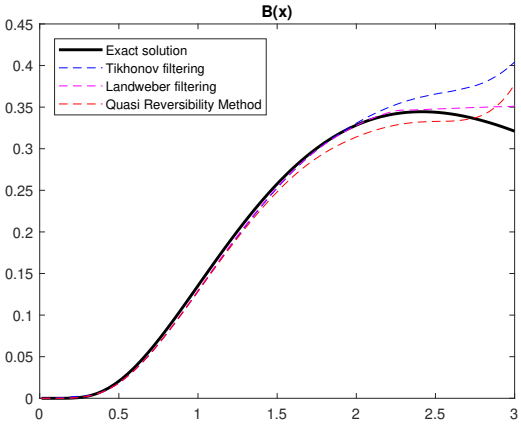


(b) Reconstruction of B

Figure 4.2: Numerical reconstruction of H and B, using the noise level $\varepsilon = 10^{-2}$.

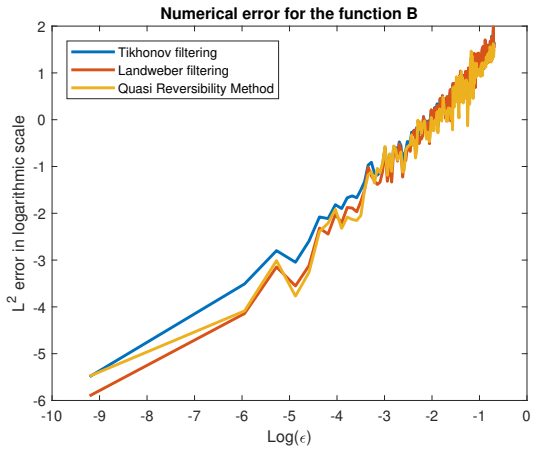


(a) Reconstruction of H

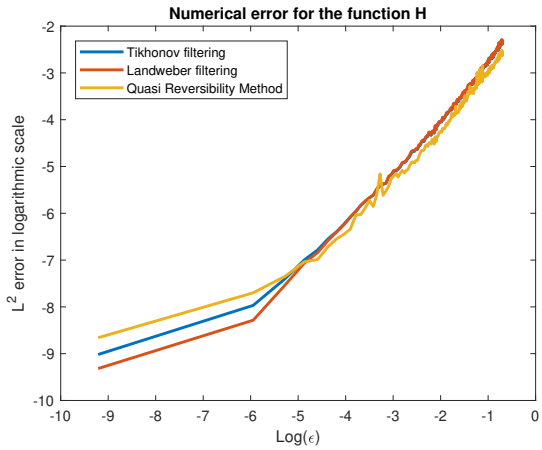


(b) Reconstruction of B

Figure 4.3: Numerical reconstruction of H and B, using the noise level $\varepsilon = 10^{-3}$.



(a) Numerical error of the reconstruction of B



(b) Numerical error of the reconstruction of H

Figure 4.4: Numerical error of the reconstruction of the functions B and H using the L^2 norm for several values of ϵ. Note that errors are on logarithmic scale

4.5.4 Numerical Error.

We compare the numerical error of the reconstructions of the functions B and H for small values of ε. Here, we use the $L^2[0,3]$ norm to estimate this error. This norm is computed using rectangular integration. For this experiment, we assume that $\epsilon \in [10^{-4}, 0.5]$. We plot the error in logarithmic scale in Figure 4.4. Observe that for small values of ϵ the Landweber filter gives a better reconstruction, which is following our theoretical results.

Chapter 5

Conclusions

This thesis consists of three main parts. In Chapters 2 and 3, we investigate some aspects related to diffusion-maps and in Chapter 4 the inverse problem of the growth-fragmentation equation. Regarding the theory of diffusion-maps, we treat the problem of dimensionality reduction of data sets whose structure is given by asymmetric kernels. Our methodology generalizes the diffusion-map framework to asymmetric kernels and computes a diffusion representation based on the kernel coordinates on a proper orthonormal basis. Our representation depends on two parameters, the first parameter defines the approximation error, and the second one the dimensionality.

In our experiments, we used the Fourier basis to represent the structure of the data set. This choice is based on the fact that the Fourier basis diagonalizes the Laplacian operator which is the main example of a diffusive process. From the numerical viewpoint, the main advantage of using the Fourier basis is that the 2d-FFT allows us a reduction from linear growth to logarithmic growth of one of the factors. The latter contributes to the computational complexity reduction when compared to traditional eigenvalue methods. If we consider that our kernel is represented by an $n \times n$ matrix, then the SVD takes $O(n^3)$ of operations to be performed, whereas the 2d-FFT decomposition is $O(n^2 \log(n))$. This

fact was confirmed in a set of experiments with randomly generated kernels. Observe that the SVD representation gives a better approximation, i.e., smaller errors. However, if we use the Fourier basis we can obtain a good approximation of the data set for a much lower computational cost. Additionally, the use of the Fourier basis allows us to see in more detail some geometric properties of the data set. This suggests that it is possible to use the Fourier basis as an alternative to the classic representation by eigenvalues, especially in computers with low performance.

We perform a few applied experiments to test the theory. In particular, we apply it to identify which regions of Brazil have presented a greater variation in the temperature *vis a vis* other ones during the last decades. In this experiment, we see that the Amazon region has presented more variations in its temperature as compared to other places. This observation indicates that further studies should be performed to investigate the possible reasons for such variations. To avoid increasing the computational cost, we did not use the Markov normalization in this kernel. Due to the high dimensionality of the kernel matrix, the SVD algorithm did not conclude in our computer for this experiment due to its complexity. However, we managed to execute the algorithm using the 2d-FFT.

We also ran some experiments with synthetic data using a wavelet basis. However, we did not obtain an improvement in the computation time or the error of the approximation when compared to the Fourier basis and the singular vector basis.

Asymmetric kernels are present in several mathematical models, for instance in weighted directed graphs. In such graphs, the transition from one node to another is measured by an asymmetric kernel. In general, asymmetric kernels are useful to represent a gain or loss of information when we move from one point to another. Weighted directed graphs are thus used to model real-world problems such as the traffic in a city, electrical network systems, water flow in hydrological basins, and commodity trading between economies. These are natural follow-up avenues to the present work.

Additionally, we use our asymmetric kernel approach to recover the gradient operator defined on Riemannian submanifolds of the Euclidean space from random samples in a neighborhood of the point of interest. Our methodology is based on the estimates of the Laplace-Beltrami operator proposed in the diffusion maps approach. The estimates do not depend on the intrinsic parametrization of the sub-manifold. This feature is useful in cases where it is not feasible to identify the sub-manifold in which the data-set is lying. A natural continuation of the present work would be to incorporate information of the cotangent bundle and deal with a duality version of our results, in this case, the aforementioned approach would be very handy. Furthermore, this circle of ideas could be conjoined with the techniques proposed in [65].

We apply our methodology in a step size algorithm as an optimization method on manifolds. This optimization method is effective in cases where it is difficult to compute the gradient of a function. As an application, we used our method to find an approximation to the sphere packing problem in dimensions 2 and 3, for the lattice packing case. Moreover, we use our approach to reconstruct tomographic images where the projected angles are unknown. The latter does not depend on *a priory* knowledge of the distribution of the angles, and its execution is computationally feasible.

Concerning the growth-fragmentation equation, we developed a new approach to regularize the inverse problem associated with the growth-fragmentation equation. Our approach is based on the Fourier transform theory for locally compact groups. We regularized the Fourier transform of differential operators using several filters in the spectral variable, such as modifications of the Landweber and Tikhonov filters. We obtain estimates for the error of each method.

For each method, we obtained their respective error estimates. The Landweber method provides an algorithm to recover the fragmentation rate B, with order $O(\frac{2m}{2m+1})$, as proved in Theorem 4.3.4. The error obtained using the Landweber method is better compared

with other methods. This fact was verified by numerical simulations.

All the codes and numerical implementations of this thesis are available at the GitHub repository [33].

Bibliography

[1] Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*, volume 78. 12 2008.

[2] Alvaro Almeida Gomez, Antônio J. Silva Neto, and Jorge P. Zubelli. Diffusion representation for asymmetric kernels. *Applied Numerical Mathematics*, 166:208–226, 2021.

[3] François Baccelli, David McDonald, and Julien Reynier. A mean-field model for multiple tcp connections through a buffer implementing red. *Performance Evaluation*, 49:77–97, 09 2002.

[4] Britta Basse, Bruce Baguley, Elaine Marshall, Wayne Joseph, Bruce van Brunt, Graeme Wake, and David Wall. A mathematical model for analysis of the cell cycle in cell lines derived from human tumors. *Journal of mathematical biology*, 47:295–312, 11 2003.

[5] S. Basu and Y. Bresler. Feasibility of tomography with unknown view angles. In *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, pages 15–19 vol.2, 1998.

[6] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, June 2003.

[7] Alain Bensoussan and Jo Menaldi. Difference equations on weighted graphs. *Journal of Convex Analysis*, 12, 2003.

[8] Joseph Frederic Bonnans, Jean Charles Gilbert, C Lemaréchal, and Claudia Sagastizabal. *Numerical Optimization – Theoretical and Practical Aspects*. Springer, 01 2006.

[9] Nicolas Boumal and Bamdev Mishra. *Manopt Toolbox*, Sep. 5, 2021. `https://github.com/NicolasBoumal/manopt`.

[10] Nicolas Boumal, Bamdev Mishra, P.-A. Absil, and Rodolphe Sepulchre. Manopt, a matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15(42):1455–1459, 2014.

[11] Thibault Bourgeron, Marie Doumic, and Miguel Escobedo. Estimating the division rate of the self-similar growth-fragmentation equation. *Inverse Problems*, 30, 01 2014.

[12] Xavier Bresson and Tony Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging*, 2:455–484, 11 2008.

[13] F. Bullo and K. Fujimoto. *Lagrangian and Hamiltonian Methods For Nonlinear Control 2006: Proceedings from the 3rd IFAC Workshop, Nagoya, Japan, July 2006*. Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 2007.

[14] Christian Chapman. *Implementation of shortest vector of a lattice using exhaustive enumeration*, 2018 (accessed December 1, 2021). `https://github.com/enthdegree/lenum.m`.

[15] Min Chen, Shiwen Mao, and Yunhao Liu. Big data: A survey. *Mobile Networks and Applications*, 19:171–209, 2014.

[16] Jeffrey Cheng. Tomographic image of a knee. `https://radiopaedia.org/cases/normal-ct-knee-1`. Accessed: 2021-11-01.

[17] Henry Cohn, Abhinav Kumar, Stephen Miller, Danylo Radchenko, and Maryna Viazovska. The sphere packing problem in dimension 24. *Annals of Mathematics*, 185, 03 2016.

[18] R. R. Coifman and M. J. Hirn. Diffusion maps for changing data. *Applied and Computational Harmonic Analysis*, 36(1):79 – 107, 2014.

[19] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5 – 30, 2006. Special Issue: Diffusion Maps and Wavelets.

[20] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21):7426–7431, 2005.

[21] Ronald R. Coifman, Yoel Shkolnisky, Fred J. Sigworth, and Amit Singer. Graph laplacian tomography from unknown random projections. *IEEE Transactions on Image Processing*, 17(10):1891–1899, 2008.

[22] K. Ch. Das. The Laplacian spectrum of a graph. *Computers & Mathematics with Applications*, 48(5):715 – 724, 2004.

[23] Ingrid Daubechies, Gerd Teschke, and Luminita Vese. Iteratively solving linear inverse problems under general convex constraints. *Inverse Problems and Imaging*, 1:29–46, 02 2007.

[24] Diario de cultura. Tom Jobim in Ipanema Beach picture. https://www.diariodecultura.com.ar/columnas/crucigrama-antonio-brasileiro/. Accessed: 2020-06-01.

[25] S.R. Deans. *The Radon Transform and Some of Its Applications*. A Wiley-Interscience publication. Wiley, 1983.

[26] Manfredo P. do Carmo. *Riemannian Geometry*. Mathematics (Boston, Mass.). Birkhäuser, 1992.

[27] Marie Doumic and Pierre Gabriel. Eigenelements of a general aggregation-fragmentation model. *Mathematical Models and Methods in Applied Sciences*, 20:757–783, 07 2009.

[28] Marie Doumic, Pedro Maia, and Jorge Zubelli. On the calibration of a size-structured population model from experimental data. *Acta biotheoretica*, 58:405–13, 12 2010.

[29] Marie Doumic, Benoit Perthame, and Jorge Zubelli. Numerical solution of an inverse problem in size-structured population dynamics,. *Inverse Problems*, 25:045008, 04 2009.

[30] Marie Doumic and Léon Tine. Estimating the division rate for the growth-fragmentation equation. *Journal of mathematical biology*, 67, 06 2012.

[31] Heinz Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands, 1996.

[32] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010.

[33] Alvaro Almeida Gomez. *Codes and numerical implementations of this thesis*, Dec. 1, 2021. https://github.com/zabu182.

[34] L. Grafakos. *Classical Fourier Analysis*. Graduate Texts in Mathematics. Springer New York, 2014.

[35] Meredith Greer, Laurent Pujo-Menjouet, and Glenn Webb . A mathematical analysis of the dynamics of prion proliferation. *Journal of theoretical biology*, 242:598–606, 11 2006.

[36] Helmut Groemer. Existenzsätze für lagerungen in metrischen räumen. *Monatshefte Fur Mathematik - MONATSH MATH*, 72:325–334, 08 1968.

[37] M.A Hajji, S Melkonian, and R Vaillancourt. Representation of differential operators in wavelet basis. *Computers & Mathematics with Applications*, 47(6):1011 – 1033, 2004.

[38] Thomas C. Hales. A proof of the kepler conjecture. *Annals of Mathematics*, 162(3):1065–1185, 2005.

[39] J. Hansen, R. Ruedy, M. Sato, and K. Lo. Global surface temperature change. *Reviews of Geophysics*, 48(4), 2010.

[40] G.T. Herman. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. Advances in Computer Vision and Pattern Recognition. Springer London, 2009.

[41] INMET. Brazilian temperature dataset. `http://www.inmet.gov.br/portal/`. Accessed: 2020-06-01.

[42] I.T. Jolliffe and Springer-Verlag. *Principal Component Analysis*. Springer, 2002.

[43] Avinash C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001.

[44] T. Kato. *Perturbation Theory for Linear Operators*. Springer Berlin Heidelberg, 1995.

[45] J. Kepner, H. Jananthan, and C.E. Leiserson. *Mathematics of Big Data: Spreadsheets, Databases, Matrices, and Graphs*. MIT Lincoln Laboratory Series. MIT Press, 2018.

[46] R. Lattès and J.-L Lions. *Méthode de quasi-réversibilité et applications*. Travaux et recherches mathématiques. Dunod, 1967.

[47] Philippe Laurençot and Benoit Perthame. Exponential decay for the growth-fragmentation/cell-division equations. *Commun. Math. Sci.*, 7(2):503–510, 06 2009.

[48] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[49] G. Malajovich and J. P. Zubelli. On the geometry of graeffe iteration. *Journal of Complexity*, 17(3):541–573, 2001.

[50] G. Malajovich and J. P. Zubelli. Tangent graeffe iteration. *Numerische Mathematik*, 89(4):749–782, 2001.

[51] Eeshan Malhotra and Ajit Rajwade. Tomographic reconstruction from projections with unknown view angles exploiting moment-based relationships. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1759–1763, 2016.

[52] N. F. Marshall and M. J. Hirn. Time coupled diffusion maps. *Applied and Computational Harmonic Analysis*, 45(3):709 – 728, 2018.

[53] Johan A. Metz and Odo. Diekmann. Formulating models for structured populations. In Johan A. Metz and Odo. Diekmann, editors, *The Dynamics of Physiologically Structured Populations*, pages 78–135. Springer Berlin Heidelberg, 1986.

[54] Philippe Michel, Stéphane Mischler, and Benoît Perthame. General entropy equations for structured population models and scattering. *Comptes Rendus Mathematique*, 338(9):697–702, 2004.

[55] Sayan Mukherjee and Qiang Wu. Estimation of gradients and coordinate covariation in classification. *Journal of Machine Learning Research*, 7:2481–2514, 12 2006.

[56] Sayan Mukherjee, Qiang Wu, and Ding-Xuan Zhou. Learning gradients on manifolds. *Bernoulli*, 16, 02 2010.

[57] F. Natterer. *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, 2001.

[58] Earth observatory. World of Change: Global Temperatures. `https://earthobservatory.nasa.gov/world-of-change/DecadalTemp`. Accessed: 2020-06-01.

[59] C.D. Olds, A. Lax, G.P. Davidoff, and G. Davidoff. *The Geometry of Numbers*. Mathematical Association of America, 2000.

[60] United Nations Framework Convention on Climate Change. The Paris Agreement. `https://unfccc.int/process-and-meetings/the-paris-agreement/d2hhdC1pcy`. Accessed: 2020-06-01.

[61] M. Pedersen. *Functional Analysis in Applied Mathematics and Engineering*. CRC Press, 1999.

[62] Benoit. Perthame. *Transport Equations in Biology*. Frontiers in Mathematics. Birkhäuser Basel, 2006.

[63] Benoit Perthame and Jorge P Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, apr 2007.

[64] Benoît Perthame and Lenya Ryzhik. Exponential decay for the fragmentation or cell-division equation. *Journal of Differential Equations*, 210(1):155–177, 2005.

[65] Dmitry Pozharskiy, Noah J. Wichrowski, Andrew B. Duncan, Grigorios A. Pavliotis, and Ioannis G. Kevrekidis. Manifold learning for accelerating coarse-grained optimization. *Journal of Computational Dynamics*, 7(2):511–536, 2020.

[66] Walter. Rudin. *Fourier Analysis on Groups*. Wiley Classics Library. 1990.

[67] M. Salhov, A. Bermanis, G. Wolf, and A. Averbuch. Approximately-isometric diffusion maps. *Applied and Computational Harmonic Analysis*, 38(3):399 – 419, 2015.

[68] M. Salhov, A. Bermanis, G. Wolf, and A. Averbuch. Diffusion representations. *Applied and Computational Harmonic Analysis*, 45(2):324 – 340, 2018.

[69] Claus-Peter Schnorr and Martin Euchner. Lattice basis reduction: Improved practical algorithms and solving subset sum problems. *Mathematical programming*, 66(1-3):181–199, 1994.

[70] Gerard L.G. Sleijpen, Peter Sonneveld, and Martin B. van Gijzen. Bi-cgstab as an induced dimension reduction method. *Applied Numerical Mathematics*, 60(11):1100–1114, 2010. Special Issue: 9th IMACS International Symposium on Iterative Methods in Scientific Computing (IISIMSC 2008).

[71] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, 1970.

[72] Gerald Teschl. Ordinary differential equations and dynamical systems. *American Mathematical Society*, 01 2008.

[73] Maryna Viazovska. The sphere packing problem in dimension 8. *Annals of Mathematics*, 185, 03 2016.

[74] Gang Wu and Fei Li. A randomized exponential canonical correlation analysis method for data analysis and dimensionality reduction. *Applied Numerical Mathematics*, 164:101–124, 2021. Special Issue on The Seventh International Conference on Numerical Algebra and Scientific Computing.

[75] J.D. Zuazo. *Fourier Analysis*. American Mathematical Soc., 2001.