Barycentric embeddings for geometric manifold learning : with application to shapes and graphs

Elodie Maignant

Plongements barycentriques pour l'apprentissage géométrique de variétés

Application aux formes et graphes

Elodie Maignant
Centre Inria d'Université Côte d'Azur, Équipe Epione

Présentée en vue de l'obtention du grade de docteur en Automatique, Traitement du Signal et des Images d'Université Côte d'Azur

Dirigée par Xavier Pennec et Alain Trouvé

Soutenue le 6 décembre 2023

Devant le jury, composé de :

Ian Dryden, Professor, Florida International University
Aasa Feragen, Professor, Danmarks Tekniske Universitet
Jean Gallier, Professor, University of Pennsylvania
Stephan Huckemann, Professor, Georg-August-Universität Göttingen
Xavier Pennec, Directeur de recherche, Centre Inria d'Université Côte d'Azur
Alice Barbara Tumpach, Maître de conférence, Université de Lille
Alain Trouvé, Professeur des universités, ENS Paris-Saclay

Plongements barycentriques pour l'apprentissage géométrique de variétés

Application aux formes et graphes

Jury

Président

Stephan Huckemann, Professor, Georg-August-Universität Göttingen

Rapporteurs

Ian Dryden, Professor, Florida International University
Aasa Feragen, Professor, Danmarks Tekniske Universitet

Examinateurs

Jean Gallier, Professor, University of Pennsylvania
Alice Barbara Tumpach, Maître de conférence, Université de Lille

Directeurs de thèse

Xavier Pennec, Directeur de recherche, Centre Inria d'Université Côte d'Azur
Alain Trouvé, Professeur des universités, ENS Paris-Saclay

Plongements barycentriques pour l'apprentissage géométrique de variétés

Application aux formes et graphes

Résumé

Une image obtenue par IRM, c'est plus de 60 000 pixels. La plus grosse protéine connue chez l'être humain est constituée d'environ 30 000 acides aminés. On parle de données en grande dimension. En réalité, la plupart des données en grande dimension ne le sont qu'en apparence. Par exemple, de toutes les images que l'on pourrait générer aléatoirement en coloriant 256 x 256 pixels, seule une infime proportion ressemblerait à l'image IRM d'un cerveau humain. C'est ce qu'on appelle la dimension intrinsèque des données. En grande dimension, apprentissage rime donc souvent avec réduction de dimension. Il existe de nombreuses méthodes de réduction de dimension, les plus récentes pouvant être classées selon deux approches.

Une première approche, connue sous le nom d'apprentissage de variétés (manifold learning) ou réduction de dimension non linéaire, part du constat que certaines lois physiques derrière les données que l'on observe ne sont pas linéaires. Ainsi, espérer expliquer la dimension intrinsèque des données par un modèle linéaire est donc parfois irréaliste. Au lieu de cela, les méthodes qui relèvent du manifold learning supposent un modèle localement linéaire.

D'autre part, avec l'émergence du domaine de l'analyse statistique de formes, il y eu une prise de conscience que de nombreuses données sont naturellement invariantes à certaines symétries (rotations, permutations, reparamétrisations...), invariances qui se reflètent directement sur la dimension intrinsèque des données. Ces invariances, la géométrie euclidienne ne peut pas les retranscrire fidèlement. Ainsi, on observe un intérêt croissant pour la modélisation des données par des structures plus fines telles que les variétés riemanniennes. Une deuxième approche en réduction de dimension consiste donc à généraliser les méthodes existantes à des données à valeurs dans des espaces non-euclidiens. On parle alors d'apprentissage géométrique. Jusqu'à présent, la plupart des travaux en apprentissage géométrique se sont focalisés sur l'analyse en composantes principales.

Dans la perspective de proposer une approche qui combine à la fois apprentissage géométrique et manifold learning, nous nous sommes intéressés à la méthode appelée locally linear embedding, qui a la particularité de reposer sur la notion de barycentre, notion a priori définie dans les espaces euclidiens mais qui se généralise aux variétés riemanniennes. C'est d'ailleurs sur cette même notion que repose une autre méthode appelée barycentric subspace analysis, et qui fait justement partie des méthodes qui généralisent l'analyse en composantes principales aux variétés riemanniennes. Ici, nous introduisons la notion nouvelle de plongement barycentrique, qui regroupe les deux méthodes. Essentiellement, cette notion englobe un ensemble de méthodes dont la structure rappelle celle des méthodes de réduction de dimension linéaires et non linéaires, mais où le modèle (localement) linéaire est remplacé par un modèle barycentrique – affine.

Le cœur de notre travail consiste en l'analyse de ces méthodes, tant sur le plan théorique que pratique. Du côté des applications, nous nous intéressons à deux exemples importants en apprentissage géométrique : les formes et les graphes. En particulier, on démontre que par rapport aux methodes standard de réduction de dimension en analyse statistique des graphes, les plongements barycentriques se distinguent par leur meilleure interprétabilité. En plus des questions pratiques liées

à l'implémentation, chacun de ces exemples soulève ses propres questions théoriques, principalement autour de la géométrie des espaces quotients. Parallèlement, nous nous attachons à caractériser géométriquement les plongements localement barycentriques, qui généralisent la projection calculée par locally linear embedding. Enfin, de nouveaux algorithmes d'apprentissage géométrique, novateurs dans leur approche, complètent ce travail.

Mots-clés : apprentissage géométrique, géométrie riemannienne et barycentrique, apprentissage de variétés, variétés quotient, espaces de formes de Kendall, analyse statistique de graphes.

Barycentric embeddings for geometric manifold learning

With application to shapes and graphs

Abstract

An MRI image has over 60,000 pixels. The largest known human protein consists of around 30,000 amino acids. We call such data high-dimensional. In practice, most high-dimensional data is high-dimensional only artificially. For example, of all the images that could be randomly generated by coloring 256 x 256 pixels, only a very small subset would resemble an MRI image of a human brain. This is known as the intrinsic dimension of such data. Therefore, learning high-dimensional data is often synonymous with dimensionality reduction. There are numerous methods for reducing the dimension of a dataset, the most recent of which can be classified according to two approaches.

A first approach known as manifold learning or non-linear dimensionality reduction is based on the observation that some of the physical laws behind the data we observe are non-linear. In this case, trying to explain the intrinsic dimension of a dataset with a linear model is sometimes unrealistic. Instead, manifold learning methods assume a locally linear model.

Moreover, with the emergence of statistical shape analysis, there has been a growing awareness that many types of data are naturally invariant to certain symmetries (rotations, reparametrizations, permutations...). Such properties are directly mirrored in the intrinsic dimension of such data. These invariances cannot be faithfully transcribed by Euclidean geometry. There is therefore a growing interest in modeling such data using finer structures such as Riemannian manifolds. A second recent approach to dimension reduction consists then in generalizing existing methods to non-Euclidean data. This is known as geometric learning.

In order to combine both geometric learning and manifold learning, we investigated the method called locally linear embedding, which has the specificity of being based on the notion of barycenter, a notion a priori defined in Euclidean spaces but which generalizes to Riemannian manifolds. In fact, the method called barycentric subspace analysis, which is one of those generalizing principal component analysis to Riemannian manifolds, is based on this notion as well. Here we rephrase both methods under the new notion of barycentric embeddings. Essentially, barycentric embeddings inherit the structure of most linear and non-linear dimension reduction methods, but rely on a (locally) barycentric – affine – model rather than a linear one.

The core of our work lies in the analysis of these methods, both on a theoretical and practical level. In particular, we address the application of barycentric embeddings to two important examples in geometric learning: shapes and graphs. In addition to practical implementation issues, each of these examples raises its own theoretical questions, mostly related to the geometry of quotient spaces. In particular, we highlight that compared to standard dimension reduction methods in graph analysis, barycentric embeddings stand out for their better interpretability. In parallel with these examples, we characterize the geometry of locally barycentric embeddings, which generalize the projection computed by locally linear embedding. Finally, algorithms for geometric manifold learning, novel in their approach, complete this work.

Keywords: geometric learning, Riemannian and barycentric geometry, manifold learning, quotient manifolds, Kendall shape spaces, statistical graph analysis.

Remerciements

Je ne vais pas faire dans l'originalité en remerciant tout d'abord mes deux directeurs de thèse, Xavier Pennec et Alain Trouvé, puisqu'après tout, ils ont chacun joué un rôle principal dans la-dite thèse. Merci à Xavier de m'avoir toujours consacré du temps, et de m'avoir offert autant de chances. À commencer par celle de travailler non pas seule, comme c'est si souvent le cas dans un domaine à cheval entre la géométrie et les statistiques, mais entourée d'une équipe digne de ce nom. Merci de m'avoir fait confiance, bien que jamais aveuglement. Merci pour sa bienveillance. Merci à Alain qui, pour reprendre ses propres mots, m'a vue grandir (scientifiquement) depuis mon premier jour à l'école, où l'on s'attachait à nous convaincre de la pertinence des mathématiques appliquées. Merci pour tout ce qu'il m'a appris, merci pour les longues discussions au tableau qui tantôt débordaient sur le déjeuner – jusqu'à ce que Fréderic Pascal vienne finalement nous interrompre – tantôt se prolongeaient jusqu'à la tombée de la nuit. Merci de toujours poser la dernière question. En bref, je souhaite remercier mes deux directeurs de thèse pour leur disponibilité et leurs enseignements, pour leur générosité. Merci de m'avoir donné le goût de la recherche.

Merci à Aasa Feragen et Ian Dryden pour leur précieux retours sur mon manuscrit qu'ils ont généreusement accepté de rapporter. Merci à eux ainsi qu'aux autres membres de mon jury, Barbara Tumpach, Jean Gallier et Stephan Huckemann, déjà d'avoir répondu présent à notre invitation, et bien sûr pour toutes les discussions intéressantes qui ont suivi. Aussi, merci à Stephan Huckemann d'avoir présidé – avec une rigueur imperturbable – ce jury, et d'avoir si méticuleusement annoté mon manuscrit.

Je prends également le temps d'un paragraphe pour remercier le petit mais riche écosystème de Math in the Black Forest, ou encore Math in the Mine, et dont les membres se reconnaîtront. En particulier, merci à Philipp Harms, qui a sans doute joué lui aussi un rôle déterminant dans mon goût pour la recherche, et à Peter Michor, pour ses réponses synthétiques mais toujours édifiantes. Pour la petite histoire, c'est à l'occasion de l'atelier en Forêt Noire auquel j'accompagnais Philipp, chez qui Alain d'ailleurs m'avait envoyée pour mon premier stage de master, que je rencontrai pour la première fois Xavier.

En guise de transition entre vie professionnelle et vie personnelle, je souhaite à présent remercier mes chers collègues de l'Inria. Ceux avec qui j'ai déjeuné de nombreuses fois au soleil à la petite terrasse du restaurant mexicain dans la vieille ville. Ceux que je croisais régulièrement en fin d'après-midi, à l'heure de ma pause, qui s'allongeait alors considérablement. Ceux qui sont partis vite mais souvent revenus donner quelques nouvelles. Lucia, avec qui j'ai passé la plupart de mes soirées à Nice cette dernière année. G-stats, la fine équipe. Je repense notamment à nos rendez-vous approximativement philosophiques au traiteur vietnamien d'Antibes. Et d'ailleurs, comment ne pas citer Anna, mon alliée précieuse aussi bien au labo qu'en dehors, que je remercie pour tout.

Merci à celles et ceux qui simplement ont été à mes côtés pendant ces trois ans, qui l'ont été avant, et le seront après, mes amies, mes amis. Merci à vous tous qui êtes venus jusqu'à Antibes pour assister à la soutenance, cela me touche énormément. Merci à ceux que j'avais perdu de vue et que j'ai retrouvés. Merci à mes colocataires du manoir de Cachan, que j'ai quitté à contre-coeur, j'ai nommé Coco, Pif, Gilou, Piops, Belette, Glier et Bastien. Merci pour cette première année haute en couleurs durant laquelle je ne me suis pas ennuyée un seul jour. Dans la lignée de cette belle brochette, merci à Charles, Alice, Pierrick, Mathis, Romane, Lucas, et d'autres... Merci au carré VIP de notre bar préféré, Lulu, Didou, Alex, Sylvain et bien sûr Ségo, avec qui j'ai planté tant d'arbres et fait germer tant d'idées. Et merci finalement à Claire et Sabrina, mes deux copines de toujours qui m'accompagnent depuis si longtemps déjà. Que le temps passe vite !

Pour clôturer ces remerciements, je m'adresse enfin à ma famille. Merci à mes parents, qui se sont toujours assurés que je ne manque de rien, surtout pas d'appétit. Le mot de la fin va naturellement à mes deux plus fidèles complices, ma petite soeur Marie et mon petit frère Jérémy. À nos sketchs surréalistes qui bien souvent animaient le dîner chaque fois que je rentrais à Paris. Merci de me faire rire, toujours.

Funding

This work was supported by the ERC grant No. 786854 G-Statistics from the European Research Council under the European Union's Horizon 2020 research and innovation program and by the French government through the 3IA Côte d'Azur Investments ANR-19-P3IA-0002 managed by the National Research Agency.

Licence

Contents

Chapter 1

Introduction

What does geometric manifold learning stand for? And what is hidden behind the notion of barycentric embedding? This chapter, modestly entitled Introduction, answers the first question, while the second will be addressed throughout this thesis. Geometric manifold learning is a composite word derived from geometric statistics and manifold learning. *Geometric statistics* entail theoretical foundations and methods for the statistical analysis of non-Euclidean data, also known as geometric data. In this thesis, among all the methods described as geometric statistics, we are more particularly interested in those for dimensionality reduction. In the last decades, the most popular approach to dimensionality reduction for geometric data, referred to as geometric learning, has been to generalize the linear model on which Principal Component Analysis is based to non-Euclidean structures. In the context of Euclidean data, however, it has been shown that linear methods sometimes struggle to capture the intrinsic dimension of a dataset. Instead, manifold learning methods rely on locally linear models. If we transpose this idea to the context of geometric data, then geometric manifold learning consists in reformulating the existing models in geometric learning locally.

This chapter is divided in three sections. The first two sections are respectively a non-exhaustive introduction to geometric statistics and a review of the state of the art in dimensionality reduction. In the last section, we outline this thesis and also summarize our different contributions chapter by chapter.

1.1 Motivations in geometric statistics

Geometric statistics is a field lying at the intersection of geometry and statistics, sometimes also referred to as non-Euclidean statistics. Geometric statistics emerged from the observation that standard statistics, that is linear statistics, fail to provide correct estimates for geometric data, such as anatomical shapes (Grenander and Miller, 1998), complex proteins (Dryden, Kim, Laughton, and Le, 2019), or electroencephalography signals (Calissano, Papadopoulo, Pennec, and Deslauriers-Gauthier, 2023). To replace the Euclidean model, a fairly popular framework to describe such non-linear data is that of Riemannian manifolds. In this context, there has been a new interest in statistics on manifolds, from the first definitions of the Riemannian mean (Fréchet, 1948; Karcher, 1977), to more recent results such as central limit theorems (Bhattacharya and Lin, 2017; Eltzner and Huckemann, 2019). Together with Riemannian geometry, we focus essentially in this section on two other specific topics in geometric statistics: the idea of invariance and statistical shape analysis. For a more exhaustive introduction to geometric statistics, we recommend the recent book edited by Pennec, Sommer, and Fletcher (2020) as well as that by Dryden and Marron (2021).

1.1.1 Invariances and symmetries

The shape of an object does not depend on how it is rotated in space. Similarly, reparametrizing a flat curve does not change its aspect. And the overall connectivity of a graph is not determined by how the nodes have been labeled. Many well-known examples of geometric data fall under the idea of invariance. To account for this phenomenon, several works in standard data analysis as well as in deep learning promote statistical methods based on invariant distances (Donnat and Holmes, 2018; Maron, Ben-Hamu, Shamir, and Lipman, 2018). In fact, the notion of invariance naturally relates to that of symmetries and group action. Indeed, equivalence classes under a group action describe perfectly the concept of a same object having multiple representations. In other words, data carrying invariance are naturally modeled in quotient spaces. For example, Calissano, Feragen, and Vantini (2023) discuss the statistical analysis of unlabeled graphs in the graph space

$$\mathbb{R}^{n\times n}/\mathfrak{S}(n)$$

resulting from the action of permutations on the nodes of a graph. Under some assumptions, it is possible to equip quotient spaces with what is called a quotient distance, defined as the minimum distance between two fibers (Younes, 2010). In this thesis, we are more particularly interested in equipping such quotient spaces with a Riemannian metric, in a similar approach to Kendall (1984). As we have illustrated, a typical example of the concept of invariance is that of the shapes. In this context, different representations of shapes as well as different metrics have been explored. Let us give here a general overview of existing approaches.

1.1.2 Statistical shape analysis

Shapes are a very well studied example of geometric data, if not the most studied. Shape analysis frameworks rely on a wide range of representations. For example, Kendall (1984) describes the shape of an object using a set of landmarks modulo a global rotation, translation and scaling of such. In his book, D'Arcy Thompson (1992) described the anatomical variability of animal species through the deformation of a template shape. More recently, S. M. Pizer et al. (2020) proposed to reconstruct a surface or a closed curve from its medial axis. Now, there are two main approaches to endowing shapes with a metric structure.

Procrustes Analysis In Procrustes analysis, the distance between two aligned objects – in rotation, translation, and eventually scaling – is measured based on a point correspondence using an extrinsic distance. This idea has been formalized in several fields such as biology, in the context of protein structural analysis (Kufareva and Abagyan, 2012). Kendall shape spaces fall for example within this approach. Srivastava and Klassen (2016) propose a similar model for the analysis of shapes represented as continuous curves under the action of reparametrizations.

Large Deformations Another approach consists in measuring the distance between two shapes as the amount of (diffeomorphic) deformation needed to transform one shape into another. In the Stationary Velocity Field framework, the deformations allowed consist in the one-parameter group of diffeomorphisms parametrized by a stationary velocity field such it has a fast implementation (Arsigny, Commowick, Pennec, and Ayache, 2006). As for the Large Deformation Diffeomorphic Metric Mapping framework, how much a deformation costs is measured through the norm of the (non stationary) vector field it integrates (Beg, Miller, Trouvé, and Younes, 2005). Within this framework, several representations of shapes have been explored such as currents (Charon, 2013).

For further reading, we refer to Younes (2010) regarding the detailed construction of shape spaces. In his thesis, Feydy (2020) provides a very well referenced overview of the different frameworks in shape analysis. Finally, the book by Dryden and Mardia (2016) covers application of shape analysis in various fields. In this context and more generally in that of geometric statistics, the notion of Riemannian manifold is central.

1.1.3 Riemannian manifolds

Riemannian geometry generalizes Euclidean geometry in that it consists in pairing a manifold with a metric, defined locally on the tangent space at every point of the manifold as an inner product

$$g_x : T_xM \times T_xM \to \mathbb{R}.$$

Such a metric provides a way of measuring on the manifold. In particular, a Riemannian metric defines a distance, which makes it possible to generalize basic concepts from Euclidean statistics, starting with the mean (Karcher, 1977). Beyond the distance, Riemannian geometry provides with a family of metric tools. Geodesics are curves of minimal length and generalize straight lines. Such a tool has played a key role in the development of geometric statistic methods. For example, geodesic models are a natural candidate for the generalization of linear models. In the same vein, geodesic regression (Fletcher, 2013) and Principal Geodesic Analysis (Fletcher, Lu, S. Pizer, and Joshi, 2004) or Geodesic Principal Component Analysis (Huckemann, Hotz, and Munk, 2010) are generalizations to Riemannian manifolds of linear regression and Principal Component Analysis respectively. Another tool, parallel transport, allows to compare information distributed in different tangent spaces, with application for example to longitudinal analysis (Lorenzi, Ayache, and Pennec, 2011).

Computational Riemannian geometry Computational Riemannian geometry encompasses automatic computations and implementations of Riemannian geometry tools, in particular integration schemes for geodesics and parallel transport (Guigui and Pennec, 2022). Bridging theoretical definition of such tools and methods in Euclidean statistics, computational geometry is currently a very active field. The Python library `geomstats` gathers implementations of geometric tools and statistical methods for various manifolds (Miolane, Guigui, et al., 2020). For a general understanding of Riemannian geometry from the computational point of view, we refer the reader to the book by Gallier and Quaintance (2020).

Optimization on Riemannian manifolds Supporting computational geometry, several works propose to generalize optimization methods to Riemannian manifolds, from gradient methods to higher order methods and constrained optimization (Absil, Mahony, and Sepulchre, 2007), together with specific libraries (Boumal, Mishra, Absil, and Sepulchre, 2014).

1.2 Dimensionality reduction

Dimensionality reduction is an old question, but because of the better resolution of data acquisitions and the increased performance of computers, it has become more and more an integral part of data analysis. In the specific context of geometric data analysis, however, dimensionality reduction has been tackled only recently, and is still an open question. We review the history of dimensionality reduction, through its approaches and methods, up to the point where this thesis begins.

1.2.1 The four quadrants of dimensionality reduction

Dimensionality reduction methods can be organized according to four different approaches: linear dimensionality reduction, to which we oppose nonlinear dimensionality reduction (or manifold learning), then moving towards geometric data analysis, geometric learning, generalizing linear dimensionality reduction to manifolds, and finally, at the crossroads of the two previous approaches, geometric manifold learning.

Linear dimensionality reduction Linear dimensionality reduction is one of the first approaches to dimensionality reduction for Euclidean data. It consists of all methods that implement a linear transformation of the data, such as a linear projection. Examples include Principal Component Analysis (Jolliffe, 2002) and Independent Component Analysis (Wang and Chang, 2006).

Nonlinear dimensionality reduction (manifold learning) In contrast, manifold learning consists of methods based on non-linear transformations of the data, again assumed to be Euclidean. It covers a wide range of methods, which we detail in the next section. The motivation for designing nonlinear methods is that some high-dimensional data, such as molecular complexes, are subject to constraints that are generally non-linear. More precisely, manifold learning methods implement locally linear transformations based on the assumption that high-dimensional data are sampled from a low-dimensional (sub)manifold. This assumption is known as the *manifold assumption*. Let us take advantage of this paragraph to make a point about vocabulary. The term *embedding* is frequently used in manifold literature. Our understanding is that it is a generic word to denote a nonlinear transformation of a dataset. In most manifold learning methods, the transformation itself is not known such that the embedding also refers to the image of the dataset by such transformation.

Non-Euclidean dimensionality reduction (geometric learning) Dimensionality reduction is a critical issue when it comes to geometric data analysis. Especially, the data modeled in the context of shape analysis – for example protein structures or anatomical shapes – are by nature high-dimensional data. Common tools for dimensionality reduction have been originally designed for Euclidean data. A first approach to process geometric data consists then in embedding them in a larger Euclidean space – or equivalently to work extrinsically. This approach has two main drawbacks. First of all, it ignores the structural information contained in the manifold model, which then may not be well recovered in areas of low sampling density. Moreover, there might be a significant gap in dimensionality between the intrinsic and the extrinsic model in some cases. Unparametrized curves illustrate well this second point as the extrinsic and the intrinsic descriptions differ by the removal of reparametrizations – diffeomorphims – which is an infinite dimensional space. Although manifold learning methods are designed to handle intrinsically non-linear data, in the case where the data is modeled on a known manifold, a more reasonable approach is to propose methods directly tailored to such a manifold. This approach is known as geometric learning. The main geometric learning methods are presented in Section 1.2.3. It is worth emphasizing that a similar approach has gained in popularity in deep learning with the emergence of the field known as geometric deep learning (Bronstein et al., 2017)

Geometric manifold learning So far, the idea behind most geometric learning methods is to generalize the linear transformations proposed in the first approach. Geometric manifold learning is part of geometric learning, but refers more specifically either to methods designed as a generalization to geometric data of manifold learning methods, or to methods reformulating locally the transformations of established geometric learning methods. To better understand the crossroads at which geometric

manifold learning finds itself, we review the main methods of its two adjacent approaches: manifold learning and geometric learning.

1.2.2 Main manifold learning methods

Here we provide a short description for the main methods in manifold learning. Note that all the methods are designed for Euclidean data.

Multidimensional Scaling (MDS) Multidimensional Scaling estimates an embedding of a dataset that preserves the original pairwise distances – or dissimilarities – between data points the best (M. A. A. Cox and T. F. Cox, 2008). In its classical formulation, its implementation relies essentially on an eigendecomposition.

Isometric Mapping (Isomap) Isometric Mapping (Tenenbaum, De Silva, and Langford, 2000) implements MDS but for another distance than the extrinsic Euclidean distance. Precisely, it relies on the distance defined for two data points as the length of the shortest path joining the two points in the neighbor graph of the data (for example the nearest neighbor graph). Such distance is expected to capture better the intrinsic distance of the data space.

Locally Linear Embedding (LLE) Locally Linear Embedding (S. T. Roweis and L. K. Saul, 2000) relies on the conservation of local barycentric relations between a data point and its nearest neighbor. Precisely, if a data point is written as a barycenter of its neighbors, then its embedding should be close to the barycenter of the embedded neighbors with the same barycentric coordinates. We review the method in more details later in the thesis.

Laplacian Eigenmaps Laplacian Eigenmaps (Belkin and Niyogi, 2003) are defined as the eigenvectors of the graph Laplacian matrix of the neighbor graph of the dataset. The first eigenmaps provide an embedding of the dataset in a low-dimensional space. The theoretical motivation for this method is the convergence of the graph Laplacian towards the Laplace-Beltrami operator of the submanifold the data points are sampled from. And the Laplace-Beltrami operator of a manifold, a second order differential operator, satisfies the nice property that its eigenvectors form a basis of the space of twice differentiable functions on this manifold.

Stochastic Neighbor Embedding (SNE) Stochastic Neighbor Embedding (Hinton and S. Roweis, 2002) estimates an embedding that preserves local distributions of data points. More precisely, if a data point has a certain probability of picking another point as its neighbor based on their respective distance, then such probability should still be accurate for the corresponding points of the embedding.

Uniform Manifold Approximation and Projection (UMAP) Finally, Uniform Manifold Approximation and Projection (McInnes, Healy, N. Saul, and Großberger, 2018) relies on a similar idea to that of Stochastic Neighbor Embedding but where distances have been locally scaled ahead of the computation of the probability matrix in such a way that the size of a neighborhood is uniform across the dataset.

1.2.3 Main geometric learning methods

Let us move now to geometric learning. As we highlighted in the introduction of this chapter, most geometric leaning methods focus on the generalization of Principal Component Analysis (PCA) to Riemannian manifolds. We identify five main generalizations of such.

Principal Geodesic Analysis (PGA) In Principal Geodesic Analysis, linear components are replaced with geodesic components, that is geodesics crossing at the Fréchet mean of the dataset. Precisely, the method estimates the first geodesic components that span a geodesic subspace onto which the projection of the dataset is of maximal variance.

Tangent Principal Component Analysis (tangent PCA) In their original paper, Fletcher, Lu, S. Pizer, and Joshi (2004) propose to approximate the projection problem as a projection onto the tangent subspace at the mean spanned by the tangent vectors encoding the first components. In this formulation, the method consists simply in a principal component analysis of the dataset, lifted in the tangent space at the mean through the exponential map. More recently, Sommer, Lauze, and Nielsen (2014) proposed an exact implementation of PGA. We refer then to the implementation by Fletcher, Lu, S. Pizer, and Joshi (2004) as tangent Principal Component Analysis.

Geodesic Principal Component Analysis (geodesic PCA) Geodesic Principal Component Analysis (Huckemann, Hotz, and Munk, 2010) is another generalization of PCA relying on the notion of geodesic component. The main difference with PGA is that geodesic Principal Component Analysis optimizes the components one by one, with the only condition that they should be orthogonal. Moreover, it does not enforce the first component to pass through the mean of the dataset.

Principal Nested Spheres (PNS) In their approach, Jung, Dryden, and Marron (2012) consider the case of a dataset lying on a high-dimensional sphere, e.g. a set of planar shapes, and propose to approximate such a dataset by a sequence of nested subspheres.

Barycentric Subspace Analysis (BSA) Barycentric Subspace Analysis (Pennec, 2018) relies on the notion of barycentric subspace rather than that of geodesic subspace. Essentially, barycentric subspaces are to affine subspaces what geodesic subspaces are to linear subspaces. We come back to this method in more details later in this thesis

1.3 General overview of the thesis

Finally, we close this introduction by unfolding the contents and the contributions of this thesis chapter by chapter.

1.3.1 Red thread

This thesis articulates mainly around the notion of barycentric embedding, introduced right from the first chapter. In their most general design, barycentric embeddings are a family of dimensionality reduction methods for manifold-valued data based on barycentric models, to be understood as affine models in the Euclidean case. In practice, barycentric embeddings seek a representation of a dataset that preserve relative positions of the data points, encoded by barycentric relations. In particular, the notion of barycentric embedding serves as a bridge between Barycentric Subspace Analysis (Pennec, 2018), a geometric learning method, and Locally Linear Embedding (S. T. Roweis and

L. K. Saul, 2000), a manifold learning method. In this thesis, we travel from theory to application, and from methods to algorithms. On one hand, we provide a general mathematical understanding of barycentric embeddings. On the other hand, in order to get a better idea of how the methods work in practice, we investigate two types of geometric data: shapes and graphs. On these examples, we illustrate a general computational approach to data analysis on quotient manifolds, and derive specific implementations. Alongside our theoretical and practical analysis, we highlight simple algorithms. At second reading, this thesis is also an introductory work on barycentric geometry in the context of dimensionality reduction.

1.3.2 Outline

Chapter 2. Learning with barycentric coordinates. In this chapter, we propose a very general introduction to barycentric geometry on Riemannian manifolds, essentially based on the work of Pennec (2018). In particular, we recall the definitions of a Riemannian barycenter and of a barycentric subspace. We formalize the notion of barycentric model, intimately related to that of barycentric subspace, in such a way that it encompasses the global model and the local model proposed by Barycentric Subspace Analysis and Locally Linear Embedding respectively. From the notion of barycentric model, we derive that of barycentric embedding. Moving towards the implementation of such methods, we provide algorithms for computing the projection of a point onto a barycentric subspace, both in the Euclidean case and in the Riemannian case, such projection being the key point in the estimation of barycentric models.

Chapter 3. Riemannian Locally Linear Embedding with design for shape spaces. This chapter is dedicated to the application of Locally Linear Embedding to Kendall shape spaces (Kendall, 1984). In particular, we propose a Riemannian formulation of the method and detail its implementation first on a general quotient manifold, and then in the specific case of Kendall shape spaces. The main technicality lies in the projection of a point onto a barycentric subspace, which requires a differentiable implementation of the parallel transport. We detail this point for Kendall shapes spaces following the work of Le (2003). We present simple experiments where we compare the performance of our method to that of standard Locally Linear Embedding. This chapter is also thought as a general example of what we are able to do in terms of computations and implementations in quotient manifolds.

Chapter 4. Barycentric Subspace Analysis of a set of graphs. This chapter covers a joint work with Anna Calissano. We investigate barycentric subspace analysis of a set of graphs, in the context of statistical graph analysis. Upstream, we introduce a new Riemannian framework for the analysis of unlabeled graphs represented by their spectrum. In this framework, we are able to describe barycentric subspaces explicitly such that the implementation of Barycentric Subspace Analysis is rather straightforward. We demonstrate the interpretability of such method on two datasets of graphs, including a real dataset from OpenFlight database (Open Flights, 2017), and compare it with another dimensionality reduction method (tangent PCA) as well as with a clustering method (hierarchical clustering). As an introduction to this chapter, we also share our initial approach to the problem of barycentric subspace analysis of a set of graphs.

Chapter 5. Locally barycentric embeddings, a gluing problem. In this chapter, we study theoretical aspects of locally barycentric embeddings, which in the Euclidean case consists in the solutions built by Locally Linear Embedding. More generally, locally barycentric embeddings rely on local barycentric models, where the local nature of such models depends for a given dataset on the

neighbor graph of the data points. We show that we can characterize certain locally barycentric embeddings as piecewise affine mappings obtained by gluing affine mappings together along the strongly connected components of such a graph. Then we derive from this results a series of lower bounds on the number of locally barycentric embeddings of a given dataset related to the connectivity of the neighbor graph of the data.

Chapter 6. New algorithms for geometric learning and geometric manifold learning. The last chapter concludes this thesis with new algorithms for geometric learning combining barycentric embeddings and isometric embeddings. Precisely, we propose to discriminate the solutions output by Barycentric Subspace Analysis and Locally Linear Embedding, a priori non unique, with a distance-based criterion. We review a first experiment on the Swiss Roll dataset supporting this combined approach.

1.3.3 Main contributions

Chapter 2 Strictly speaking, the new contributions of this chapter are rather limited. Essentially, the role of this chapter was to formalize properly all the notions for the rest of the thesis. We should just point out the algorithm for the projection on a barycentric subspace in the Riemannian case, which is a contribution borrowed to Chapter 3.

Chapter 3 The contributions of this chapter are mainly three. First, an implementation of Kendall shape spaces in the Python library `geomstats` as well as visualization modules, reviewed as part of the *ICLR Computational Geometry & Topology Challenge 2021*. Then, the automatic computation of parallel transport on Kendall shape spaces, presented at *GSI'21*. Finally, the generalization of Locally Linear Embedding to manifold-valued data with application to Kendall shape spaces, presented at *GSI'23*.

- E. Maignant, A. Trouvé, and X. Pennec (2023). "Riemannian locally linear embedding with application to Kendall shape spaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 12–20. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_2
- N. Guigui, E. Maignant, A. Trouvé, and X. Pennec (2021). "Parallel transport on Kendall shape spaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer International Publishing, pp. 103–110. ISBN: 978-3-030-80209-7. DOI: 10.1007/978-3-030-80209-7_12
- N. Miolane, M. Caorsi, et al. (2021). *ICLR 2021 Challenge for computational geometry & topology: Design and results*. arXiv: 2108.09810[cs,math]

Chapter 4 The first contribution of this chapter is a new Riemannian framework for the analysis unlabeled graphs represented by their spectrum. To our knowledge, although the space we construct has already been studied for its geometric and algebraic structure, it has not been introduced in the context of graph analysis before. In this framework, we provide an explicit description of the barycentric subspaces. Finally, the last contribution consists in the implementation and the experimentation of Barycentric Subspace Analysis in the context of graph analysis. The work presented in this chapter is to be submitted soon as a journal publication. A preliminary work has been presented already at *GSI'23*.

– A. Calissano, E. Maignant, and X. Pennec (2023). "Towards quotient barycentric subspaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 366–374. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_36

Chapter 5 This chapter entails a series of new results on locally barycentric embeddings. The two main results are the description of certain locally barycentric mappings as piecewise affine mappings, and a lower bound on the number of locally barycentric embeddings. In particular, these results allow to reinterpret Locally Linear Embedding and answer partially the question of the uniqueness of the solutions output by the method.

Chapter 6 Finally in the last chapter, the main contribution is a new method for geometric manifold learning combining Locally Linear Embedding and Isomap, relying on the theoretical results of the previous chapter. We also propose to extend Barycentric subspace analysis as a visualization method. The first algorithm together with the main results of the previous paragraph are to be submitted soon as a journal publication.

References

Absil, P.-A., R. Mahony, and R. Sepulchre (2007). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. 240 pp. ISBN: 978-0-691-13298-3.

Arsigny, V. et al. (2006). "A log-Euclidean framework for statistics on diffeomorphisms". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*. Ed. by R. Larsen, M. Nielsen, and J. Sporring. Lecture Notes in Computer Science. Springer, pp. 924–931. ISBN: 978-3-540-44708-5. DOI: 10.1007/11866565_113.

Beg, M. F. et al. (2005). "Computing large deformation metric mappings via geodesic flows of diffeomorphisms". In: *International Journal of Computer Vision* 61.2, pp. 139–157. ISSN: 1573-1405. DOI: 10.1023/B:VISI.0000043755.93987.aa.

Belkin, M. and P. Niyogi (2003). "Laplacian eigenmaps for dimensionality reduction and data representation". In: *Neural Computation* 15.6, pp. 1373–1396. ISSN: 0899-7667. DOI: 10.1162/089976603321780317.

Bhattacharya, R. and L. Lin (2017). "Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces". In: *Proceedings of the American Mathematical Society* 145.1, pp. 413–428. ISSN: 0002-9939, 1088-6826. DOI: 10.1090/proc/13216.

Boumal, N. et al. (2014). "Manopt, a Matlab toolbox for optimization on manifolds". In: *Journal of Machine Learning Research* 15.42, pp. 1455–1459. URL: https://www.manopt.org.

Bronstein, M. M. et al. (2017). "Geometric deep learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4, pp. 18–42. ISSN: 1558-0792. DOI: 10.1109/MSP.2017.2693418.

Calissano, A., A. Feragen, and S. Vantini (2023). "Populations of unlabelled networks: Graph space geometry and generalized geodesic principal components". In: *Biometrika*, asad024. ISSN: 1464-3510. DOI: 10.1093/biomet/asad024.

Calissano, A., E. Maignant, and X. Pennec (2023). "Towards quotient barycentric subspaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 366–374. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_36.

Calissano, A. et al. (2023). "Graph alignment exploiting the spatial organisation improves the similarity of brain networks". In: *Human Brain Mapping*. DOI: 10.1002/hbm.26554.

Charon, N. (2013). "Analysis of geometric and functional shapes with extensions of currents : Applications to registration and atlas estimation". Thèse de doctorat. École Normale Supérieure de Cachan.

Cox, M. A. A. and T. F. Cox (2008). "Multidimensional scaling". In: *Handbook of Data Visualization*. Ed. by C.-H. Chen, W. Härdle, and A. Unwin. Springer Handbooks Comp.Statistics. Springer, pp. 315–347. ISBN: 978-3-540-33037-0. DOI: 10.1007/978-3-540-33037-0_14.

D'Arcy Thompson, W. (1992). *On Growth and Form*. Ed. by J. T. Bonner. Canto. Cambridge University Press. DOI: 10.1017/CBO9781107325852.

Donnat, C. and S. Holmes (2018). "Tracking network dynamics: A survey using graph distances". In: *The Annals of Applied Statistics* 12.2, pp. 971–1012. ISSN: 1932-6157, 1941-7330. DOI: 10.1214/18-AOAS1176.

Dryden, I. L. and K. V. Mardia (2016). *Statistical Shape Analysis, with Applications in R*. John Wiley & Sons, Ltd. ISBN: 978-1-119-07249-2. DOI: 10.1002/9781119072492.

Dryden, I. L. and J. S. Marron (2021). *Object Oriented Data Analysis*. Chapman and Hall/CRC. 436 pp. ISBN: 978-1-351-18967-5. DOI: 10.1201/9781351189675.

Dryden, I. L. et al. (2019). "Principal nested shape space analysis of molecular dynamics data". In: *The Annals of Applied Statistics* 13.4, pp. 2213–2234. ISSN: 1932-6157, 1941-7330. DOI: 10.1214/19-AOAS1277.

Eltzner, B and S. Huckemann (2019). "A smeary central limit theorem for manifolds with application to high-dimensional spheres". In: *The Annals of Statistics* 47.6, pp. 3360–3381. ISSN: 00905364, 21688966.

Feydy, J. (2020). "Analyse de données géométriques, au delà des convolutions". These de doctorat. université Paris-Saclay.

Fletcher, T. (2013). "Geodesic regression and the theory of least squares on Riemannian manifolds". In: *International Journal of Computer Vision* 105.2, pp. 171–185. ISSN: 1573-1405. DOI: 10.1007/s11263-012-0591-y.

Fletcher, T. et al. (2004). "Principal geodesic analysis for the study of nonlinear statistics of shape". In: *IEEE Transactions on Medical Imaging* 23.8, pp. 995–1005. ISSN: 1558-254X. DOI: 10.1109/TMI.2004.831793.

Fréchet, M. (1948). "Les éléments aléatoires de nature quelconque dans un espace distancié". In: *Annales de l'institut Henri Poincaré* 10.4, pp. 215–310. ISSN: 0365-320X.

Gallier, J. and J. Quaintance (2020). *Differential Geometry and Lie groups: A Computational Perspective*. Vol. 12. Geometry and Computing. Springer International Publishing. ISBN: 978-3-030-46039-6, 978-3-030-46040-2. DOI: 10.1007/978-3-030-46040-2.

Grenander, U. and M. I. Miller (1998). "Computational anatomy: An emerging discipline". In: *Quarterly of Applied Mathematics* 56.4, pp. 617–694. ISSN: 0033-569X.

Guigui, N. and X. Pennec (2022). "Numerical accuracy of ladder schemes for parallel transport on manifolds". In: *Foundations of Computational Mathematics* 22.3, pp. 757–790. ISSN: 1615-3383. DOI: 10.1007/s10208-021-09515-x.

Guigui, N. et al. (2021). "Parallel transport on Kendall shape spaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer International Publishing, pp. 103–110. ISBN: 978-3-030-80209-7. DOI: 10.1007/978-3-030-80209-7_12.

Hinton, G. E. and S. Roweis (2002). "Stochastic neighbor embedding". In: *Advances in Neural Information Processing Systems* 15.

Huckemann, S., T. Hotz, and A. Munk (2010). "Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions". In: *Statistica Sinica* 20.1, pp. 1–58. ISSN: 1017-0405.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag. ISBN: 978-0-387-95442-4. DOI: 10.1007/b98835.

Jung, S, I. L. Dryden, and J. S. Marron (July 2012). "Analysis of principal nested spheres". In: *Biometrika* 99.3, pp. 551–568. ISSN: 0006-3444. DOI: 10.1093/biomet/ass022.

Karcher, H. (1977). "Riemannian center of mass and mollifier smoothing". In: *Communications on Pure and Applied Mathematics* 30.5, pp. 509–541. ISSN: 1097-0312. DOI: 10.1002/cpa.3160300502.

Kendall, D. G. (1984). "Shape manifolds, Procrustean etrics, and complex projective spaces". In: *Bulletin of the London Mathematical Society* 16.2, pp. 81–121. ISSN: 1469-2120. DOI: 10.1112/blms/16.2.81.

Kufareva, I. and R. Abagyan (2012). "Methods of protein structure comparison". In: *Homology Modeling: Methods and Protocols*. Ed. by A. J. W. Orry and R. Abagyan. Methods in Molecular Biology. Humana Press, pp. 231–257. ISBN: 978-1-61779-588-6. DOI: 10.1007/978-1-61779-588-6_10.

Le, H. (2003). "Unrolling shape curves". In: *Journal of the London Mathematical Society* 68.2, pp. 511–526. ISSN: 1469-7750, 0024-6107. DOI: 10.1112/S0024610703004393.

Lorenzi, M., N. Ayache, and X. Pennec (2011). "Schild's ladder for the parallel transport of deformations in time series of images". In: *Information Processing in Medical Imaging*. Ed. by G. Székely and H. K. Hahn. Lecture Notes in Computer Science. Springer, pp. 463–474. ISBN: 978-3-642-22092-0. DOI: 10.1007/978-3-642-22092-0_38.

Maignant, E., A. Trouvé, and X. Pennec (2023). "Riemannian locally linear embedding with application to Kendall shape spaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and

F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 12–20. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_2.

Maron, H. et al. (2018). "Invariant and equivariant graph networks". In: *International Conference on Learning Representations*.

McInnes, L. et al. (2018). "UMAP: Uniform manifold approximation and projection". In: *Journal of Open Source Software* 3.29, p. 861. ISSN: 2475-9066. DOI: 10.21105/joss.00861.

Miolane, N., M. Caorsi, et al. (2021). *ICLR 2021 Challenge for computational geometry & topology: Design and results*. arXiv: 2108.09810[cs,math].

Miolane, N., N. Guigui, et al. (2020). "Geomstats: A Python package for Riemannian geometry in machine learning". In: *Journal of Machine Learning Research* 21.223, pp. 1–9. ISSN: 1533-7928.

Open Flights (2017). *Open Flight Database*. Accessed: 2023-05-16.

Pennec, X. (2018). "Barycentric subspace analysis on manifolds". In: *The Annals of Statistics* 46.6, pp. 2711–2746. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/17-AOS1636.

Pennec, X., S. Sommer, and T. Fletcher, eds. (2020). *Riemannian geometric statistics in medical image analysis*. Academic Press. ISBN: 978-0-12-814725-2. DOI: 10.1016/C2017-0-01561-6.

Pizer, S. M. et al. (2020). "6 - Object shape representation via skeletal models (s-reps) and statistical analysis". In: *Riemannian Geometric Statistics in Medical Image Analysis*. Ed. by X. Pennec, S. Sommer, and T. Fletcher. Academic Press, pp. 233–271. ISBN: 978-0-12-814725-2. DOI: 10.1016/B978-0-12-814725-2.00014-5.

Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding". In: *Science* 290.5500, pp. 2323–2326. DOI: 10.1126/science.290.5500.2323.

Sommer, S., F. Lauze, and M. Nielsen (2014). "Optimization over geodesics for exact principal geodesic analysis". In: *Advances in Computational Mathematics* 40.2, pp. 283–313. ISSN: 1572-9044. DOI: 10.1007/s10444-013-9308-1.

Srivastava, A. and E. P. Klassen (2016). *Functional and Shape Data Analysis*. Springer Series in Statistics. Springer. ISBN: 978-1-4939-4018-9, 978-1-4939-4020-2. DOI: 10.1007/978-1-4939-4020-2.

Tenenbaum, J. B., V. De Silva, and J. C. Langford (2000). "A Global geometric framework for nonlinear dimensionality reduction". In: *Science* 290.5500, pp. 2319–2323. DOI: 10.1126/science.290.5500.2319.

Wang, J. and C.-I. Chang (2006). "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis". In: *IEEE Transactions on Geoscience and Remote Sensing* 44.6, pp. 1586–1600. ISSN: 1558-0644. DOI: 10.1109/TGRS.2005.863297.

Younes, L. (2010). *Shapes and Diffeomorphisms*. Vol. 171. Applied Mathematical Sciences. Springer. ISBN: 978-3-642-12054-1, 978-3-642-12055-8. DOI: 10.1007/978-3-642-12055-8.

Chapter 2

Learning with barycentric coordinates

Back to the early stages of dimensionality reduction, at the very beginning of the chronology we outlined in Chapter 1. Assuming Euclidean data, the most straightforward approach to reducing the dimension of a dataset is to estimate a low-dimensional linear model – that is a linear subspace – which fits to the data points (Jolliffe, 2002). In the same vein, we discuss in this chapter the more general notion of barycentric subspace (Pennec, 2018) and we describe some other approach to dimensionality reduction. In particular, we investigate learning methods for non-Euclidean data.

The first section covers basic concepts of barycentric geometry on Riemannian manifolds. It mainly defines the notions of barycenter, barycentric coordinates and barycentric subspace. In the second section, we derive step by step from the notion of barycentric subspace a general approach to dimensionality reduction which we refer to as barycentric embeddings. In particular, we detail how to compute the projection onto a barycentric subspace in practice. In the third and last section, we review two methods in the literature which implement barycentric models, Barycentric Subspace Analysis (Pennec, 2018) and Locally Linear Embedding (Roweis and Saul, 2000). The first one is a dimensionality reduction method which generalizes Principal Component Analysis to manifolds. The second one is a manifold learning method based on local barycentric models.

2.1 Starter pack for barycentric geometry on manifolds

Here barycentric geometry is to be understood as the geometry of barycentric subspaces, which were first introduced by Pennec (2018). Since we do not have yet all the notions required to define properly what is a barycentric subspace, let us start by giving some intuition on Euclidean spaces. There are two different approaches to defining subspaces of a Euclidean space. In the standard approach, a k-dimensional subspace is determined by one point, the origin, and a family of k independent vectors, called a frame. Each point of the subspace is then referenced by coordinates called Cartesian coordinates. A slightly different approach consists instead in fixing $k+1$ (affinely) independent points $a_0, \ldots, a_k$. There is then a unique subspace of dimension k which contains the points $a_0, \ldots, a_k$ and it is the set of all the weighted barycenters

$$\mathrm{bar}(a_i, w_i)_{0 \leq i \leq k} = \sum_{i=0}^{k} w_i a_i \tag{2.1}$$

for weights $w_0, \ldots, w_k$ that sum to 1. Such a subspace is usually referred to as the affine span of the points, themselves called an affine basis or barycentric frame. For each point x of the subspace, written as $x = \mathrm{bar}(a_i, w_i)_{0 \leq i \leq k}$, the weights $w_0, \ldots, w_k$ are called the barycentric coordinates of x. Unlike

Cartesian coordinates, barycentric coordinates characterize properties which do not depend explicitly on distances and angles such as point alignments. Although these two constructions generate the same subspaces in a Euclidean space, they generalize differently on a generic Riemannian manifold. Barycentric subspaces extend the second construction to non-Euclidean spaces. Let us then move forward with a first definition.

2.1.1 Barycenters

There is a more canonical approach to defining barycenters than writing them in terms of weighted arithmetic means, as in the equation above. Consider again $k+1$ points $a_0, \dots, a_k$ in $\mathbb{R}^n$ and let $w_0, \dots w_k$ be $k+1$ real numbers summing to 1. Then the barycenter $\mathrm{bar}(a_i, w_i)_{0 \leq i \leq k}$ of the points $a_0, \dots, a_k$ with corresponding weights $w_0, \dots w_k$ is the point in $\mathbb{R}^n$ which minimizes the weighted sum of squared errors

$$\mathrm{bar}(a_i, w_i)_{0 \leq i \leq k} = \operatorname*{arg\,min}_{x \in \mathbb{R}^n} \sum_{i=0}^{k} w_i \|a_i - x\|^2. \tag{2.2}$$

Weighted Fréchet means This definition generalizes naturally to points $a_0, \dots, a_k$ in any metric space (M, d) and corresponds in fact to the notion of weighted Fréchet means. If the points $a_0, \dots, a_k$ are assigned weights $w_0, \dots w_k$, then the weighted Fréchet means are exactly those points m on M that minimize the weighted sum of squared distances

$$m \in \operatorname*{arg\,min}_{x \in M} \sum_{i=0}^{k} w_i d(a_i, x)^2. \tag{2.3}$$

Exponential barycenters Now this optimization problem does not necessarily have an explicit solution when M is not a Euclidean space and therefore the definition is not straightforward to implement. Instead, we consider the notion of exponential barycenter on Riemannian manifolds. The notion was first introduced by Emery and Mokobodzki (1991) and Corcuera and Kendall (1999) for continuous weights measures, and then reformulated more recently for discrete weights by Pennec and Arsigny (2013) as follows

Definition 2.1.1. *Let (M, g) be a Riemannian manifold. Let $a_0, \dots, a_k \in M$ and let $w_0, \dots w_k \in \mathbb{R}$ sum to 1. Then $x \in M$ is an exponential barycenter of $a_0, \dots, a_k$ with corresponding weights $w_0, \dots w_k$ if it belongs to the set*

$$\mathrm{bar}(a_i, w_i)_{0 \leq i \leq k} = \left\{ x \in M^*(a_0, \dots, a_k) \mid \sum_{i=0}^{k} w_i \log_x a_i = 0 \right\} \tag{2.4}$$

where $\log_x$ denotes the Riemannian logarithm – the inverse map of the exponential map – of M at x and $M^*(a_0, \dots, a_k)$ is such that for all i, there is a unique minimizing geodesic connecting any $x \in M^*(a_0, \dots, a_k)$ to a_i (see Pennec, 2018).

Weighted Fréchet means are exponential barycenters This definition is a relaxed version of the Fréchet definition. In fact, it corresponds exactly to the first order condition (critical condition) for a point to be a solution of the minimization Problem (2.3). In the case where M is a Euclidean space, this condition is equivalent to Equation (2.2) since the function is convex and one can check that it defines a unique solution given by Equation (2.1). In the general case, Fréchet weighted means are exponential barycenters but the reverse is not necessarily true. Moreover, there is a priori no

unique solution when M is not a Euclidean space for either exponential barycenters or Fréchet means. Note that exponential barycenters are also well defined on affine connection spaces or geodesic spaces where the log map is still defined. In what follows, we only consider exponential barycenters of a Riemannian manifold and we simply refer to it as a barycenter without ambiguity.

Affine mappings and barycenters Let us comment on the relationship between barycenters and affine mappings. It is well known that affine mappings between two affine spaces are exactly those maps f which preserve barycenters, that is such that

$$f\big(\operatorname{bar}(a_i, w_i)_{0\leq i\leq k}\big) \in \operatorname{bar}(f(a_i), w_i)_{0\leq i\leq k}. \tag{2.5}$$

More generally, Kobayashi and Nomizu (1996, Chapter 6) define affine mappings between two Riemannian manifolds to be differentiable maps that preserve geodesics. Deriving the observation that such mappings commute with the exponential map, we can still show that these also preserve barycenters. However, the reverse does not hold anymore a priori.

Computing barycenters Finally, coming back to the question of implementation, let us comment on the computation in practice of a barycenter. We already mentioned that Fréchet weighted means are not comfortable to manipulate as it involves solving an optimization problem. Instead, we have defined barycenters in Definition 2.1.1 as the solution of a critical equation. This implicit formulation is as we see in Section 2.2 easy to implement within larger problems like the projection problem. Now if we wish to compute explicitly the barycenter of some points with given weights, then we might implement Newton-like methods (Absil, Mahony, and Sepulchre, 2007) to solve Equation (2.4).

2.1.2 Barycentric subspaces and barycentric coordinates

Affinely independent points Before moving to the definition of barycentric subspaces, let us extend the notion of affinely independent points to manifolds. In a Euclidean space, some points $a_0, \ldots, a_k$ are said to be affinely independent if and even if the vectors $a_1 - a_0, \ldots, a_k - a_0$ are linearly independent, or equivalently if this is the case for any of the sets $(a_j - a_i)_{0\leq j\leq k,\, j\neq i}$. On manifolds, Pennec (2018) defines the following

Definition 2.1.2. *Let (M, g) be a Riemannian manifold of dimension n. Then the points $a_0, \ldots, a_k \in M$, with $k \leq n$, are said to be affinely independent if and only if none of them is in the cut-locus of one other and if for all $0 \leq i \leq k$, the subset $\big(\log_{a_i}(a_j)\big)_{0\leq j\leq k,\, j\neq i}$ of the tangent space of M at a_i is linearly independent.*

Exponential barycentric subspaces Now we are interested in generalizing the notions of affine span and barycentric frame. It is only natural then to consider the set of all barycenters of a set of points. Implementing this idea, exponential barycentric subspaces – derived from the notion of exponential barycenter – were introduced on Riemannian manifolds by Pennec (2018) as

Definition 2.1.3. *Let (M, g) be a Riemannian manifold. Let $a_0, \ldots, a_k \in M$ be affinely independent points. The exponential barycentric subspace of $a_0, \ldots, a_k$ is the set*

$$\operatorname{EBS}(a_0, \ldots, a_k) = \bigcup_{\substack{w_0,\ldots,w_k\in\mathbb{R} \\ w_0+\ldots+w_k=1}} \operatorname{bar}(a_i, w_i)_{0\leq i\leq k} \tag{2.6}$$

of all barycenters of $a_0, \ldots, a_k$. The points $a_0, \ldots, a_k$ are then referred to as the reference points of such subspace.

About other definitions of barycentric subspaces There are as many ways of defining a barycentric subspace as they are of defining barycenters. Pennec also introduces for example in his paper the notion of Fréchet barycentric subspaces. In this thesis, we focus only on the previous definition. We then simply refer to exponential barycentric subspaces as barycentric subspaces and denote the barycentric subspace of $a_0, \dots, a_k$ by $\mathrm{BS}(a_0, \dots, a_k)$.

Dimension of a barycentric subspace The barycentric subspace of $k+1$ affinely independent points is a stratified space of dimension k, provided we removed some degenerate points (Pennec, 2018). The reader would easily check than when k reaches the dimension n of M, it then coincides with the set $M^*(a_0, \dots, a_k)$. In the case where M is a Euclidean space, the barycentric subspace of $k+1$ affinely independent points is an affine subspace of dimension k and it corresponds exactly to the affine span of the points. In general, barycentric subspaces are not necessarily connected nor complete. In order to ensure such properties, Pennec further consider the closure of a barycentric subspace in the manifold M which he also refers to as an affine span. Let us give two other examples taken from Pennec's paper.

Example 2.1.1 (Barycentric subspaces of the sphere)**.** *The affine span of $k+1$ affinely independent points on the n-dimensional sphere S^n is the great subsphere of dimension k which contains all the reference points. Their barycentric subspace is the same except that it does not contain the cut locus of the reference points.*

Example 2.1.2 (Barycentric subspaces of the hyperbolic space)**.** *The barycentric subspace of $k+1$ affinely independent points on the n-dimensional hyperbolic space H^n is the hyperbolic subspace of dimension k which contains all the reference points.*

A link with geodesic subspaces In the Euclidean case as well as in the two previous constant curvature examples, affine spans coincide with geodesic subspaces as introduced in Principal Geodesic Analysis (Fletcher, Lu, Pizer, and Joshi, 2004). In the general case, one can check that 2-dimensional affine spans coincide locally with geodesics. The comparison does not hold anymore however for higher dimensional subspaces. Investigating further the connection between the two constructions, Pennec (2018, p. 24) also demonstrates that the affine span of a set of points converges towards some restricted geodesic subspace when those reference points collapse.

Barycentric coordinates Finally a barycentric subspace is naturally equipped with a frame which consists simply of the reference points. Such a frame defines a coordinate system and any point of a barycentric subspace admit barycentric coordinates defined as

Definition 2.1.4. *Let (M, g) be a Riemannian manifold. Let $a_0, \dots, a_k \in M$ be affinely independent points. Then a point $x \in \mathrm{BS}(a_0, \dots, a_k)$ has barycentric coordinates $w_1, \dots, w_k \in \mathbb{R}$ with respect to the reference points $a_0, \dots, a_k$ if the weights $w_0, \dots, w_k$ satisfy*

$$\sum_{i=0}^{k} w_i \log_x a_i = 0 \tag{2.7}$$

Additionally, we ask for barycentric coordinates to be normalized, that is to sum to 1*.*

Remark 2.1.1. *The normalization condition is necessary for the barycentric coordinates of a point to be unique but it is however not sufficient. Whether such coordinates are unique also depends on the dimension of the barycentric subspace at the point (Pennec, 2018, p. 22). On a computational aspect,*

the barycentric coordinates of a point x are the solutions $w = (w_0, \dots, w_k)$ of the linear equation $\varphi(x)w = 0$ up to normalization, where $\varphi(x)$ is the matrix whose i-th column is the vector $\log_x a_i$ written in local coordinates. The previous remark states that the matrix $\varphi(x)$ might have a rank less than k and therefore the dimension of the solution set might be greater than 1.

2.2 Barycentric projection and barycentric embeddings

Having barycentric subspaces now in hand, let us go back to the problem of dimensionality reduction. Following the approach discussed at the very beginning of this chapter, a reasonable method for computing a k-dimensional embedding $y_1, \dots, y_N$ of a dataset $x_1, \dots, x_N$ consists in estimating a barycentric subspace $\mathrm{BS}(a_0, \dots, a_k)$ which fits to the data points and take their projection onto the subspace, or more precisely onto its closure

$$y_i \in \mathrm{proj}(x_i) \subset \overline{\mathrm{BS}}(a_0, \dots, a_k). \tag{2.8}$$

Theoretically, this method is well defined on any manifold M assuming that $a_0, \dots, a_k$ are such that the dataset lies in $\in M^*(a_0, \dots, a_k)$. In practice however, the computation of the projection may be tricky when the manifold is not a Euclidean space. Moreover, since barycenters are defined by an implicit equation, it is more natural to compute the barycentric coordinates of the projected points than the points themselves. Recovering the embedding from the barycentric coordinates is therefore a problem in its own that we examine together with the projection problem in what follows.

2.2.1 General design of barycentric embeddings

Let us fix the setting. As input, we have a dataset $x_1, \dots, x_N$ on a n-dimensional Riemannian manifold (M, g). As output, we want an embedding $y_1, \dots, y_N$ of the dataset in a k-dimensional space with $k \leq n$. The method we describe to build this embedding consists mainly of two steps. First, we estimate a barycentric subspace that fits the data points and compute their projection onto the subspace. In a second step, we recover the embedding.

Barycentric projection The projections of a point $x \in M$ onto the barycentric subspace of affinely independent points $a_0, \dots, a_k \in M$ are the points of the subspace which minimize the Riemannian distance to x

$$\mathrm{proj}_{\mathrm{BS}(a_0,\dots,a_k)}(x) = \operatorname*{arg\,min}_{y \in \overline{\mathrm{BS}}(a_0,\dots,a_k)} d(x, y)^2. \tag{2.9}$$

sometimes abbreviated to $\mathrm{proj}(x)$ when there is no ambiguity on $a_0, \dots, a_k$. When M is a general manifold, the projection is a priori not unique. It is unique almost surely however when M is a Euclidean space, a sphere or a hyperbolic space (Pennec, 2018). Such question also relates closely to that of identifying the focal points (Huckemann, Hotz, and Munk, 2010, Appendix A) of the barycentric subspace. Now points of the barycentric subspace are exactly barycenters of the reference points. Therefore, the previous projection problem can also be formulated as an optimization problem on the barycentric coordinates of the projection

$$\begin{aligned} &\underset{w_0,\dots,w_k \in \mathbb{R}}{\text{minimize}} \\ &w_0+\dots+w_k=1 \end{aligned} \quad d\big(x, \mathrm{bar}(a_i, w_i)_{0 \leq i \leq k}\big)^2. \tag{2.10}$$

The question of uniqueness here is even more complex as given weights can define several barycenters as we saw in the previous section. The distance above is then more precisely the distance from x to

the set $\mathrm{bar}(w_i, a_i)_{0\leq i\leq k}$. In practice, we simply take for the projection the output of the optimization algorithm, which should be close to one of the minima. Let us leave the implementation question for later and move on to the first step of the method.

Step 1. Barycentric model This first step mainly consists in estimating k-dimensional barycentric subspace $\mathrm{BS}(a_0, \ldots, a_k)$ which fits to the data points and compute their projection onto the subspace. Ideally, the subspace should fit as best as possible to the point, that is minimize the projection error for each data point

$$\underset{a_0,\ldots,a_k\in M}{\text{minimize}} \quad \sum_{i=1}^{N} d\big(x_i, \mathrm{proj}_{\mathrm{BS}(a_0,\ldots,a_k)}(x_i)\big)^2. \tag{2.11}$$

This estimation relies in fact itself on the computation of the projection such that both the subspace and the projection have to be optimized simultaneously. Now we saw that the projection problem has two possible formulations. If we substitute the second formulation in the previous problem for the projection, we obtain the following joint optimization problem

$$\underset{\substack{a_0,\ldots,a_k\in M\\ w_{ij}\in\mathbb{R}\\ w_{i0}+\ldots+w_{ik}=1}}{\text{minimize}} \quad \sum_{i=1}^{N} d(x_i, \mathrm{bar}(a_j, w_{ij})_{0\leq j\leq k})^2 \tag{2.12}$$

The projection of each data point on the optimal subspace is then parameterized by the optimal weights. Another way to see our problem is model fitting. Let $x_1, \ldots, x_N$ be N observations on M. Then the barycentric model of parameters $a_0, \ldots, a_k$ (the reference points) and latent variables w_{ij} (the weights) estimates x_i with $\mathrm{bar}(a_j, w_{ij})_{0\leq j\leq k}$. When M is the Euclidean space $\mathbb{R}^n$, such model has a nice and simple formulation. The relation between the observations and the latent weights is

$$x_i = \sum_{j=0}^{k} w_{ij} a_j + \varepsilon_i. \tag{2.13}$$

where ε_i is the error of the model at x_i.

Local barycentric models Now let us switch from dimensionality reduction to manifold learning, that is from a global approach to a local approach. We recall that manifold learning methods usually assume that the dataset is sampled from some submanifold of a linear space and therefore rely on local linear fits to construct an embedding. Essentially, the mathematical rationale behind this approach is that normal coordinates define a diffeomorphism between the submanifold and an open set of a linear subspace. Dyer, Vegter, and Wintraecken (2016) proved a similar result for barycentric coordinates. Following this result, we may as well estimate a collection of N local barycentric models parametrized by local frames of reference

$$\underset{\substack{a_0^{(i)},\ldots,a_{k_i}^{(i)}\in M\\ w_{j0}^{(i)},\ldots,w_{jk_i}^{(i)}\in\mathbb{R}\\ w_{j0}^{(i)}+\ldots+w_{jk_i}^{(i)}=1}}{\text{minimize}} \quad \sum_{j\in N(i)} d(x_j, \mathrm{bar}(a_l^{(i)}, w_{jl}^{(i)})_{0\leq l\leq k_i})^2 \qquad \text{for } 1\leq i\leq N. \tag{2.14}$$

where $j \in N(i)$ if and only if x_j is in the neighborhood of x_i, fixed prior to the model estimation, and where k_i is the dimension of the model at x_i and may differ from one local model to another so that the whole model can fit to a dataset across which the intrinsic dimension is not uniform. The

main issue in this local construction is that we need to know how to transition from one frame to another. Locally Linear Embedding (Roweis and Saul, 2000) gets round this problem however by choosing very specific local frames as we see in the last section.

What happens if we stop at this step Once the optimal reference points $a_0, \dots, a_k$ and weights w_{ij} have been computed, recovering the projection $\mathrm{bar}(a_i, w_{ij})_{0 \leq j \leq k}$ requires to solve the critical equation (2.4), except when the manifold M is a Euclidean space. This can be achieved using Newton's method for manifolds (Absil, Mahony, and Sepulchre, 2007). However, assuming the points have been successfully projected and embedded in the barycentric subspace of $a_0, \dots, a_k$, it may not be obvious to interpret the subspace as a k-dimensional space, nor to perform intrinsic computations therein.

Step 2. Barycentric mapping Instead, we propose to build an embedding of the data points in a k-dimensional Riemannian manifold M' with known metric g'

$$y_i = f(x_i) \in M' \tag{2.15}$$

For this purpose, instead of computing the embedding map f explicitly, we are going to build each of the points y_i. Let us detail the construction in the case where M is the Euclidean space $\mathbb{R}^n$. The barycentric subspace of $a_0, \dots, a_k$ is then simply the affine span $\mathrm{aff}(a_0, \dots, a_k)$. Therefore, we can consider affine mappings

$$\begin{aligned} \phi : \mathrm{aff}(a_0, \dots, a_k) &\longrightarrow \mathbb{R}^k \\ \mathrm{proj}(x_i) &\longmapsto y_i. \end{aligned} \tag{2.16}$$

But because affine mappings are fully determined by their image in a barycentric frame, it is equivalent to fix the image b_j of each of the reference points and map the projection $\mathrm{proj}(x_i)$ onto the following barycenter in $\mathbb{R}^k$

$$\phi\Big(\sum_{j=0}^{k} w_{ij} a_i\Big) = \sum_{j=0}^{k} w_{ij} b_j. \tag{2.17}$$

To ensure that our embedding is k-dimensional, we need to fix the embedded reference points b_j in such a way that they are affinely independent. In the context of surface triangulation, such mapping is also referred to as a barycentric mapping (Floater, 2015). In this case, our embedding map is then exactly $f = \phi \circ \mathrm{proj}$. Now technically, we can generalize this construction to any Riemannian data space M and any Riemannian embedding space M', simply by taking

$$y_i \in \mathrm{bar}(b_j, w_{ij})_{0 \leq j \leq k} \tag{2.18}$$

where $b_0, \dots, b_k$ have been fixed in M'. In practice, since the computation of a barycenter involves an iterative scheme as we saw before, we choose M' to be relatively simple, that is either a Euclidean space or a constant curvature space (spherical or hyperbolic), where almost everything is known in closed form.

Counting barycentric mappings Assume a barycentric model. We saw already that in the case where both the data space M and the embedding space M' are Euclidean, barycentric mappings identify with affine mappings. In the general case, we saw that affine mappings between M and M' preserve barycenters and therefore they define barycentric mappings. However, there is a priori no reason for all barycentric mappings to be derived from an affine mapping. If the embedding space is Euclidean, since there are as many choices of embeddings $y_1, \dots, y_N$ as there are of embedded

reference points $b_0, \dots, b_k$, then the set of barycentric embeddings identifies with $\mathbb{R}^{k(k+1)}$. Now assume M' is a k-dimensional non-Euclidean constant curvature space. First notice that because it preserves barycenters, any affine transformation of M' – that is an isomorphic affine mapping of M' onto himself – of a barycentric embedding is still a barycentric embedding. Moreover, according to Kobayashi's theorem (Kobayashi, 1955), affine transformations of M' are exactly isometries of M'. Finally, since the group of isometries of a k-dimensional constant curvature space is of dimension $\frac{1}{2}k(k+1)$ (see Kobayashi (1995) for example), then there exist at least as much embeddings $y_1, \dots, y_N$.

Towards locally barycentric embeddings Finally, following the same idea, we can reconstruct an embedding from local barycentric models. Locally Linear Embedding (Roweis and Saul, 2000) proposes such a construction which we detail in the last section of this chapter and refer to as locally barycentric embedding. How much locally barycentric embeddings one is able to construct from given local barycentric models and how it relates to affine mappings is addressed to some extent in Chapter 5.

2.2.2 Computing the projection onto a barycentric subspace

Here we detail how to solve the barycentric projection problem in practice, both in the Euclidean case and in the general Riemannian case. Let again (M, g) be a Riemannian manifold. Let then $a_0, \dots a_k$ be affinely independent points on M and let x another point on M. We want to solve the following problem

$$\underset{\substack{w_0,\dots,w_k\in\mathbb{R}\\ w_0+\dots+w_k=1}}{\text{minimize}} \quad d(x, \text{bar}(a_i, w_i)_{0\leq i\leq k})^2. \tag{2.19}$$

where d denotes the Riemannian distance on M. In the Euclidean case, the projection problem can be solved explicitly but this is not the case in general. Therefore, we discuss the two cases separately.

Euclidean barycentric projection When M is a Euclidean space, the projection problem is simply a constrained least squares problem and its solution can be computed with the method of Lagrange multipliers as in Locally Linear Embedding (Saul and Roweis, 2003). Let us detail the method. Let $a_0, \dots a_k \in \mathbb{R}^n$ be affinely independent points and let $x \in \mathbb{R}^n$. The barycentric projection problem is the following

$$\underset{\substack{w_0,\dots,w_k\in\mathbb{R}\\ w_0+\dots+w_k=1}}{\text{minimize}} \quad \left\| x - \sum_{i=1}^{k} w_i a_i \right\|^2 \tag{2.20}$$

If $w = (w_0, \dots, w_k)$, then we can rewrite this problem as

$$\underset{\substack{w\in\mathbb{R}^{k+1}\\ \mathbf{1}^T w=1}}{\text{minimize}} \quad w^T G w \tag{2.21}$$

where G is a positive semi-definite matrix defined by $G = \varphi(x)^T\varphi(x)$ and $\varphi(x)$ the matrix given by

$$\varphi(x) = \begin{bmatrix} a_{0,0} - x_0 & \dots & a_{k,0} - x_0 \\ a_{0,1} - x_1 & \dots & a_{k,1} - x_1 \\ \vdots & & \vdots \\ a_{0,n} - x_n & \dots & a_{k,n} - x_n \end{bmatrix}. \tag{2.22}$$

Now let us write the Lagrangian function of the problem

$$\mathcal{L}(w, \lambda) = w^T G w + \lambda\big(\mathbf{1}^T w - 1\big). \tag{2.23}$$

An optimal couple (w^*, λ^*) is then a solution of

$$\begin{cases} \partial_w \mathcal{L}(w^*, \lambda^*) = 0 \\ \partial_\lambda \mathcal{L}(w^*, \lambda^*) = 0 \end{cases} \tag{2.24}$$

that is

$$\begin{cases} 2Gw^* = -\lambda^* \mathbf{1} \\ \mathbf{1}^T w^* = 1 \end{cases} \tag{2.25}$$

Since we assumed that the points $a_0, \dots, a_k$ are affinely independent, then this linear system has a unique solution. In practice however, this assumption might not always be satisfied, whether we optimize the reference points or fix them before solving the projection problem.

Riemannian barycentric projection When M is a general Riemannian manifold, the formulation of Problem (2.19) is not as straightforward as barycenters can not be written in closed-form. Instead, we introduce an auxiliary variable $z \in M$ satisfying Equation (2.6) and reformulate Problem (2.19) as a constrained optimization problem on manifolds. More explicitly, for affinely independent point $a_0, \dots a_k \in M$ and for $x \in M$, we solve

$$\begin{aligned} \underset{\substack{z \in M \\ w_0, \dots, w_0 \in \mathbb{R} \\ w_0 + \dots + w_k = 1}}{\text{minimize}} \quad & d(x, z)^2 \\ \text{subject to} \quad & \sum_{i=0}^{k} w_i \log_z(a_i) = 0. \end{aligned} \tag{2.26}$$

It is not clear however how to solve this problem in practice. Especially because the constraint lies in the tangent space $T_z M$ which depends itself on the value of z which we wish to optimize. Rather, we propose to look at the equivalent translated problem

$$\begin{aligned} \underset{\substack{z \in M \\ w_0, \dots, w_0 \in \mathbb{R} \\ w_0 + \dots + w_k = 1}}{\text{minimize}} \quad & d(x, z)^2 \\ \text{subject to} \quad & \sum_{i=0}^{k} w_i P_{z,x}\left(\log_z(a_i)\right) = 0. \end{aligned} \tag{2.27}$$

where $P_{z,x}$ denotes the parallel transport map of M along the geodesic joining z and x. Since the parallel transport is an isometric map, Problems (2.26) and (2.27) are equivalent. In this new formulation however, the constraint lies in the tangent space of M at x, which is independent of the optimization state. In its current formulation, Problem (2.27) can be solved using Lagrangian methods for constrained optimization on manifolds using specific optimization tools like the ones implemented in the `Manopt` toolbox (Boumal, Mishra, Absil, and Sepulchre, 2014). However, it can be also be formulated alternatively as a vector-valued optimization problem in the tangent space of M at x.

Tangent barycentric projection Precisely, we keep track of the estimate z with the tangent vector $v \in T_x M$ such that

$$\exp_x(v) = z \tag{2.28}$$

where $\exp_x$ denotes the exponential map of M at x. Additionally, we set

$$u_i = P_{z,x}\left(\log_z(a_i)\right). \tag{2.29}$$

We derive the following optimization problem

$$\begin{aligned} \underset{\substack{v \in T_x M \\ w_0,\ldots,w_k \in \mathbb{R} \\ w_0+\ldots+w_k=1 \\ u_0,\ldots,u_k \in T_x M}}{\text{minimize}} \quad & g_x(v,v) \\ \text{subject to} \quad & \sum_{i=0}^{k} w_i u_i = 0 \\ & \exp_{[\exp_x(v)]}\left(P_{x,[\exp_x(v)]}(u_i)\right) = a_i \quad (\forall i). \end{aligned} \tag{2.30}$$

where g_x denotes the inner product of the tangent space at x defined by the Riemannian metric g on M. Problem (2.30) is a priori a vector-valued optimization problem on the product space $(T_x M)^{k+2} \times \mathbb{R}^{k+1}$. Now, provided that a basis of $T_x M$ can been explicitly computed, then the search space is the Euclidean space $\mathbb{R}^{kn+2n+k+1}$, where n is the dimension of M, and the optimization task is performed using standard Lagrangian methods implemented in most libraries. As a reference, we use the `SLSQP` solver from `scipy`. We detail this point more concretely on the example of Kendall shape spaces in Chapter 3. Note that in practice, the complexity of the algorithm strongly depends on whether one knows the exponential map and the parallel transport in closed-form. In any case, it requires an implementation of both methods which is compatible with automatic differentiation so that we can provide the chosen algorithm with the gradient of the constraints.

2.3 Barycentric embeddings in the literature

Finally, we propose in this section to review two major methods in the literature that we classify as barycentric embeddings. The first method, which we have already mentioned several times in this chapter, is Barycentric subspace Analysis (Pennec, 2018). It is precisely as part of this method that the notion of barycentric subspace has been formalized for the first time and used for dimensionality reduction. The second method, Locally Linear Embedding (Roweis and Saul, 2000), was introduced before Barycentric Subspace Analysis but can now be reinterpreted in terms of local barycentric models. In fact, our journey started with generalizing Locally Linear Embedding to manifolds, but taking interest in Barycentric Subspace Analysis, we spotted common patterns. We then thought about a general formalism that would encompass both methods, and that is exactly how this chapter came about.

2.3.1 Barycentric Subspace Analysis

Barycentric Subspace Analysis (BSA) is a dimensionality reduction method for manifold-valued data introduced by Pennec (2018) in an attempt to generalize Principal Component Analysis (PCA) to manifolds. At that time, PCA already had several generalizations on manifolds, all relying on the notion of geodesic components. We could cite for example PGA (Fletcher, Lu, Pizer, and Joshi, 2004) or geodesic PCA (Huckemann, Hotz, and Munk, 2010). The idea behind BSA was then to generalize

the notion of subspace rather than that of component. Earlier in this chapter, we discussed why the notion of barycentric subspaces was a good one for this purpose. Now in its simplest form, BSA introduces and estimates the global barycentric model (2.12) we described in the previous section. That said, one of the objectives of BSA was also to generalize PCA as a nested sequence of subspaces, that is an ordered sequence for inclusion. In his paper, Pennec (2018) proposes several ways of doing so.

Nested barycentric models Given a dataset $x_1, \dots, x_N$ on a Riemannian manifold (M, g), the idea is to build a sequence of nested barycentric subspaces

$$a_0 = \mathrm{BS}(a_0) \subset \dots \subset \mathrm{BS}(a_0, \dots, a_k) \dots \subset \mathrm{BS}(a_0, \dots, a_n) = M \tag{2.31}$$

that fit to the data points x_i in such a way that the approximation improves as much as possible as the sequence progresses. A natural criterion for assessing the accuracy of the approximation is the unexplained variance or equivalently the projection error

$$\sigma^2(a_0, \dots, a_k) = \sum_{i=1}^{N} d\big(x_i, \mathrm{proj}_{\mathrm{BS}(a_0,\dots,a_k)}(x_i)\big)^2. \tag{2.32}$$

The author details then three methods to build a sequence of nested barycentric subspaces that optimizes this criterion.

Forward Barycentric Subspace Analysis The first method consists in a forward construction, starting with a unique reference point a_0. One can check that such point that optimizes the criterion above is exactly the Fréchet mean. Then at step k, the subspace $\mathrm{BS}(a_0, \dots, a_{k-1})$ has been constructed already and we then look for the new reference point a_k that minimizes the criterion $\sigma(a_0, \dots, a_k)$ before adding $\mathrm{BS}(a_0, \dots, a_k)$ to the sequence. In practice the search stops when the intrinsic dimension of the data has been reached (there are several algorithms to know when this happens). By construction, the sequence has to contain the Fréchet mean, which is the main limit of this method. In the Euclidean case, Pennec (2018) showed that the sequence build by forward BSA is the same as the one computed by PCA.

Backward Barycentric Subspace Analysis The second method consists on the other hand in a backward analysis. It starts by estimating a k-dimensional barycentric model of parameters $a_0, \dots, a_k$. Then at each step, the point a_i that increases the criterion the least is removed. At the end of the iteration, the reference points are reordered in such a way that the the criterion (2.32) decreases when the sequence progresses. While backward BSA is in its construction less constrained than forward BSA, it is obviously quite more expensive, especially for a large k. Moreover the estimation of a k-dimension model is not a well-parameterized problem neither is it a well-conditioned problem as explained by Pennec (2018). Note that in the case where $k = 1$, the geodesic model optimized by backward BSA coincides with the first component of geodesic PCA.

Accumulated Unexplained Variances Finally both forward BSA and backward BSA provide a sequence that is sub-optimal in the sense that for a given dimension k, the k-th barycentric subspace does not minimize the projection error a priori. To answer this problem, the author introduces the accumulated unexplained variance

$$\mathrm{AUV}(a_0, \dots, a_k) = \sum_{i=0}^{k} \sigma^2(a_0, \dots, a_i) \tag{2.33}$$

which proposes to optimize in one step on the whole sequence. We refer to the paper for more details on this method. All the three methods require optimization on manifolds (or on flags of manifolds). Again, optimization on manifolds may be achieved using the `Manopt` toolbox (Boumal, Mishra, Absil, and Sepulchre, 2014), while optimization on flags of manifolds might be implemented following the work of Huckemann and Eltzner (2018). Note that in its current formulation, BSA consists mainly in the model step of barycentric embeddings and does not implement a mapping step. Let us now detail some last variant of BSA which is of specific interest to us.

Sample-limited Barycentric Subspace Analysis In this formulation of BSA, the reference points are constrained to belong to the dataset

$$\underset{\substack{1\leq i_0<\ldots<i_k\leq N \\ w_{ij}\in\mathbb{R} \\ w_{i0}+\ldots+w_{ik}=1}}{\text{minimize}} \quad \sum_{i=1}^{N} d(x_i, \text{bar}(x_{i_j}, w_{ij})_{0\leq j\leq k})^2. \tag{2.34}$$

By construction, the barycentric model is then straightforward to interpret with respect to the data. In a way, sample-limited BSA is mix of the k-means method (Lloyd, 1982) and the k-medoids method (Schubert and Rousseeuw, 2021), as it computes quadratic estimators, but drawn from within the data. Note that this formulation is also compatible with the nested construction of all three previous methods.

Convex sample-limited Barycentric Subspace Analysis We may also enforce the reference points to be extreme points by constraining the weights to be positive when computing the projection

$$\underset{\substack{1\leq i_0<\ldots<i_k\leq N \\ w_{ij}\in\mathbb{R}^+ \\ w_{i0}+\ldots+w_{ik}=1}}{\text{minimize}} \quad \sum_{i=1}^{N} d(x_i, \text{bar}(x_{i_j}, w_{ij})_{0\leq j\leq k})^2. \tag{2.35}$$

We refer to this method as convex sample-limited BSA. In this formulation, BSA allows to perform archetypal analysis (Cutler and Breiman, 1994). The main disadvantage of sample-limited BSA is that it is a combinatorial optimization problem. Precisely, the computational complexity of a k-dimensional model is proportional to $(k+1)!$. In Chapter 4, we illustrate both sample-limited BSA and convex sample-limited BSA in the context of graph analysis, where interpretability is a particularly critical issue. Such intrinsic methods would also be very interesting to investigate for shape analysis.

2.3.2 Locally Linear Embedding

Locally Linear embedding (LLE) is a manifold learning method introduced by Roweis and Saul (2000) that computes local coordinate charts based on linear weights reconstructing each data point as a weighted sum of its neighbors (that is, as their barycenter). The method is based on the assumption that the dataset lies on a submanifold of $\mathbb{R}^n$, and that therefore there should exist a global chart that consists of aligning local chart with linear maps (or rather affine mappings, in fact). Now, since barycenters are invariant with respect to affine mappings, LLE retrieves the global chart as the embedding that preserves the weights it computed in the first step. Here we propose to review the method from the point of view of barycentric embeddings. Additionally, we raise the question of the uniqueness of the [locally linear] embedding. This question is discussed in Chapter 5, where in particular we investigate the link between such embedding and local affine mappings.

Step 1. Local barycentric model Let us consider again some dataset $x_1, \dots, x_N$, this time in $\mathbb{R}^n$. The first step of LLE can be thought of as estimating local barycentric models, but where the parameters (reference points) of the i-th model have been fixed prior to estimation and are the data points that are the closest to x_i

$$a_0^{(i)}, \dots, a_{k_i}^{(i)} = x_{j_0}, \dots, x_{j_{k_i}}, \tag{2.36}$$

and only the latent weights remain to be optimized

$$\underset{\substack{w_{ij_0},\dots,w_{ij_{k_i}} \in \mathbb{R} \\ w_{ij_0}+\dots+w_{ij_{k_i}}=1}}{\text{minimize}} \quad \left\| x_i - \sum_{j \in N(i)} w_{ij} x_j \right\|^2 \qquad \text{for } 1 \leq i \leq N \tag{2.37}$$

where $N(i) = \{j_0, \dots, j_{k_i}\}$ stores the indices for the neighbors of x_i. There are several ways to select such neighbors. The two most common ways are either to look for the K nearest neighbors of x_i and fix $k_i = K - 1$, or to select all the data points within a ball of fixed radius r around x_i. In the original algorithm, such search is performed using the Euclidean distance in $\mathbb{R}^n$. In ISOLLE (Varini, Degenhard, and Nattkemper, 2006), the neighbors are selected according to the geodesic distance defined in ISOMAP (Tenenbaum, De Silva, and Langford, 2000). Now Problem (2.37) amounts exactly to computing a Euclidean projection. In the case where the reference neighbors are not (or almost not) affinely independent, then Roweis and Saul (2000) and later Modified LLE (Zhang and Wang, 2006) propose solutions to get around uniqueness issues based on a penalized version of the projection problem. Finally, if the data points are not too much spread in $\mathbb{R}^n$, then the i-th model estimated by LLE should be close to the optimal local barycentric model solving (2.14), where the local reference points are not fixed in advance.

Step 2. Locally barycentric mapping The second step of LLE consists then in building the embedding $y_1, \dots, y_n \in \mathbb{R}^k$ in such a way that locally, it coincides with the barycentric mapping

$$y_i = \sum_{j \in N(i)} w_{ij} y_j. \tag{2.38}$$

Whether this problem has an exact solution is not straightforward. Instead, the method proposes to find the best non-trivial solution solving

$$\underset{y_1,\dots,y_N \in \mathbb{R}^k}{\text{minimize}} \quad \sum_{i=1}^{N} \left\| y_i - \sum_{j \in N(i)} w_{ij} y_j \right\|^2. \tag{2.39}$$

We refer to such solution as a locally barycentric mapping. According to Saul and Roweis (2003), solving this problem is equivalent to computing the lowest eigenvalues of the matrix $(Id-w)^T(Id-w)$ following the Rayleitz-Ritz theorem. We detail here this equivalence and derive conditions for the optimal solution to be unique.

Optimal mapping(s) Let y be the $k \times N$ matrix whose i-th row is the vector $y_i = \begin{bmatrix} y_{i1} \dots y_{ik} \end{bmatrix}$ and let w be a $N \times N$ matrix with entries $w_{ij} = 0$ whenever x_j is not a neighbor of x_i (which includes x_i itself, that is $w_{ii} = 0$). Then Problem (2.39) can be posed slightly differently as

$$\begin{aligned} \underset{y_1,\dots,y_N \in \mathbb{R}^k}{\text{minimize}} \quad & \operatorname{tr}(y^T(Id-w)^T(Id-w)y) \\ \text{subject to} \quad & \sum_{i=1}^{N} y_i = 0 \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N} y_i y_i^T = I_k. \end{aligned} \tag{2.40}$$

where, in order for the problem to be well-posed, we constraint the embedding to be centered and of unit covariance matrix. The matrix $(Id - w)^T(Id - w)$ is a symmetric positive semi-definite matrix. Let $\lambda_0 \leq \ldots \leq \lambda_{N-1}$ be its eigenvalues with multiplicity arranged in non-decreasing order. Note that since the weights $w_{i1}, \ldots, w_{iN}$ sum to 1, we have $(Id - w)^T(Id - w)\mathbb{1} = 0$. Let us fix then $\lambda_0 = 0$. Because of the first constraint, then for all $1 \leq j \leq k$, the vector $\begin{bmatrix} y_{1j} \ldots y_{Nj} \end{bmatrix}$ has to be in the orthogonal of $\mathbb{1}$. It follows from the Courant-Fisher theorem (Horn and Johnson, 2012) that

$$\begin{aligned} \operatorname{tr}(y(Id - w)^T(Id - w)y^T) &= \sum_{j=1}^{k} \begin{bmatrix} y_{1j} \ldots y_{Nj} \end{bmatrix}^T (Id - w)^T(Id - w) \begin{bmatrix} y_{1j} \ldots y_{Nj} \end{bmatrix} \\ &\geq N \sum_{j=1}^{k} \lambda_j \end{aligned} \tag{2.41}$$

Let us assume that y satisfies the second constraint of Problem (2.40]. Then the columns of $y/\sqrt{N}$ form an orthonormal basis. As a consequence of the Rayleigh theorem (Horn and Johnson, 2012), the inequality (2.41) becomes an equality if and only if, up to some permutation of the eigenvalues, the unit vector $1/\sqrt{N} \begin{bmatrix} y_{1j} \ldots y_{Nj} \end{bmatrix}$ is an eigenvector of $(Id - w)^T(Id - w)$ for the eigenvalue λ_j. Finally, provided that $\lambda_{k+1} > \lambda_k$, the optimal embedding is unique up to a global orthogonal transformation of the embedding space. How this condition relates to the configuration of the points and the choice of neighbors is however a rather difficult question. In chapter 5, we study such relationship in the case where each of the local models is perfect, that is when each data point is a barycenter of its neighbor, and estimate the number of embeddings, which we refer to as locally barycentric embeddings.

Riemannian Locally Linear Embedding Finally, using the concepts and tools of the two first sections, there is not much missing to generalize Locally Linear Embedding to manifold-valued data. Essentially, the first step of the method, which consist in estimating the latent weights, can be performed the same way but using the Riemannian projection. This is exactly the idea implemented by Riemannian Locally Linear Embedding (Maignant, Trouvé, and Pennec, 2023) which we detail in the case of Kendall shape spaces in the next chapter.

References

Absil, P.-A., R. Mahony, and R. Sepulchre (2007). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. 240 pp. ISBN: 978-0-691-13298-3.

Boumal, N. et al. (2014). "Manopt, a Matlab toolbox for optimization on manifolds". In: *Journal of Machine Learning Research* 15.42, pp. 1455–1459. URL: `https://www.manopt.org`.

Corcuera, J. M. and W. S. Kendall (1999). "Riemannian barycentres and geodesic convexity". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 127.2, pp. 253–269. ISSN: 1469-8064, 0305-0041. DOI: `10.1017/S0305004199003643`.

Cutler, A. and L. Breiman (1994). "Archetypal analysis". In: *Technometrics* 36.4, pp. 338–347. ISSN: 0040-1706. DOI: `10.2307/1269949`.

Dyer, R., G. Vegter, and M. Wintraecken (2016). *Barycentric coordinate neighbourhoods in Riemannian manifolds*. DOI: `10.48550/arXiv.1606.01585`.

Emery, M. and G. Mokobodzki (1991). "Sur le barycentre d'une probabilité dans une variété". In: *Séminaire de Probabilités XXV*. Ed. by J. Azéma, M. Yor, and P. A. Meyer. Lecture Notes in Mathematics. Springer, pp. 220–233. ISBN: 978-3-540-38496-0. DOI: 10.1007/BFb0100858.

Fletcher, T. et al. (2004). "Principal geodesic analysis for the study of nonlinear statistics of shape". In: *IEEE Transactions on Medical Imaging* 23.8, pp. 995–1005. ISSN: 1558-254X. DOI: 10.1109/TMI.2004.831793.

Floater, M. S. (2015). "Generalized barycentric coordinates and applications". In: *Acta Numerica* 24, pp. 161–214. ISSN: 0962-4929, 1474-0508. DOI: 10.1017/S0962492914000129.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis*. 2nd ed. Cambridge University Press. 643 pp. ISBN: 978-0-521-83940-2.

Huckemann, S. and B. Eltzner (2018). "Backward nested descriptors asymptotics with inference on stem cell differentiation". In: *The Annals of Statistics* 46.5, pp. 1994–2019. DOI: 10.1214/17-AOS1609. URL: https://doi.org/10.1214/17-AOS1609.

Huckemann, S., T. Hotz, and A. Munk (2010). "Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions". In: *Statistica Sinica* 20.1, pp. 1–58. ISSN: 1017-0405.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag. ISBN: 978-0-387-95442-4. DOI: 10.1007/b98835.

Kobayashi, S. (1955). "A theorem on the affine transformation group of a Riemannian manifold". In: *Nagoya Mathematical Journal* 9 (none), pp. 39–41. ISSN: 0027-7630.

Kobayashi, S. (1995). *Transformation Groups in Differential Geometry*. Classics in Mathematics. Springer. ISBN: 978-3-642-61981-6. DOI: 10.1007/978-3-642-61981-6_2.

Kobayashi, S. and K. Nomizu (1996). *Foundations of Differential Geometry*. Wiley classics library ed. Wiley classics library. Wiley. 2 pp. ISBN: 978-0-471-15733-5, 978-0-471-15732-8.

Lloyd, S. (1982). "Least squares quantization in PCM". In: *IEEE Transactions on Information Theory* 28.2, pp. 129–137. ISSN: 1557-9654. DOI: 10.1109/TIT.1982.1056489.

Maignant, E., A. Trouvé, and X. Pennec (2023). "Riemannian locally linear embedding with application to Kendall shape spaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 12–20. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_2.

Pennec, X. (2018). "Barycentric subspace analysis on manifolds". In: *The Annals of Statistics* 46.6, pp. 2711–2746. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/17-AOS1636.

Pennec, X. and V. Arsigny (2013). "Exponential barycenters of the canonical cartan connection and invariant means on Lie groups". In: *Matrix Information Geometry*. Ed. by F. Nielsen and R. Bhatia. Springer, pp. 123–166. ISBN: 978-3-642-30232-9. DOI: 10.1007/978-3-642-30232-9_7.

Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding". In: *Science* 290.5500, pp. 2323–2326. DOI: 10.1126/science.290.5500.2323.

Saul, L. K. and S. T. Roweis (2003). "Think globally, fit locally: Unsupervised learning of low dimensional manifolds". In: *Journal of Machine Learning Research* 4 (Jun), pp. 119–155. ISSN: ISSN 1533-7928.

Schubert, E. and P. J. Rousseeuw (2021). "Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms". In: *Information Systems* 101, p. 101804. ISSN: 0306-4379. DOI: 10.1016/j.is.2021.101804.

Tenenbaum, J. B., V. De Silva, and J. C. Langford (2000). "A Global geometric framework for nonlinear dimensionality reduction". In: *Science* 290.5500, pp. 2319–2323. DOI: 10.1126/science.290.5500.2319.

Varini, C., A. Degenhard, and T. W. Nattkemper (2006). "ISOLLE: LLE with geodesic distance". In: *Neurocomputing*. Blind Source Separation and Independent Component Analysis 69.13, pp. 1768–1771. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2005.12.120.

Zhang, Z. and J. Wang (2006). "MLLE: Modified locally linear embedding using multiple weights". In: *Advances in Neural Information Processing Systems*. Vol. 19. MIT Press.

Chapter 3

Riemannian Locally Linear Embedding with design for shape spaces

"Shape is all the geometrical information that remains when location, scale and rotational effects are removed from an object."

– Dryden and Mardia (2016)

Shapes may be the best-known example of geometric data. Since the shape of an object is invariant to certain symmetries of that object, it cannot be faithfully described in a linear space. Statistical shape analysis has had an important impact in several fields such as medical image analysis or computational anatomy (Dryden and Mardia, 2016). The shape of some complex objects can only be accurately described in large dimension. Synthesizing and visualizing shape analysis is all the more critical when it comes to exchanging with other fields. After discussing with biologists, we realized that, although Principal Component Analysis is well established among bio-statisticians, manifold learning methods are beginning to gain in popularity. Hence, there is a real need to provide learning methods that take into account the formalization and modeling work carried out in the context of shape analysis. We first came across Kendall shape spaces (Kendall, 1977) while investigating some models to analyze protein configurations. Kendall shape spaces have been used for multiple applications, from archaeology to epidemiology (Nava-Yazdani, Hege, Sullivan, and Von Tycowicz, 2020). In fact, we are not the first either to take an interest in dimensionality reduction in such spaces (Huckemann, Hotz, and Munk, 2010). Now due to its invariant nature, the shape of an object is often modeled by an equivalence class, so that most shape analysis frameworks, in particular Kendall shape spaces, are quotient spaces. In this thesis, we focused exclusively on Kendall shape spaces among other shape spaces, as this is a well-studied and relatively simple example, but nevertheless rich in terms of structure. However, we provide keys for computing and learning in quotient spaces beyond the case of Kendall shape spaces, even in contexts other than shape analysis, such as graph analysis.

In the first section, we describe through the example of Kendall Shape Spaces a general procedure for computing on quotient manifold. In particular, we detail how to derive from O'Neill theorems on Riemannian submersions a new and simpler formulation of parallel transport in Kendall shape spaces. The second section focuses on the Riemannian Locally Linear Embedding method (Maignant, Trouvé, and Pennec, 2023). We propose a version of the projection algorithm in quotient manifolds and provide all the tools needed to implement it on Kendall shape spaces. Finally in the last section, we illustrate the method on two simple examples in Kendall shape spaces and analyze its computational

complexity. Perspectives for our method and future work are also discussed, in terms of both theory and application.

3.1 Computing on quotient manifolds

The case considered here is that described by Le and Kendall (1993) in the context of Kendall shape spaces, where a smooth Lie group G acts on a differentiable manifold M

$$g, x \mapsto g \cdot x. \tag{3.1}$$

The quotient manifold theorem states that if the action of G is both proper and free, then the quotient space M/G is a differentiable manifold (Lee, 2012). Now let us assume that M is equipped with a Riemannian metric g that is invariant to the action of G. Then such action induces a Riemannian metric on M/G, in the same way that a G-invariant distance on M induces, under certain conditions, a quotient distance on M/G (Younes, 2010). Let us take a closer look.

3.1.1 Riemannian submersions

In this section, we assume a smooth, proper and free action of a Lie group G on a Riemannian manifold (M, g) such that the metric g is G-invariant. A natural way of equipping the quotient manifold M/G with a Riemannian metric is to make the quotient map $\pi : M \to M/G$ a Riemannian submersion. Such construction is the one chosen by Kendall and proves rather well suited to computations and implementation, but it is neither the only nor the most general way of building a quotient metric, as explained by Tumpach and Preston (2023). Roughly speaking, a Riemannian submersion is a submersion between two Riemannian manifolds that put in correspondence the tangent bundle of the *base manifold* with a subbundle of the tangent bundle of the *top manifold* in an isometric way. Let us start by characterizing such a subbundle.

Vertical and horizontal bundles In finite dimension, the rank-nullity theorem suggests that a good candidate would be the orthogonal complement of the kernel of the tangent map $d\pi$. The vertical space of M at a point x is precisely defined as the kernel of $d_x\pi$. If the tangent space of M at x admits an orthogonal decomposition into the vertical space of M at x and its complement, then the latter is referred to as the horizontal space of M at x. Let us write proper definitions.

Definition 3.1.1 (Vertical space)**.** *Let $x \in M$. The vertical space of M at x is*

$$\mathrm{Ver}_x M = \ker d_x\pi \tag{3.2}$$

Moreover, the vertical tangent vectors at x are also exactly those tangent vectors which are tangent to the fiber $\pi^{-1}(\pi(x))$.

Definition 3.1.2 (Horizontal space)**.** *Let $x \in M$. Assume that the tangent space of M at x admits an orthogonal decomposition*

$$T_x M = \mathrm{Ver}_x M \oplus (\mathrm{Ver}_x M)^{\perp}. \tag{3.3}$$

Then the orthogonal complement $(\mathrm{Ver}_x M)^{\perp}$ of the vertical space of M at x is called the horizontal space at x and is denoted by $\mathrm{Hor}_x M$. In particular, the horizontal bundle is always defined in finite dimension.

Riemannian submersions Now a Riemannian submersion is a smooth submersion π between two Riemannian manifolds such that the tangent map $d\pi$ is an isometry between the tangent bundle of the base manifold and the horizontal bundle of the top manifold. Precisely, citing Le and Kendall (1993), a smooth submersion π from one Riemannian manifold onto another Riemannian manifold is said to be Riemannian if at any point x of the top manifold, the tangent map $d_x\pi$ maps any two horizontal vectors at x onto two tangent vectors at $\pi(x)$ that have the same inner product.

Induced Riemannian metric Then let us come back to our initial purpose and the action of G on M. Assume that M is a n-dimensional manifold and consider the decomposition of its tangent bundle into the vertical and horizontal bundles. According to the Riemannian submersion theorem, there exists a unique metric g^* on the manifold M/G that makes the quotient map π a Riemannian manifold (Tumpach and Preston, 2023). Each vector field on M/G identifies with a unique horizontal vector field on M through the tangent map $d\pi$, referred as its horizontal lift (Lee, 1997). According to the previous paragraph on Riemannian submersions, then the inner product of two vector fields on M/G with horizontal lifts u and v is given by

$$g^*(d\pi(u), d\pi(v)) = g(u, v) \tag{3.4}$$

where the right member is well defined because it is constant on the fibers for the action of G. This construction of the metric corresponds to the one detailed by Le and Kendall (1993). In practice, any computation on M/G can then be lifted to the manifold M using the previous identity. This approach has two advantages. First, in terms of implementation. Indeed, quotient classes are a priori abstract objects and are therefore difficult to encode when there is no canonical representative. If all the computations on the quotient manifold are performed in the top manifold, then there is no need to encode explicitly the quotient classes anymore. Second, in the case where M is relatively simple such that most geometric tools have a closed expression, and G is a matrix Lie group, we might be able to derive nice formulas for such tools on M/G as well. The case of Kendall shape spaces, on which next section focuses on, illustrates perfectly this second point.

O'Neill theorem Let us now give a result from O'Neill (1966) that carries on the correspondence to a higher order. Given a Riemannian submersion between two Riemannian manifolds, The O'Neill theorem relates the Levi-Civitá connection of the base manifold to the one of the top manifold. We detail such a result in the case of a quotient submersion.

Theorem 3.1.1. *Let ∇ and ∇^* denote the Levi-Civitá connection on M and M/G, respectively. Then two vector fields on M/G with horizontal lifts u and v satisfy the identity*

$$\nabla^*_{d\pi(u)} d\pi(v) = d\pi(\nabla_u v). \tag{3.5}$$

Horizontal geodesics The previous theorem allows us to lift the main differential equations of Riemannian geometry to the top manifold. In particular, the parallel transport equation, to which we come back later, and more simply the geodesic equation. Precisely, let $\pi(\gamma(s))$ be a geodesic on M/G and let v be the horizontal lift of its derivative with respect to s. Then following Theorem 3.1.1, the vector field $\nabla_v v$ has to be vertical. But since such vector field is also always horizontal (O'Neill, 1966), then this is equivalent to it vanishing. Now let $\gamma^h(s)$ be the curve on M with initial value $\gamma(0)$ and whose derivative with respect to s is v. Then we just showed that $\pi(\gamma^h)$ is a geodesic on M. Moreover, since the two geodesics $\pi(\gamma^h)$ and $\pi(\gamma)$ share the same initial conditions, they coincide everywhere. Finally, geodesics on M/G correspond exactly to the image by π of horizontal geodesics on M, that is the geodesics spanned by horizontal tangent vectors. The distance between x

and y is then the length of the minimizing horizontal geodesic that connects x to y and this definition coincides with the one of Younes (2010).

3.1.2 Focus on Kendall shape spaces

We now apply the previous results to the case of Kendall shape spaces. In particular, we make a point of detailing the calculations so that they can easily be reproduced for other spaces, whether by us or others. First, let us recall the definition of Kendall shape spaces. In the definition of Kendall (1984), a shape is the configuration in the space of a set of landmarks up to any translation, scaling and rotation. It is convenient to build the shape space in two steps.

Pre-shape space More precisely, let us look at the configuration of a set of points $x_1, \dots x_q$ in the Euclidean space $\mathbb{R}^p$ represented by the matrix $x \in \mathbb{R}^{p \times q}$. The space of $p \times q$ matrices is equipped with the Frobenius metric such that the distance between corresponding points of two configurations is measured by the canonical metric of $\mathbb{R}^p$. Let the translations and scaling act on such matrices. It defines a quotient space referred to as the Kendall pre-shape space of order (q, p) and given by

$$S_p^q = \Big\{ x \in \mathbb{R}^{p \times q} \;\Big|\; \sum_{i=1}^{q} x_i = 0 \text{ and } \|x\| = 1 \Big\}. \tag{3.6}$$

where $\|\cdot\|$ denotes the Frobenius norm. The pre-shape space is a sphere of the space of matrices, equipped with its canonical metric, that is the restriction of the Frobenius metric $\langle \cdot, \cdot \rangle$ to the tangent bundle of the sphere, which makes it a Riemannian manifold. The induced distance is simply the arc length

$$d(x, y) = \arccos\langle x, y \rangle \tag{3.7}$$

Kendall shape spaces Then let the group of rotations act on the pre-shape space by left multiplication. The shape space of order (q, p) is defined as the quotient space

$$\Sigma_p^q = \mathcal{S}_p^q / \operatorname{SO}(p). \tag{3.8}$$

It is a differentiable manifold with singularities at the points where the action is not free. Such singularities correspond exactly to degenerate shapes, that is shapes with some landmarks aligned or collapsing, and therefore admitting more symmetries than just the translations, scaling and rotations. In practice however, we restrict ourselves to the regular stratum of the shape space, that is the complement of the singular set in the shape space.

Vertical and horizontal bundles of the pre-shape space Now the Kendall shape space has been equipped with a Riemannian structure on its regular stratum in Le and Kendall (1993), exactly in the way we described above. First, we need to compute the vertical and the horizontal bundles of the pre-shape space. In theory, it should be possible to adapt the proofs we provide to redo the computations with another manifold M and matrix Lie group G.

Proposition 3.1.1. *Let $x \in S_p^q$. Then the vertical space of S_p^q at x is*

$$\operatorname{Ver}_x S_p^q = \operatorname{Skew}(p) x \tag{3.9}$$

where $\operatorname{Skew}(p)$ *denotes the space of skew-symmetric matrices of size p.*

Proof. We saw that the vertical space at x consists of those tangent vectors at x that are also tangent to the fiber at $\pi(x)$. Let us then consider a smooth curve γ of the fiber at $\pi(x)$ with $\gamma(0) = x$. Such curve can be written

$$\gamma(t) = R(t)x \tag{3.10}$$

where R is a smooth curve in $\mathrm{SO}(p)$ with $R(0) = Id$. The derivative of γ at $t = 0$ is given by $\dot{\gamma}(0) = \dot{R}(0)x$. Now let us that recall that $\mathrm{SO}(p)$ is a Lie group and that its Lie algebra $\mathfrak{so}(p)$ identifies with the space of skew-symmetric matrices $\mathrm{Skew}(p)$. Hence there exists a skew symmetric matrix A such that $\dot{R}(0) = A$. The result follows. □

Proposition 3.1.2. *Let $x \in S_p^q$. Then the horizontal space of S_p^q at x is*

$$\mathrm{Hor}_x\, S_p^q = \left\{ v \in \mathbb{R}^{p\times q} \;\middle|\; \sum_{i=1}^{q} v_i = 0,\ \langle x, v\rangle = 0 \text{ and } vx^T = xv^T \right\} \tag{3.11}$$

Proof. The proof is straightforward. Let v be horizontal vector at x. First of all, v has to be a tangent vector of S_p^q at x, that is v has to be centered and orthogonal to x (remember that the pre-shape space is a sphere). Then v has to be orthogonal to all the vertical vectors at x. In other words, v has to satisfy $\langle v, Ax\rangle = 0$ for any skew-symmetric matrix A. Rearranging the different terms within the trace, we obtain the equivalent condition $\langle vx^T, A\rangle = 0$. This is satisfied for any skew-symmetric matrix A if and only if the matrix vx^T is symmetric. □

Shape geodesics Now the Frobenius metric is invariant to the action of rotations, so it induces a Riemannian metric on the (main stratum of the) shape space. The geodesics of the pre-shape space are great circles. Therefore a geodesic of the shape space is a projection of a great circle tangent to the horizontal bundle and the exponential map of the tangent space of the shape space at $\pi(x)$ is the projection by the quotient map π of the exponential map of the horizontal space of the pre-shape space at x

$$\exp_x(v) = \cos(\|v\|)x + \sin(\|v\|)\frac{v}{\|v\|}. \tag{3.12}$$

At this stage, the last piece missing before being able to implement Riemannian Locally Linear Embedding on Kendall shape spaces is the parallel transport. Let us move then to the last part of this section.

3.1.3 Lifting the parallel transport

Here, we review a result by Le (2003), later supplemented by Kim, Dryden, Le, and Severn (2021), which allows to compute the parallel transport on Kendall shape spaces as the solution of a first order ordinary differential equation. Given a Riemannian submersion, there is a priori no clear correspondence between the parallel transport in the base manifold and the parallel transport in the top manifold. Still, O'Neill's theorem (see Theorem 3.1.1) holds and we can lift the parallel transport equation to the pre-shape space.

Proposition 3.1.3. *Let $\gamma : I \to S_p^q$ be a horizontal C^1-curve on such that* $\mathrm{rank}(\gamma(s)) \geq m - 1$ *for all $s \in I$ and let v_0 be a horizontal vector at $\gamma(0)$. Then the horizontal vector field v defined along γ is the horizontal lift of the parallel transport of the tangent vector $d\pi(v_0)$ along the curve $\pi(\gamma)$ if and only if it solves*

$$\dot{v}(s) = -\langle \dot{\gamma}(s), v(s)\rangle \gamma(s) + A(s)\gamma(s), \qquad v(0) = v_0 \tag{3.13}$$

where for all s, the matrix $A(s) \in Skew(m)$ is the unique solution of the Sylvester equation

$$A(s)\gamma(s)\gamma(s)^T + \gamma(s)\gamma(s)^T A(s) = \dot{\gamma}(s)v(s)^T - v(s)\dot{\gamma}(s)^T. \tag{3.14}$$

In this case, we say that the vector field v is the horizontal parallel transport of v_0 along the curve γ and we denote it by $P^h_\gamma(v_0)$.

Proof. Here again, we detail the proof as much as possible so that it is reproducible. Indeed, I believe that this type of equation could be established in several spaces other than Kendall shape spaces.[1] Let us come back to the proof. Let then v be a horizontal vector field along γ with initial value v_0 at $\gamma(0)$. By definition, the vector field $d\pi(v)$ is the parallel transport of the tangent vector $d\pi(v_0)$ along the curve $\pi(\gamma)$ if and only if it solves the equation

$$\nabla^*_{d\pi(\dot{\gamma})} d\pi(v) = 0. \tag{3.15}$$

where ∇^* is now the Levi-Civitá connection on the shape space. Following O'Neill theorem (see Theorem 3.1.1), it is equivalent to

$$\nabla_{\dot{\gamma}} v \in \operatorname{Ver} S^q_p \tag{3.16}$$

with ∇ the Levi-Civitá connection of the pre-shape space. Remember that the vertical space of the pre-shape space at x is $\text{Skew}(p)x$. The previous condition may then be rewritten in terms of an equation

$$\nabla_{\dot{\gamma}(s)} v(s) = A(s)\gamma(s) \tag{3.17}$$

where $A(s)$ is a skew-symmetric matrix for all $s \in I$. Moreover, the Levi-Civitá connection of the pre-shape space is non other than the Levi-Civitá connection of the sphere

$$\nabla_{\dot{\gamma}(s)} v(s) = \dot{v}(s) + \langle \dot{\gamma}(s), v(s) \rangle \gamma(s). \tag{3.18}$$

and Equation (3.13) follows. Now let us characterize $A(s)$ as a function of v, γ and $\dot{\gamma}$. To do so, let us compute another equation on $\dot{v}$. Since the vector field v is horizontal, it satisfies

$$v(s)\gamma(s)^T = \gamma(s)v(s)^T. \tag{3.19}$$

for all $s \in I$. We derive this expression

$$\dot{v}(s)\gamma(s)^T + v(s)\dot{\gamma}(s)^T = \dot{\gamma}(s)v(s)^T + \gamma(s)\dot{v}(s)^T. \tag{3.20}$$

Finally, replacing here $\dot{v}$ by its expression in Equation (3.13) yields exactly Equation (3.14). □

3.2 Riemannian Locally Linear Embedding

Now that we have defined all the tools required, let us come back to the Riemannian Locally Linear Embedding method (Maignant, Trouvé, and Pennec, 2023) briefly introduced in the Chapter 2, but this time specifically tailored for quotient manifolds, and let us detail then how to implement the method on Kendall shape spaces. First, we recall the general outline of Riemannian Locally Linear Embedding.

[1] In fact, Yann Thanwerdas (2022) came to me once with a very similar equation on a matrix space he was studying.

3.2.1 Outline of the method

Riemannian Locally Linear Embedding (RLLE) follows the same structure as Locally Linear Embedding (LLE) and implements two main steps, a model fitting step and a mapping step. The main difference lies in LLE taking as input a vector-valued dataset, while RLLE applies in a more general framework where the dataset can be manifold-valued. Essentially, it is the model fitting step that changes.

Step 1. Local Riemannian barycentric model Consider a dataset $x_1, \dots, x_N$ on a n-dimensional Riemannian manifold (M, g). The first step of RLLE consists in estimating local barycentric models that fit to the dataset. We recall that such a model is given as a pair of first reference points and then barycentric coordinates for the data points it approximates. Now in LLE as well as in RLLE, the reference points of the i-th model are fixed prior to estimation and are the neighbors $a_{i0}, \dots, a_{ik_i} = x_{j_0}, \dots, x_{j_{k_i}}$ of the point x_i within the dataset. As explained in Chapter 2, there are several ways to select such neighbors, the most common being the K-Nearest Neighbors algorithm and fixed-radius neighborhoods. In this case, this preliminary step is performed using the Riemannian distance on M. Then only the latent weights of each data point x_i in the corresponding local frame $x_{j_0}, \dots, x_{j_{k_i}}$ remain to be optimized

$$\begin{aligned} &\underset{\substack{w_{ij_0},\dots,w_{ij_{k_i}} \in \mathbb{R} \\ w_{ij_0}+\dots+w_{ij_{k_i}}=1}}{\text{minimize}} \quad d\big(x, \operatorname{bar}(x_j, w_{ij})_{j\in N(i)}\big)^2 \qquad \text{for } 1 \leq i \leq N \end{aligned} \tag{3.21}$$

where $N(i) = \{j_0, \dots, j_{k_i}\}$. Provided the data points are not too much spread in M, then the i-th model estimated by RLLE should be close to the optimal local barycentric model that would fit the best to the data points.

Step 2. Locally barycentric mapping The second step of RLLE is then exactly the same as the second step of LLE and consists in building an embedding $y_1, \dots, y_n$ in the dataset in $\mathbb{R}^k$ in such a way that locally, it coincides with the barycentric mapping (2.38). An optimal solution of this problem is found by solving

$$\underset{y_1,\dots,y_N \in \mathbb{R}^k}{\text{minimize}} \quad \sum_{i=1}^{N} \Big\| y_i - \sum_{j\in N(i)} w_{ij} y_j \Big\|^2.$$

where $N(i) = \{j_0, \dots, j_{k_i}\}$ stores the indices for the neighbors of x_i. We refer the reader to the Chapter 2 for more details on how to perform this second step and we now propose a specific algorithm for solving the first step when M is a quotient manifold.

3.2.2 Computing the barycentric projection in quotient manifolds

We consider here the case where the dataset lies on a Riemannian quotient manifold M/G resulting from the smooth, free, proper and isometric action of a Lie group G on a Riemannian manifold M. As we mentioned already in the first section, for there does not always exist an explicit description for quotient objects, computations in quotient manifolds are generally rather performed in the top manifold. In practice, this approach also makes the computations more comfortable. In the same spirit, also Problem (3.21) is designed for any Riemannian manifold and can be solved using one of the projection algorithms described in Chapter 2, we propose here a new algorithm specifically designed for quotient manifolds leveraging the results of the Section 3.1.

Horizontal barycentric projection First we can assume that the dataset is represented by N points $x_1, \dots, x_N$ in the top manifold M. Remember that the first step of LLE is equivalent to computing N barycentric projections. The same goes for RLLE. In Chapter 2, we proposed a tangent formulation for the barycentric projection problem (see Problem (2.30)). We can easily lift such formulation to the horizontal bundle of M. Precisely, let $x \in M$ and $a_0, \dots, a_k \in M$. Then the projection of $\pi(x)$ onto the barycentric subspace of $\pi(a_0), \dots, \pi(a_k)$ solves

$$\begin{aligned} \underset{\substack{v \in \mathrm{Hor}_x M \\ w_0,\dots,w_k \in \mathbb{R} \\ w_0+\dots+w_k=1 \\ u_0,\dots,u_k \in \mathrm{Hor}_x M \\ g_0,\dots,g_k \in G}}{\text{minimize}} \quad & g_x(v, v) \\ \text{subject to} \quad & \sum_{i=0}^{k} w_i u_i = 0 \\ & g_i \cdot \exp_{[\exp_x(v)]} \left(P^h_{x,[\exp_x(v)]}(u_i) \right) = a_i \quad (\forall i) \end{aligned} \tag{3.22}$$

where $P^h_{x,y}$ denotes the horizontal parallel transport of M along the geodesic joining x and y. We recall that the horizontal tangent vector v unrolls the geodesic connecting the point x to its projection and the horizontal tangent vector u_i encodes the logarithm of each reference point a_i at the projection. The optimization variables g_i are the elements of the group G that lift the equality constraint linking u_i and a_i

$$\pi(\exp_{[\exp_x(v)]} \left(P^h_{x,[\exp_x(v)]}(u_i) \right)) = \pi(a_i) \tag{3.23}$$

to the top manifold M. If G is an exponential Lie group, then it identifies with its Lie algebra $\mathfrak{g}$ through the matrix exponential such that the search space can still be written as a vector space.

3.2.3 Implementation in Kendall shapes spaces

Now consider the case of a dataset on some Kendall shape space, that is let M be the pre-shape space S_p^q and let G be the group of rotations $\mathrm{SO}(p)$. The implementation of RLLE on Kendall shape spaces depends essentially on the implementation of the horizontal barycentric projection on these spaces. Let us explicit then how to solve the projection problem in practice. At this point, we would like to mention that all the following methods and programs rely on our implementation of Kendall shape spaces in the Python library `geomstats` (Miolane, Guigui, et al., 2020). The corresponding code is available at geomstats/geometry (2023).

Barycentric projection problem Take a point $x \in S_p^q$ and points $a_0, \dots, a_k \in S_p^q$. Deriving Problem (3.22), the projection of $\pi(x) \in \Sigma_p^q$ onto the barycentric subspace of $\pi(a_0), \dots, \pi(a_k) \in \Sigma_p^q$ solves the following optimization problem

$$\begin{aligned} \underset{\substack{v \in \mathrm{Hor}_x S_p^q \\ w_0,\dots,w_k \in \mathbb{R} \\ w_0+\dots+w_k=1 \\ u_0,\dots,u_k \in \mathrm{Hor}_x S_p^q \\ A_0,\dots,A_k \in \mathrm{Skew}(p)}}{\text{minimize}} \quad & \|v\|^2 \\ \text{subject to} \quad & \sum_{i=0}^{k} w_i u_i = 0 \\ & \exp(A_i) \exp_{[\exp_x(v)]} \left(P^h_{x,[\exp_x(v)]}(u_i) \right) = a_i \quad (\forall i). \end{aligned} \tag{3.24}$$

Our implementation of the RLLE relies on the `SLSQP` method of the `scipy` library to solve the projection problem. This method requires a gradient function for both the functional to optimize and the constraints. We cannot hope to compute these gradients in closed-form. However, it is perfectly possible to compute them numerically using automatic differentiation, for example with the library `pytorch` (used here). Therefore, we will take care to propose an implementation of each of the functions to derive (especially the second constraint) that it is compatible with automatic differentiation.

Initialization Let $\varphi(\pi(x))$ be the matrix introduced in Chapter 2 (in Remark 2.1.1) whose i-th column is the logarithm of $\pi(a_i)$ at $\pi(x)$ written in local coordinates. Intuitively, since the optimal weights of the projection problem satisfy the critical condition $\sum w_i \log_{\text{proj}(\pi(x))} \pi(a_i) = 0$.

- We set then $w_0, \dots, w_k$ so as to minimize – for the metric at $\pi(x)$ – the quantity $\sum w_i \log_{\pi(x)} \pi(a_i)$ or equivalently we take them to be the coordinates of the first eigenvector of $\varphi(\pi(x))$. If $\pi(x)$ is not too far from its projection (which is be the case in a model fitting task), the quantity $\sum w_i \log_{\pi(x)} \pi(a_i)$ should still be relatively small and our initialization should provide a good approximation of the optimal weights.
- Equivalently, we take the barycenter m of the reference points $a_0, \dots a_k$ with corresponding coordinates $w_0, \dots, w_k$ as an initial estimation of the projection and we retrieve the corresponding value of v by unrolling the geodesic between x and m.
- We compute $u_0, \dots, u_k$ as the horizontal parallel transport from the mean m to x of the logarithm of a_i at m.
- Finally, to fix the initial value of $A_0, \dots A_k$, we solve the alignment problem encoded by the second equality constraint in the previous optimization problem. This amounts to solve a singular-value decomposition as explained by Nava-Yazdani, Hege, Sullivan, and Von Tycowicz (2020).

Construction of an orthonormal basis of the horizontal subspace The main advantage of the tangent and horizontal projection algorithms is that they can be solved as vector-valued optimization problems. In practice, this assumes that we know a basis of the horizontal space at x. Let us detail then a simple procedure for computing an orthonormal basis of the horizontal space at x. The shape space of order (p, q) has dimension $n = p(q - \frac{1}{2}(p-1)) - 1$. Start then with building a set of n linearly independent horizontal vectors at x incrementally. Once such set has been built, apply a Gram Schmidt orthonormalization process. The previous step requires to implement the following simple program that returns a random horizontal vector. First, take a random centered matrix u. Second, take its projection $v = u - \langle x, u \rangle x$ onto the tangent space at x. Third, take the horizontal component of such projection by discarding its vertical part Ax where the skew-symmetric matrix A satisfies the Sylvester equation $(v - Ax)x^T = x(v - Ax)^T$. We recall a bit later how to solve this equation, as well as the one involved in horizontal parallel transport, which is the subject of the next paragraph.

Computation of the horizontal parallel transport The horizontal parallel transport is the solution of Equation (3.13). Since it is a first-order differential equation, it can be solved using any iterative method like the Runge-Kutta method. This method has been implemented in the library `geomstats` as part of our work with Nicolas Guigui (Guigui, Maignant, Trouvé, and Pennec, 2021) where we compare it on Kendall shape spaces with the pole ladder method (Guigui and Pennec,

2022). Additionally, let us not forget that the functional in Equation (3.13) is defined implicitly by the solution of a Sylvester equation (precisely Equation (3.14)). This motivates our next paragraph.

Solution of Sylvester equation Both the Sylvester equation defining the horizontal projection and the Sylvester equation parametrizing the horizontal parallel transport equation are of the form $SX + XS = C$ where S is a symmetric positive matrix and C a skew-symmetric matrix. The reader may easily check that such equation has a unique skew-symmetric solution given by $X = PYP^T$ where $Y_{ij} = \frac{1}{d_{ii}+d_{jj}}(P^TCP)_{ij}$ for $i \neq j$ and $S = PDP^T$ is an eigendecomposition of S.

Parametrization of the Lie group of rotations The group of rotations $SO(p)$ is an exponential Lie group and its Lie algebra is $\mathfrak{so}(p) = \mathrm{Skew}(p)$. Therefore any element g of the group is encoded as the matrix exponential of a skew-symmetric matrix $A \in \mathrm{Skew}(p)$.

3.3 Application to Kendall shape spaces

In this last section, we first review two simple experiments on Kendall shape spaces that illustrate the gain in accuracy of RLLE, whose specific implementation for these spaces has been just detailed, over its original version, LLE, in a scenario where input shapes would be treated simply as $p \times q$ matrices. We also compare the computational complexity of the two methods. As a conclusion to this chapter, we sketch RLLE use cases and discuss future work on this method. This section essentially reproduces the experimental section of our paper (Maignant, Trouvé, and Pennec, 2023), but is intended to be a little more detailed. In particular, we provide a slightly reworked version of the two experiments originally proposed.

3.3.1 Benchmark experiments

We propose two initial experiments on the space Σ_3^3 of 3D triangles. Although this may appear as a simplistic example, it provides a way of visually interpreting the experiments. Indeed, the shape space Σ_3^3 is isometric to a 2-dimensional hemisphere (Le and Kendall, 1993), which can be visualized flat as a disk, a visualization we have implemented in `geomstats` (see Figure 3.2 for example) as part of a contribution to the ICLR Computational Geometry & Topology Challenge 2021 (Miolane, Caorsi, et al., 2021). Essentially, both this visualization and the better-known Kendall's sphere of 2D triangles use the formulas calculated by Kendall (1984). We suggest the reader look directly at the code for implementation details (geomstats/visualization, 2023).

Large scale accuracy of RLLE In this first experiment, we wish to illustrate the ability of RLLE to better recover dispersed datasets than LLE. The underlying idea is that the Euclidean model implemented by LLE is a first-order approximation and should therefore yield a significant error if the distances observed in the data are too large. Thus, we experiment with both methods for embedding in $\mathbb{R}^2$ two different datasets on the shape space Σ_3^3, one sampled from a small-scale distribution and another one sampled from a large-scale distribution (see Figure 3.1). We chose each distribution to be a mixture of two Gaussian on Σ_3^3, reflecting a case where the use of a local method seems of particular interest compared to that of a global method like PCA or BSA. What we understand here by Gaussian distribution is in fact the projection onto Σ_3^3 of such a distribution on the ambient space of 3×3 matrices.

– The first dataset is sampled from a mixture of two Gaussian of standard deviation $\sigma_1 = 1 \times 10^{-3}$ with respective means close in shape. We expect LLE to perform reasonably well on this dataset.

– The second dataset is sampled from a mixture of two of two Gaussian of standard deviation $\sigma_2 = 1 \times 10^{-2}$ with easily distinguishable means. On this dataset, RLLE should perform significantly better than LLE. Note that from the interpretation point of view, such a large deviation correspond to higher variability in shape.

In both methods, we choose a fixed number K of neighbors and select the neighbors according to the Riemannian distance on Σ_3^3. Since we seek for a 2-dimensional embedding of a dataset, it would seem natural to set the number of neighbors to 3. In fact, such a value would also be enough for the local models computed by RLLE to coincide with the shape space Σ_3^3 – which we recall is a 2-dimensional manifold– and therefore to fit exactly to the dataset. Yet, we are forced to chose a higher number of neighbors, precisely $K = 5$, to ensure that the embedding is unique according to eigenvalue criterion detailed in Section 2.3 of the previous chapter. This behavior is later explained in Chapter 5, but the intuition is that if the number is lower, then the graph is divided into several components, and there exists independent ways of mapping the different components. In the last chapter, a new algorithmic approach to this question is proposed. Such a choice of K essentially benefits only to LLE as in the case of RLLE, it does not increase the dimension of the model, bounded from above by the dimension of the shape manifold, that is $n = 2$, but rather results in overfitted local models. However, it makes in fact the computations much easier. Indeed, if the i-th model contains the data point x_i, then the weights of this model are simply the barycentric coordinates of x_i and instead of optimizing the projection problem, we only need to invert the critical condition on x_i being a barycenter to retrieve the weights (see Remark 2.1.1). In practice however, as the intrinsic dimension of the data is never known, we always have to solve the projection problem. That being said, we could think of a way of estimating such an intrinsic dimension based on the critical condition being satisfied or not.

Finally, to evaluate the accuracy of each of the two methods, we propose to compare the embedding they compute to the visualization of the original dataset on the disk. To allow for such a comparison, we need to align RLLE and LLE on this reference embedding in translation, scaling and orthogonal transformation, as we recall that the locally barycentric mapping problem is formulated in a way that it is invariant to such transformations. In fact, on both datasets, RLLE and LLE decompose in an exact locally barycentric mapping, that is achieve a zero final cost. In that specific case, the optimal embedding is unique only up to an affine transformation of the embedding space, and thus we perform here an affine alignment rather than one in translation, scaling and orthogonal transformation. Additionally, our implementation of LLE performs first a Procrustean alignment step before solving each Euclidean projection problem (Problem (2.20)). Precisely, we align each neighbor x_j of the point x_i onto x_i in rotation. Note that without this alignment step, the accuracy of LLE drops significantly.

A Swiss Roll example The second experiment is deigned to compare the accuracy of RLLE with that of LLE numerically. For this purpose, we study the embedding in $\mathbb{R}$ of a dataset sampled from a curve, that is a 1-dimensional submanifold, such that the natural parameterization of the curve in $[0, 1]$ by its arc-length provides an embedding of reference to evaluate RLLE and LLE. More precisely, we generate a logarithmic spiral curve on Σ_3^3 inspired by the "Swiss Roll" example. We sample a set of 20 shapes along this curve in a way that they are approximately equidistant (see Figure 3.2). We embed the dataset in $\mathbb{R}$ using RLLE and LLE. To avoid dealing with uniqueness issues like in the first experiment, we fix the neighbor graph on behalf for $K = 2$ in such a way that the each

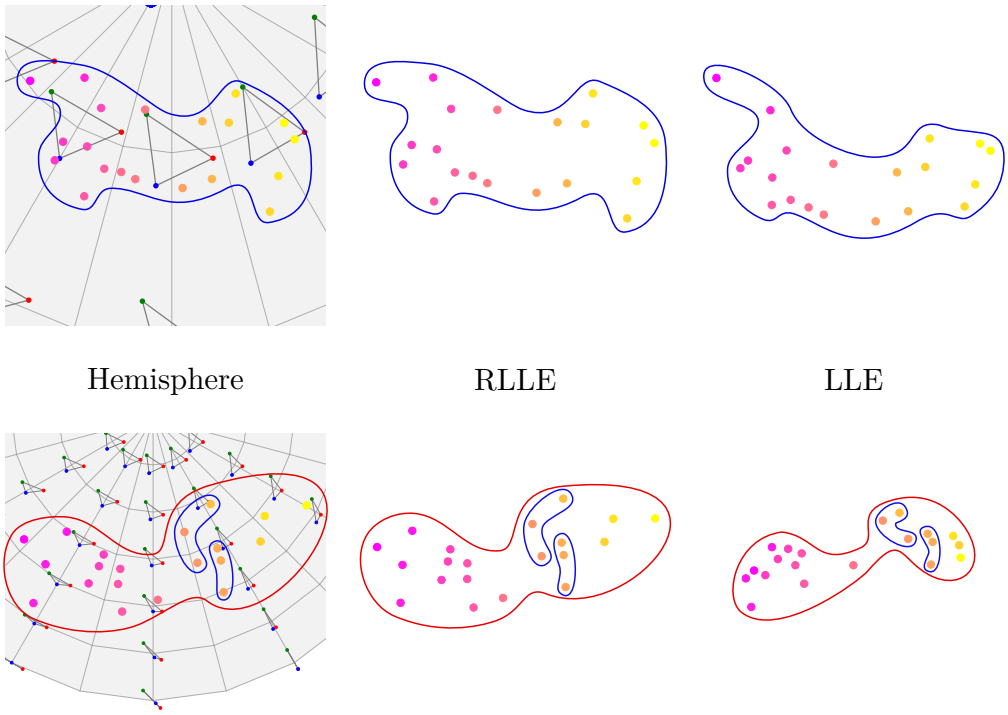

Figure 3.1: First experiment. (First row) On the first dataset, both RLLE and LLE provide an faithful embedding with respect to the representation of the original dataset on the hemisphere (flatten into a disk). In fact, after alignment, the embedding computed with RLLE coincides almost perfectly with the representation on the disk. It is important to stress however that it is not an isometric representation, but only homeomorphic to the representation on the hemisphere (which itself is isometric). (Second row) On the second dataset, the accuracy of LLE drops. Essentially, we notice that LLE is still able to recover local patterns faithfully but fails at a larger scale. Local patterns are reorganized differently than they were in the original data set, and the overall distribution appears slightly distorted. This experiment confirm our prediction that LLE accuracy decreases as the dataset spreads while RLLE performs well both at small and large scales.

data point is connected to the previous point and to the next point on the curve according to the order defined by the arc length parametrisation. We measure the error made by either method as the L^2 distance between the 20-dimensional vector they output and the arc-length parameterization of the data points, after a proper rescaling of the two output embeddings to $[0, 1]$. We generate 100 datasets, for each of which we measure the error made by RLLE and LLE, and we compare the overall accuracy of each of the two methods. In this example, everything has been designed to ensure that we have an exact way of evaluating the methods. In general however, especially in higher dimension, there is no canonical way of evaluating RLLE and LLE as there does not exist an embedding of reference. We suggest then to use distance preservation as an accuracy criterion. Given the output of either method, we can compare the embedded distance matrix with the distance matrix of the original distribution. Let us just remember that, since the locally barycentric mapping problem is made scale-invariant by enforcing a unitary covariance matrix, we would need to measure the two matrices up to a rescaling.

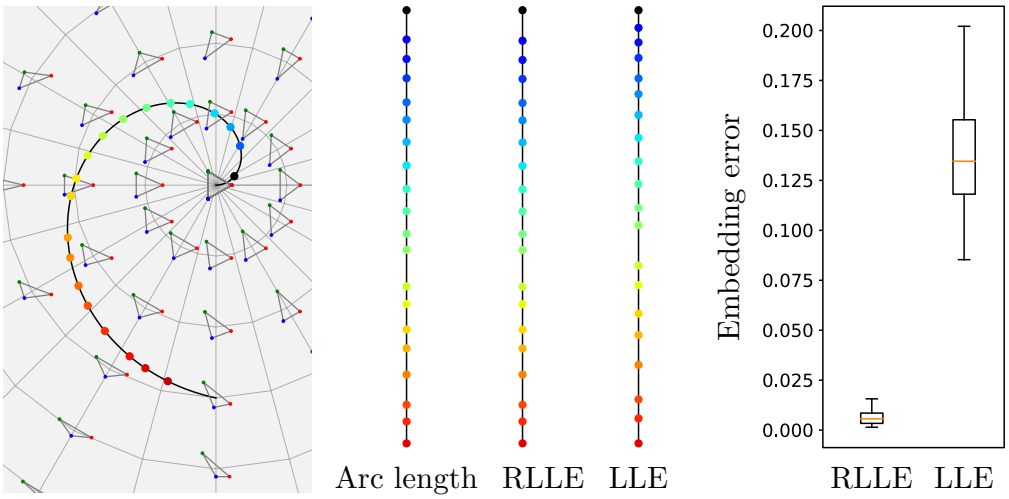

Figure 3.2: Second Experiment. (Left) A typical dataset in the experiment. (Center) Output of RLLE and LLE for this dataset. Of the 100 datasets generated for the experiment, it is in fact the one for which LLE performed the worst. We can explain this by the following observation. As it implements geodesic models, RLLE underestimates all the distances in the dataset. Therefore, the approximation error at the blue end of the curve and that at the red end of the curve compensate each other. On the other hand, LLE generally underestimates the distances in the blue area while it overestimates them in the red area. Therefore the errors do not compensate anymore but rather accumulate. In this dataset, only a few points were sampled in the blue area, where the linear model is the most inaccurate. As a consequence, LLE substantially underestimates the very first distance such that the entire embedding recovered by the method is shifted towards the blue end. (Right) Overall accuracy of RLLE and LLE on the 100 datasets generated, summarized in a box plot. We recall that the embedding error corresponds to the L^2 distance between the embedding computed by either method and the arc-length parameterization of the dataset, each of them corresponding to a vector in $\mathbb{R}$. We observe that RLLE performs significantly better and moreover is more stable. Note however that RLLE is very sensitive to the initialization, most likely due to the existence of local minima. Precisely, the same experiment but where we initialize RLLE with equal weights $w_0 = \ldots = w_k = \frac{1}{k+1}$, or the weights found by LLE, produces significantly worse results.

3.3.2 About the computational complexity

RLLE shares the main drawback of intrinsic manifold learning methods: it is computationally quite expensive. Let us detail this point. We mainly focus on the model fitting step as the mapping step is common to the LLE method. We consider that either the horizontal or the tangent projection algorithm is used depending on whether the dataset lies on a quotient manifold or not. We also assume that the neighbors selected for each data point are its K nearest neighbors. First, the search space of each local projection problem is a space of either dimension $n + K + Kn$ if the tangent projection algorithm is used, or dimension $n + K + Kn + Km$ if the horizontal projection algorithm is used, where we recall that n is the dimension of the manifold or quotient manifold the dataset lies on, and were m denotes the dimension of the Lie group G in the second case. For the Kendall shape space of parameters p and q, we have $n = p(q - \frac{1}{2}(p-1)) - 1$. Then, we need to take into account the computational cost of the exponential and the parallel transport methods. In the case of Kendall shape spaces, while the first one is for free, the latter performs in $\mathcal{O}(sp^3)$, where s is the number of integration steps (we recall that each call of the functional involves solving a Sylvester equation

which roughly consists in an eigendecomposition). Each evaluation of the constraint – and so each step of the optimization task – costs the same. Finally, our implementation uses a SLSQP method to solve each optimization problem. Finally, since there are as many projection problems to solve as there are data points, the overall cost of RLLE is the cost of one projection problem multiplied size N of the dataset. As a comparison, the Euclidean projection method is equivalent to a matrix inversion of dimension k. Therefore, LLE performs in roughly $\mathcal{O}(Nk^3)$, such that its complexity does not depend on the dimension n of the data space.

3.3.3 Perspectives

Given the computational cost of RLLE, it is important to understand for which type of data the method is particularly suited. Typically, the locally linear assumption made by LLE may be valid for large and well-concentrated data sets. In these cases, we extrapolate from the first experiment both methods should perform more or less the same. Moreover, non-local methods like PCA, PGA or BSA provide a good estimation whenever the data is sufficiently concentrated. Finally, LLE and RLLE seem of particular use in cases where a distance-based method like MDS or Isomap does not perform well. These three remarks suggest that RLLE is more specifically designed for small sample size data sets with large dispersion, and provides an embedding which might allow to observe more informative patterns than the ones characterized by the distance only.

As for now, RLLE has been implemented for Kendall shape spaces only. We are contemplating a general implementation of the method into the library `geomstats` for various manifolds and quotient manifolds – starting with the simplest manifolds and manifolds that are already implemented in `geomstats`. Knowing that the parallel transport on symmetric spaces can be computed in one step thanks to pole ladder (Guigui and Pennec, 2022), it would then be relevant to focus onto those specific manifolds. In fact, our implementation of RLLE on Kendall shape space is not yet part of `geomstats` as it relies on automatic differentiation modules that were not available in the library at the time, but we expect to add it soon. Following our first two experiments, it would also be useful to try higher-dimensional examples, on Kendall spaces but on manifold spaces where everything is known in closed form, such as the sphere. Applications to real data sets will also be developed in future work. In particular, it would be interesting to investigate further the analysis of protein conformations using Kendall's framework as an extension of an early work (Stolyarchuk et al., 2021). Finally, theoretical improvements for this work would involve understanding better how the stratification of Kendall shape spaces – especially of non-triangular shapes – interfere with our projection algorithm. Regarding this question, the work from Huckemann, Hotz, and Munk (2010) on designing PCA for stratified G-manifolds, and in particular Kendall shape spaces, appears to be a particularly relevant reference.

References

Dryden, I. L. and K. V. Mardia (2016). *Statistical Shape Analysis, with Applications in R*. John Wiley & Sons, Ltd. ISBN: 978-1-119-07249-2. DOI: 10.1002/9781119072492.

geomstats/geometry (2023). URL: https://github.com/geomstats/geomstats/blob/master/geomstats/geometry/pre_shape.py.

geomstats/visualization (2023). URL: https://github.com/geomstats/geomstats/blob/master/geomstats/visualization/pre_shape.py.

Guigui, N. and X. Pennec (2022). "Numerical accuracy of ladder schemes for parallel transport on manifolds". In: *Foundations of Computational Mathematics* 22.3, pp. 757–790. ISSN: 1615-3383. DOI: 10.1007/s10208-021-09515-x.

Guigui, N. et al. (2021). "Parallel transport on Kendall shape spaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer International Publishing, pp. 103–110. ISBN: 978-3-030-80209-7. DOI: 10.1007/978-3-030-80209-7_12.

Huckemann, S., T. Hotz, and A. Munk (2010). "Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions". In: *Statistica Sinica* 20.1, pp. 1–58. ISSN: 1017-0405.

Kendall, D. G. (1977). "The diffusion of shape". In: *Advances in Applied Probability* 9.3, pp. 428–430. ISSN: 0001-8678. DOI: 10.2307/1426091.

Kendall, D. G. (1984). "Shape manifolds, Procrustean etrics, and complex projective spaces". In: *Bulletin of the London Mathematical Society* 16.2, pp. 81–121. ISSN: 1469-2120. DOI: 10.1112/blms/16.2.81.

Kim, K.-R. et al. (2021). "Smoothing splines on Riemannian manifolds, with applications to 3D shape space". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.1, pp. 108–132. ISSN: 1369-7412. DOI: 10.1111/rssb.12402.

Le, H. (2003). "Unrolling shape curves". In: *Journal of the London Mathematical Society* 68.2, pp. 511–526. ISSN: 1469-7750, 0024-6107. DOI: 10.1112/S0024610703004393.

Le, H. and D. G. Kendall (1993). "The Riemannian structure of Euclidean shape spaces: A novel environment for statistics". In: *The Annals of Statistics* 21.3, pp. 1225–1271. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1176349259.

Lee, J. M. (1997). *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics. Springer-Verlag. ISBN: 978-0-387-98271-7. DOI: 10.1007/b98852.

Lee, J. M. (2012). *Introduction to Smooth Manifolds*. Vol. 218. Graduate Texts in Mathematics. Springer. ISBN: 978-1-4419-9981-8, 978-1-4419-9982-5. DOI: 10.1007/978-1-4419-9982-5.

Maignant, E., A. Trouvé, and X. Pennec (2023). "Riemannian locally linear embedding with application to Kendall shape spaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 12–20. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_2.

Miolane, N., M. Caorsi, et al. (2021). *ICLR 2021 Challenge for computational geometry & topology: Design and results*. arXiv: 2108.09810[cs,math].

Miolane, N., N. Guigui, et al. (2020). "Geomstats: A Python package for Riemannian geometry in machine learning". In: *Journal of Machine Learning Research* 21.223, pp. 1–9. ISSN: 1533-7928.

Nava-Yazdani, E. et al. (2020). "Geodesic analysis in Kendall's shape space with epidemiological applications". In: *Journal of Mathematical Imaging and Vision* 62.4, pp. 549–559. ISSN: 1573-7683. DOI: 10.1007/s10851-020-00945-w.

O'Neill, B. (1966). "The fundamental equations of a submersion." In: *Michigan Mathematical Journal* 13.4, pp. 459–469. ISSN: 0026-2285, 1945-2365. DOI: 10.1307/mmj/1028999604.

Stolyarchuk, M. et al. (2021). "Identification of the primary factors determining the specificity of the human VKORC1 recognition by thioredoxin-fold proteins". In: *International Journal of Molecular Sciences* 22.2, p. 802. DOI: 10.3390/ijms22020802.

Thanwerdas, Y. (2022). "Géométries riemanniennes et stratifiées des matrices de covariance et de corrélation". Thèse de doctorat. Université Côte d'Azur.

Tumpach, A. B. and S. C. Preston (2023). "Three methods to put a Riemannian metric on shape space". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 3–11. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_1.

Younes, L. (2010). *Shapes and Diffeomorphisms*. Vol. 171. Applied Mathematical Sciences. Springer. ISBN: 978-3-642-12054-1, 978-3-642-12055-8. DOI: 10.1007/978-3-642-12055-8.

Chapter 4

Barycentric Subspace Analysis of a set of graphs

In collaboration with Anna Calissano, currently at Imperial College, London.

This is a story that started in October 2022 at the Institut Henri Poincaré in Paris, where a semester around geometry and statistics was taking place:

> While I was wandering around in the first level of the institute, I came across Anna working on a black board. Taking a glance at the board, I noticed an optimization problem that looked a lot like the one I had encountered a year before when trying to implement Locally Linear Embedding on Kendall shape spaces. A bit curious, I went to her and asked what she was working on. She explained she was trying to figure out an algorithm to perform Barycentric Subspace Analysis for sets graphs, more precisely unlabeled graphs. At the time, I was not very familiar with Barycentric Subspace Analysis and I had not yet made the link with Locally Linear Embedding. Still I realized that, apart from the notations, the problem Anna was studying and the weight calculation problem in Locally Linear Embedding were the same, that is the barycentric projection problem. However, I soon understood that her problem was way more tricky.

Indeed, unlabeled graphs are usually modeled in a quotient space over the action of permutations. Since such an action is not free, the corresponding quotient space is unfortunately not a manifold. Still, it is possible to a certain extent to define statistical methods like Principal Component Analysis on such space using specific tools, and in theory, the same tools also allow to define Barycentric Subspace Analysis for sets of unlabeled graphs. How to implement the method in practice, however, is a different matter, mainly due to the discrete nature of the action, as we shall see. And so, stuck in a kind of technical impasse, we decided to rethink the modeling of unlabeled graphs, with the desire to this time fall into the case of Riemannian quotient manifolds. This chapter is precisely structured around this new modeling approach and its application to barycentric subspace analysis.

The first section serves as a motivation for the whole chapter. Supplementing the short story you just read, it provides a brief overview of some key issues in graph analysis and reviews our initial approach to tackling the problem of Barycentric Subspace Analysis of a set of graphs, that proved rather unsuccessful but nonetheless instructive. In the second section, we introduce a new framework for the study of graphs, which nicely echoes spectral graph theory and proves to have simple and beautiful geometric properties. The last section implements Barycentric Subspace Analysis in this

framework and demonstrates their joint potential through two experiments, one on a simulated dataset and another one on a real dataset consisting of airlines route maps data.

4.1 Why learning graphs is challenging

Here, we want to emphasize the specifics of working with graphs and highlight the main issues that have motivated and shaped this work. Mainly, on one hand we want to demonstrate some of the limitations of current modeling and motivate the use of a Riemannian model, on the other hand we want to raise the critical question of the interpretability of dimensionality reduction methods when dealing with graph data.

4.1.1 A modest state of the art in statistical graph analysis

In recent years, there has been growing interest in the analysis of a set of graphs, owing to the various applications in which graph data arise: brain connectivity (Simpson et al., 2013; Durante, Dunson, and Vogelstein, 2017; Calissano, Papadopoulo, Pennec, and Deslauriers-Gauthier, 2023), brain arterial networks (Guo, Srivastava, and Sarkar, 2021), anatomical trees (Wang and Marron, 2007; Feragen et al., 2013), mobility networks (Von Ferber, T. Holovatch, Y. Holovatch, and Palchykov, 2009). The field of graph analysis has therefore developed significantly, essentially around the design of statistical tools for carrying out various tasks such as regression or classification (Tsuda and Saigo, 2010). In some cases, however, the complexity of building a statistical framework for graph analysis lies not only in defining a metric structure or designing a suitable distance function, but also in dealing with the fact that there may be no clear correspondence between nodes in the observed graphs. Such graphs are often referred to as unlabeled graphs, and the ambiguity of correspondence between nodes can vary from totally unlabeled settings like social networks connecting different groups of individuals, to partially labeled settings as for example molecules involving different types of atoms. In this chapter, we focus the totally unlabeled setting, which in fact also includes the partially labeled one up to some reformulation.

Graph spaces A number of studies have therefore focused on defining statistical tools specifically designed for unlabeled graphs. In statistical analysis, the first tool to be defined is the distance. There are several distances available in the literature, the key point being that they must be invariant to nodes permutations in order to compare graphs independently of their labeling. The so-called spectral distances (Jurman, Visintainer, and Furlanello, 2011; Donnat and Holmes, 2018) are a well-known and widely used example. Still in the context of the invariant (or equivariant) approach, the analysis of unlabeled graphs has also been tackled by geometric deep learning (Maron, Ben-Hamu, Shamir, and Lipman, 2018) which proved particularly successful for tasks such as clustering or classification. Finally, to go a step further, some work has focused on embedding graphs in a quotient space resulting from the action of node permutations on graphs, represented for example by their adjacency matrix (Jain and Obermayer, 2009; Calissano, Feragen, and Vantini, 2023), explicitly

$$\mathbb{R}^{p\times p}/\mathfrak{S}(p), \tag{4.1}$$

or by the ordered set of their edges (Kolaczyk et al., 2020). The reader might want to check that the action by conjugation of a permutation $\sigma \in \mathfrak{S}(p)$ on a graph – represented by its adjacency matrix – consists exactly in relabeling the nodes of the graph. The graph space thus defined is then endowed with a metric structure given by the quotient distance induced by that of the top space (Younes, 2010). Although this last approach is more intrinsic than the previous ones, it entails two

major issues. From a theoretical point of view, as we have already pointed out, since the action of permutations is not free, the techniques from differential geometry described in the previous chapter do not apply a priori, at least not to the whole graph space. Instead, Kolaczyk et al. (2020) proposed to study such a quotient space from the point set topology point of view. From a practical point of view, this approach suffers from computational limitations beyond a certain number of nodes, resulting from the computation of the quotient distance being worth a graph matching problem (Conte, Foggia, Sansone, and Vento, 2004). Essentially, as we hope to illustrate throughout this chapter, one good way to get around these issues is to rethink modeling outside the action of a discrete group. In this regard, it is precisely important to mention that Severn, Dryden, and Preston (2022) have recently introduced a new model for graph analysis that allows the study of unlabeled graphs as elements of a Procrustes manifold. We shall come back to it in a little more detail in Section 4.2 which focuses on our new Riemannian framework for the analysis of graphs.

Dimensionality reduction for graphs A significant proportion of the work on dimension reduction for sets of graphs revolve around Principal Component Analysis (PCA). Let us mention for example Aydın et al. (2009) in the specific case of trees. In the case of unlabeled graphs, several methods focus on defining PCA for a set of graphs embedded in the graph space $\mathbb{R}^{p\times p}/\mathfrak{S}(p)$. For example, Guo, Srivastava, and Sarkar (2021) proposed a method which consists in linearizing the graph data by aligning each graph x_i at the mean graph m

$$v_i = P_i(\sigma) \cdot x_i - m \tag{4.2}$$

before applying a standard PCA. A link may be drawn between this approach and tangent PCA (a priori defined in the case of Riemannian manifolds only) in that the vector v_i thus computed is similar to the Riemannian logarithm as it generates a straight line in $\mathbb{R}^{p\times p}$ of minimum length. Mentioning the tangent PCA, Severn, Dryden, and Preston (2022) also proposed an implementation based on their modeling. However, tangent PCA as well as the previous method are not intrinsic. To address this problem, Calissano, Feragen, and Vantini (2023) have proposed more recently an implementation of geodesic PCA (Huckemann, Hotz, and Munk, 2010) on the graph space $\mathbb{R}^{p\times p}/\mathfrak{S}(p)$.

Interpretability in statistical graph analysis When it comes to reducing the dimensionality of a set of graphs, interpretability is generally a recurring issue. The two methods we have just described are not exempt from this. Indeed, one is not entirely intrinsic, and the other allows the data set to be described only along one-dimensional components, which in the case of graphs have proven to be quite difficult to interpret in practice. Therefore, there has been in graph analysis a growing interest in archetypal analysis (Cutler and Breiman, 1994) and sampled-limited statistics (Feragen et al., 2013; Zhai, 2016). Now, we saw precisely in Chapter 2 that Barycentric Subspace Analysis (BSA) may be formulated as such. For this reason, in order to retrieve high-dimensional features, and because barycentric geometry only requires a priori a notion of logarithm, which can be defined in a general quotient space as we explain in the next section, BSA appears to be a particularly relevant dimensionality reduction tool for the analysis of graph sets. Note that the method's great interpretability has already been put to good use in other applications (Rohé, Sermesant, and Pennec, 2018).

4.1.2 Barycentric Subspace Analysis under a finite group action

Our first approach to implement BSA for sets of graphs relies on the graph space $\mathbb{R}^{p\times p}/\mathfrak{S}(p)$ introduced above. As we highlighted, it is not a manifold. Therefore, barycentric subspaces are a priori not defined. Instead, we investigate a new definition that applies to the case of any quotient

space, including that of the graph space, and coincides with the usual one in the case of quotient manifolds. This section takes up and summarizes the more detailed work we presented at the 6th International Conference on Geometric Science of Information (Calissano, Maignant, and Pennec, 2023).

Horizontal logarithm We recall that the general idea of BSA is to fit a barycentric subspace to a dataset. To extend the notions of barycenter and barycentric subspace to a non-manifold quotient space, we need an analogue of the Riemannian logarithm. We propose here a definition based on the alignment of two points in top space, reminiscent of the PCA of graphs introduced by Guo, Srivastava, and Sarkar (2021) and described in the previous section. Intuitively, the logarithm should correspond to the vector that generates the minimum-length geodesic in the top space according to the quotient distance. Note that this idea coincides with that of the horizontal lift of the Riemannian logarithm when the space is a quotient manifold introduced in Chapter 3. From now on and throughout this section, we consider then the case of a quotient space M/T, where M is a smooth manifold equipped with a Riemannian metric g and T is a finite group acting on M in such a way that the distance d on M is T-invariant. Additionally, we denote by π the canonical projection map from M to M/T. Let us now give a precise definition of the logarithm of such a space.

Definition 4.1.1. *Let x and $y \in M$ such that there exists a unique element $t_{x,y} \in T$ that minimizes the quantity $d(x, t \cdot y)$ over T, which is the case if and only if y is not equidistant to any two representatives of $\pi(x)$. Then x and $t_{x,y} \cdot y$ are said to be optimally positioned with respect to each other and the horizontal logarithm of y at x is the tangent vector*

$$\log_x^h(y) = \log_x(t_{x,y} \cdot y) \in T_x M \tag{4.3}$$

where $\log_x$ denotes the Riemannian logarithm at x on M. Since the set of all points which are equidistant to at least two elements of the orbit $\pi(x)$ is a set of null measure, the horizontal logarithm at x is defined almost everywhere. Note however that it may have discontinuities at those specific points.

Quotient barycentric subspaces Following this definition, we can extend that of barycenters and barycentric subspaces.

Definition 4.1.2. *Let $a_0, \dots, a_k \in M$ be affinely independent points and let $w_0, \dots, w_k \in \mathbb{R}$ sum to 1. Then $\pi(x) \in M/T$ is a barycenter of $\pi(a_0), \dots, \pi(a_k)$ with corresponding weights $w_0, \dots, w_k$ if it belongs to the set*

$$\mathrm{bar}(\pi(a_i), w_i)_{0 \le i \le k} = \Big\{ \pi(x) \in M^*(a_0, \dots, a_k)/T \mid \sum_{i=0}^{k} w_i \log_x^h(a_i) = 0 \Big\}. \tag{4.4}$$

where $M^*(a_0, \dots, a_k)$ is the set of all points x in M for which a given reference point a_i is not equidistant to any two representatives of x or equivalently, for which x is not equidistant to any two representatives of a_i. The set of all such barycenters is called the quotient barycentric subspace of $\pi(a_0), \dots, \pi(a_k)$.

Discontinuity of quotient barycentric subspaces Now moving towards BSA, we are still missing how to compute the barycentric projection. Indeed, none of the algorithms we proposed so far are valid in this context, mainly because they rely on continuous optimization. Therefore, we need to approach the problem differently. There are two main obstacles to implementing an algorithm

in our context here. First, since the horizontal logarithm relies on an alignment procedure, which does not have a closed form solution a priori, the projection problem suffers from combinatorial complexity. Moreover, since the horizontal logarithm is discontinuous, then so are the barycentric subspaces. The problem is that wherever these discontinuities occur is not known in closed form either, since it also depends directly on the alignment problem. More importantly, the projection problem therefore raises what we call geometric complexity. To illustrate such, let us simplify the problem a little bit. In the specific case of graphs, the top manifold M is a Euclidean space equipped with its canonical metric (precisely, the Frobenius metric). If this is the case, we are able to write the Riemannian logarithm of M explicitly as a subtraction of vectors and we have that $\pi(x)$ is a barycenter of $\pi(a_0), \dots, \pi(a_k)$ with corresponding weights $w_0, \dots, w_k$ if and only if

$$x = \sum_i w_i t_i \cdot a_i \quad \text{where} \quad t_i = \underset{t \in T}{\arg\min}\ d(x, t \cdot a_i). \tag{4.5}$$

Although this new formulation is more explicit, it remains implicitly related to the alignment problem. As we have no closed form solution for such in the case of graphs, we propose instead in Figure 4.1 to illustrate the double complexity of the projection problem on the simple example of the action on the plane of the point reflection across the origin. In this specific case, we are able to compute the discontinuities of the barycentric subspaces explicitly. This is not true anymore in higher dimensions or in more complex quotient spaces. Therefore, locating the discontinuities and computing the barycentric projection require in the general case an exhaustive search. Instead, we explored two algorithms that approximate the solution of such a problem. We do not present them here, but which the reader may find in our paper (Calissano, Maignant, and Pennec, 2023). Essentially, although one of the two algorithms proved accurate in most cases, both failed drastically for some specific reference point distributions, even in our simple example. Faced with this conclusion, we decided to take up the problem of implementing BSA for sets of graphs from the beginning and investigated a new graph space, one that would be more regular and therefore easier to compute with.

4.2 Spectral graph spaces

Drawing lessons from our first approach, we set out to build a model that would be compatible with continuous Riemannian geometry and therefore with the work of Chapter 2 and Chapter 3. As a relaxation of the previous graph spaces, we introduce spectral graph spaces to be the quotient spaces resulting from the action of rotations, rather than the one of permutations, on the set of adjacency matrices. This approach consists somehow in a symmetric version of the one by Severn, Dryden, and Preston (2022), where the orthogonal group acts on the right only, more similarly to a Procrustean approach than to the original invariant approach for graphs we described in Section 4.1. Now, beyond the idea of relaxation, our approach takes on a natural interpretation from the point of view of spectral graph theory, that is the study of graphs through their spectrum, a descriptor that has proved to encode numerous properties (Chung, 1996; Brouwer and Haemers, 2012). Indeed, the action of rotations by conjugation preserves the eigenvalues of a given graph – represented by its adjacency matrix – and the resulting equivalence classes are precisely the sets of cospectral graphs, that is graphs sharing the same set of eigenvalues (Haemers and Spence, 1995). As we shall see, our framework also coincides with one of the spectral distances mentioned in Section 4.1. Finally, the main theoretical interest of spectral graph spaces is that they are naturally endowed with the very specific Riemannian structure of a Weyl chamber. In practice, this means that, first, BSA is well defined on the whole spaces, and moreover, that it comes with simple implementations, in particular because it turns out that the barycentric subspaces can be described explicitly, such that

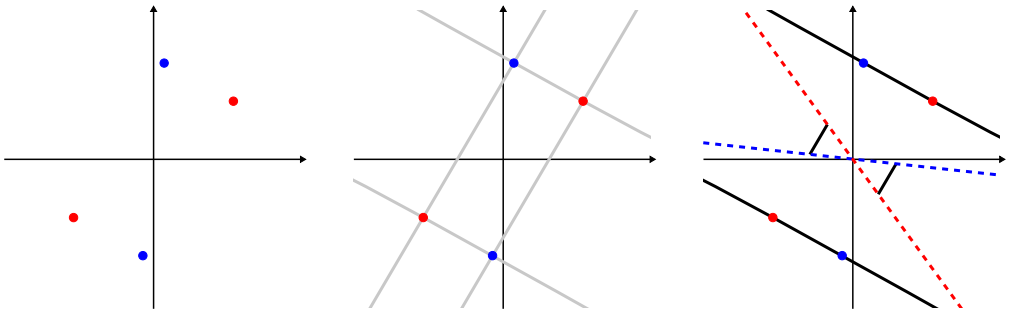

Figure 4.1: Geometry of a quotient barycentric subspace. (Left) We consider the orbits of two points in the plane (in red and blue) under action of the point reflection across the origin $x \mapsto -x$. (Center) According to Equation (4.5), a barycenter of the two orbits is always written in top space as a weighted sum of a representative of the first point and a representative of the second point (in gray). In total, all these weighted sums generate a priori 4 different lines in the plane. (Right) Now among these weighted sums, only a subset (in black) satisfy the alignment condition in Equation (4.5). For this specific quotient space, we can invert such an alignment condition to obtain an explicit condition on the barycentric weights for the weighted sum to project on a valid barycenter. It divides the lines generated by all the weighted sums into valid and non valid segments such that even in this very simplistic example, the quotient barycentric subspace consists already in 4 disconnected components. Additionally, we may also solve the problem geometrically. The red (respectively blue) dashed line separates the plane into the points that are aligned on one representative of the red (respectively blue) orbit on one hand and the points that are aligned on the other representative on the other hand. These two lines parametrize exactly the valid segments.

the computation of the barycentric projection is straightforward. Through several examples, we are able to illustrate that BSA is a powerful dimensionality reduction tool particularly suited to the analysis of a set of graphs.

Spectral graph spaces We consider undirected weighted graphs with a finite set of nodes (vertices). Such a graph is given as a triplet $G = (V, E, a)$, where $V = \{1, \dots, p\}$ denotes the set of nodes, E is a set of paired nodes referred to as edges, and $a : E \to \mathbb{R}$ is a scalar attribute (weight) function defined on the edges. We allow for graphs to have self-loops or node attributes, that is for $a\{i, i\}$ to be non-zero. A weighted graph can be represented by its weighted adjacency matrix $x \in \mathbb{R}^{p \times p}$ defined by

$$x_{ij} = x_{ji} = \begin{cases} a\{i,j\} & \text{if } \{i,j\} \in E \\ 0 & \text{else} \end{cases} \tag{4.6}$$

and embedded as a point in the space $\mathrm{Sym}(p)$ of symmetric matrices of size p:

$$\mathrm{Sym}(p) = \left\{ x \in \mathbb{R}^{p \times p} \mid x^T = x \right\} \tag{4.7}$$

In an effort to relax the action of the permutation group while sticking to the original idea of classifying graphs according to their structure, we consider the conjugation action of the orthogonal group $\mathrm{O}(p)$ on $\mathrm{Sym}(p)$

$$(R, x) \mapsto RxR^T \tag{4.8}$$

which transforms one graph into another cospectral (or isospectral) graph, that is a graph whose adjacency matrix shares the same spectrum. The orbit of a graph consists then exactly of all the other cospectral graphs. As the spectrum of a graph contains a lot of information about its topological

structure like connectivity or regularity (Haemers and Spence, 1995; Halbeisen and Hungerbühler, 2000; Van Mieghem, 2010), identifying a graph with its orbit amounts to identify graphs with a similar structure. In particular, a given graph is identified with all the other graphs obtained by permuting the nodes as illustrated in Figure 4.2. We then represent a graph as a point in the quotient space Γ_p defined by

$$\Gamma_p = \mathrm{Sym}(p)/\,\mathrm{O}(p) \tag{4.9}$$

It is a differentiable manifold with singularities at the points x where the action of $\mathrm{O}(p)$ is not free. We can show that these points correspond exactly to the symmetric matrices with at least two equal eigenvalues. Let then $\mathrm{Sym}^{\mathrm{reg}}(p)$ be the subset of symmetric matrices whose eigenvalues are pairwise distinct. It is an open subset of $\mathrm{Sym}(p)$ and it is dense in $\mathrm{Sym}(p)$. On $\mathrm{Sym}^{\mathrm{reg}}(p)$, the action of $\mathrm{O}(p)$ is proper and free such that the quotient space $\Gamma_p^{\mathrm{reg}} = \mathrm{Sym}^{\mathrm{reg}}(p)/\,\mathrm{O}(p)$ is a differentiable manifold.

Definition 4.2.1. *The quotient space Γ_p is called the spectral graph space of size p.*

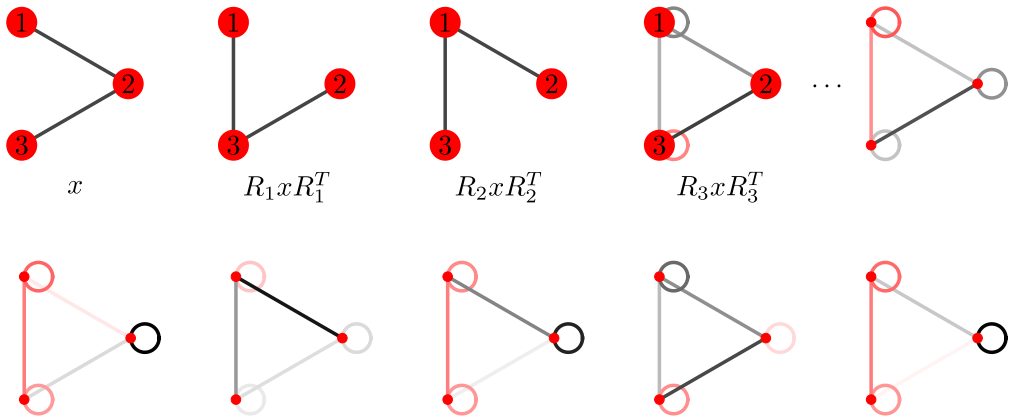

Figure 4.2: Equivalence class of a graph. The nodes of the graphs are represented by red dots (or filled circles) and the edge connecting two nodes is represented by the segment that joins the two corresponding dots. The nodes are always arranged clockwise but we added the labels on the first graphs to make it easier to understand. Edges with positive (respectively negative) weight are colored in black (respectively red) while the opacity of one edge is proportional to the absolute value of its weight. As for the length of the edges being constant, it is only a choice of layout and it does not encode any information. Self loops (node attributes) are represented by circular edges using the same color code. The same color code and layout hold for all the examples of this chapter. Here the first graph is given by its adjacency matrix x and the other graphs correspond each to the action of a rotation $R_i \in \mathrm{O}(p)$ on x by conjugation. The two first correspond more specifically to the action of a permutation. Self-loops occur almost every time a rotation – that is not a permutation – acts on the original graph.

4.2.1 The L^2 spectral distance

A straightforward way of providing spectral graph spaces with a distance function is to define what is called a quotient distance (Younes, 2010). Let us fix $p \in \mathbb{N}^*$ from now on. The space $\mathrm{Sym}(p)$ carries naturally a distance function induced by the Frobenius norm. One might check easily that this distance function is invariant with respect to the action of $\mathrm{O}(p)$ by conjugation. Therefore, if $\pi : \mathrm{Sym}(p) \to \Gamma_p$ denotes the canonical projection map, then the function

$$d_*\left(\pi(x), \pi(y)\right) = \inf_{R \in \mathrm{O}(p)} \left\| RyR^T - x \right\| \tag{4.10}$$

is well defined and satisfies all the axioms of a distance function which we can equip the graph space with. Now we have the following

Proposition 4.2.1. *Let $x, y \in \mathrm{Sym}(p)$. Let $\lambda_1 \leq \ldots \leq \lambda_p$ (respectively $\mu_1 \leq \ldots \leq \mu_p$) denote the eigenvalues of x (respectively y). Then the distance from $\pi(x)$ to $\pi(y)$ can be reformulated as*

$$d_* \left(\pi(x), \pi(y)\right) = \left(\sum_{i=1}^{p} \left|\mu_i - \lambda_i\right|^2 \right)^{1/2}, \tag{4.11}$$

such that it belongs to the family of spectral distances (Donnat and Holmes, 2018). We refer more precisely to the distance function d_* as the L^2 spectral distance. The optimal orthogonal matrix aligning y on x is given by

$$R_x(y) = R_1 R_2^T \tag{4.12}$$

where $x = R_1 \operatorname{diag}\left(\lambda_1, \ldots, \lambda_p\right) R_1^T$ and $y = R_2 \operatorname{diag}\left(\mu_1, \ldots, \mu_p\right) R_2^T$.

Proof. Consider the functional

$$d_{x,y} : R \mapsto \|RyR^T - x\|^2 \tag{4.13}$$

Its gradient at $R \in \mathrm{O}(p)$ with respect to the Frobenius metric on $T_R\,\mathrm{O}(p) = \mathrm{Skew}(p) \cdot R$ is given by

$$\operatorname{grad} d_{x,y}(R) = 2RyR^T x - 2xRyR^T \in \mathrm{Skew}(p) \tag{4.14}$$

Now the functional $d_{x,y}$ is minimized only if the first order condition

$$\operatorname{grad} d_{x,y}(R) \in \mathrm{Skew}(p)^{\perp} = \mathrm{Sym}(p) \tag{4.15}$$

applies. This condition is satisfied if and only if x and RyR^T commute with each other. Assume then that R is such that it is the case. Then x and RyR^T may be diagonalised in a common orthonormal basis

$$x = R_{x,y} D_x R_{x,y} \qquad \text{and} \qquad RyR^T = R_{x,y} D_y R_{x,y} \tag{4.16}$$

and we have

$$d_{x,y}(R) = \|D_x - D_y\|^2 \tag{4.17}$$

Up to a permutation σ, this amounts exactly to

$$d_{x,y}(R) = \sum_i \left|\lambda_i(x) - \lambda_{\sigma(i)}(y)\right|^2 \tag{4.18}$$

Finally, the rearrangement inequality states that

$$\sum_{i=1}^{p} \lambda_i \mu_{\sigma(i)} \leq \sum_{i}^{p} \lambda_i \mu_i \tag{4.19}$$

for any permutation $\sigma \in \mathfrak{S}(p)$. Therefore, the functional $d_{x,y}$ is minimal only if $\sigma = id$ and its minimum is the quantity

$$d_{x,y}(R) = \sum_i \left|\lambda_i(x) - \lambda_i(y)\right|^2. \tag{4.20}$$

Moreover, the reader may wan to check that such a choice of σ corresponds exactly to taking $R = R_1 R_2^T$. Reversely, if $R = R_1 R_2^T$ then the previous optimal equality is satisfied. □

4.2.2 The Riemannian structure of spectral graph spaces

Now moving towards a more sophisticated analysis of graphs, we propose to equip spectral graph spaces with a Riemannian structure which is easy to derive and implement. Precisely, we construct a Riemannian metric on Γ_p^{reg} in such a way that the quotient map π is a Riemannian submersion, similarly to what we described for Kendall shape spaces in Chapter 3. Once again, we leverage such a construction to compute the Riemannian structure of Γ_p^{reg} from the structure of the total space $\text{Sym}^{\text{reg}}(p)$ which is easier to operate. Finally, we discuss how this Riemannian structure extends to the whole graph space Γ_p.

Vertical and horizontal bundles We recall that the vertical subspace of $\text{Sym}^{\text{reg}}(p)$ at a point x consists of all the tangent vector at x that are tangent to the fiber $\pi^{-1}(x)$ while the horizontal subspace at x is the orthogonal complement of such in the tangent space at x with respect to the Riemannian metric of $\text{Sym}^{\text{reg}}(p)$. We choose this metric to be that induced by the Frobenius inner product $\langle \cdot, \cdot \rangle$. This choice is motivated later when constructing the Riemannian metric on $\text{Sym}^{\text{reg}}(p)/\,\text{O}(p)$. We derive an explicit description for both the vertical bundle and the horizontal bundle of $\text{Sym}^{\text{reg}}(p)$.

Proposition 4.2.2. *Let $x \in \text{Sym}^{\text{reg}}(p)$. The vertical subspace of $\text{Sym}^{\text{reg}}(p)$ at x is*

$$\text{Ver}_x \,\text{Sym}^{\text{reg}}(p) = \{Ax - xA,\ A \in \text{Skew}(m)\}\,. \tag{4.21}$$

Its dimension is $p(p-1)/2$.

Proof. $\text{Ver}_x \,\text{Sym}^{\text{reg}}(p)$ consists of exactly the tangent vectors at x which are also tangent to the fiber F of π at $\pi(x)$. Consider a curve γ in the fiber F such that $\gamma(0) = x$. It can be written as

$$\gamma(s) = R(s)xR(s)^T \tag{4.22}$$

where R is a curve in $\text{O}(p)$ with $R(0) = Id$. Then let us compute the derivative of γ at 0. We have

$$\dot{\gamma}(0) = \dot{R}(0)xR(0)^T + R(0)x\dot{R}(0)^T. \tag{4.23}$$

Since $\text{O}(p)$ is a Lie group of Lie algebra $\text{Skew}(p)$, then there exists $A \in \text{Skew}(p)$ such that

$$\dot{R}(0) = A. \tag{4.24}$$

Therefore, we have

$$\dot{\gamma}(0) = Ax + xA^T. \tag{4.25}$$

□

Proposition 4.2.3. *Let $x \in \text{Sym}^{\text{reg}}(p)$. The horizontal subspace of $\text{Sym}^{\text{reg}}(p)$ at x is*

$$\text{Hor}_x \,\text{Sym}^{\text{reg}}(p) = \{u \in \text{Sym}(p) \mid xu = ux\}\,. \tag{4.26}$$

If $x = R\,\text{diag}(\lambda_1, \ldots, \lambda_p)R^T$ with $R \in \text{O}(p)$, then it identifies with

$$\text{Hor}_x \,\text{Sym}^{\text{reg}}(p) = \left\{R\,\text{diag}(\mu_1, \ldots, \mu_p)R^T \mid (\mu_1, \ldots, \mu_p) \in \mathbb{R}^p\right\}. \tag{4.27}$$

Proof. The horizontal subspace at x is defined as the orthogonal of the vertical space at x in the tangent space at x

$$\operatorname{Hor}_x \operatorname{Sym}^{\mathrm{reg}}(p) + \operatorname{Ver}_x \operatorname{Sym}^{\mathrm{reg}}(p) = T_x \operatorname{Sym}^{\mathrm{reg}}(p). \tag{4.28}$$

Let $u \in \operatorname{Hor}_x \operatorname{Sym}^{\mathrm{reg}}(p)$. Then for all $A \in \operatorname{Skew}(m)$ we have

$$\langle u, Ax - xA \rangle = 0 \tag{4.29}$$

that is

$$\langle xu - ux, A \rangle = 0. \tag{4.30}$$

Finally, since $xu - ux$ is skew-symmetric, this is equivalent to

$$xu - ux = 0. \tag{4.31}$$

□

Riemannian metric Since the vertical space of $\operatorname{Sym}^{\mathrm{reg}}(p)$ at a point x identifies with $\ker d_x\pi$, then the map $d_x\pi$ defines an isomorphism between the horizontal space at x and its image $T_{\pi(x)}\Gamma_p^{\mathrm{reg}}$. Now assuming that we equipped Γ_p^{reg} with a Riemannian metric g, then the quotient map π is a Riemannian submersion if and only if each of the maps $d_x\pi$ is an isometry. In other words, π is a Riemannian submersion if an only if $d\pi$ maps any two horizontal vectors at x to two tangent vectors $d\pi_x(u)$ and $d\pi_x(v)$ at $\pi(x) \in \Gamma_p^{\mathrm{reg}}$ with the same inner product. Let then $x = R\operatorname{diag}(\lambda_1, \ldots, \lambda_p)R^T$ with $R \in \mathrm{O}(p)$. According to the previous proposition, we can write $u = R\operatorname{diag}(\mu_1, \ldots, \mu_p)R^T$ and $v = R\operatorname{diag}(\nu_1, \ldots, \nu_p)R^T$. The Frobenius inner product of u and v is then $\langle u, v \rangle = \operatorname{tr}\left(R\operatorname{diag}(\mu_1, \ldots, \mu_p)R^T R\operatorname{diag}(\nu_1, \ldots, \nu_p)R^T\right)$ which simply reduces to $\sum_{i=1}^p \mu_i\nu_i$ such that it does not depend on the location of x on the fiber indicated by R. Therefore the inner product

$$g^*_{\pi(x)}\left(d_x\pi(u), d_x\pi(v)\right) = \sum_{i=1}^{p} \mu_i\nu_i \tag{4.32}$$

is well defined and we have the following

Proposition 4.2.4. *The tensor field g^* defined by (4.32) is a Riemannian metric on Γ_p^{reg}. Moreover, away from the singularities, the quotient map π is a Riemannian submersion.*

Exponential map and Riemannian logarithm Proposition 4.2.4 allows us to leverage O'Neill's theorems (O'Neill, 1966). In particular, one of these tells us that locally the geodesics of Γ_p^{reg} are the projections by the submersion π of horizontal geodesics – that is the geodesics spanned by horizontal vectors – of the top space Γ_p^{reg}. It allows to write explicit formulas for the exponential map and the logarithm map in Γ_p^{reg}.

Proposition 4.2.5. *Let $x \in \operatorname{Sym}^{\mathrm{reg}}(p)$ and $u \in \operatorname{Hor}_x \operatorname{Sym}^{\mathrm{reg}}(p)$. Provided that u is sufficiently small such that the local assumption mentioned above holds, the Riemannian exponential of $d_x\pi(u)$ at $\pi(x)$ is*

$$\exp^*_{\pi(x)}(d_x\pi(u)) = \pi(x + u). \tag{4.33}$$

Let $y \in \operatorname{Sym}^{\mathrm{reg}}(p)$. The Riemannian logarithm of $\pi(y)$ at $\pi(x)$ is

$$\log^*_{\pi(x)}(\pi(y)) = d\pi_x\pi(R_x(y)yR_x(y)^T - x) \tag{4.34}$$

where $R_x(y)$ is the orthogonal matrix aligning y on x as given in Proposition 4.2.1.

Graph spectral spaces are isometric to cones Finally, let us formulate a fundamental statement on spectral graph spaces linking together all the observations made so far. We derive a series of identities.

Theorem 4.2.1. *Now let $\mathcal{C}_p$ be the convex cone of ordered p-tuples of $\mathbb{R}$ and let $\lambda : \mathrm{Sym}(p) \to \mathcal{C}_p$ be the map defined by*

$$\lambda : x \mapsto (\lambda_1(x), \ldots, \lambda_p(x)) \tag{4.35}$$

where $\lambda_1(x) \leq \ldots \leq \lambda_p(x)$ are the eigenvalues of x. There exists a unique diffeomorphism $f : \Gamma_p^{\mathrm{reg}} \to \operatorname{int} \mathcal{C}_p$ which satisfies $\lambda = f \circ \pi$ and its inverse map is given by

$$f^{-1}(\lambda_1, \ldots, \lambda_p) = \pi\left(\mathrm{diag}(\lambda_1, \ldots, \lambda_p)\right). \tag{4.36}$$

Then $f : (\Gamma_p^{\mathrm{reg}}, g) \to (\operatorname{int} \mathcal{C}_p, \langle \cdot, \cdot, \rangle)$ is an isometry and we have the following

1. *Let $x \in \mathrm{Sym}^{\mathrm{reg}}(p)$. Let $u, v \in \mathrm{Hor}_x \mathrm{Sym}^{\mathrm{reg}}(p)$. Then the metric tensor at $\pi(x)$ is given by*

$$g^*_{\pi(x)}\left(d_x\pi(u), d_x\pi(v)\right) = \langle \lambda(u), \lambda(v) \rangle. \tag{4.37}$$

2. *Let $x, y \in \mathrm{Sym}^{\mathrm{reg}}(p)$. The Riemannian distance from $\pi(x)$ to $\pi(y)$ is given by*

$$d_*\left(\pi(x), \pi(y)\right) = \|\lambda(y) - \lambda(x)\|. \tag{4.38}$$

   and it corresponds exactly to the L^2 spectral distance from x to y.

3. *Let $x \in \mathrm{Sym}^{\mathrm{reg}}(p)$ and $u \in \mathrm{Hor}_x \mathrm{Sym}^{\mathrm{reg}}(p)$. The exponential map at $\pi(x)$ is given by*

$$\exp^*_{\pi(x)}(d_x\pi(u)) = f^{-1}\left(\lambda(x) + \lambda(u)\right). \tag{4.39}$$

4. *Let $x, y \in \mathrm{Sym}^{\mathrm{reg}}(p)$. The Riemannian logarithm of $\pi(y)$ at $\pi(x)$ is given by*

$$\log^*_{\pi(x)}(\pi(y)) = df^{-1}_{\lambda(x)}\left(\lambda(y) - \lambda(x)\right). \tag{4.40}$$

4.2.3 About an enlightening discussion with Peter Michor

As a final note to this second section, we are reporting here on a discussion we had with Peter Michor just a few months ago, and which sheds new light on spectral graph spaces. On reading our definition of spectral graph spaces, Peter Michor immediately declared: "it is a Weyl chamber". What a Weyl chamber is and what it means in the case of orbit spaces is very well detailed by Alekseevsky, Kriegl, Losik, and Michor (2003) in their paper available on Peter Michor's personal page. In particular, we refer to Proposition 3.2 and Example 4.5. Let us just comment on this example in the specific case of spectral graph spaces. In a nutshell, the vector space of diagonal matrices $\mathrm{Diag}(p)$ meets all the orbits of the action of the orthogonal group $\mathrm{O}(p)$ on the space $\mathrm{Sym}(p)$ of symmetric matrices orthogonally. Therefore, the quotient space $\Gamma_p = \mathrm{Sym}(p)/\mathrm{O}(p)$ is isometric (as a metric space) to the Weyl chamber $\mathrm{Diag}(p)/W$ where W denotes the Weyl group of $\mathrm{Diag}(p)$. In fact, we can show that W is the group $\mathfrak{S}(p)$ of permutations such that the space $\mathrm{Diag}(p)/W$ is exactly the cone $\mathcal{C}_p$ we introduced in the theorem just above. In other words, f defines a isometry between the whole graph space (Γ_p, d_*) and the closed cone $\mathcal{C}_p$. Now an important result following these statements is that any two points of Γ_p are always joined by a unique minimal geodesic segment, coinciding with the minimizing geodesic of Γ^{reg}, which identifies with the straight segment joining the two corresponding points in $\mathcal{C}_p$. The case of geodesics of $\mathrm{Sym}(p)/O(p)$ is also specifically covered by Example 34.17 of Peter Michor's book *Topics in differential geometry* (Michor, 2008). Based on

this result and working towards the next section on barycentric subspace analysis, we extend our definition of the Riemannian logarithm of a graph y at a graph x to the boundaries of Γ_p as the vector spanning the minimal segment from x to y. Peter Michor would have called such a logarithm the restricted logarithm because it corresponds to restricting the Riemannian logarithm of $\mathrm{Diag}(p)$ to the subbundle of the tangent bundle which consists of vectors that point inside the chamber only. We refer to it simply as the logarithm and keep the same notation as before. Similarly, we extend the tangent map df to the boundary of Γ_p as the differential of f only in these directions that point inside the chamber. In practice, the main consequence of this discussion is that we are now able to account for singular graphs such as star-shaped graphs within our framework. Finally, we could only recommend to read the paper of Peter Michor and his co-authors and we sincerely thank him for his precious help on this project.

4.3 Barycentric subspace analysis on spectral graph spaces

Take N graphs $\pi(x_1), \dots, \pi(x_N)$ in the graph space Γ_p. Barycentric Subspace Analysis (BSA) consists in finding reference graphs $\pi(a_0), \dots \pi(a_k)$ whose barycentric subspace minimizes the distance to the data points. Thanks to the results of the previous section, BSA has a very nice formulation on spectral graph spaces. As an example, let us start with the simplest case where $k = 0$. In this case, BSA consists exactly in computing the Fréchet mean of the dataset. Indeed since the barycentric subspaces of dimension 0 are just points, the minimization problem simply consists in minimizing the sum of squared distances to the data points. Leveraging the closed-form of L^2 spectral distance d_*, we can show easily that this problem has a unique solution given by

$$m = \frac{1}{N} \sum_{i=1}^{N} \mathrm{diag}(\lambda(x_i)). \tag{4.41}$$

Now to the general case, and let us show that we may still derive a similar simplification of the barycenters and the BSA.

4.3.1 Barycentric subspaces of spectral graph spaces

Let us first recall the definition of barycentric subspaces when the space is the graph space Γ_p. We just saw in the previous section that any two points of Γ_p are always joined by a unique minimal geodesic such that the logarithm may be extended to the whole space. Therefore, the critical condition for a point to be a barycenter of a given set of reference points is defined everywhere on Γ_p and we have

Definition 4.3.1. *Let $a_0, \dots a_k \in \mathrm{Sym}(p)$. The barycentric subspace of $\pi(a_0), \dots \pi(a_k) \in \Gamma_p$ is defined as follows*

$$\mathrm{BS}\left(\pi(a_0), \dots, \pi(a_k)\right) = \bigcup_{\substack{w_0,\dots,w_k \in \mathbb{R} \\ w_0+\dots+w_k=1}} \mathrm{bar}(\pi(a_i), w_i)_{0 \leq i \leq k} \tag{4.42}$$

where we recall that the set $\mathrm{bar}(\pi(a_i), w_i)_{0 \leq i \leq k}$ consists of all the barycenters of $\pi(a_0), \dots \pi(a_k)$ with corresponding weights $w_0, \dots, w_k \in \mathbb{R}$ and is given by

$$\mathrm{bar}(\pi(a_i), w_i)_{0 \leq i \leq k} = \Big\{ \pi(x) \in \Gamma_p \mid \sum_{i=0}^{k} w_i \log^*_{\pi(x)} \pi(a_i) = 0 \Big\}. \tag{4.43}$$

Finally, we also recall that the points $\pi(a_0), \dots \pi(a_k)$ are then called reference points of the subspace $\mathrm{BS}\left(\pi(a_0), \dots, \pi(a_k)\right)$.

Barycentric subspaces of spectral graph spaces are convex polytopes We can show that the geometry of barycentric subspaces of spectral graph spaces is exactly that of convex polytopes. Moreover the equations defining the polytope that corresponds to a given barycentric subspace may be explicitly computed from the reference points of the subspace themselves. We illustrate our result on a concrete example.

Theorem 4.3.1. *Let $a_0, \dots a_k \in \mathrm{Sym}(p)$. Then the barycentric subspace of $\pi(a_0), \dots \pi(a_k) \in \Gamma_p$ is isometric in $\mathcal{C}_p$ to the set*

$$\mathcal{P}(a_0, \dots, a_k) = \Big\{ \sum_{i=0}^{k} w_i \lambda(a_i) \mid \sum_{i=0}^{k} w_i = 1;\ \forall 1 \leq q \leq p-1,\ \sum_{i=0}^{k} w_i \lambda_q(a_i) \leq \sum_{i=0}^{k} w_i \lambda_{q+1}(a_i) \Big\} \quad (4.44)$$

which is a convex polytope defined as the intersection of the affine space spanned by $\lambda(a_0), \dots, \lambda(a_k)$ with the $p-1$ half-spaces

$$\alpha_q^T \lambda(x) \geq \beta_q, \quad 1 \leq q \leq p-1 \quad (4.45)$$

where for $1 \leq q \leq p-1$, the pair (α_q, β_q) solves

$$\begin{bmatrix} \lambda_1(a_0) & \dots & \lambda_p(a_0) & 1 \\ \vdots & & \vdots & \vdots \\ \lambda_1(a_k) & \dots & \lambda_p(a_k) & 1 \end{bmatrix} \begin{bmatrix} \alpha_{q,1} \\ \vdots \\ \alpha_{q,p} \\ -\beta_q \end{bmatrix} = \begin{bmatrix} \lambda_{q+1}(a_0) - \lambda_q(a_0) \\ \vdots \\ \lambda_{q+1}(a_k) - \lambda_q(a_k) \end{bmatrix} \quad (4.46)$$

Moreover, it contains the convex hull of $a_0, \dots, a_k$.

Proof. Let $x \in \mathrm{Sym}(p)$. Then $\pi(x)$ belongs to the barycentric subspace of $\pi(a_0), \dots, (a_k)$ if and only if there exists $w_0, \dots, w_n \in \mathbb{R}$ summing to 1 such that

$$\sum_{i=0}^{k} w_i \log^*_{\pi(x)} (\pi(a_i)) = 0. \quad (4.47)$$

Following Theorem 4.2.1, this condition translates into

$$\sum_{i=0}^{k} w_i df^{-1}_{\lambda(x)} (\lambda(a_i) - \lambda(x)) = 0. \quad (4.48)$$

Now since $df^{-1}_{\lambda(x)}$ is a bijective linear map, it is equivalent to

$$\sum_{i=0}^{k} w_i (\lambda(a_i) - \lambda(x)) = 0. \quad (4.49)$$

Adding the constraint that $w_0, \dots, w_k$ have to sum to 1, we get

$$\lambda(x) = \sum_{i=0}^{k} w_i \lambda(a_i). \quad (4.50)$$

Now we recall that by definition of λ, we have $\lambda(x) \in \mathcal{C}_p$. Therefore, $w_0, \dots, w_k$ should satisfy

$$\forall 1 \leq q \leq p-1, \quad \sum_{i=0}^{k} w_i \lambda_q(a_i) \leq \sum_{i=0}^{k} w_i \lambda_{q+1}(a_i). \quad (4.51)$$

Let us now derive from Equations (4.50) and (4.51) a simple geometric constraint on $\lambda(x)$. Let us denote

$$\Lambda = \begin{bmatrix} \lambda_1(a_0) & \dots & \lambda_1(a_k) \\ \vdots & & \vdots \\ \lambda_p(a_0) & \dots & \lambda_p(a_k) \\ 1 & \dots & 1 \end{bmatrix} \quad \text{and} \quad \theta_q = \begin{bmatrix} \lambda_{q+1}(a_0) - \lambda_q(a_0) \\ \vdots \\ \lambda_{q+1}(a_k) - \lambda_q(a_k) \end{bmatrix} \tag{4.52}$$

On one hand, Equation (4.51) may be rearranged in

$$\forall 1 \leq q \leq p-1, \quad \begin{bmatrix} w_0 & \dots & w_k \end{bmatrix}^T \theta_q \geq 0 \tag{4.53}$$

and on the other hand, Equation (4.50) together with the constraint $\sum_{i=0}^{k} w_i = 1$ is expressed as

$$\begin{bmatrix} \lambda_1(x) & \dots & \lambda_p(x) & 1 \end{bmatrix} = \Lambda \begin{bmatrix} w_0 & \dots & w_k \end{bmatrix}. \tag{4.54}$$

Notice that $\theta_q \in \ker(\Lambda)^{\perp}$. It follows that

$$\theta_q \in \mathrm{im}(\Lambda^T). \tag{4.55}$$

For all $1 \leq q \leq p-1$, let then $\alpha_{q,1}, \dots, \alpha_{q,p}$ and β_q solve

$$\Lambda^T \begin{bmatrix} \alpha_{q,1} & \dots & \alpha_{q,p} & -\beta_q \end{bmatrix} = \theta_q. \tag{4.56}$$

Then Equation (4.51) is equivalent to $\lambda(x)$ satisfying

$$\alpha_q^T \lambda(x) \geq \beta_q. \tag{4.57}$$

□

Example 4.3.1. *Let us consider 3 graphs with adjacency matrices a_0, a_1 and $a_2 \in \mathrm{Sym}(p)$ and let us position ourselves in the plane spanned by the 3 p-dimensional vectors $\lambda(a_0)$, $\lambda(a_1)$ and $\lambda(a_2)$. The barycentric subspace of $\pi(a_0), \pi(a_1)$ and $\pi(a_2)$ is isometric in the cone to the convex polygon $\mathcal{P}(a_0, a_1, a_2)$ of at most $p-1$ sides (see Figure 4.3). Now let us fix y_0, y_1 and $y_2 \in \mathbb{R}^2$ such that $\|y_i - y_j\| = \|\lambda(a_i) - \lambda(a_j)\|$. In other words, let us assign intrinsic coordinates to $\lambda(a_0)$, $\lambda(a_1)$ and $\lambda(a_2)$ within the plane they span. In practice, for $k \geq 4$, such coordinates are computed using multi-dimensional scaling. Then the equation of the half-plane (4.45) translates in $\mathbb{R}^2$ into the intrinsic equation $\gamma_q^T y \geq \delta_q$ where γ_q and δ_q solve*

$$\begin{bmatrix} y_{0,1} & y_{0,2} & 1 \\ y_{1,1} & y_{1,2} & 1 \\ y_{2,1} & y_{2,2} & 1 \end{bmatrix} \begin{bmatrix} \gamma_{q,1} \\ \gamma_{q,2} \\ -\delta_q \end{bmatrix} = \begin{bmatrix} \lambda_{q+1}(a_0) - \lambda_q(a_0) \\ \lambda_{q+1}(a_1) - \lambda_q(a_1) \\ \lambda_{q+1}(a_2) - \lambda_q(a_2) \end{bmatrix}. \tag{4.58}$$

To derive this equation, the proof is similar to the one of Theorem 4.3.1. It might be useful to introduce a the linear projection $L : \mathbb{R}^p \to \mathbb{R}^2$ onto the plane spanned by $\lambda(a_0)$, $\lambda(a_1)$ and $\lambda(a_2)$such that $L(\lambda(a_i)) = y_i$.

4.3.2 Sample-limited barycentric subspace analysis

In the context of statistical graph analysis, we are more particularly interested in sampled-limited Barycentric Subspace Analysis (sample-limited BSA) that we introduced in Chapter 2 following Pennec, 2018. Indeed, the specificity of this method consists in looking for reference points within the samples. In other words, it allows to retrieve descriptors which live by nature within our data

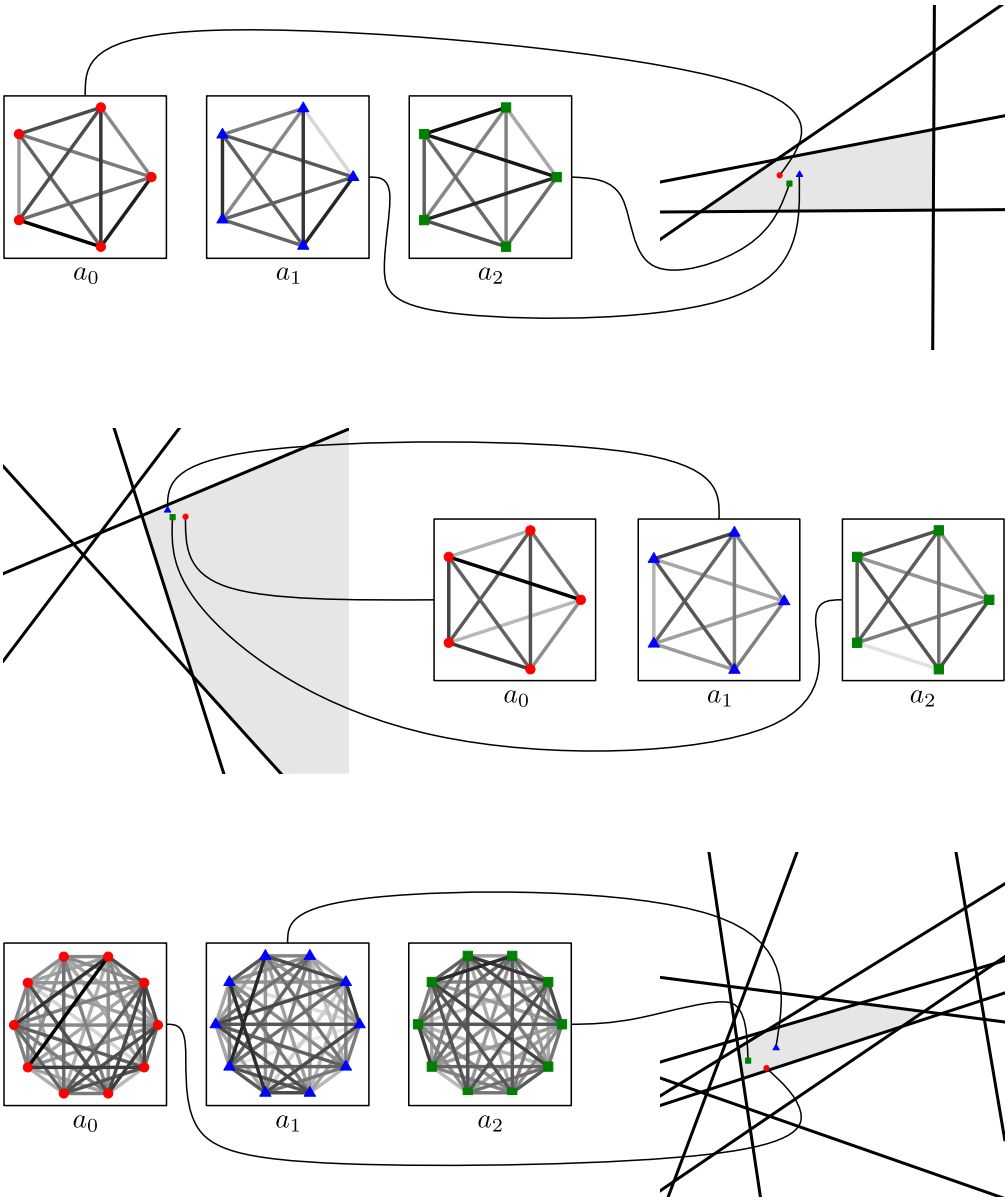

Figure 4.3: Barycentric subspace of 3 graphs. (Top) The barycentric subspace of the 3 graphs is isometric to a closed tetragon. It corresponds to the situation where the subspace has the maximum number of sides, that is one less than the number of nodes of the reference graphs. (Center) The second example illustrates a situation where the barycentric subspace is isometric to a polygon that is not closed. (Bottom) In the last example, where we doubled the number of nodes, the intersection of the half-planes is not minimal and the corresponding polygon is an hexagon.

space, therefore improving significantly the interpretability of such. Let us motivate sample-limited BSA on the simple example of the mean. We simulated a dataset which consists of two types of graphs such that none of the two types is included in the other (see Figure 4.4). We demonstrate that in this case, the Fréchet mean does not belong to any of the two types and is therefore hard to interpret. For $k = 0$, sample-limited BSA also provides a descriptor for the dataset, precisely the data point that minimizes the sum of squared distances to the other data points. But as it belongs to the dataset, this new descriptor is very straightforward to interpret.

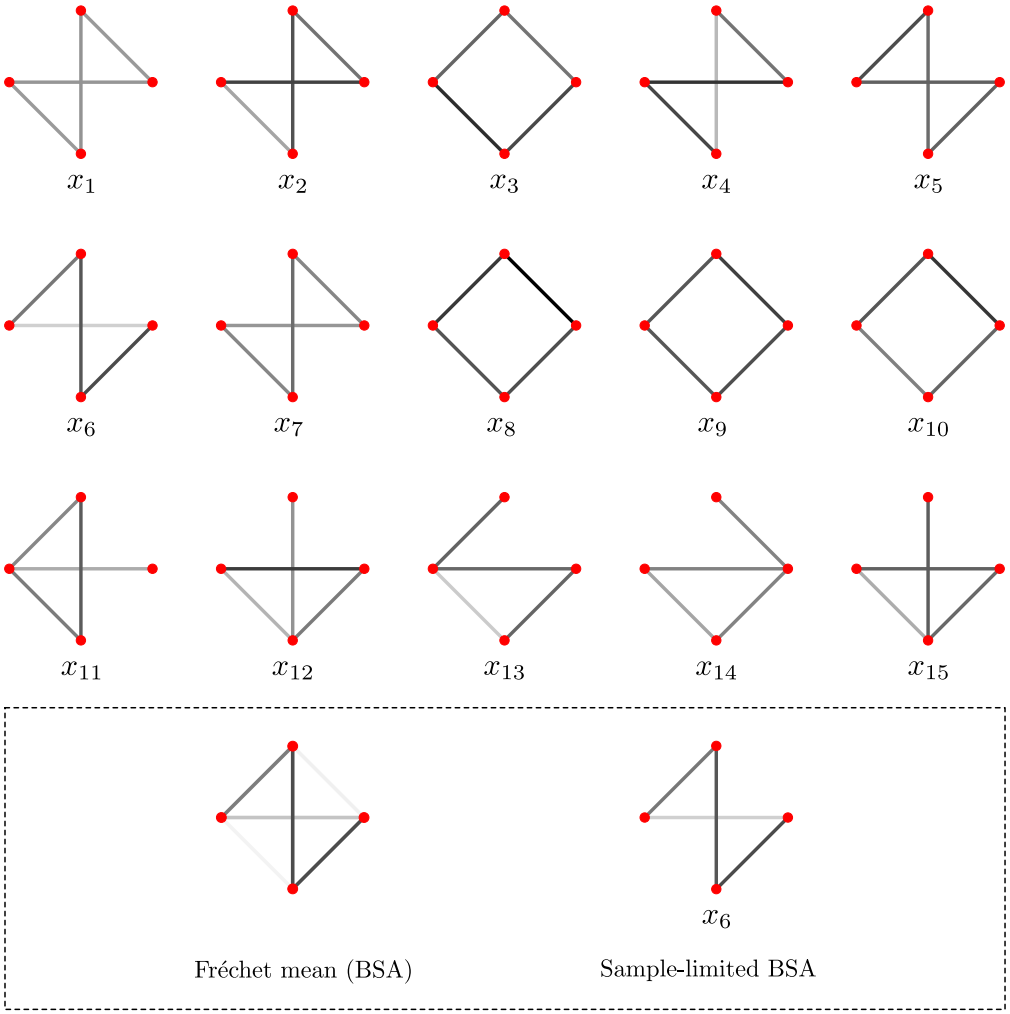

Figure 4.4: Fréchet mean and sample-limited BSA in dimension 0. The dataset consists of cyclic graphs (two first rows) and graphs that consist of edge glued to a cycle (third row). In the bottom of the figure, we show the Fréchet mean of the dataset and the graph picked in the data set by sample-limited BSA with $k = 0$. As we expected, we recover with the Fréchet mean a graph whose structure can not be identified exactly as one of the two structures that we observe within the original dataset. Sample-limited BSA on the other hand provides with a more meaningful descriptor. In particular, we observe that it retrieves without ambiguity the most represented graph structure within the dataset, that is the cyclic structure. Note that the Fréchet mean as it is given by Equation (4.41) is a diagonal matrix. Such a matrix would represent a graph with only self loops and therefore be challenging to interpret. Instead, we draw the graph that corresponds to the projection of the Fréchet mean onto the closest adjacency matrix with no weights on the diagonal. In practice, this projection problem may have multiple optimal solutions. Here it is initialized with the graph found by sample-limited BSA to achieve the most realistic vizualisation. We come back to such a question of the choice of a representative in more details in appendix.

Barycentric model Now for an arbitrary value of k, we recall that sample-limited BSA consist in identifying the subset $x_{i_0}, \dots, x_{i_k}$ of the dataset which minimizes the squared distance (here the L^2 spectral distance) to the data points (see Chapter 2 Equation (2.34)). Leveraging Theorem 4.3.1 on the geometry of barycentric subspaces of graph spectral spaces, we can formulate the optimization

problem posed by sample-limited BSA on the graph space Γ_p as an equivalent but simpler problem on the cone $\mathcal{C}_p$

$$\begin{aligned} \underset{\substack{1\leq i_0<\ldots<i_k\leq N \\ w_{ij}\in\mathbb{R} \\ w_{i0}+\ldots+w_{ik}=1}}{\text{minimize}} \quad & \sum_{i=1}^{N}\sum_{q=1}^{p}\left|\lambda_q(x_i)-\sum_{j=0}^{k} w_{ij}\lambda_q(x_{i_j})\right|^2 \\ \text{subject to} \quad & \sum_{j=0}^{k} w_{ij}\lambda_q(x_{i_j}) < \sum_{j=0}^{k} w_{ij}\lambda_{q+1}(x_{i_j}) \quad (\forall q, i). \end{aligned} \tag{4.59}$$

Such a problem can be split into the problem of picking the reference points and the usual barycentric projection problem. While the first problem is a combinatorial optimization problem, the projection problem itself is in this case a Euclidean convex optimization problem. Therefore any local minimum is a global minimum and the problem can be solved by usual gradient descent methods on $\mathbb{R}^k$. Finally, the main computational challenge of sample-limited BSA lies in the combinatorial search.

Convex barycentric subspace analysis To increase the interpretability of the representation, we may enforce the reference points – still taken within the samples – to be as extremal as possible. More precisely, we restrict the projection to the convex hull of the reference points and solve Problem (4.59) with the additional constraint that the barycentric weights must be positive. We refer to this method as convex sample-limited BSA. It is particularly suited for the identification of archetypal points. Another aspect is that convex sample-limited BSA should be more robust to data noise than standard BSA as the volume of the convex hull of given reference points is significantly smaller than that of their barycentric subspace (see Figure 4.3).

A comparison with tangent principal component analysis Let us look then for two descriptors of our dataset and the corresponding one dimensional representation. The barycentric coordinates obtained by projecting each sample onto the barycentric subspace of the two reference points allow to embed our samples in a line on which we have fixed the coordinates of the two reference points beforehand (see Figure 4.5). To underline further the interpretability of sample-limited BSA, we compare it with tangent Principal Component Analysis (tangent PCA) which we recall generalizes PCA to manifold-valued samples and consists in performing standard PCA in the tangent space at the Fréchet mean. As it requires only to compute Riemannian logarithms and the Fréchet mean, it is straightforward to implement on spectral graph spaces. When comparing tangent PCA and sample-limited BSA, we may want to emphasize two aspects. Regarding interpretability, we have seen that sample-limited BSA is very powerful. Essentially, the graphs selected as reference points allow to explain the projection space in terms of the original data. Such interpretability increases with the convex formulation. On the other hand, the components build by tangent PCA together with the mean are not always straightforward to interpret with respect to the sampled graphs. Moreover, as the projection space optimized by tangent PCA always includes the mean by construction, it is less likely to adapt to a non-Gaussian sampled distribution. Computationally speaking however, since sampled-limited BSA consists in a combinatorial search, it is a quite expensive procedure compared to tangent PCA.

4.3.3 Experiments and perspectives

This section presents two experiments we carried out on BSA on graph spectral spaces, with the underlying objective of illustrating the method's ability to classify graph data. The first experiment

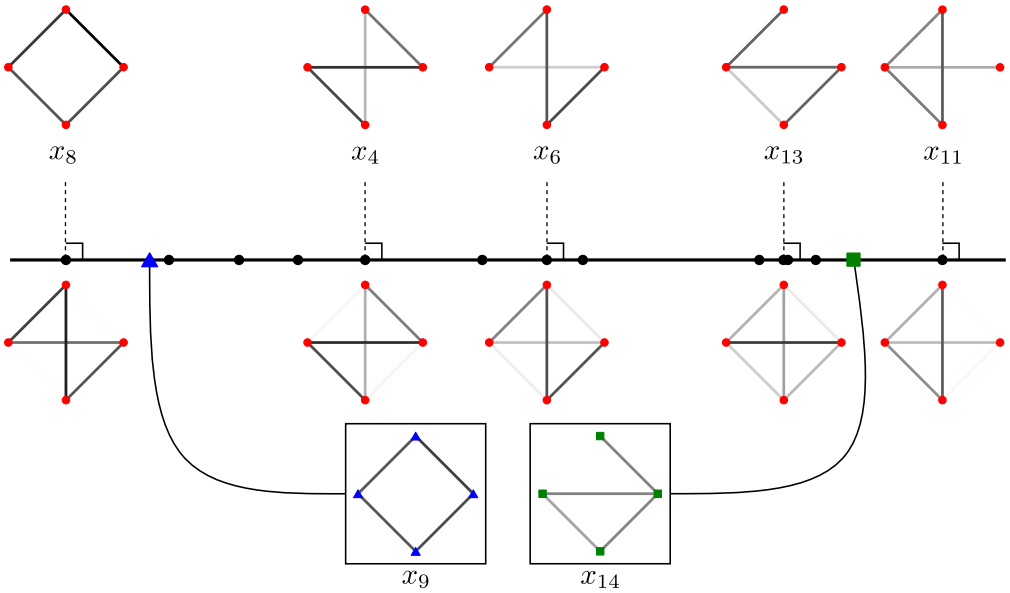

Figure 4.5: Sample-limited BSA in dimension 1. Graphs above the line correspond to graphs of the dataset while graphs below the line represent their projection on the barycentric subspace of the two reference points (in the middle). More precisely, they correspond to graphs without self-loops whose spectrum is that of the barycentric projection (computed thanks the procedure described in appendix initialized with the corresponding data point). The plots in the next two figures are generated in the same way. Sample-limited BSA naturally picks one graph of each structure as reference points. Such reference points are determining for the interpretability of the projection as the structure of the original graph data is not well preserved by projection.

was carried out on a simulated dataset for which classification is evident, and the second on a real dataset, for which we compare the result of BSA with the classification achieved by a standard clustering method.

Benchmark experiment on a simulated dataset In the first experiment, the dataset is made up of 3 different groups: fully connected, star-shaped and cyclic graphs (see Figure 4.8). Each of the groups consists in 5 graphs with 10 nodes, generated by assigning random attributes on the edges sampled from a Gaussian of mean 1 and standard deviation 1/2 while keeping the general connectivity structure fixed. We perform backward sample-limited BSA and its convex counterpart. Precisely, we build a decreasing nested sequence of barycentric subspaces that solve at each step the minimization problem posed by (convex) sample-limited BSA. We retrieve the squared projection error across the sequence and we trace the error profile, where the error is normalized dividing it by the size N of the dataset times the square of the maximal attribute observed in the dataset. From such profile, we can draw a rough prediction of the number of classes in the dataset. By stopping at precisely 3 reference points, we are able to capture the 3 different graph topologies within the dataset. Together with the 3 reference graphs selected, the visualization obtained by projecting the dataset on the corresponding 2-dimensional barycentric subspace provides a very rich description of the dataset. Practically, the visualization consists in first embedding the reference graphs in the plane in such a way that the pairwise distances are preserved (using multidimensional scaling for example) and then embedding the rest of the graph data according to their barycentric weights. It is exactly the same procedure as the one we used in Figure 4.3.

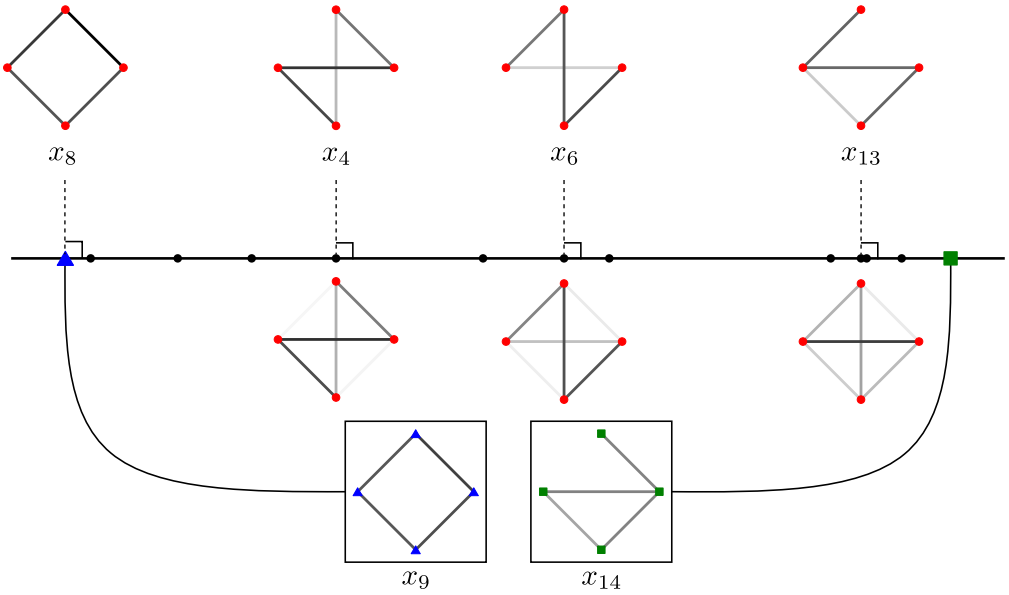

Figure 4.6: Convex sample-limited BSA in dimension 1. The fact that the same reference points are picked is only specific to this example. The extremal graph x_8 is projected on the reference point x_9 such that the two graphs are not distinct in the embedding space. Similarly, x_{11} is projected on the reference point x_{14}. Still, this allows to capture the structure of x_8 and x_{11} exactly. In comparison, the projection of x_{11} by non-convex sample-limited BSA carries a negative weighted edge.

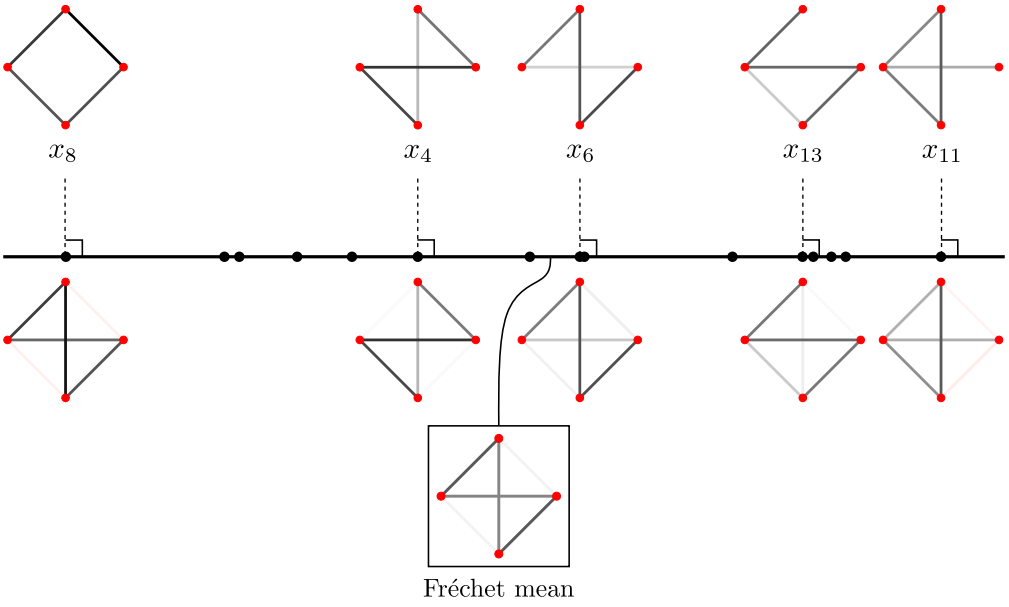

Figure 4.7: Tangent PCA in dimension 1. Except for the graph x_{13}, which is slightly better projected by this method, most projected graphs are not straightforward to interpret without reference points, especially because their structure is not well preserved. Now, regarding numerical performance, the overall projection squared error output by tangent PCA, divided by the size N of the dataset and the square of the maximal attribute over dataset, is $3,8 \times 10^{-2}$. On this aspect, tangent PCA is comparable to sample-limited BSA $(4,6 \times 10^{-2})$ and convex sample-limited BSA $(4,9 \times 10^{-2})$

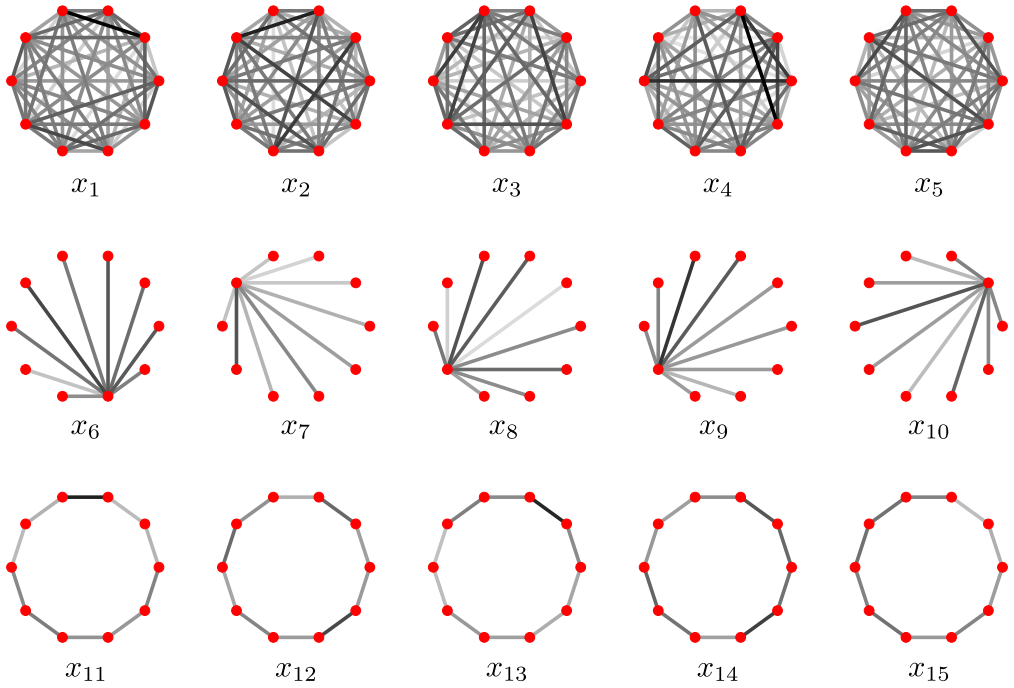

Figure 4.8: A dataset of different graph topologies. We expect sample-limited BSA to be able to classify the dataset into these three groups and in particular to retrieve in dimension 2 one graph of each topology.

Application to European airlines route maps In the second experiment, the dataset consists of airlines route maps. Precisely, we consider the graphs describing the route map of 12 different European companies (see Figure 4.10) commuting between the main 6 of the 24 sub-regions according to which we classified the world's airports in a preliminary pre-processing step. Each node represents one airport while each edge weights the number of flights operated between the two corresponding countries during a fixed time frame. All the data come from the OpenFlight data repository (Open Flights, 2017). Although the graphs are naturally labeled (by the sub-regions), since we are essentially interested in comparing the general commuting strategy of the different companies, we fall within the unlabeled setting. On this dataset, we perform only forward convex sample-limited BSA (see Figure 4.12) as it has proven to carry even more interpretability than simply sample-limited BSA. This time, it is not clear how many classes the dataset consist of. Instead, we observe that the error profile relates very well to the classification obtained by hierarchical clustering using Ward linkage (Müllner, 2011) and the L^2 spectral distance (see Figure 4.11). We retrieve through BSA a visualization of the dataset for 2, 3 and 4 reference graphs.

A note on future projects Through these two examples, we have been able to demonstrate that BSA combined with our framework is a relevant tool for graph analysis that advances important issues in statistical graph analysis, especially that of interpretability. There are now several directions in which we would like to take our work further. On theoretical aspects, it would be interesting to investigate the stratified geometry of spectral graph spaces, which we did not explicitly define until now, and to supplement such an exploration with concrete examples of singular graphs. This question is directly related to convex BSA and archetypal analysis because singular graphs are exactly those graphs that lie on the boundary of spectral graph spaces. Now back to the practical aspects, we plan to experiment with our framework on larger data sets and graphs with higher number of nodes,

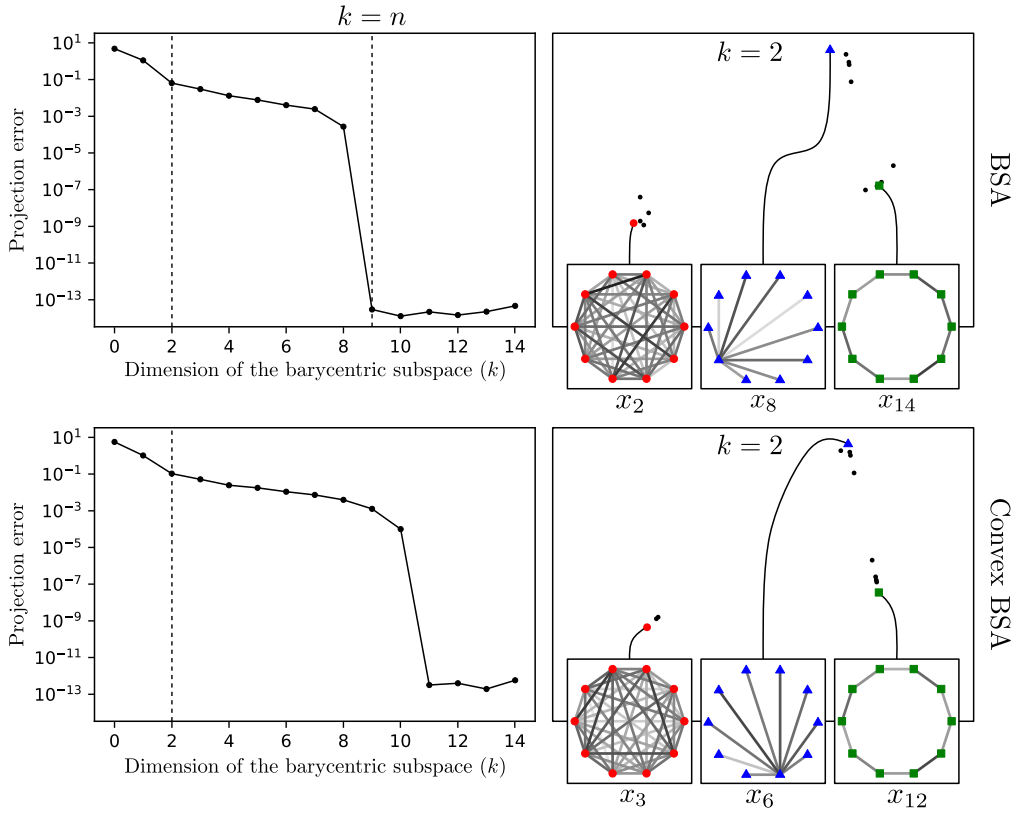

Figure 4.9: Sample-limited BSA and convex sample-limited BSA of the simulated dataset. (Left) Evolution of the squared projection error across the nested sequence of barycentric subspaces produced by forward BSA. The very first observation is that sample-limited BSA and convex sample-limited BSA perform similarly. In sample-limited BSA, the error drops to zero when the barycentric subspace reaches the dimension of the data space because it then coincides exactly with the data space. A prori, the data space is the graph space of size $p = 10$ that has dimension 10. In fact, because they do not have any node attributes, all the graphs have a null diagonal such that the dataset lies in the subspace of Γ_{10} of dimension 9 which consists of all the graphs with trace 0 (a property preserved by the action of the orthogonal group as we see later). In convex sample-limited BSA, the error drops to zero later, supposedly when the number of reference points is enough for them to cover the whole dataset with their convex hull. By applying the *elbow method* (James, Witten, Hastie, and Tibshirani, 2013) on both error profiles, we are able to predict the fact that the dataset can be classified in 3 groups. (Right) The dataset (in black) is projected on the barycentric subspace of the 3 reference points (colored) picked by either method. The classification of the dataset into three groups is visually very strong. BSA also suggests here that the cyclic graphs and star-shaped graphs are topologically more similar than the fully connected graphs. Hierarchical clustering, which we do not present on this dataset, provides with a similar analysis.

and to explore other types of data, for example related to brain connectivity. In relation to the stratification of graph spectral spaces, we may also question how well our framework and BSA would perform on graphs with several connected components.

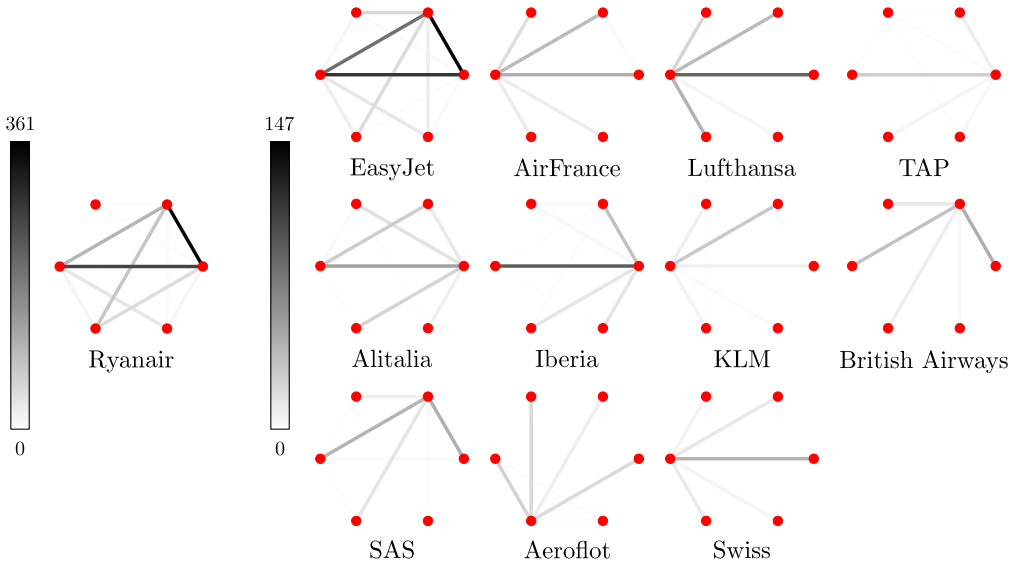

Figure 4.10: An airlines route maps dataset. Due to significantly higher edge values, the graph representing Ryanair route map is plotted with a different scale. Note that we use the same two scales in the two other figures. Although the classification of the dataset is much more unclear in this example than in the simulated one, we can at least observe that the graphs corresponding to the two low-cost airlines (Ryanair and EasyJet) have a fairly symmetrical structure, while those representing the other airlines have a more or less star-shaped structure. This reflects the fact that low-cost airlines have no preferred hub, and tend to be scattered across different airports. Note

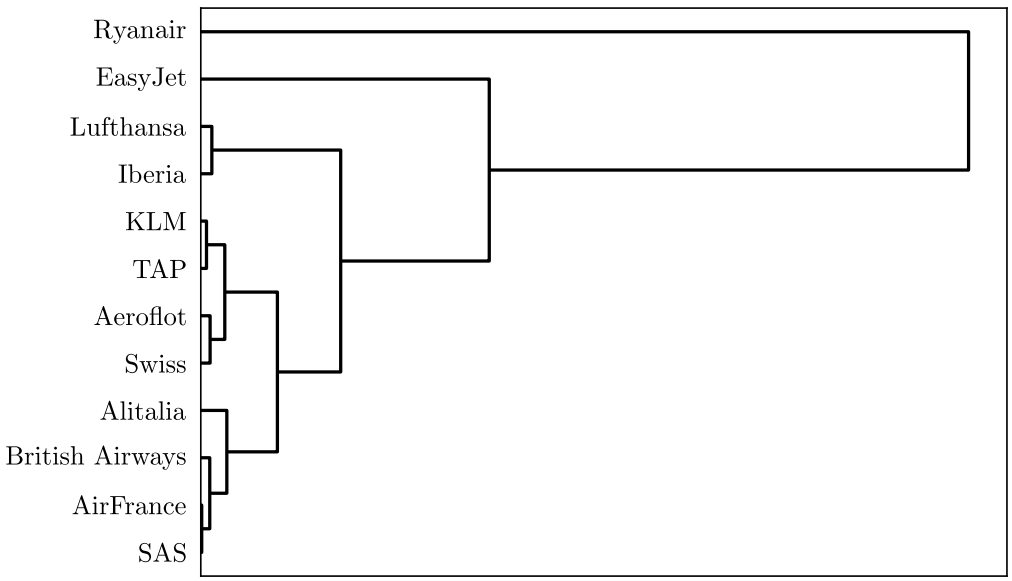

Figure 4.11: Hierarchical clustering. We visualize the result of hierarchical clustering in a dendogram. The method provides a classification very similar to that of BSA. What's more, if we push the number of clusters beyond 4, classification consists only in odd numbers of clusters (5, 7...), echoing the error profile in the previous figure, where we observe a break in slope at even dimensions.

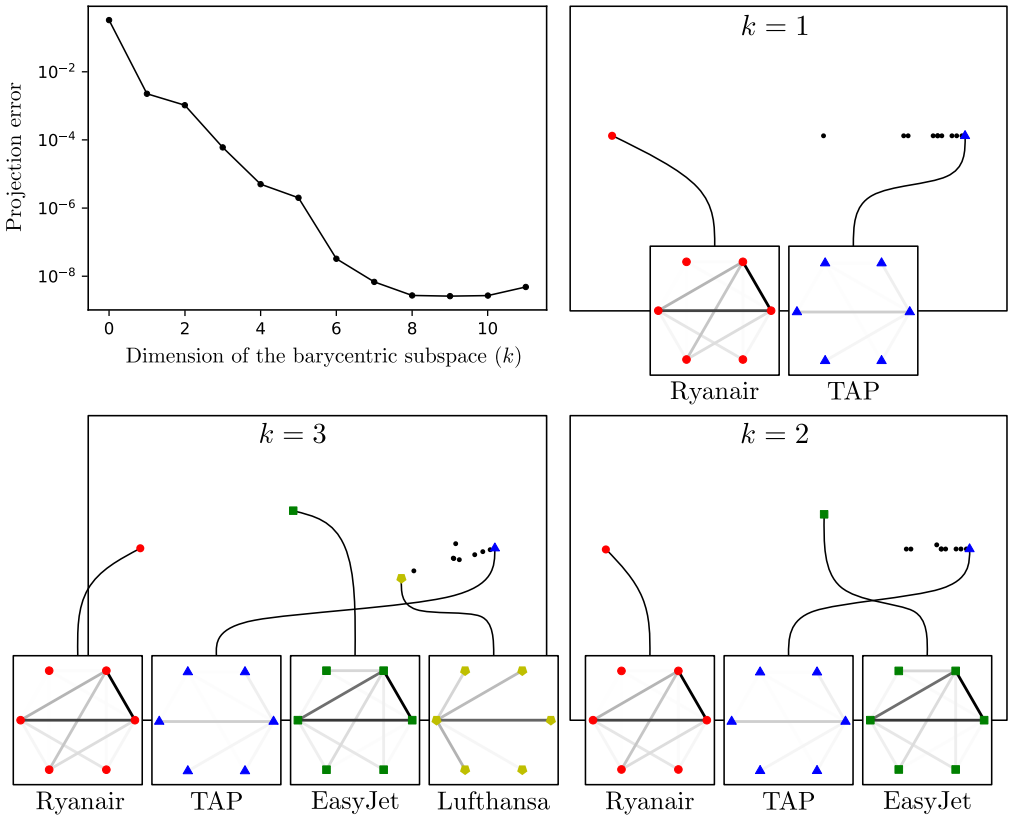

Figure 4.12: Convex sample-limited BSA for the airlines dataset. The method classifies Ryanair as an outlier, which is quite logical since it carries attributes twice as large as the other graphs. Going from dimension 0 to the dimension 1, the error is significantly reduced. Increasing the dimension even further allows us to identify more groups. EasyJet, which is the third reference graph, may also be considered an outlier. The fourth reference graph picked by BSA is Lufthansa, which, with another point corresponding to Iberia, appears (bottom left) a little outside the main cluster. If we go back to the dataset (see Figure 4.10), we observe indeed that both airlines include a strong edge.

Appendix. On the choice of a representative

We are adding this extra section to address the question of the choice of a representative in the equivalence class of a graph. As we saw earlier in this chapter, the equivalence class of even the simplest graph includes graphs with node attributes (self loops) in the framework of spectral graph spaces. The problem is that in some applications, attributes on the nodes may have no meaning at all, or at least carry no information that we would be interested in taking into account. Now, even when the input graphs have no node attributes, computations or statistical methods may still output graphs with node attributes, decreasing significantly the interpretability of such. We illustrate this phenomenon on the example of geodesics in Figure 4.13. It is interesting to note that this issue did not occur when performing sample-limited BSA in the previous examples. The reason is twofold. First, spectral graph spaces identify with cones, which are Euclidean manifolds, although with boundaries. In particular, barycenters of graphs spectral spaces are weighted sums. Second, assume that the dataset consists of graphs without node attributes. Then so do the reference points picked by sample-limited BSA. Finally, as a linear combination of graphs without nodes attributes – or equivalently with a null diagonal – is still a graph without node attributes, then by construction,

sample-limited BSA outputs only such graphs. Now, considering future works and the use of other methods, we need to be able to automatically chose a meaningful representative within the equivalence class of an output graph. Section 4.2 might suggest representing the classes of spectral graph spaces by diagonal matrices. Although such a representation is very convenient from a computational point of view, it has a no clear meaning from an application point of view and moreover it does not solve the problem of node attributes. Instead, we propose a simple procedure to transform a graph with node attribute into an equivalent graph but with no more node attribute. More precisely, we consider the case of graphs that have been generated by applying simple operations of graph spectral spaces to some input graphs, all having no node attributes. A interesting observation is that a graph – represented by its adjacency matrix – with no node attributes has trace zero, a property that is invariant both by the action of the rotations and by linear operations, which is the case of all the operations of graph spectral spaces. As a consequence, graphs that have been generated from some input graphs without node attributes should also satisfy such property. Now inversely, for any symmetric matrix of trace zero, then there always exists at least one equivalent symmetric matrix of diagonal zero. Given a graph of trace zero, let us then simply compute the closest equivalent graph with diagonal zero, that is no node attributes. Back to the example of the geodesic, this procedure actually outputs a visually very convincing result.

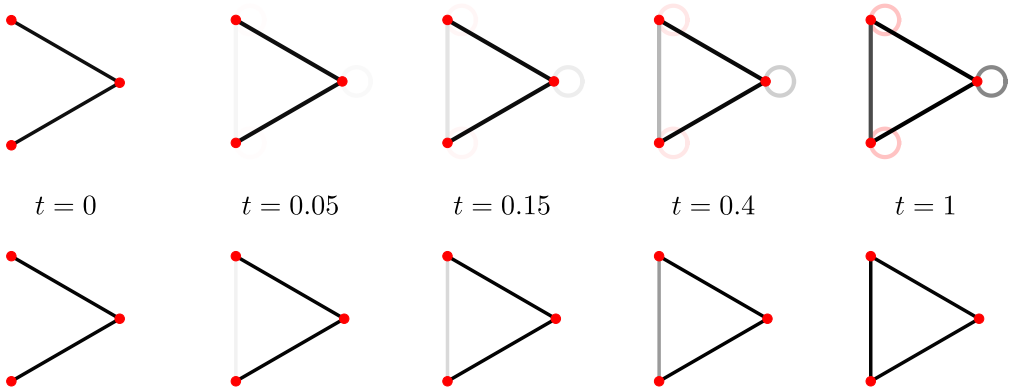

Figure 4.13: Geodesic joining two graphs. We propose two representations of the same geodesic joining the equivalent classes of a binary tree x (bottom left) and a cyclic graph y (bottom right) in the graph space Γ_3. (Top) The geodesic from the class $\pi(x)$ to the class $\pi(y)$ naturally identifies through the tangent map $d\pi$ with the horizontal geodesic from x to $R_x(y)yR_x(y)^T$. However such a horizontal geodesic contains graphs with node attributes (self-loops). (Bottom) An equivalent representation consists in replacing the graphs in the horizontal geodesic by their closest equivalent graph without nodes attributes. By construction, such a curve is also projected onto the geodesic from $\pi(x)$ to $\pi(y)$, but corresponds better than the horizontal geodesic to how one might picture the geodesic in the graph space. Note however that this second curve is a priori not a geodesic in the top manifold of symmetric matrices

References

Alekseevsky, D. et al. (2003). "The Riemannian geometry of orbit spaces – the metric, geodesics, and integrable systems". In: *Publicationes Mathematicae Debrecen* 62.3, pp. 247–276. ISSN: 00333883. DOI: 10.5486/PMD.2003.2821.

Aydın, B. et al. (2009). "A principal component analysis for trees". In: *The Annals of Applied Statistics* 3.4, pp. 1597–1615. ISSN: 1932-6157, 1941-7330. DOI: 10.1214/09-AOAS263.

Brouwer, A. E. and W. H. Haemers (2012). *Spectra of Graphs*. Universitext. Springer. ISBN: 978-1-4614-1938-9, 978-1-4614-1939-6. DOI: 10.1007/978-1-4614-1939-6.

Calissano, A., A. Feragen, and S. Vantini (2023). "Populations of unlabelled networks: Graph space geometry and generalized geodesic principal components". In: *Biometrika*, asad024. ISSN: 1464-3510. DOI: 10.1093/biomet/asad024.

Calissano, A., E. Maignant, and X. Pennec (2023). "Towards quotient barycentric subspaces". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 366–374. ISBN: 978-3-031-38271-0. DOI: 10.1007/978-3-031-38271-0_36.

Calissano, A. et al. (2023). "Graph alignment exploiting the spatial organisation improves the similarity of brain networks". In: *Human Brain Mapping*. DOI: 10.1002/hbm.26554.

Chung, F. (1996). *Spectral Graph Theory*. Vol. 92. CBMS Regional Conference Series in Mathematics. American Mathematical Society. ISBN: 978-0-8218-0315-8, 978-0-8218-8936-7, 978-1-4704-2452-7. DOI: 10.1090/cbms/092.

Conte, D. et al. (2004). "Thirty years of graph matching in pattern recognition". In: *International Journal of Pattern Recognition and Artificial Intelligence* 18.3, pp. 265–298. ISSN: 0218-0014. DOI: 10.1142/S0218001404003228.

Cutler, A. and L. Breiman (1994). "Archetypal analysis". In: *Technometrics* 36.4, pp. 338–347. ISSN: 0040-1706. DOI: 10.2307/1269949.

Donnat, C. and S. Holmes (2018). "Tracking network dynamics: A survey using graph distances". In: *The Annals of Applied Statistics* 12.2, pp. 971–1012. ISSN: 1932-6157, 1941-7330. DOI: 10.1214/18-AOAS1176.

Durante, D., D. B. Dunson, and J. T. Vogelstein (2017). "Nonparametric Bayes modeling of populations of networks". In: *Journal of the American Statistical Association* 112.520, pp. 1516–1530. ISSN: 0162-1459. DOI: 10.1080/01621459.2016.1219260.

Feragen, A. et al. (2013). "Tree-Space statistics and approximations for large-scale analysis of anatomical trees". In: *Information Processing in Medical Imaging*. Ed. by J. C. Gee et al. Lecture Notes in Computer Science. Springer, pp. 74–85. ISBN: 978-3-642-38868-2. DOI: 10.1007/978-3-642-38868-2_7.

Guo, X., A. Srivastava, and S. Sarkar (2021). "A quotient space formulation for generative statistical analysis of graphical data". In: *Journal of Mathematical Imaging and Vision* 63.6, pp. 735–752. ISSN: 1573-7683. DOI: 10.1007/s10851-021-01027-1.

Haemers, W. H. and E. Spence (1995). "Graphs cospectral with distance-regular graphs". In: *Linear and Multilinear Algebra* 39.1, pp. 91–107. ISSN: 0308-1087. DOI: 10.1080/03081089508818382.

Halbeisen, L. and N. Hungerbühler (2000). "Reconstruction of weighted graphs by their spectrum". In: *European Journal of Combinatorics* 21.5, pp. 641–650. ISSN: 0195-6698. DOI: 10.1006/eujc.1999.0410.

Huckemann, S., T. Hotz, and A. Munk (2010). "Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions". In: *Statistica Sinica* 20.1, pp. 1–58. ISSN: 1017-0405.

Jain, B. J. and K. Obermayer (2009). "Structure spaces". In: *Journal of Machine Learning Research* 10.93, pp. 2667–2714. ISSN: 1533-7928.

James, G. et al. (2013). *An Introduction to Statistical Learning*. Vol. 103. Springer Texts in Statistics. Springer. ISBN: 978-1-4614-7137-0, 978-1-4614-7138-7. DOI: 10.1007/978-1-4614-7138-7.

Jurman, G., R. Visintainer, and C. Furlanello (2011). "An introduction to spectral distances in networks". In: *Neural Nets WIRN10*. IOS Press, pp. 227–234. DOI: 10.3233/978-1-60750-692-8-227.

Kolaczyk, E. D. et al. (2020). "Averages of unlabeled networks: Geometric characterization and asymptotic behavior". In: *The Annals of Statistics* 48.1, pp. 514–538. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/19-AOS1820.

Maron, H. et al. (2018). "Invariant and equivariant graph networks". In: *International Conference on Learning Representations*.

Michor, P. (2008). *Topics in Differential Geometry*. Vol. 93. Graduate Studies in Mathematics. American Mathematical Society. ISBN: 978-0-8218-2003-2, 978-0-8218-8410-2, 978-0-8218-8721-9, 978-1-4704-1161-9. DOI: 10.1090/gsm/093.

Müllner, D. (2011). *Modern hierarchical, agglomerative clustering algorithms*. DOI: 10.48550/arXiv.1109.2378.

O'Neill, B. (1966). "The fundamental equations of a submersion." In: *Michigan Mathematical Journal* 13.4, pp. 459–469. ISSN: 0026-2285, 1945-2365. DOI: 10.1307/mmj/1028999604.

Open Flights (2017). *Open Flight Database*. Accessed: 2023-05-16.

Pennec, X. (2018). "Barycentric subspace analysis on manifolds". In: *The Annals of Statistics* 46.6, pp. 2711–2746. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/17-AOS1636.

Rohé, M.-M., M. Sermesant, and X. Pennec (2018). "Low-dimensional representation of cardiac motion using Barycentric Subspaces: A new group-wise paradigm for estimation, analysis, and reconstruction". In: *Medical Image Analysis* 45, pp. 1–12. ISSN: 1361-8415. DOI: 10.1016/j.media.2017.12.008.

Severn, K. E., I. L. Dryden, and S. P. Preston (2022). "Manifold valued data analysis of samples of networks, with applications in corpus linguistics". In: *The Annals of Applied Statistics* 16.1, pp. 368–390. ISSN: 1932-6157, 1941-7330. DOI: 10.1214/21-AOAS1480.

Simpson, S. et al. (2013). "A permutation testing framework to compare groups of brain networks". In: *Frontiers in Computational Neuroscience* 7. ISSN: 1662-5188.

Tsuda, Koji and Hiroto Saigo (2010). "Graph classification". In: *Managing and Mining Graph Data*. Ed. by C. C. Aggarwal and H. Wang. Advances in database systems. Springer US, pp. 337–363. ISBN: 978-1-4419-6045-0. DOI: 10.1007/978-1-4419-6045-0_11.

Van Mieghem, P. (2010). *Graph Spectra for Complex Networks*. Cambridge University Press. ISBN: 978-0-521-19458-7. DOI: 10.1017/CBO9780511921681.

Von Ferber, C. et al. (2009). "Public transport networks: Empirical analysis and modeling". In: *The European Physical Journal B* 68.2, pp. 261–275. ISSN: 1434-6036. DOI: 10.1140/epjb/e2009-00090-x.

Wang, H. and J. S. Marron (2007). "Object oriented data analysis: Sets of trees". In: *The Annals of Statistics* 35.5, pp. 1849–1873. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/009053607000000217.

Younes, L. (2010). *Shapes and Diffeomorphisms*. Vol. 171. Applied Mathematical Sciences. Springer. ISBN: 978-3-642-12054-1, 978-3-642-12055-8. DOI: 10.1007/978-3-642-12055-8.

Zhai, H. (2016). "Principal component analysis in phylogenetic tree space". PhD thesis. The University of North Carolina at Chapel Hill University Libraries. DOI: 10.17615/TX71-2Y54.

Chapter 5

Locally barycentric embeddings, a gluing problem

This chapter takes us back to Chapter 2, where we parted with the conclusion that the optimal mapping problem solved by locally barycentric embeddings does not necessarily defines a unique embedding. In fact, this had already been highlighted by earlier work as the consequence of two main phenomena. On one hand, as pointed out by Roweis and Saul (2000) in their original paper *Nonlinear Dimensionality Reduction by Locally Linear Embedding*, the existence of multiple connected components in the nearest neighbor graph of the data points generates just as many embeddings. Such a problem has been addressed by Polito and Perona (2001). There is a strong connection between this phenomenon and the matrix tree theorem, which relates the number of connected components of a graph to the dimension of the kernel of its graph Laplacian, which we shall see is directly linked to the optimal mapping problem. On the other hand, as locally barycentric embeddings rely on the conservation of barycentric coordinates, then affine mappings should provide good candidates to the optimal mapping problem. In particular, Lin (2021) proved that, under the assumption that each data point has been exactly described as a barycenter of its neighbors, any orthogonal projection of a vector-valued dataset onto a lower-dimensional subspace is a locally barycentric embedding of such. Following on from these results, we show that a larger subset of the solutions of the optimal mapping problem may be specified by *gluing* affine mappings together along the connected components of the neighbor graph of the data points. In other words, we show that locally affine mappings are locally barycentric embeddings.

In the first section of this chapter, we propose a proper definition of the notion of locally barycentric mapping as a mapping of a weighted graph of the data space into the embedding space. Then, in the second section, geometric conditions are stated for the mapping obtained by gluing affine mappings together along the strongly connected components of a weighted graph to be a locally barycentric mapping of such. Finally, we derive from such results in the last section a series of lower bounds on the number of solutions of the optimal mapping problem when the embedding space is Euclidean. We supplement these with some conjectures in the more general case and draw several directions for further analysis.

5.1 Locally barycentric mappings

Let us recall that locally barycentric embeddings consist in two steps, a model fitting step and the optimal mapping problem. The local barycentric model estimates the barycentric coordinates of

the projection of each data point on the barycentric subspace of its neighbor points. Based on these coordinates, the optimal mapping problem is then the problem of finding a set of points in the embedding space in that matches back those coordinates the best. Counting locally barycentric embeddings of a dataset is then equivalent to counting the solutions of the optimal mapping problem. In this chapter, we make the same assumption as Lin (2021) that each data point has been exactly described as a barycenter of its neighbors, that is that the local barycentric model is exact. We discuss at the end of the chapter how to lift such an assumption. Although the optimal mapping problem a priori depends only on the barycentric coordinates estimated by the model, we saw that under this assumption, manipulations on the data points only also yield to solutions to such a problem. To account for both, we define locally barycentric mappings as the correspondence between a pair made of a dataset and corresponding barycentric coordinates, and a solution of the optimal mapping problem.

5.1.1 Local barycentric models are weighted graphs

Together with the neighbor graph of the dataset, fixed ahead of the two steps of locally barycentric embeddings, the barycentric coordinates estimated by the local barycentric model describe a directed weighted graph. Precisely, let each data point be represented by a vertex and draw an arc from such to all the vertices that represent the neighbors of the point. Then the graph thus defined is a directed graph and the arc connecting one vertex to another can be assigned the barycentric coordinate of the point represented by the first vertex with respect to the point represented by the second.

Barycentric weighted graphs Let now M be a Riemannian manifold of dimension n. We define a barycentric weighted graph of M as the union of a dataset in M and the graph defined by an exact local barycentric model of such dataset.

Definition 5.1.1. *A barycentric weighted set of M is a pair (X, w) where $X = \{x_1, \dots x_N\}$ is a set of points in M and $w = (w_{ij})$ is a real square matrix of size N with each line summing to 1 that satisfies*

$$x_i \in \mathrm{bar}(x_j, w_{ij})_{1 \leq j \leq N} \qquad \forall 1 \leq i \leq N. \tag{5.1}$$

In practice, we do not make the distinction between a barycentric weighted set (X, w) and the corresponding directed weighted graph $(V = \{1, \dots, N\}, w)$.

5.1.2 Locally barycentric mappings of a barycentric weighted graph

Now once the barycentric weights of model had been estimated, we had defined roughly in Chapter 2 locally barycentric mappings in the context of Locally Linear Embedding as the solutions of the optimal mapping problem

$$\underset{y_1, \dots, y_N \in \mathbb{R}^k}{\text{minimize}} \quad \sum_{i=1}^{N} \left\| y_i - \sum_{j \in N(i)} w_{ij} y_j \right\|^2 \tag{5.2}$$

where we recall that $j \in N(i)$ if and only if x_j has been assigned as a neighbor of x_i. In this chapter we refer only to the exact solutions of the problem as locally barycentric mappings, that is embeddings of the dataset in which the image of data point is still exactly the barycenter of the image of its neighbors. More generally, we seek for a solution $y_1, \dots, y_N$ in a manifold M' and define locally barycentric mappings in the most general case.

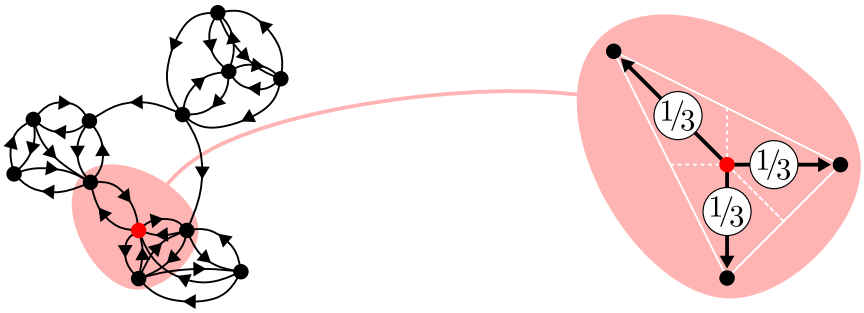

Figure 5.1: Barycentric weighted graph. The 3-nearest neighbor graph of a set of points in the plane. Since the barycentric subspaces of 3 affinely independent points in the plane is the plane itself, any point of the set can be described exactly as a barycenter of its neighbors. The weights assigned to the arc leaving the red point correspond to the barycentric coordinates of the latter with respect to its neighbors.

Locally barycentric mappings Let M' be a Riemannian manifold of dimension k. We define locally barycentric mappings of a barycentric weighted set of M in M' as follows

Definition 5.1.2. *Let (X, w) be a barycentric weighted set of M. A locally barycentric mapping of (X, w) in M' is a set of points $y_1, \dots, y_N \in M'$ such that*

$$y_i \in \mathrm{bar}(y_j, w_{ij})_{1 \leq j \leq N} \qquad \forall 1 \leq i \leq N. \tag{5.3}$$

Note that the pair (Y, w) is then a barycentric weighted set of M'. In practice, we refer both to the set Y and the map $f(x_i) = y_i$ as a locally barycentric mapping.

Affine mappings are locally barycentric mappings Reformulated through these two new definitions, the result from Lin (2021) states than any orthogonal projection of a barycentric weighted set of $\mathbb{R}^n$ onto $\mathbb{R}^k$ is a locally barycentric mapping. In fact, this result takes into account the two constraints on the solutions of the original optimal mapping problem to be centered and have unit covariance matrix. Mainly in order to simplify the analysis of the dimensionality of locally barycentric mappings, we did not account for these constraints in our definition. Therefore, we can state the more general result

Proposition 5.1.1. *Let (X, w) be a barycentric weighted set of M. Then any affine mapping f from M to M' (Kobayashi and Nomizu, 1996, Chapter 6) defines a locally barycentric mapping $f(x_1), \dots, f(x_N)$ of (X, w) in M'.*

Essentially, in the case where we add the constraints for the solutions of the optimal mapping problem to be centered and have unit covariance matrix, all the non orthogonal affine mappings as well as translations are discarded such that the only affine mappings left are the orthogonal projections.

A first count of locally barycentric mappings The previous result provides a first bound on the number of locally barycentric mappings of a barycentric weighted set. More precisely, the image of $x_1, \dots, x_N \in \mathbb{R}^n$ by an affine mapping f from $\mathbb{R}^n$ to $\mathbb{R}^k$ is also their image by the restriction of f to their affine span. Moreover such a restriction is an affine mapping from the affine span of the points to $\mathbb{R}^k$. Now let us recall that an affine mapping is completely determined by its value in an affine basis of the original space. A locally barycentric mapping $f(x_1), \dots, f(x_N)$ is then completely

determined by the value of f in an affine basis of the affine span of $x_1, \dots, x_N$. Therefore, the set locally barycentric mappings of a barycentric weighted set (X, w) in $\mathbb{R}^n$, which we prove later in the last section of this chapter is a vector space, is at least of dimension $k\,\mathrm{rank}(X)$, where the notation $\mathrm{rank}(X)$ corresponds here to the affine rank of X, that is the cardinality of a maximal affinely independent subset of X.

In the general case of affine mappings between any two Riemannian manifold M and M', we cannot say much a priori. However, observe that if f_0 is an affine mapping from M to M' and Φ is an affine mapping of M', then $\Phi \circ f_0$ is still an affine mapping from M to M'. Let us consider the case where M' is either the Euclidean space $\mathbb{R}^k$, the sphere S^k, or the hyperbolic space H^k. The first case is included in the case we just solved. As for the two other cases, the dimension of the set of affine transformations of M' is known to be $\frac{1}{2}k(k+1)$ (Kobayashi, 1995) and we conjecture that we can derive from such a result similar estimations on the dimensionality of locally barycentric mappings. In the same vein, we are now going to investigate a more local version of Proposition 5.1.1 and derive corresponding lower bounds on the dimension of locally barycentric embeddings.

5.2 Untangling locally barycentric embeddings

The main result of this section is a decomposition theorem for locally barycentric mappings that states that under some conditions, gluing affine mappings together along the strongly connected components of a barycentric weighted set defines a locally barycentric mapping of such.

5.2.1 Components of a barycentric weighted graph

Let us recall that a barycentric weighted set (X, w) can be interpreted as a directed weighted graph of adjacency matrix w. We discussed at the beginning of this chapter how the number of locally barycentric embeddings of X proves to be related to the connectivity of the neighbor graph of X, or equivalently of the barycentric weighted set (X, w), in particular to the number of connected components. In fact, since a barycentric weighted set is by construction a directed graph, then a more natural notion to study the connectivity of such graphs is the notion of strongly connected components. Before defining such a notion, let us just introduce the following notation

Definition 5.2.1. *Let (X, w) be a barycentric weighted set. Then a point $x_j \in X$ is said to be an neighbor of a point $x_i \in X$ if $w_{ij} \neq 0$. We denote such a relation by $x_i \to x_j$, and if $x_i \in X$ has a neighbor in $X' \subset X$, then we write $x_i \to X'$.*

Strongly connected components The notion of strong connectivity relies on that of directed paths. A directed path from a vertex $x_i \in X$ to a point $x_j \in X$ is a sequence $x_{i_0} = x_i \to x_1 \dots x_{i_{m-1}} \to x_{i_m} = x_j$ such that for all $1 \leq l \leq m$, the point x_{i_l} is a neighbor of $x_{i_{l-1}}$.

Definition 5.2.2. *Let (X, w) be a barycentric weighted set. Two points $x_i \in X$ and $x_j \in X$ are said to be strongly connected if there exist a directed path from x_i to x_j and a directed path from x_j to x_i. Strong connectivity of points defines an equivalence relation on (X, w) and its equivalence classes are referred to as the strongly connected components of (X, w).*

Now assume that (X, w) has t strongly connected components. Then we can show such components may be reordered as $X_1, \dots, X_t$ in such a way that

$$\forall 1 \leq q < p \leq t,\ \forall x_i \in X_p,\ \forall x_j \in X_q,\ w_{ij} = 0. \tag{5.4}$$

We refer the reader to (Brualdi et al., 2018) for more details on how to prove this result as well as for the previous definitions.

Isolated components The components that contain no points with neighbors in other strongly connected components are called isolated components (Bauer, 2012). The way the strongly connected components are ordered above forces the isolated components to appear at the end of the sequence $X_1, \ldots, X_t$. We will see later in this chapter that we can derive from such a notion a similar but stronger result on the number of locally barycentric embeddings than the one based on the notion of connected component.

5.2.2 Decomposition of locally barycentric embeddings

Now the idea is that if two components of a barycentric weighted set are not connected or are poorly connected with each other, then it is possible to transform them differently up to some degree of freedom while preserving the preserving the barycentric coordinates of each point with respect to its neighbors. Let us illustrate this idea with the simplest example of a barycentric weighted set (X, w) with two connected components X_1 and X_2. The fact that X_1 and X_2 are not connected means that any point of each set has its neighbors in the same set. If so, consider then an affine mapping f_1 acting on X_1 and an affine mapping f_2 acting only on X_2. Then f_1 preserves the barycentric relationships in X_1 without affecting X_1 and f_2 preserves the barycentric relationships on X_2 without affecting x_1, then gluing the set obtained by gluing $f_1(x_1)$ and $f_2(x_2)$ is a locally barycentric mapping of X. We illustrate such a mapping in Figure 5.2.

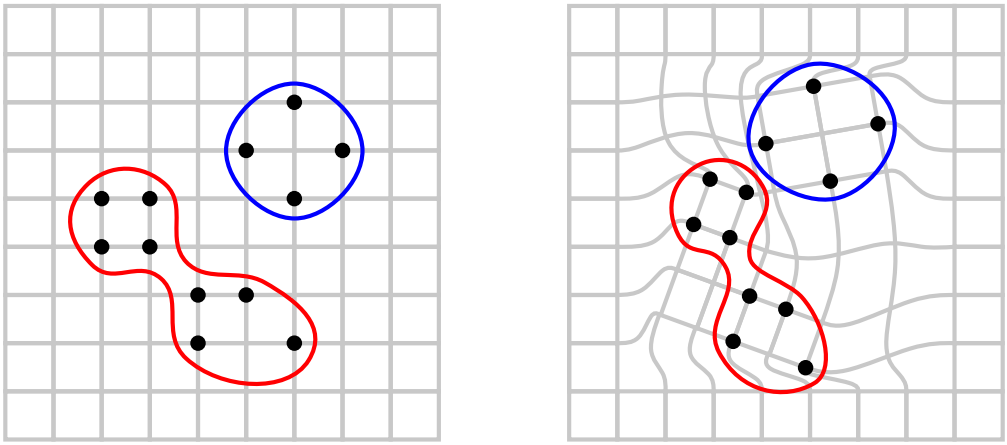

Figure 5.2: Locally barycentric mapping of a barycentric weighted set of the plane with two components. The mapping is obtained by gluing together an affine transformation of the first component and another affine transformation of the second component. We visualize each mapping as a deformation of the background grid.

Pasting condition In the case of two components are not totally disconnected, the action of each affine mapping may interfere with each other. To ensure that the barycentric relationships are still preserved, the two affine mappings have to obey some condition on the points where their action interfere. We refer to such condition as the pasting condition (or gluing condition). In the case where the data space is a Riemannian manifold M of dimension n and the embedding space is a Riemannian manifold M' of dimension k, the pasting condition is the following

Lemma 5.2.1. *Let (X, w) be a barycentric weighted graph of M. Let $X = X_1 \cup X_2$ be a partition of X and let f_1 and f_2 be two affine mappings from M to M' such that*

1. For all $x_i \in X_1$ with at least one neighbor in X_2

$$\sum_{x_j \in X_2} w_{ij} \log_{f_1(x_i)} f_1(x_j) = \sum_{x_j \in X_2} w_{ij} \log_{f_1(x_i)} f_2(x_j) \tag{5.5}$$

2. For all $x_i \in X_2$ with at least one neighbor in X_1

$$\sum_{x_j \in X_1} w_{ij} \log_{f_2(x_i)} f_1(x_j) = \sum_{x_j \in X_1} w_{ij} \log_{f_2(x_i)} f_2(x_j) \tag{5.6}$$

Then the piecewise affine mapping f obtained by gluing f_1 and f_2 together

$$f|_{X_1} = f_1 \qquad f|_{X_2} = f_2 \tag{5.7}$$

defines a locally barycentric mapping of (X, w) in M'.

Proof. The map f is a locally barycentric mapping of (X, w) if and only if

$$\sum_{j=1}^{N} w_{ij} \log_{f(x_i)} f(x_j) = 0 \qquad \forall 1 \leq i \leq N. \tag{5.8}$$

Let $x_i \in X_1$. We have

$$\sum_{j=1}^{N} w_{ij} \log_{f(x_i)} f(x_j) = \sum_{x_j \in X_1} w_{ij} \log_{f_1(x_i)} f_1(x_j) + \sum_{x_j \in X_2} w_{ij} \log_{f_1(x_i)} f_2(x_j) \tag{5.9}$$

such that the previous condition is true for i if and only if

$$\sum_{x_j \in X_2} w_{ij} \log_{f_1(x_i)} f_2(x_j) = - \sum_{x_j \in X_1} w_{ij} \log_{f_1(x_i)} f_2(x_j) \tag{5.10}$$

Now by definition of a barycentric weighted set, the point x_i is the barycenter of $x_1, \ldots, x_N$ with corresponding weights $w_{i1}, \ldots, w_{iN}$. Moreover, because f_1 is an affine mapping, it preserves barycenters. Therefore, we have

$$\sum_{j=1}^{N} w_{ij} \log_{f_1(x_i)} f_1(x_j) = 0 \tag{5.11}$$

and Equation (5.10) is equivalent to

$$\sum_{x_j \in X_1} w_{ij} \log_{f_2(x_i)} f_2(x_j) = \sum_{x_j \in X_1} w_{ij} \log_{f_1(x_i)} f_1(x_j) \tag{5.12}$$

Suppose that for all $x_j \in X_2$, we have $w_i j = 0$. Then by combining these two last equations, we see that the condition is automatically satisfied. We derive the second pasting condition with the same reasoning on $x_i \in X_2$. □

Pasting condition for affine mappings from $\mathbb{R}^n$ to $\mathbb{R}^k$ In the Euclidean case, we show that the pasting condition has a nice geometric interpretation. Assume then that $M = \mathbb{R}^n$ and $M' = \mathbb{R}^k$. The mapping f defines a locally barycentric mapping of (X, w) in $\mathbb{R}^k$ if the two following condition are satisfied

1. For all $x_i \in X_1$ with at least one neighbor in X_2

$$\sum_{x_j \in X_2} w_{ij}(f_1(x_j) - f_1(x_i)) = \sum_{x_j \in X_2} w_{ij}(f_2(x_j) - f_1(x_i)) \tag{5.13}$$

2. For all $x_i \in X_2$ with at least one neighbor in X_1

$$\sum_{x_j \in X_1} w_{ij} f_1((x_j) - f_2(x_i)) = \sum_{x_j \in X_1} w_{ij}(f_2(x_j) - f_2(x_i)). \tag{5.14}$$

The two equations simplify to

$$\sum_{x_j \in X_2} w_{ij} f_1(x_j) = \sum_{x_j \in X_2} w_{ij} f_2(x_j) \tag{5.15}$$

and

$$\sum_{x_j \in X_1} w_{ij} f_1(x_j) = \sum_{x_j \in X_1} w_{ij} f_2(x_j). \tag{5.16}$$

If $\sum_{x_j \in X_2} w_{ij} \neq 0$, then let us leverage the fact that f_1 and f_2 are affine mappings to write

$$\begin{aligned} f_1\Big(\sum_{x_j \in X_2} \frac{w_{ij}}{\sum_{x_j \in X_2} w_{ij}} x_j\Big) &= \sum_{x_j \in X_2} \frac{w_{ij}}{\sum_{x_j \in X_2} w_{ij}} f_1(x_j) && (5.17)\\ &= \sum_{x_j \in X_2} \frac{w_{ij}}{\sum_{x_j \in X_2} w_{ij}} f_2(x_j) && (5.18)\\ &= f_2\Big(\sum_{x_j \in X_2} \frac{w_{ij}}{\sum_{x_j \in X_2} w_{ij}} x_j\Big). && (5.19) \end{aligned}$$

Otherwise, let $L_{f_1} = f_1 - f_1(0)$ and $L_{f_2} = f_2 - f_2(0)$ denote the linear part of f_1 and f_2 respectively. Then we have

$$\begin{aligned} L_{f_1}\Big(\sum_{x_j \in X_2} w_{ij} x_j\Big) = \sum_{x_j \in X_2} w_{ij} L_{f_1}(x_j) &= \sum_{x_j \in X_2} w_{ij} f_1(x_j) && (5.20)\\ &= \sum_{x_j \in X_2} w_{ij} f_2(x_j) && (5.21)\\ &= \sum_{x_j \in X_2} w_{ij} L_{f_2}(x_j) = L_{f_2}\Big(\sum_{x_j \in X_2} w_{ij} x_j\Big) && (5.22) \end{aligned}$$

We can follow exactly the same reasoning for the second conditions. Now for all $i \in X_1$ with at least one neighbor in X_2, we introduce the point x_i^2 defined by

$$x_i^2 = \begin{cases} \sum_{j \in X_2} \frac{w_{ij}}{\sum_{x_j \in X_2} w_{ij}} x_j & \text{if} \quad \sum_{x_j \in X_2} w_{ij} \neq 0 \\ \sum_{x_j \in X_2} w_{ij} x_j & \text{otherwise.} \end{cases} \tag{5.23}$$

Similarly, for all $i \in X_2$ with at least one neighbor in X_1, we introduce the point x_i^1 defined by

$$x_i^1 = \begin{cases} \sum_{x_j \in X_1} \frac{w_{ij}}{\sum_{x_j \in X_1} w_{ij}} x_j & \text{if} \quad \sum_{x_j \in X_1} w_{ij} \neq 0 \\ \\ \sum_{x_j \in X_1} w_{ij} x_j & \text{otherwise.} \end{cases} \tag{5.24}$$

Then the pasting conditions apply specifically at those points. The mapping f defines a proper locally barycentric mapping if either two affine mappings f_1 and f_2 or their linear parts coincide at the pasting points. Precisely, we have

Proposition 5.2.1. *Let (X, w) be a barycentric weighted graph of $\mathbb{R}^n$. Let $X = X_1 \cup X_2$ be a partition of X and let f_1 and f_2 be two affine mappings from $\mathbb{R}^n$ to $\mathbb{R}^k$ such that*

1. *For all $x_i \in X_1$ with at least one neighbor in X_2*

$$f_1(x_i^2) = f_2(x_i^2) \quad \textit{if} \quad \sum_{x_j \in X_2} w_{ij} \neq 0 \quad \textit{otherwise} \quad L_{f_1}(x_i^2) = L_{f_2}(x_i^2) \tag{5.25}$$

2. *For all $x_i \in X_2$ with at least one neighbor in X_1*

$$f_1(x_i^1) = f_2(x_i^1) \quad \textit{if} \quad \sum_{x_j \in X_1} w_{ij} \neq 0 \quad \textit{otherwise} \quad L_{f_1}(x_i^1) = L_{f_2}(x_i^1) \tag{5.26}$$

Then the piecewise affine mapping f obtained by gluing f_1 and f_2 together along the partition $X = X_1 \cup X_2$ defines a locally barycentric mapping of (X, w) in $\mathbb{R}^k$.

Pasting condition for affine mappings from M to $\mathbb{R}^k$ Finally, in the case where only the embedding space is Euclidean, say $M' = \mathbb{R}^k$, if we assume that there exists at least one affine mapping f_0 from M to $\mathbb{R}^k$, then for any affine mappings Φ_1 and Φ_2 of $\mathbb{R}^k$, the maps $\Phi_1 \circ f_0$ and $\Phi_2 \circ f_0$ are affine mapping from M to $\mathbb{R}^k$ and it is equivalent to apply the pasting lemma to $\Phi_1 \circ f_0$ and $\Phi_2 \circ f_0$ on (X, w) and to apply the previous proposition to Φ_1 and Φ_1 on $(f_0(X), w)$.

Decomposition of locally barycentric mappings Now we derive from the pasting lemma a similar result but for a barycentric weighted set partitioned in t strongly connected components. Precisely, we state conditions for a piecewise affine mapping obtained by gluing affine mappings together along the strongly connected components of the barycentric weighted set to define a locally barycentric mapping of such set. We refer to this result as the decomposition of locally barycentric mappings after the underlying decomposition of the graph. Note however that this is not a result that allows to decompose any locally affine mapping.

Decomposition theorem In the most general case where the data space is a Riemannian manifold of dimension n and the embedding space is a Riemannian manifold M' of dimension k, decomposition of locally barycentric mappings states the following

Theorem 5.2.1. *Let (X, w) be a barycentric weighted graph of M. Let $X_1, \ldots, X_t$ denote its ordered strongly connected components and let $f_1, \ldots, f_t$ be affine mappings from M to M' such that for all $1 \leq p \leq t-1$, the affine mapping f_p satisfies*

$$\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} \log_{f_p(x_i)} f_p(x_j) = \sum_{q>p} \sum_{x_j \in X_q} w_{ij} \log_{f_p(x_i)} f_q(x_j) \tag{5.27}$$

for all $x_i \in X_p$ with at least one neighbor in at least one of the greater components X_q. Then the map f obtained by gluing $f_1, \dots, f_t$ together

$$f|_{X_p} = f_p \tag{5.28}$$

defines a locally barycentric mapping of (X, w) in M'.

Remark 5.2.1. *Before moving to the proof let us make a useful observation. Looking back at the proof of the pasting lemma carefully, it is enough to ask for f_1 to preserve the barycenters $\mathrm{bar}(x_j, w_{ij})_{1\leq j\leq N}$ for all $1 \leq i \leq N$. The same applies to f_2. Therefore, the pasting lemma still holds for f_1 and f_2 two locally barycentric mappings of (X, w). This is an important remark, first because it will make the proof of the theorem easier, but more importantly because it eventually allows to glue together two locally barycentric mappings that have been generated thanks to the decomposition theorem.*

Proof. This statement is proven by induction using Remark 5.2.1. Let $1 \leq p \leq t-1$ and assume the map f obtained by gluing $f_{p+1}, \dots, f_t$ together along $X_{p+1}, \dots, X_t$ is a locally barycentric embedding. Let us apply the remark to f_p and f. Then the map obtained by gluing f_p and f along X_p and $\bigcup_{q>p} X_q$ is a locally barycentric embedding if for all x_i with at least one neighbor in $\bigcup_{q>p} X_q$, the mapping f_p satisfies the pasting condition

$$\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} \log_{f_p(x_i)}(f_p(x_i)) = \sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} \log_{f_p(x_i)}(f(x_i)) \tag{5.29}$$

$$= \sum_{q>p} \sum_{x_j \in X_q} w_{ij} \log_{f_p(x_i)}(f_q(x_i)) \tag{5.30}$$

Since there exists no $x_i \bigcup_{q>p} X_q$ that has neighbors in X_p by definition of strongly connected components, there is no pasting condition on f. □

Decomposition of locally barycentric mappings in $\mathbb{R}^k$ Again, in the Euclidean case, the decomposition theorem has a nice geometric interpretation in terms of pointwise pasting. For all $x_i \in X_p$ with at least one neighbor in at least one of the greater components X_q, let us introduce for all $q > p$ the point

$$x_i^q = \begin{cases} \sum_{x_j \in X_q} \frac{w_{ij}}{\sum_{x_j \in X_q} w_{ij}} x_j & \text{if} \quad \sum_{x_j \in X_q} w_{ij} \neq 0 \\ \sum_{x_j \in X_q} w_{ij} x_j & \text{otherwise} \end{cases} \tag{5.31}$$

and the weight

$$w_i^q = \begin{cases} \frac{\sum_{x_j \in X_q} w_{ij}}{\sum_{x_j \in \bigcup_{r>p} X_r} w_{ij}} & \text{if} \quad \sum_{x_j \in \bigcup_{r>p} X_r} w_{ij} \neq 0 \quad \text{and} \quad \sum_{x_j \in X_q} w_{ij} \neq 0 \\ \frac{1}{\sum_{x_j \in \bigcup_{r>p} X_r} w_{ij}} & \text{if} \quad \sum_{x_j \in \bigcup_{r>p} X_r} w_{ij} \neq 0 \quad \text{and} \quad \sum_{x_j \in X_q} w_{ij} = 0 \\ \sum_{x_j \in X_q} w_{ij} & \text{if} \quad \sum_{x_j \in \bigcup_{r>p} X_r} w_{ij} = 0 \quad \text{and} \quad \sum_{x_j \in X_q} w_{ij} \neq 0 \\ 1 & \text{if} \quad \sum_{x_j \in \bigcup_{r>p} X_{qr}} w_{ij} = 0 \quad \text{and} \quad \sum_{x_j \in X_q} w_{ij} = 0. \end{cases} \tag{5.32}$$

Then we have the following decomposition result

Proposition 5.2.2. *Let (X, w) be a barycentric weighted set of $\mathbb{R}^n$. Let $X_1, \ldots, X_t$ denote its ordered strongly connected components and let $f_1, \ldots, f_t$ be affine mappings from $\mathbb{R}^n$ to $\mathbb{R}^k$ such that for all $1 \leq p \leq t-1$, for all $x_i \in X_p$ with at least one neighbor in at least one of the greater components X_q, the affine mapping f_p satisfies, if $\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} \neq 0$,*

$$f_p\Big(\sum_{q>p} w_i^q x_i^q\Big) = \sum_{q>p} w_i^q \begin{cases} f_q(x_i^q) & \text{if} \quad \sum_{x_j \in X_q} w_{ij} \neq 0 \\ \\ L_{f_q}(x_i^q) & \text{otherwise,} \end{cases} \tag{5.33}$$

and otherwise

$$L_{f_p}\Big(\sum_{q>p} w_i^q x_i^q\Big) = \sum_{q>p} w_i^q \begin{cases} f_q(x_i^q) & \text{if} \quad \sum_{x_j \in X_q} w_{ij} \neq 0 \\ \\ L_{f_q}(x_i^q) & \text{otherwise.} \end{cases} \tag{5.34}$$

Then the map f obtained by gluing $f_1, \ldots, f_t$ together along the decomposition $X = X_1 \cup \ldots \cup X_t$ defines a locally barycentric mapping of (X, w) in $\mathbb{R}^k$.

In the last case where the data space is a Riemannian manifold M and the embedding space is $\mathbb{R}^k$, if we assume that there exists at least one affine mapping f_0 from M to $\mathbb{R}^k$, then for any affine mappings Φ_1 and Φ_2 of $\mathbb{R}^k$, the maps $\Phi_1 \circ f_0, \ldots, \Phi_t \circ f_0$ are affine mapping from M to $\mathbb{R}^k$ and it is equivalent to apply the pasting lemma to $\Phi_1 \circ f_0, \ldots, \Phi_t \circ f_0$ on (X, w) and to apply the previous proposition to $\Phi_1, \ldots, \Phi_t$ on $(f_0(X), w)$.

Example Although the formulation of the decomposition theorem is quite heavy in terms of notations and may be hard to digest from the first reading, the interpretation behind is very simple. Let us take an example in the plane again to illustrate the theorem. We consider this time the case of a barycentric weighted set (X, w) decomposed into two strongly connected components X_1 and X_2 such that there exists a unique point x_1 in X_1 with at least one neighbor in X_2 as illustrated in Figure 5.3. The point x_1 has neighbors both in X_1 and X_2 and its barycentric coordinates are the following

$$x_1 = \frac{1}{2}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_4 \tag{5.35}$$

Now let f_1 and f_2 be two affine transformations of the plane and let f_1 acts on X_1 and f_2 acts on X_2. Then the corresponding transformation of X defines a locally affine mapping of (X, w) provided that f_1 and f_2 coincide at the point x_1^2

$$f_1(x_1^2) = f_2(x_1^2) \quad \text{where} \quad x_1^2 = \tfrac{1}{2}x_3 + \tfrac{1}{2}x_4 \tag{5.36}$$

Finally, another way of looking at the pasting condition and the decomposition theorem is as precious recipes for generating new locally barycentric mappings from the known set of affine mappings, such that the decomposition theorem is rather a composition theorem. Based this observation, let us now derive bounds on the number of locally barycentric mappings of a barycentric weighted set or equivalently the number locally barycentric embeddings of the corresponding dataset.

5.3 Dimensionality of locally barycentric embeddings

This last section is of particular interest regarding the next chapter. We propose to bound the number of locally barycentric mappings of a barycentric weighted set from below by the number of locally barycentric mappings defined by piecewise affine mappings satisfying the decomposition theorem. We restrict ourselves to the case where the embedding space is $\mathbb{R}^k$. In that case, the set of locally barycentric mappings of a barycentric weighted set is a vector space and so we study more precisely the dimension of such set.

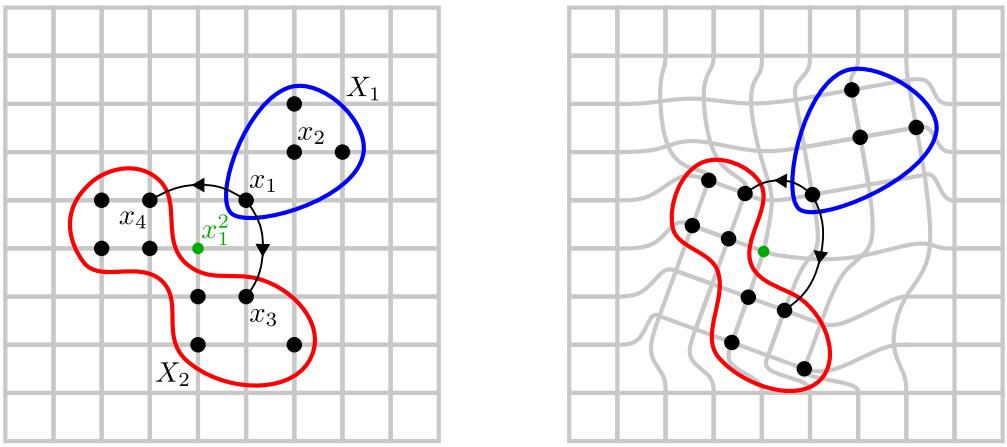

Figure 5.3: Locally barycentric mapping of a barycentric weighted set with two strong connected components. The affine mapping acting on the first component (blue) is constrained to coincide with that of acting on the second component (red) on one point (green). Essentially, such a constraint is a constraint in translation. As a consequence, the deformation of the grid induced by the global piecewise affine mapping is smaller than that of Figure 5.2.

5.3.1 The vector space $\mathrm{B}(X, w, k)$

Coming back to Definition 5.1.2 of a locally affine mapping in $\mathbb{R}^k$ of a barycentric set of M, we check easily that the set of all locally barycentric mappings of a given barycentric weighted set. Let us define the following

Definition 5.3.1. *Let (X, w) be a barycentric weighted graph of M. Then the set of locally barycentric mappings of (X, w) in $\mathbb{R}^k$ is a vector space and is denoted by $\mathrm{B}(X, w, k)$.*

Graph characterization Let us take advantage of this section to make another link between locally barycentric mappings and the graph structure of barycentric weighted sets. As we discussed in Chapter 2, Saul and Roweis (2003) proved that the optimal barycentric mapping problem can be formulated as an eigenvalue problem. In fact, Belkin and Niyogi (2003) highlighted that such an eigenvalue problem is that of finding the first eigenvectors of the graph Laplacian matrix of the neighbor graph, defined as follows

Definition 5.3.2. *Let (X, w) be a barycentric weighted set of M. The symmetrized (normalized) graph Laplacian matrix of (X, w) is the matrix*

$$L_w = \Delta_w^T \Delta_w = (Id - w)^T (Id - w) \tag{5.37}$$

It is a symmetric positive semi-definite matrix.

This way of looking at the problem allows to compare Locally Linear Embedding with other methods based on the graph Laplacian and to study the convergence of optimal solutions using Laplace Beltrami operator (H.-T. Wu and N. Wu, 2018). Now note that we have the following identity

Proposition 5.3.1. *Let (X, w) be a barycentric weighted set of M. Then*

$$\mathrm{B}(X, w, k) = \ker(\Delta_w)^k = \ker(L_w)^k. \tag{5.38}$$

In fact, all the decomposition results we present in Section 5.2 can be formulated equivalently in terms of the decomposition of the graph Laplacian matrix. However, we prefer the geometric approach to such results because we think it is easier to interpret. And the Laplacian matrix is mainly of computational interest to us, as we see in the Chapter 6.

Now in the case of a graph with positive weights, the dimension of the kernel of the graph Laplacian corresponds to the number of connected components of the graph. In our case, since barycentric weighted sets may carry negative weights, such a result is no longer true, and only one of the two inequalities remains valid. In fact, there exists a slightly more precise result

Proposition 5.3.2. *Let (X, w) be a barycentric weighted graph of M. Assume that (X, w) has s isolated components. Then we have the inequality*

$$\dim \mathrm{B}(X, w, k) \geq ks \tag{5.39}$$

This bound is the consequence of a decomposition result on the normalized graph Laplacian (Bauer, 2012). In our case, since the weights w reflects some information on the dataset X, then we are able to say more than this inequality.

5.3.2 Lower bounds on the dimension of $\mathrm{B}(X, w, k)$

In this section, we derive from the decomposition theorem a series of lower bounds on the dimension of the space $\mathrm{B}(X, w, k)$ of locally barycentric mappings in $\mathbb{R}^k$ of a barycentric weighted set of M. The point here is not to compute the dimension of $\mathrm{B}(X, w, k)$ – which is easy to do from the Laplacian matrix of (X, w) anyway – but rather to predict it based on the data itself. It is also a question of showing that the mappings we have characterized through the decomposition theorem form a significant subset of $\mathrm{B}(X, w, k)$.

Let (X, w) be a barycentric weighted set of $\mathbb{R}^n$. In the first section of this chapter we leverage the fact that any affine mapping from $\mathbb{R}^n$ to $\mathbb{R}^k$ defines a locally barycentric mapping of (X, w) to show that

$$\dim \mathrm{B}(X, w, k) \geq k \operatorname{rank}(X), \tag{5.40}$$

where we recall that the notation $\operatorname{rank}(X)$ corresponds here to the affine rank of X, that is the cardinality of a maximal affinely independent subset of X. Similarly, in the case where (X, w) is a barycentric weighted set of a Riemannian manifold M, we show that under the assumption that there exists an affine mapping f_0 from M to $\mathbb{R}^k$, we have the following inequality

$$\dim \mathrm{B}(X, w, k) \geq k \operatorname{rank}(f_0(X)). \tag{5.41}$$

Now more generally, we can derive better bounds from Theorem 5.2.1 on the decomposition of locally barycentric mappings. First, let us assume that the dataset lies in a Euclidean space. Then we have the following

Theorem 5.3.1. *Let (X, w) be a barycentric weighted set of $\mathbb{R}^n$ and let $X_1, \ldots, X_t$ denote its strongly connected components. Assume that for all $1 \leq p \leq t-1$, for all $x_i \in X_p$ with at least one neighbor in $\bigcup_{q>p} X_q$, we have $\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} \neq 0$. Assume that for all $1 \leq p \leq t-1$, the pasting conditions at X_p are independent, that is that the points $\sum_{q>p} w_i^q x_i^q$, for $x_i \in X_p$ with at least one neighbor in $\bigcup_{q>p} X_q$, are affinely independent. Then we have the inequality*

$$\dim \mathrm{B}(X, w, k) \geq k \operatorname{rank}(X_t) + k \sum_{p=1}^{t-1} (\operatorname{rank}(X_p) - K_p). \tag{5.42}$$

where K_p denotes the number of pasting conditions at X_p

$$K_p = \mathrm{card}\Big\{ \sum_{q>p} w_i^q x_i^q \mid x_i \in X_p,\ x_i \to \bigcup_{q>p} X_q \Big\} \tag{5.43}$$

Proof. The set of all piecewise affine locally barycentric mappings $f(X)$ obtained by gluing together t affine mappings $f_1, \ldots, f_t$ from $\mathbb{R}^n$ to $\mathbb{R}^k$ along $X_1, \ldots, X_t$ according to Theorem 5.2.1 is a vector subspace of $\mathrm{B}(X, w, k)$. Let f be such a mapping.

Let $1 \leq p \leq t$. Let us then introduce the set

$$X_p^> = \Big\{ \sum_{q>p} w_i^q x_i^q \mid x_i \in X_p,\ x_i \to \bigcup_{q>p} X_q \Big\}. \tag{5.44}$$

First, since we assumed that $\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} \neq 0$, observe that for all $\sum_{q>p} w_i^q x_i^q \in X_p^>$, we have the identity

$$\begin{aligned}
\sum_{q>p} w_i^q x_i^q &= \frac{1}{\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij}} \sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} x_j && (5.45)\\
&= \frac{1}{\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij}} \Big(x_i - \sum_{x_j \in X_p} w_{ij} x_j \Big) && (5.46)\\
&= \frac{1}{1 - \sum_{x_j \in X_p} w_{ij}} \Big(x_i - \sum_{x_j \in X_p} w_{ij} x_j \Big) && (5.47)
\end{aligned}$$

such that $\sum_{q>p} w_i^q x_i^q$ is in the affine span of X_p, which we denote by $\mathrm{aff}(X_p)$. We assumed that $X_p^>$ is affinely independent. According to the previous observation, this means in particular that $K_p = \mathrm{card}(X_p^>) \leq \mathrm{rank}(X_p)$. Let us complete $X_p^>$ in an affine basis $B_p = X_p^> \cup B_p'$ of $\mathrm{aff}(X_p)$. Then $f_p(X_p)$ is completely determined by $f_p(B_p)$. Now, assume that for all $q > p$, the set $f_q(X_q)$ has already been fixed. The decomposition theorem states that f_p satisfies the pasting condition

$$f_p\Big(\sum_{q>p} w_i^q x_i^q \Big) = \sum_{q>p} w_i^q \begin{cases} f_q(x_i^q) & \text{if} \quad \sum_{x_j \in X_q} w_{ij} \neq 0 \\ \\ L_{f_q}(x_i^q) & \text{otherwise.} \end{cases} \tag{5.48}$$

For all $q > p$, if the first case applies, because $x_i^q \in \mathrm{aff}(X_q)$ and because f_q is an affine mapping, the value of f_q at x_i^q has already been fixed. If the second case applies, notice that we have $L_{f_q}(x_i^q) = \sum_{x_j \in X_q} w_{ij} f_q(x_j)$ such that the value of L_{f_q} at x_i^q has also been fixed. because Therefore, the value of f_p at every $\sum_{q>p} w_i^q x_i^q \in X_p^>$ is fixed such that $f_p(X_p)$ is completely determined by $f_p(B_p')$ only.

Unrolling our reasoning, we show by induction that the set $f(X)$ is completely determined by $f_t(X_t)$ and $f_1(B_1'), \ldots, f_{t-1}(B_{t-1}')$. Now as we already discussed in Section 5.1, the set of all $f_t(X_t)$ for f_t an affine mapping from $\mathbb{R}^n$ to $\mathbb{R}^k$ is a vector space of dimension $k\,\mathrm{rank}(X_t)$. And for $1 \leq p \leq t-1$, the set of all $f_p(B_p')$ for f_p an affine mapping from $\mathbb{R}^n$ to $\mathbb{R}^k$ is a vector space of dimension

$$k\,\mathrm{rank}(B_p') = k\mathrm{card}(B_p') = k(\mathrm{card}(B_p) - \mathrm{card}(X_p^>)) = k(\mathrm{rank}(X_p) - K_p) \tag{5.49}$$

and the result follows. □

Let us then move to the general case where the dataset lies in a Riemannian manifold M. We assume that there exists at least one affine mapping f_0 from M to $\mathbb{R}^k$. Then we can show that the dimension of $\mathrm{B}(f_0(X), w, k)$ satisfies Equation (5.42). Moreover, since the composition of f_0 by any affine mapping of $\mathbb{R}^k$ defines a locally barycentric mapping of (X, w), then we have that $\mathrm{B}(f_0(X), w, k) \subset \mathrm{B}(X, w, k)$ and we derive the following

Corollary 5.3.1. *Let (X, w) be a barycentric weighted set of M and let $X_1, \ldots, X_t$ denote its strongly connected components. Assume that there exists at least one affine mapping f_0 from M to $\mathbb{R}^k$. For all $1 \leq i \leq N$, let $y_i = f_0(x_i)$ and for all $1 \leq p \leq t$, let $Y_p = f_0(X_p)$. Assume that for all $1 \leq p \leq t-1$, for all $y_i \in Y_p$ with at least one neighbor in $\bigcup_{q>p} Y_q$, we have $\sum_{x_j \in \bigcup_{q>p} Y_q} w_{ij} \neq 0$. Assume that for all $1 \leq p \leq t-1$, the pasting conditions at Y_p are independent, that is that the points $\sum_{q>p} w_i^q y_i^q$, for $xy_i \in Y_p$ with at least one neighbor in $\bigcup_{q>p} Y_q$, are affinely independent. Then we have the inequality*

$$\dim \mathrm{B}(X, w, k) \geq k \operatorname{rank}(Y_t) + k \sum_{p=1}^{t-1} (\operatorname{rank}(Y_p) - K_p) \tag{5.50}$$

where K_p denotes the number of pasting conditions at Y_p

$$K_p = \operatorname{card}\Big\{ \sum_{q>p} w_i^q y_i^q \mid y_i \in Y_p,\ y_i \to \bigcup_{q>p} Y_q \Big\} \tag{5.51}$$

Finally let us derive from these two bounds two other lower but simpler bounds based on counting the number of isolated components within the barycentric weighted set.

Corollary 5.3.2. *Let (X, w) be a barycentric weighted set of $\mathbb{R}^n$ and let $X_1, \ldots, X_t$ denote its strongly connected components. Assume that for all $1 \leq p \leq t-1$, for all $x_i \in X_p$ with at least one neighbor in $\bigcup_{q>p} X_q$, we have $\sum_{x_j \in \bigcup_{q>p} X_q} w_{ij} \neq 0$. Assume that for all $1 \leq p \leq t-1$, the pasting conditions at X_p are independent, that is that the points $\sum_{q>p} w_i^q x_i^q$, for $x_i \in X_p$ with at least one neighbor in $\bigcup_{q>p} X_q$, are affinely independent. Assume that (X, w) has s isolated components $X_{p=t-s+1}, \ldots, X_t$. Then we have the inequality*

$$\dim \mathrm{B}(X, w, k) \geq k \sum_{p=t-s+1}^{t} \operatorname{rank}(X_p) \tag{5.52}$$

Corollary 5.3.3. *Let (X, w) be a barycentric weighted set of M and let $X_1, \ldots, X_t$ denote its strongly connected components. Assume that there exists at least one affine mapping f_0 from M to $\mathbb{R}^k$. For all $1 \leq p \leq t$, let $Y_p = f_0(X_p)$. Assume that for all $1 \leq p \leq t-1$, for all $y_i \in Y_p$ with at least one neighbor in $\bigcup_{q>p} Y_q$, we have $\sum_{x_j \in \bigcup_{q>p} Y_q} w_{ij} \neq 0$. Assume that for all $1 \leq p \leq t-1$, the pasting conditions at Y_p are independent, that is that the points $\sum_{q>p} w_i^q y_i^q$, for $xy_i \in Y_p$ with at least one neighbor in $\bigcup_{q>p} Y_q$, are affinely independent. Assume that (X, w) has s isolated components $X_{p=t-s+1}, \ldots, X_t$. Then we have the inequality*

$$\dim \mathrm{B}(X, w, k) \geq k \sum_{p=t-s+1}^{t} \operatorname{rank}(Y_p) \tag{5.53}$$

Remark 5.3.1. *In fact, the two previous inequalities still hold without any hypothesis on the independency of the pasting conditions. Similar to that of Proposition 5.3.2, the proof relies on the decomposition of the normalized graph Laplacian.*

Proof. Let (X, w) be a barycentric weighted set of $\mathbb{R}^n$ and let $X_1, \dots, X_t$ denote its strongly connected components. Then up to some permutation of its rows and columns, the normalized graph Laplacian matrix of (X, w) may be written

$$\Delta_w = \begin{bmatrix} \Delta_1 & \Delta_{12} & \dots & \Delta_{1t} \\ 0 & \Delta_2 & \dots & \Delta_{2t} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \Delta_t \end{bmatrix} \tag{5.54}$$

This is referred as the Frobenius normal form of such (Brualdi et al., 2018). We recall that one has

$$\mathrm{B}(X, w, k) = \ker(\Delta_w)^k \tag{5.55}$$

Now assume that the components $X_{p=t-s+1}, \dots, X_t$ are isolated. Then we have

$$\Delta_{pq} = 0 \qquad \forall q > p > t - s \tag{5.56}$$

Writing Δ_w as a sum of $s + 1$ sparse matrices in such a way that each submatrix Δ_p for $p > s - t$ appears exactly in one of these, we get the following inequality on the rank of Δ_w

$$\mathrm{rank}(\Delta_w) \leq \mathrm{rank}\left(\begin{bmatrix} \Delta_1 & \Delta_{12} & & \dots & & \Delta_{1t} \\ 0 & \Delta_2 & & \dots & & \Delta_{2t} \\ \vdots & \ddots & \ddots & & & \vdots \\ 0 & \dots & 0 & \Delta_{t-s} & \dots & \Delta_{t-s\,t} \end{bmatrix} \right) + \sum_{p=t-s+1}^{t} \mathrm{rank}(\Delta_p). \tag{5.57}$$

Equivalently, the dimension of $\ker(\Delta_w)$ satisfies the inequality

$$\dim \ker(\Delta_w) \geq \sum_{p=t-s+1}^{t} \dim \ker(\Delta_p). \tag{5.58}$$

Finally, notice that for all $p > t - s$, the submatrix Δ_p is the graph Laplacian of the barycentric weighted set $(X_p, w|_{X_p})$, where $w|_{X_p}$ denotes the submatrix of w which consists of the rows and columns indexed by X_p. Following Equation (5.40), we have then

$$\begin{aligned} \dim \mathrm{B}(X, w, k) = k \dim \ker(\Delta_w) &\geq \sum_{p=t-s+1}^{t} k \dim \ker(\Delta_p) && (5.59) \\ &= \sum_{p=t-s+1}^{t} \dim \ker(\Delta_p)^k && (5.60) \\ &= \sum_{p=t-s+1}^{t} \dim \mathrm{B}(X_p, w|_{X_p}, k) && (5.61) \\ &\geq \sum_{p=t-s+1}^{t} k \,\mathrm{rank}(X_p). && (5.62) \end{aligned}$$

□

In practice, we can assume that k has been well chosen such that the dataset X is at least of local intrinsic dimension k. In other words, we can assume that the affine rank of any subset of X that consist of more than $k + 1$ points is at least $k + 1$. Such assumption allows to estimate more easily the several bounds we listed.

5.3.3 Asymptotic behavior of the dimension of $\mathrm{B}(X, w, k)$

In this section, we propose to estimate the asymptotic behavior of the dimension of $\mathrm{B}(X, w, k)$ when the size of X increases. More precisely, we study the asymptotic behavior of the bound introduced in the last corollary of the previous section. We show in particular that the number of isolated components of a barycentric weighted set (X, w) is asymptotically proportional to the size of X and so does the corresponding bound, although this is not necessarily intuitive in the first place.

Now let C be a compact domain of $\mathbb{R}^n$ and assume that C has non empty interior. Let X be a random finite subset of C such that for all measurable set $A \subset C$, the random variable $n_A = |X \cap A|$ has a Poisson distribution $\mathcal{P}(\lambda\mu(A))$, where μ is a measure on C which is absolutely continuous with respect to the Lebesgue measure λ_n on $\mathbb{R}^n$, and $X \cap A | n_A$ follows the law of a set of n_A i.i.d. random vectors sampled from the density of μ on A.

Proposition 5.3.3. *Assume that $\mu = \lambda_n$. Then we have the following inequality*

$$\mathbb{E}\big[\dim \mathrm{B}(X, w, k)\big] \geq c_n k(n+1) \left\lfloor \frac{(a\sqrt{\pi})^n}{2^n \Gamma\left(\frac{n}{2}+1\right)} \lambda \right\rfloor . \tag{5.63}$$

where $c_n = \big(e(n+2)!3^{n(n+2)}\big)^{-1}$ is a constant which depends only on n and $a > 0$ is the side length of the largest cubic domain $\mathcal{C}(x, a)$ that fits into C.

Proof. We know that $\dim \mathrm{B}(X, w, k)$ is bounded from below by

$$\dim \mathrm{B}(X, w, k) \geq k \sum_{p=t-s+1}^{t} \operatorname{rank}(X_p) \tag{5.64}$$

where s denotes the number of isolated components of (X, w).

Now let us make a simple observation. Suppose that w has been computed along the $n+1$-nearest neighbor graph of X so as to ensure that the barycentric model fits exactly to X (which is a priori n-dimensional). Consider a ball of radius r in C and assume that it contains exactly $n+2$ points of C. Then if these points lie within the concentric ball of radius $r/3$, they create an isolated component of (X, w).

Fix $r > 0$ and let $\{\mathcal{B}_i = \mathcal{B}(x_i, r)\,,\ 1 \leq i \leq s\}$ be a finite family of disjoints balls of radius r included in C. Let also $\mathcal{B}'_i = B(x_i, r/3)$ be the ball with B_i but with a radius three time smaller. Following the previous observation, we have the inequality

$$\mathbb{E}[\dim \mathrm{B}(X, w, k)] \geq \mathbb{E}\left[\sum_{i=1}^{s} k \operatorname{rank}(X \cap \mathcal{B}_i) \mathbb{1}_{\left\{n_{\mathcal{B}_i} = n+2\right\} \cap \left\{X \cap \mathcal{B}_i \subset \mathcal{B}'_i\right\}}\right] \tag{5.65}$$

$$= k(n+1) \sum_{i=1}^{s} \mathbb{E}\left[\mathbb{1}_{\left\{n_{\mathcal{B}_i} = n+2\right\} \cap \left\{X \cap \mathcal{B}_i \subset \mathcal{B}'_i\right\}}\right]. \tag{5.66}$$

Let us compute each expectation in the previous formula. For all $A \in C$, the random variable $X | n_A$

is a set of n_A i.i.d. random vectors sampled from the uniform distribution on A such that we have

$$\mathbb{E}\left[\mathbb{1}_{\left\{n_{\mathcal{B}_i}=n+2\right\}\cap\left\{X\cap\mathcal{B}_i\subset\mathcal{B}_i'\right\}}\right] = \mathbb{P}\left[X\cap\mathcal{B}_i\subset\mathcal{B}_i' \,\middle|\, n_{\mathcal{B}_i}=n+2\right]\mathbb{P}\left[n_{\mathcal{B}_i}=n+2\right] \tag{5.67}$$

$$= \left(\frac{\lambda_n\left(\mathcal{B}_i'\right)}{\lambda_n(\mathcal{B}_i)}\right)^{n+2}\frac{\left(\lambda\lambda_n(\mathcal{B}_i)\right)^{n+2}}{(n+2)!}\exp\left(-\lambda\lambda_n(\mathcal{B}_i)\right) \tag{5.68}$$

$$= \frac{1}{3^{n(n+2)}}\frac{\left(\lambda r^n\lambda_n\left(\mathcal{B}(0,1)\right)\right)^{n+2}}{(n+2)!}\exp\left(-\lambda r^n\lambda_n\left(\mathcal{B}(0,1)\right)\right). \tag{5.69}$$

Set $r_\lambda > 0$ such that

$$\exp\left(-\lambda r_\lambda^n\lambda_n\left(\mathcal{B}(0,1)\right)\right) = 0, \tag{5.70}$$

that is

$$r_\lambda = \left(\lambda_n\left(\mathcal{B}(0,1)\right)\lambda\right)^{-\frac{1}{n}} = \frac{\Gamma\left(\frac{n}{2}+1\right)^{\frac{1}{n}}}{\sqrt{\pi}}\lambda^{-\frac{1}{n}} \tag{5.71}$$

Accordingly, let $\{\mathcal{B}_i\,,\ 1\leq i\leq s\}$ be a finite family of disjoints balls of radius r_λ. The expression above reduces to

$$\mathbb{E}\left[\mathbb{1}_{\left\{n_{\mathcal{B}_i}=n+2\right\}\cap\left\{X\cap\mathcal{B}_i\subset\mathcal{B}_i'\right\}}\right] = \frac{1}{\mathrm{e}(n+2)!3^{n(n+2)}}. \tag{5.72}$$

Finally, let $a>0$ be the side length of the largest cubic domain $\mathcal{C}(x,a)$ that fits into C. Then the number of balls of radius r_λ that can fit into $\mathcal{C}(x,a)$ is $\left\lfloor\left(\frac{a}{2r_\lambda}\right)^n\right\rfloor$. Let us take then $s=\left\lfloor\left(\frac{a}{2r_\lambda}\right)^n\right\rfloor$. We have the inequality

$$\mathbb{E}\big[\dim\mathrm{B}(X,w,k)\big] \geq \frac{1}{\mathrm{e}(n+2)!3^{n(n+2)}}k(n+1)\left\lfloor\frac{(a\sqrt{\pi})^n}{2^n\Gamma\left(\frac{n}{2}+1\right)}\lambda\right\rfloor. \tag{5.73}$$

□

Proposition 5.3.4. *Now assume that $d\mu=\rho d\lambda_n$ with $\rho_1\leq\rho\leq\rho_2$. Then we have the following inequality*

$$\mathbb{E}\big[\dim\mathrm{B}(X,w,k)\big] \geq \left(\frac{\rho_1}{\rho_2}\right)^{2(n+2)}c_nk(n+1)\left\lfloor\frac{(a\sqrt{\pi})^n}{2^n\Gamma\left(\frac{n}{2}+1\right)}\rho_2\lambda\right\rfloor. \tag{5.74}$$

where $c_k'=c_n\left(\frac{\rho_1}{\rho_2}\right)^{2(n+2)}$ is a constant which depends only on n.

Proof. This proof is similar to the one of Theorem 5.3.3. Fix $r>0$ and let $\{\mathcal{B}_i=\mathcal{B}(x_i,r)\,,\ 1\leq i\leq s\}$ be a finite family of disjoints balls of radius r included in C. Let also $\mathcal{B}_i'=B(x_ir/3)$ be the ball with B_i but with a radius three time smaller. We have the inequality

$$\mathbb{E}[\dim\mathrm{B}(X,w,k)] \geq k(n+1)\sum_{i=1}^{s}\mathbb{E}\left[\mathbb{1}_{\left\{n_{\mathcal{B}_i}=n+2\right\}\cap\left\{X\cap\mathcal{B}_i\subset\mathcal{B}_i'\right\}}\right]. \tag{5.75}$$

Now let us compute each expectation in the previous formula. The random variable $X|n_A$ is a set of n_A i.i.d. random vectors sampled from the probability distribution of density $\rho/\rho(A)$ such that we

have

$$\mathbb{E}\left[\mathbb{1}_{\left\{n_{\mathcal{B}_i}=n+2\right\}\cap\left\{X\cap\mathcal{B}_i\subset\mathcal{B}'_i\right\}}\right] = \mathbb{P}\left[X\cap\mathcal{B}_i\subset\mathcal{B}'_i \,\middle|\, n_{\mathcal{B}_i}=n+2\right]\mathbb{P}\left[n_{\mathcal{B}_i}=n+2\right] \tag{5.76}$$

$$= \left(\frac{\mu\left(\mathcal{B}'_i\right)}{\mu(\mathcal{B}_i)}\right)^{n+2}\frac{\left(\lambda\mu(\mathcal{B}_i)\right)^{n+2}}{(n+2)!}\exp\left(-\lambda\mu(\mathcal{B}_i)\right) \tag{5.77}$$

$$\geq \left(\frac{\rho_1\lambda_n\left(\mathcal{B}'_i\right)}{\rho_2\lambda_n(\mathcal{B}_i)}\right)^{n+2}\frac{\left(\lambda\rho_1\lambda_n(\mathcal{B}_i)\right)^{n+2}}{(n+2)!}\exp\left(-\lambda\rho_2\lambda_n(\mathcal{B}_i)\right) \tag{5.78}$$

$$= \left(\frac{\rho_1}{3\rho_2}\right)^{n+2}\frac{\left(\lambda\rho_1 r^n\lambda_n\left(\mathcal{B}(0,1)\right)\right)^{n+2}}{(n+2)!}\exp\left(-\lambda\rho_2 r^n\lambda_n\left(\mathcal{B}(0,1)\right)\right). \tag{5.79}$$

Recall that we set

$$r_\lambda = \left(\lambda_n\left(\mathcal{B}(0,1)\right)\lambda\right)^{-\frac{1}{n}} = \frac{\Gamma\left(\frac{n}{2}+1\right)^{\frac{1}{n}}}{\sqrt{\pi}}\lambda^{-\frac{1}{n}} \tag{5.80}$$

such that

$$\exp\left(-\lambda r_\lambda^n\lambda_n\left(\mathcal{B}(0,1)\right)\right) = 0 \tag{5.81}$$

and let $\{\mathcal{B}_i\,,\ 1\leq i\leq p\}$ be a finite family of disjoints balls of radius $\rho_2^{-\frac{1}{n}}r_\lambda$. The expression above reduces to

$$\mathbb{E}\left[\mathbb{1}_{\left\{n_{\mathcal{B}_i}=n+2\right\}\cap\left\{X\cap\mathcal{B}_i\subset\mathcal{B}'_i\right\}}\right] = \left(\frac{\rho_1}{\rho_2}\right)^{2(n+2)}\frac{1}{\mathrm{e}(n+2)!3^{n(n+2)}}. \tag{5.82}$$

Finally, let $a>0$ be the side length of the largest cubic domain $\mathcal{C}(x,a)$ that fits into C. Then the number of balls of radius r_λ that can fit into $\mathcal{C}(x,a)$ is $\left\lfloor\left(\frac{a}{r_\lambda}\right)^n\right\rfloor$. Let us take then $s=\left\lfloor\left(\frac{a}{r_\lambda}\right)^n\right\rfloor$. We have the inequality

$$\mathbb{E}\big[\dim \mathrm{B}(X,w,k)\big] \geq \left(\frac{\rho_1}{\rho_2}\right)^{2(n+2)} c_n k(n+1)\left\lfloor\frac{(a\sqrt{\pi})^n}{2^n\Gamma\left(\frac{n}{2}+1\right)}\rho_2\lambda\right\rfloor. \tag{5.83}$$

where we recall that $c_n = \left(\mathrm{e}(n+2)!3^{n(n+2)}\right)^{-1}$. □

Essentially what these two results tell us is that in expectation, the dimension of the space of locally barycentric mapping in $\mathbb{R}^k$ of a barycentric weighted set (X,w) of $\mathbb{R}^n$ increases proportionally to the density of the set X in the compact domain C, represented by the parameter λ.

5.3.4 Conjectures and avenues for further investigation

There are many paths of exploration for this work. First, in practice, the assumption that the optimal local barycentric model fits the data exactly is a priori not satisfied. Therefore, it would be important to study the continuity of our results in the case where the model is close to the data but is not exact, or when the data are noisy. In particular, it is only natural to ask if our solutions would still minimize the cost of the optimal mapping problem, and to what extent the bounds on the dimension of $\mathrm{B}(X,w,k)$ are still true. Another path we want to explore is the generalization of our the results to the case of constant curvature embedding spaces. Especially, we conjecture that the space of locally barycentric mappings of a barycentric weighted set of size N in $\mathbb{R}^k$ is a submanifold of $\mathbb{R}^{N\times k}$ and that we can bound its dimension leveraging the results of Kobayashi (1995). Finally, the description of locally barycentric mappings we propose in this chapter is not exhaustive. Exploring other notions of connectivity for directed graphs should provide new decomposition theorems, allow to characterize other locally barycentric mappings mappings, and yield better bounds. In particular, the notion of vertex cut and higher order vertex connectivity (Diestel, 2017) look promising.

References

Bauer, F. (2012). "Normalized graph Laplacians for directed graphs". In: *Linear Algebra and its Applications* 436.11, pp. 4193–4222. ISSN: 0024-3795. DOI: 10.1016/j.laa.2012.01.020.

Belkin, M. and P. Niyogi (2003). "Laplacian eigenmaps for dimensionality reduction and data representation". In: *Neural Computation* 15.6, pp. 1373–1396. ISSN: 0899-7667. DOI: 10.1162/089976603321780317.

Brualdi, R. A. et al. (2018). *Combinatorial Matrix Theory*. Ed. by A. M. Encinas and M. Mitjana. Advanced Courses in Mathematics - CRM Barcelona. Springer International Publishing. ISBN: 978-3-319-70952-9, 978-3-319-70953-6. DOI: 10.1007/978-3-319-70953-6.

Diestel, R. (2017). *Graph Theory*. Vol. 173. Graduate Texts in Mathematics. Springer. ISBN: 978-3-662-53621-6, 978-3-662-53622-3. DOI: 10.1007/978-3-662-53622-3.

Kobayashi, S. (1995). *Transformation Groups in Differential Geometry*. Classics in Mathematics. Springer. ISBN: 978-3-642-61981-6. DOI: 10.1007/978-3-642-61981-6_2.

Kobayashi, S. and K. Nomizu (1996). *Foundations of Differential Geometry*. Wiley classics library ed. Wiley classics library. Wiley. 2 pp. ISBN: 978-0-471-15733-5, 978-0-471-15732-8.

Lin, L. (2021). *Avoiding unwanted results in locally linear embedding: A new understanding of regularization*. DOI: 10.48550/arXiv.2108.12680.

Polito, M. and P. Perona (2001). "Grouping and dimensionality reduction by locally linear embedding". In: Advances in neural information processing systems. Vol. 14.

Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding". In: *Science* 290.5500, pp. 2323–2326. DOI: 10.1126/science.290.5500.2323.

Saul, L. K. and S. T. Roweis (2003). "Think globally, fit locally: Unsupervised learning of low dimensional manifolds". In: *Journal of Machine Learning Research* 4 (Jun), pp. 119–155. ISSN: ISSN 1533-7928.

Wu, H.-T. and N. Wu (2018). "Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding". In: *The Annals of Statistics* 46.6, pp. 3805–3837. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/17-AOS1676.

Chapter 6

By way of epilogue

New algorithms for geometric learning and geometric manifold learning

In Chapter 2, we highlight the multiplicity of barycentric embeddings. In Chapter 5, we unveil a family of locally barycentric embeddings. In this chapter, we leverage such a multiplicity. Initially, we regarded the non-uniqueness of the solution computed by Locally Linear Embedding (S. T. Roweis and Saul, 2000) as a problem to be solved. In fact, the work we presented in Chapter 5 was initially intended to identify uniqueness conditions. At a certain point, however, we realized that non-uniqueness was not a weakness but rather a strength. Indeed, being able to choose means being able to choose the best. To counterbalance the non-isometric nature of barycentric models, we propose here to discriminate between barycentric embeddings (respectively locally barycentric embeddings) according to a distance-based criterion. We redesign Barycentric Subspace Analysis (Pennec, 2018) and Locally Linear Embedding under this approach and derive two new algorithms for geometric learning and geometric manifold learning.

Each of the two sections focuses on one algorithm. In the section, we describe a dimensionality reduction method which computes an isometric embedding of a dataset based on the model estimated by Barycentric Subspace Analysis. In fact, the idea is already introduced informally in the examples of Chapter 4 on barycentric subspace analysis of graphs. The second section presents a manifold learning method combining Locally Linear Embedding and Isomap (Tenenbaum, De Silva, and Langford, 2000). We experiment this new method on the standard example of the Swiss Roll and compare its performance with both Locally Linear Embedding and Isomap.

6.1 From barycentric subspace analysis to barycentric embedding

Let us recall the difference between Barycentric Subspace Analysis (BSA) and barycentric embeddings. From a dataset, Barycentric Subspace Analysis extracts a k-dimensional barycentric model, that is a set of $k+1$ reference points together with barycentric weights. The reference points are those of the barycentric subspace that fits the dataset the best and the weights are those of the barycentric projection on each data point onto such a barycentric subspace. Barycentric embeddings are a step ahead. The barycentric subspace estimated by the model is embedded in a known space of same dimension, for example $\mathbb{R}^k$, by mapping the reference points onto as many points in the embedding

space. The projection of each data point is then embedded itself as the barycenter of the new reference points with corresponding coordinates the weights estimated by the model. This step is referred to in Chapter 2 as a barycentric mapping. By construction, barycentric embeddings are then not unique, and in the case where the embedding space is $\mathbb{R}^k$, we saw that there are as many embeddings as they are choices of $k+1$ new reference points in the embedding space. The following algorithm, which we refer to as Isometric Barycentric Subspace Analysis (IsoBSA), picks the unique configuration of these new reference points that respects the pairwise distances of the original set of reference points.

6.1.1 Isometric barycentric subspace analysis

Consider a dataset $x_1 \dots, x_N$ on a manifold M. The first step of Algorithm 1 corresponds to the original formulation of BSA and consists in fitting a k-dimensional barycentric model to the dataset. IsoBSA is formulated a priori for manifold-valued data and this first step relies on the several implementations of the barycentric projection we proposed in Chapter 2 and Chapter 3. The purpose of the second step is to select the image of the reference points estimated by the model in the embedding space in such a way that the distance between two reference points, measured with respect to the metric of M, is the same as the distance between their images. If the embedding space is $\mathbb{R}^k$, then there exists an exact solution to such a problem which is unique up to a rotation. In practice, we can compute this solution thanks to Multidimensional Scaling (MDS), which we recall is a dimensionality reduction method that optimizes the conservation of pairwise distances of a dataset (M. A. A. Cox and T. F. Cox, 2008). Although, it has been originally designed for Euclidean distances, the algorithm can be reformulated rather efficiently for Riemannian distances (Harms, Maignant, and Schlager, 2019). Finally, the last step of the algorithm simply retrieves the embedding of the dataset in $\mathbb{R}^k$ accordingly to the new reference points selected. Together with the embedding, the algorithm also outputs the original reference points that provide multiple estimators of the dataset.

Algorithm 1 IsoBSA

Input: $x_1, \dots, x_N \in M$
$1 \leq k \leq n, N-1$

1: **Compute** $a_0, \dots, a_k, w_{ij} = \texttt{barycentric\_model}(x_1, \dots, x_N, k)$

2: **Compute** $b_0, \dots, b_k = \texttt{MDS}(a_0, \dots, a_k, k)$

3: **For all** $i = 1$ **to** N **set** $y_i = \sum_{j=0}^{k} w_{ij} b_j$

Output: $a_0, \dots, a_k \in \mathbb{R}^k$
$y_1, \dots, y_N \in \mathbb{R}^k$

Naturally, IsoBSA is still compatible with any variant of BSA, especially sample-limited BSA and convex sample-limited BSA. The criterion is also perfectly exchangeable with any other criterion, for example based on a different distance function. Note that designing a similar algorithm to embed a dataset in a constant curvature space, in such a way that it remains computationally efficient, may be quite challenging, especially due to the increasing complexity of step 3 as soon as the barycenters are not formulated explicitly anymore.

6.1.2 Coming back to an example in graph analysis

Although it had not been formalized yet as that point, we already experimented IsoBSA. Precisely in Chapter 4, on several datasets of spectral graph spaces. In fact, in the case of spectral graph spaces, the method is more specific than just IsoBSA. Indeed, barycentric subspaces of spectral graph spaces are flat. Therefore, the barycentric mapping implemented by the algorithm is simply a Euclidean isometry. In general, this is not the case and only the distances between the reference points are conserved. We reuse here one the examples of Chapter 4 to illustrate how IsoBSA works (see Figure 6.1). Specifically, the algorithm used here is its sample-limited variant.

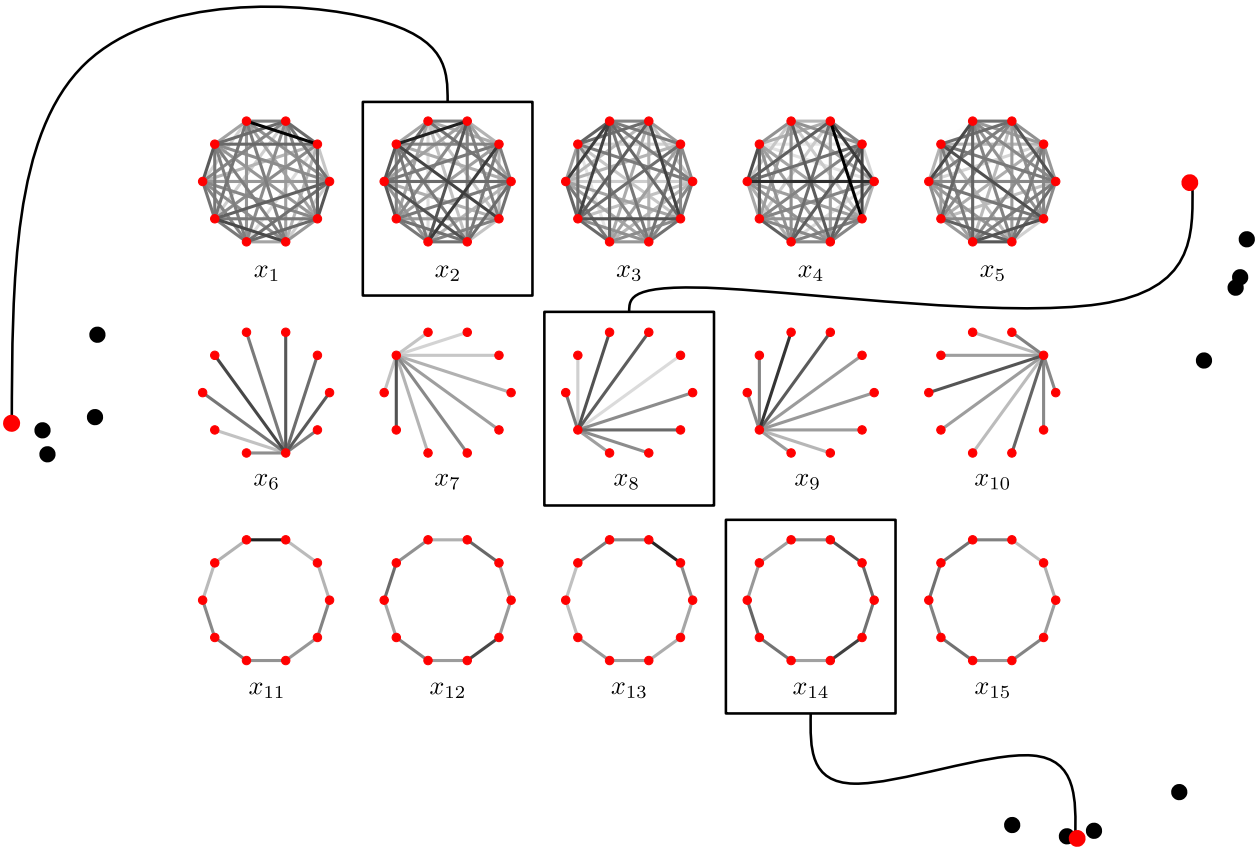

Figure 6.1: Sample-limited IsoBSA of a set of graphs. The dataset, originally of dimension 9, is embedded in the plane with respect to the three reference points picked with the dataset. As we already discussed, the providing of the method of reference points in addition to the embedding is a very powerful feature in terms of interpretability

6.2 A combined approach to geometric manifold learning

The second algorithm reflects the same philosophy as IsoBSA. As well as barycentric embeddings, locally barycentric embeddings are structured in two steps. The dataset is first approximated by a local barycentric model, distributed along the neighbor graph of the dataset. The projection of each data point on the barycentric subspace of its neighbors yield local barycentric weights. Then, from these barycentric weights, we need to solve the optimal mapping problem, that is find the embedding of the dataset in the embedding space, for example $\mathbb{R}^k$ in the case of Locally Linear Embedding (LLE). This step is refered to as a locally barycentric mapping. Now provided that the local barycentric

model fits to the dataset perfectly, we show in Chapter 2 that there exists multiple solutions to the optimal mapping problem on $\mathbb{R}^k$. In practice, although this assumption is not satisfied exactly, if the model is close enough to the dataset then there still exists multiple mappings almost equivalently optimal. The following algorithm, which we refer to as Isometric Locally Barycentric Embedding (IsoLBE) picks among all the equivalent solutions the one that optimizes the criterion proposed in Isomap (Tenenbaum, De Silva, and Langford, 2000) and based on the conservation of the intrinsic distance across the dataset, approximated by the shortest path distance on the neighbor graph of the data points. IsoLBE is not to be confused with the method introduced as ISOLLE by Ghojogh, Ghodsi, Karray, and Crowley (2020) in a survey on *Locally Linear Embedding and its Variants*.

6.2.1 Isometric locally barycentric embedding

Consider a dataset $x_1 \dots, x_N$ on a manifold M. The first step of Algorithm 2 is also the first step of LLE and consist in fitting a local barycentric model to the dataset. More specifically, each data point x_i is modeled by its projection onto the barycentric subspace of its $k+1$ nearest neighbors and we retrieve the barycentric coordinates $w_{ij_0}, \dots, w_{ij_k}$ of such projection. Let us now detail the second step. The final goal of this step is to find the optimal locally barycentric mapping $y_1, \dots y_N \in bR^k$ that minimizes the cost

$$C(y) = \sum_{i,j} \left| \|y_j - y_i\|^2 - d_{ij}^2 \right|^2 \tag{6.1}$$

where d_{ij} denotes the shortest path distance between x_i and x_j in the r-ball neighbor graph implemented by Isomap. Now the key of this algorithm is that a locally barycentric mapping in $\mathbb{R}^k$ can be formulated in terms of the normalized Laplacian matrix L_w of the nearest neighbor graph. Precisely, the set of points $y_1, \dots, y_N \in \mathbb{R}^k$ is a locally barycentric mapping of the local barycentric model (w_{ij}) if and only if, for any $1 \leq l \leq k$, its l-th component belongs to the kernel of the Laplacian matrix

$$\begin{bmatrix} y_1^l & \dots & y_N^l \end{bmatrix} \in \ker L_w. \tag{6.2}$$

In practice, whether a vector belongs to the kernel of L_w or not is decided according to some threshold value ε. Now the kernel of Laplacian matrix is a vector space, say of dimension D, and the D first eigenvectors $e_1, \dots, e_D$ of L_w form a basis of such space. Therefore the set of points $y_1, \dots, y_N \in \mathbb{R}^k$ is a barycentric mapping of the local barycentric model (w_{ij}) if and only if their exists a $k \times D$ matrix (v_{lm}) such that

$$\begin{bmatrix} y_1^l & \dots & y_N^l \end{bmatrix} = \sum_{m=1}^{D} v_{lm} e_m. \tag{6.3}$$

Under this new formulation, the optimization of the isometric criterion is simply a vector-valued optimization problem and is compatible with the standard solvers of the library `scipy` such as BFGS. Again, the criterion is perfectly exchangeable with any other criterion, even borrowed from another dimensionality reduction method such as t-SNE (Hinton and S. Roweis, 2002). Note however that a distance-based criterion preserves the neighborhoods of the original dataset, which is not guaranteed at all by locally barycentric mappings.

6.2.2 Unfolding the Swiss Roll, a benchmark experiment

We compare IsoLBE to Isomap and LLE on the Swiss Roll dataset (see Figure 6.2) that serves as a benchmark to most manifold learning methods. Precisely, we compare the 2-dimensional embedding of the dataset generated by each method with the exact unrolling of the dataset in the plane. We

Algorithm 2 IsoLBE

Input: $x_1, \dots, x_N \in M$
$1 \leq k \leq n, N-1$
$\varepsilon > 0$
$r > 0$

1: **Compute** $G = \texttt{nearest\_neighbors\_graph}(x_1, \dots, x_N, k+1)$
For all $i = 1$ **to** N
Set $j_0, \dots, j_k = \texttt{nearest\_neighbors}(i, G)$
compute $w_{ij_0}, \dots, w_{ij_k} = \texttt{barycentric\_projection}(x_i, x_{j_0}, \dots, x_{j_k})$
if j **not in** $j_0, \dots j_k$ **then set** $w_{ij} = 0$

2: **Set** $L_w = (Id - w)^T (Id - w)$
Compute $e_1, \dots, e_D = \texttt{kernel}(L_w, \varepsilon)$
Compute $d_{ij} = \texttt{isomap\_distance}(x_1, \dots, x_N, r)$
Define $\begin{bmatrix} \texttt{y1}(v) & \dots & \texttt{yN}(v) \end{bmatrix} = \begin{bmatrix} \sum v_{1j} e_j & \dots & \sum v_{kj} e_j \end{bmatrix}^T \in \mathbb{R}^{k \times N}$
Define $\texttt{cost}(v) = \sum \left| \|\texttt{yi}(v) - \texttt{yj}(v)\|^2 - d_{ij}^2 \right|^2$
Compute $v = \texttt{minimize}(\texttt{cost})$
Set $y_1 \dots, y_N = \texttt{y1}(v), \dots, \texttt{yN}(v)$

Output: $y_1, \dots, y_N \in \mathbb{R}^k$

expect IsoLBE to inherit from the characteristics of both methods Isomap and LLE and correct the unwanted behaviors of each. Let us comment on some practical settings.

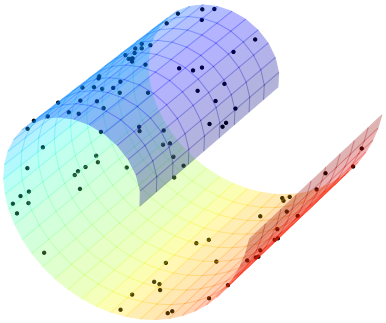

Figure 6.2: Dataset on the Swiss Roll.

Fitting parameters First of all, which radius r to fix for the construction of the graph distance in Isomap has been well studied for example by Balasubramanian and Schwartz (2002). Essentially, we should select the largest radius such that the graph is as connected as possible while all the edges are still close to the underlying submanifold on which the dataset is sampled. In the case where such a submanifold is the Swiss Roll, this optimal value corresponds to the last largest radius such that

the blue end of the Swiss Roll is not connected to its red end, which would results in a embedding that resemble a one-dimensional Swiss Roll. In Figure 6.3, we fix the same radius for both IsoLBE and Isomap. In our example, it corresponds to a value $r = 10$.

For LLE as well, the choice of good parameters, namely the number K of neighbors and the regularization factor δ, has been the subject of extensive study (Daza-Santacoloma, Acosta-Medina, and Castellanos-Domínguez, 2010). Especially, the choice an appropriate regularization is very crucial in order to retrieve accurate solutions (Lin, 2021). In our example, we take $K = 10$ and $\delta = 1 \times 10^{-3}$.

Finally the last parameter to fit is the threshold value ε for embeddings to be in the kernel of the Laplacian matrix L_w. More precisely, we fix a threshold value on the ratio between the norm of a given vector and the maximal eigenvalue of L_w. This is a trade-off between values that are too large to ensure the barycentric nature of the embedding and that do not preserve the local barycentric model at all anymore, and values that are too low to enable embeddings for which the cost of the optimal mapping problem would vanish if the dataset was exactly modeled. At one extreme, the kernel of the Laplacian matrix L_w contains enough solutions for Isomap to be one of them such that IsoLBE yields exaclty the same embedding as Isomap, and at the other, the isometric criterion is ignored at a local scale and IsoLBE does not perform well in regions where the $k+1$-nearest neighbor graph of the data points is poorly connected. In practice, we propose two ways to fit the threshold value. First, assuming that the model is exact, we may compute the lower bounds listed in Chapter 5 based on a decomposition of the $k+1$-nearest neighbor graph. These allow to bound the threshold value from above. Another approach consists in applying the elbow method to the eigenvalue profile of the Laplacian matrix L_w to separate noisy locally barycentric embeddings from the embeddings that are not locally barycentric. In our example, the performance of IsoLBE as well as the one of the two other methods is evaluated by measuring the pointwise distance between the isometric unrolling of the dataset in the plane and the embedding generated by the method up to a rotation. The optimal threshold value is then chosen to optimize the performance of IsoLBE and we fix $\varepsilon = 1 \times 10^{-5}$. Such a value generates a kernel of size $D = 70$.

6.2.3 Conclusion and future works

Let us start this last section with some perspectives for this chapter specifically. Essentially, we plan to improve our method on two aspects. First, we would like to explore other criteria for the selection of a barycentric embedding and to understand how they influence the optimal embedding. Then, we are currently investigating the generalization of the two algorithms to constant curvature embedding spaces. Finally, we would like to experiment more with the two algorithms, especially on datasets where the underlying structure is not known in advance.

Let us now move on to the final word of this thesis. First of all, I would like to emphasize that barycentric geometry is a very interesting subject, which nicely blends Riemannian geometry and straightedge-and-compass constructions, which I had not practiced for several years. In parallel, I have had several opportunities to explore quotient spaces, which I have been familiar with for some time and still enjoy working with today. In its final form, this thesis is the reflection of Xavier's intuition that put me on the right track, Alain's persistence that did not let me give up on the open questions I came across before I had spent enough time on them, the fruitful encounter with Anna, not to mention inspiring cross-disciplinary discussions, the valuable advice from wise mathematicians, and a few fortunate coincidences. At the moment, there are still a lot of questions that I have opened without closing them, carefully written down in my notebook.

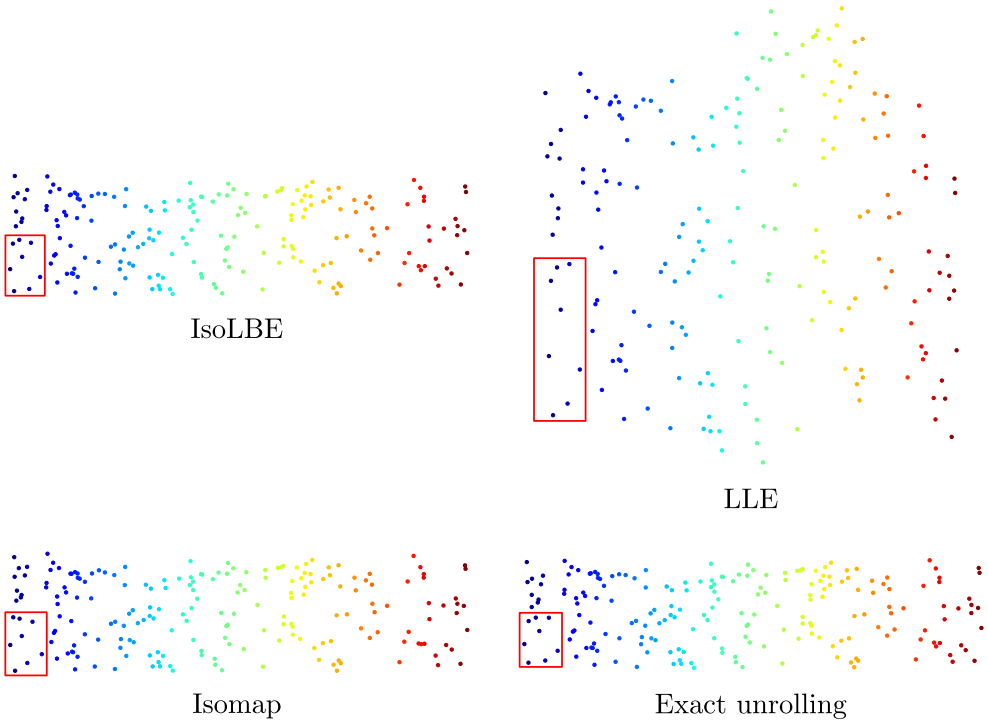

Figure 6.3: Comparison of IsoLBE, LLE, and Isomap. Numerically, that is according to the error measure described previously, the performance of IsoLBE and Isomap is comparable although IsoLBE performs slighly better (after normalizing by the largest distance computed by Isomap on the r-ball neigbor graph, we measure an error of $8,9 \times 10^{-4}$ and $9,0 \times 10^{-4}$ respectively). As expected, LLE perfoms significantly worse as the embedding is not well scaled. Now visually, we higlight on the subset contained in the red box how IsoLBE combines the information retrieved by LLE, that is the relative positions of the data points in the subset, and that retrieved by Isomap, that is the scale of the subset, to provide with the most accurate embedding of such subset among the three methods.

Let me summarize these questions. On theoretical aspects of barycentric embeddings, I want to refine the analysis locally barycentric embeddings of Chapter 5, starting by the conjectures we discussed at the end of the chapter, and understand how our results are altered when lifting some assumptions, especially when the model does not fit to the dataset exactly or when the data are noisy (Wang, Wong, and Lee, 2019). Studying the geometry of barycentric subspaces of spectral graph spaces is also on my list. On the computational aspects of barycentric embeddings, I already mentioned above two ways of generalizing IsoBSA and IsoLBE. Finally, I would like to experiment the two algorithms on datasets modeled in Kendall shapes spaces such as protein structure, as well our new statistical graph analysis framework for other applications such as the study of brain connectivity.

References

Balasubramanian, M. and E. L. Schwartz (2002). "The Isomap algorithm and topological stability". In: *Science* 295.5552, pp. 7–7. DOI: 10.1126/science.295.5552.7a.

Cox, M. A. A. and T. F. Cox (2008). "Multidimensional scaling". In: *Handbook of Data Visualization*. Ed. by C.-H. Chen, W. Härdle, and A. Unwin. Springer Handbooks Comp.Statistics. Springer, pp. 315–347. ISBN: 978-3-540-33037-0. DOI: 10.1007/978-3-540-33037-0_14.

Daza-Santacoloma, G., C. D. Acosta-Medina, and G. Castellanos-Domínguez (2010). "Regularization parameter choice in locally linear embedding". In: *Neurocomputing*. Subspace Learning / Selected papers from the European Symposium on Time Series Prediction 73.10, pp. 1595–1605. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2009.11.038.

Ghojogh, B. et al. (2020). *Locally linear embedding and its variants: Tutorial and survey*. DOI: 10.48550/arXiv.2011.10925.

Harms, P., E. Maignant, and S. Schlager (2019). *Approximation of Riemannian distances and applications to distance-based learning on manifolds*. DOI: 10.48550/arXiv.1904.11860.

Hinton, G. E. and S. Roweis (2002). "Stochastic neighbor embedding". In: *Advances in Neural Information Processing Systems* 15.

Lin, L. (2021). *Avoiding unwanted results in locally linear embedding: A new understanding of regularization*. DOI: 10.48550/arXiv.2108.12680.

Pennec, X. (2018). "Barycentric subspace analysis on manifolds". In: *The Annals of Statistics* 46.6, pp. 2711–2746. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/17-AOS1636.

Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding". In: *Science* 290.5500, pp. 2323–2326. DOI: 10.1126/science.290.5500.2323.

Tenenbaum, J. B., V. De Silva, and J. C. Langford (2000). "A Global geometric framework for nonlinear dimensionality reduction". In: *Science* 290.5500, pp. 2319–2323. DOI: 10.1126/science.290.5500.2319.

Wang, J., R. K. W. Wong, and T. C. M. Lee (2019). "Locally linear embedding with additive noise". In: *Pattern Recognition Letters* 123, pp. 47–52. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2019.02.030.

"Je déjeunais sous un oranger, et là, une orange tombe dans mon assiette. Elle était délicieuse."

– Alain Trouvé. Antibes, le 31 mai 2023