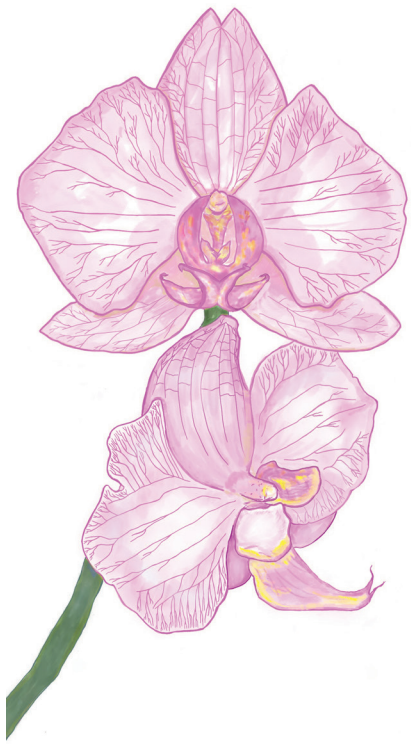
Doctoral Thesis in Computer Science

Hyperbolic Geometric Inference

Preserve, Analyze, Extract, and Interpret Structures in Latent Data Representations

ANISS AIMAN MEDBOUHI

KTH ROYAL INSTITUTE OF TECHNOLOGY



Hyperbolic Geometric Inference

Preserve, Analyze, Extract, and Interpret Structures in Latent Data Representations

ANISS AIMAN MEDBOUHI

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Thursday the 12th of November 2026, at 13:00 in F3 (Flodis), Lindstedtsvägen 26, Stockholm.

Doctoral Thesis in Computer Science
KTH Royal Institute of Technology
Stockholm, Sweden 2026

AI-Acknowledgment: In line with the requirement from the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology and the doctoral programmes, the author acknowledges the use of Artificial Intelligence (AI) tools, in particular ChatGPT serie 5 from Open AI, to get feedback and assist with language editing and text structure. The author has critically reviewed and revised all generated material and bears full responsibility for the content of the thesis.

© Aniss Aiman Medbouhi

Cover page photo: Sara E. Medbouhi

TRITA-EECS-AVL-2026:89
ISBN 978-91-8106-733-0

Printed by: Universitetsservice US-AB, Sweden 2026

I dedicate this work to the students who could not finish their studies
the teachers, lecturers and professors who could not continue to teach
the researchers who could not pursue their research
the administrators and staff who could not support education anymore
to all these innocents who have been killed during the time of my PhD
in what is sometimes referred to as a ***scholasticide****.*

Abstract

High-dimensional data are often represented in lower-dimensional latent spaces for analysis. The geometry chosen for such a representation determines how distances, neighborhoods, and other relations between observations are expressed. Hyperbolic geometry is well suited to hierarchical and tree-like data because of its negative curvature and exponential growth. Once a hyperbolic representation has been constructed, a further question arises: what information can be inferred? This thesis studies this question from the perspective of **hyperbolic geometric inference**: the analysis and extraction of structural information encoded in hyperbolic latent representations, together with their interpretation. We first consider the upstream problem of preserving selected structure during representation learning, using geometric and topological constraints in variational auto-encoders. We then develop methods for geometric inference from hyperbolic latent representations, with applications in single-cell biology and olfactory perception. Hyperbolic Delaunay Geometric Alignment compares predefined subsets through their Delaunay connectivity. A complementary study of olfactory descriptor data investigates radial, directional, and regional organization. Finally, HyperSteiner and Randomized HyperSteiner (RHS) construct approximate solutions to the hyperbolic Steiner minimal tree problem using Delaunay triangulation. RHS broadens the exploration of candidate tree topologies and refines Steiner-point locations through Riemannian optimization. The main contribution of this thesis is a set of methods for geometric inference from hyperbolic latent representations: quantifying geometric relations, analyzing geometric organization, and extracting parsimonious tree structures.

Keywords
Hyperbolic data analysis, computational geometry, optimization

Sammanfattning

Högdimensionella data representeras ofta i latenta rum med lägre dimension för analys. Den geometri som väljs för en sådan representation avgör hur avstånd, grannskap och andra relationer mellan observationer uttrycks. Hyperbolisk geometri är väl lämpad för hierarkiska och trädliknande data på grund av dess negativa krökning och exponentiella tillväxt. När en hyperbolisk representation har konstruerats uppstår frågan: vilken information kan härledas? Denna avhandling studerar denna fråga utifrån perspektivet **hyperbolisk geometrisk inferens**: analys och extraktion av strukturell information som är kodad i hyperboliska latenta representationer, tillsammans med tolkningen av denna information. Vi behandlar först det föregående problemet att bevara utvald struktur under representationsinlärning genom geometriska och topologiska begränsningar i variational auto-encoders. Därefter utvecklar vi metoder för geometrisk inferens från hyperboliska latenta representationer, med tillämpningar inom encellsbiologi och olfaktorisk perception. Hyperbolic Delaunay Geometric Alignment jämför fördefinierade delmängder genom deras Delaunay-konnektivitet. En kompletterande studie av olfaktoriska deskriptordata undersöker radiell, riktningsberoende och regional organisation. Slutligen konstruerar HyperSteiner och Randomized HyperSteiner (RHS) approximativa lösningar till det hyperboliska Steiner-minimalträdsproblemet med hjälp av Delaunay-triangulering. RHS breddar utforskningen av möjliga trädtopologier och förfinar Steinerpunkternas positioner genom Riemannsk optimering. Avhandlingens bidrag är metoder för geometrisk inferens från hyperboliska latenta representationer: kvantifiering av geometriska relationer, analys av organisation och extraktion av parsimoniska trädstrukturer.

Nyckelord
Hyperbolisk dataanalys, beräkningsgeometri, optimering

Preface

In The Name of God, The Most Gracious, The Dispenser of Grace

Read in the name of thy Sustainer, who has created
created man out of a germ-cell!
Read - for thy Sustainer is the Most Bountiful One
who has taught [man] the use of the pen
taught man what he did not know!

– From ”The Message of The Quran”, 96:1–5, translated by Muhammad Asad.

Reading, understanding the universe, pursuing knowledge and science, are, I believe, a divine injunction. This principle is powerfully captured in these opening verses, which, according to Islamic tradition, were the first revealed by God to the Prophet Muhammad. These words resonate deeply with me, drawing my thoughts toward the fascinating processes of biological development. In fact, although my PhD is in computer science, the captivating nature of living systems has been an important source of inspiration and applications. Furthermore, the symbol of the ”pen” in these verses serves as a humbling reminder: I do believe that whatever modest insights we may have discovered on this journey were acquired thanks to God’s grace, while all errors and shortcomings in my thesis work remain entirely my own. I sincerely hope that this thesis benefits society and proves useful to future researchers, engineers, and anyone who might build upon this work or put it to *good* use.

Acknowledgment

First and foremost, I thank God for giving me the strength to complete this PhD – working continuously hard for several years, spending countless nights and weekends, was not an easy task. The true difficulty lay not merely in the heavy workload itself, but in the profound sacrifices it demanded.

I would like to thank some people who were key in my PhD journey, starting with my supervisors and co-authors. Thank you, Dani, for giving me the freedom to work on topics I found interesting and for ensuring that we had the necessary funding. I would like to express my deepest gratitude toward Giovanni, who was my co-supervisor *de facto* and without whom this thesis would not have been possible. Thank you, Giovanni, for sharing your expertise, for your guidance and feedback, and for caring so much. The scientific discussions and brainstorming we had in the beginning of each new project are the most memorable moments of the scientific journey. Thank you very much for challenging my ideas all the time! Thank you for being a great mentor, colleague and friend. Thank you, Anastasia, Vlad, Alex and Petra, for getting me on the right track at the very beginning of this PhD. I would like to thank my co-authors who brought a new dimension to my work. Thank you, Alejandro, for your attention to detail and for sharing the amazing hyperbolic Steiner tree journey! And thank you, Farzaneh, for the olfactory flavor!

I would like to thank my advance reviewer, Mårten Björkman, for his valuable feedback on this thesis. I am also very grateful to my opponent, Georgios Arvanitidis, and the grading committee members — Mariangeles Serrano, Fredrik Ohlsson, and Pawel Herman — for accepting the invitation to assess my work for the doctoral defence.

Next, my thanks go to my other office-mates in L24, Nona and Mohammed, for being such great colleagues. Mohammed, I will not forget our nightly philosophical discussions! Thank you, Yonk, Sanaz, Marcel and Marco, for also being great office-mates but in the old RPL in T14. Marco, your startup journey was truly inspiring! Thank you, Gustaf and Alfredo, for the amazing discussions! Alfredo, I have always been impressed by your erudition, not only in machine learning, and your hard work and dedication were contagious! Thank you, Alberta and Miguel, for organizing nice group activities to make the PhD life less painful. Thank you to the other members of the group for also sharing some nice times, lunches and discussions: Ciwan, Zehang, Michael, Jens, Zheyu, Yuchong, Hafoei, Katharina, Yufei, Lakshadeep, Nick, Diluka... Thank you, Grace, for helping me out with all the administrative stuff. Thank you to all the nice members of other groups that I met at RPL, among them Parag, Youssef, Robert, Rafael, Amir, Matteo, Federico, Loizos, Zahid, Leo, Stefan, Sean, Jesús, Timon, Ajinkya, Nicolas... While I cannot name everyone individually, I would like to thank the entire Robotics Perception and Learning Division (now Department) and its amazing people, as well as express my gratitude for the generous resources and equipment made available to us. Thank you to the members of our Geometry and Machine Learning reading group for the very interesting discussions! And I wish all the best to the new batch of PhD students in our group: Ziqi, Thoranna, Arda, Jiacheng, Jiajie and Junseo. Good luck and stay strong!

I am grateful to the Swedish and European taxpayers for their financial contributions, provided through the Swedish and European Research Councils. I would also like to acknowledge the Knut and Alice Wallenberg Foundation for supporting this work.

I would like to express my deepest gratitude to all my teachers and professors, as well as all my students. During my PhD, I had the chance to teach not only at KTH but also at DIS Study Abroad, which offered me an amazing and very enriching experience. There, I learned a lot about teaching from Natalia and Christopher but also from Henry and my amazing students! Thank you all, it was a new and complementary approach compared to what I could do at KTH! Thank you to all the DIS faculty and staff for the warm welcome. Speaking about teaching, I would also like to take this opportunity to thank the professors and instructors involved in the most engaging and interesting courses of my PhD at KTH: FDD3401 Neuroscience, FDT3306 Critical and Interdisciplinary Perspectives in AI Music Studies, and M2585 Artificial Intelligence in Society.

Thank you, Sumeyra and Emma, Hannes and the whole KTH Innovation team, for being part of the amazing MoMente adventure! Thanks to you all, I have learned so much about entrepreneurship!

I was in a robotics lab, but since I am a human and not a robot, I also have a life outside the lab, and unfortunately this is sometimes forgotten.

I would like to thank my friends and family for their incredible support and understanding and I hope that you forgive me for not having spent enough time with you during these years.

There are no words to describe my gratitude toward my parents who educated me, je vous remercie du fond du cœur et c'est grâce à vous que j'ai pu en arriver là. Et mes chers frères et sœur, je vous remercie profondément, car en réalité j'apprends toujours de vous, sans arrêt, et je vous souhaite tout le meilleur, force à vous !

Enfin, la meilleure décision que j'ai prise pendant mon doctorat a été de me marier, et je remercie Dieu tous les jours pour m'avoir fait rencontrer ma tendre femme. Sara, il n'y a pas de mots pour exprimer mon amour, et je te remercie infiniment pour le tien, ton soutien et ta patience.

I wish you all a pleasant reading!

Best regards,

Aniss Aiman Medbouhi
Stockholm, October 2026

Academic Ancestors

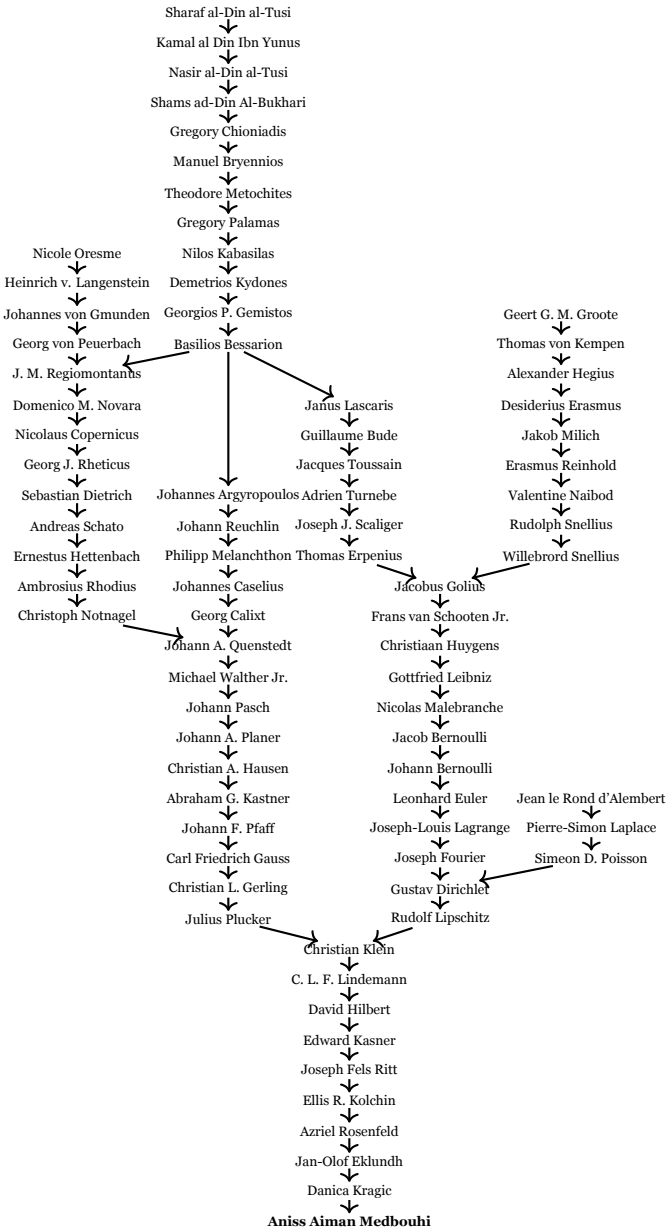


Figure .0.1: Simplified academic genealogy tree, extracted mainly from: https://j2kun.github.io/math-genealogy. Note that the oldest known academic ancestor, whose work is related to this thesis, is Nasir al-Din al-Tusi, and he is mentioned in Chapter 2.

List of Included Papers and Contributions by the Author

Contributions This doctoral dissertation consists of two parts. The first part provides an overview of the research field on which I focused and a summary of my contributions to it. The second part comprises the contributions that I have made to this research field through included publications and manuscript.

Paper A **InvMap and Witness Simplicial Variational Auto-Encoders**

Aniss Aiman Medbouhi, Vladislav Polianskii, Anastasia Varava, and Danica Kragic

In Machine Learning and Knowledge Extraction, 2023

Contributions by the author: Formulated the problem, conducted the literature review, designed and implemented the methods and experiments, analyzed the results, wrote the entire paper.

Paper B **Hyperbolic Delaunay Geometric Alignment**

Aniss Aiman Medbouhi, Giovanni Luca Marchetti, Vladislav Polianskii, Alexander Kravberg, Petra Poklukar, Anastasia Varava, and Danica Kragic

In Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), 2024

Contributions by the author: Formulated the problem, conducted the literature review, designed and implemented the methods and experiments, analyzed the results, wrote the entire paper, and presented the work orally and as a poster at ECML PKDD.

Paper C **Geometric organization of olfactory descriptor data in the Poincaré disk**

Aniss Aiman Medbouhi, Farzaneh Taleb, Giovanni Luca Marchetti, and Danica Kragic

arXiv preprint, 2026

Contributions by the author: Proposed and formulated the problem, co-conducted the literature review, designed and implemented most of the methods and experiments, analyzed the results, wrote the entire paper.

Paper D **HyperSteiner: Computing Heuristic Hyperbolic Steiner Minimal Trees**

Alejandro García-Castellanos[*], **Aniss Aiman Medbouhi**[*], Giovanni Luca Marchetti, Erik J. Bekkers, and Danica Kragic

In Proceedings of the 27th Society for Industrial and Applied Mathematics (SIAM) Symposium on Algorithm Engineering and Experiments (ALENEX), 2025

Contributions by the author: Proposed the problem, co-conducted the literature review, designed and implemented part of the method, designed and implemented the experiments on real-life data and analyzed their results, wrote most of the main paper, and presented the work orally at ALENEX.

Paper E **Randomized HyperSteiner: A Stochastic Delaunay Triangulation Heuristic for the Hyperbolic Steiner Minimal Tree**

Aniss Aiman Medbouhi[*], Alejandro García-Castellanos[*], Giovanni Luca Marchetti, Daniël M. Pelt, Erik J. Bekkers, and Danica Kragic

In The 29th International Conference on Artificial Intelligence and Statistics (AISTATS), 2026

Contributions by the author: Proposed the problem, co-conducted the literature review, designed and implemented parts of the method, co-designed the experiments, implemented the experiments described in 5.1 and "Approaching the Theoretical Limit" in 5.2 and analyzed their results, wrote most of the main paper and parts of the appendix, and presented the work as a poster at AISTATS.

Other contributions by the author not included in the thesis.

Paper F **Machine-Learning-Driven Reconstruction of Organic Aerosol Sources across Dense Monitoring Networks in Europe**

A. Jouanny, A. Upadhyay, J. Jiang, P. Vasilakos, M. Via, Y. Cheng, B. Flueckiger, G. Uzu, J. Jaffrezo, C. Voiron, O. Favez, H. Chebaicheb, A. Bourin, A. Font, V. Riffault, E. Freney, N. Marchand, B. Chazeau, S. Conil, J. Petit, J.C. De La Rosa, A. Sanchez De La Campa Verdona, D. Sanchez-Rodas Navarro, S. Castillo, A. Alastuey, X. Querol, C. Reche, M.C. Minguillón, M. Maasikmets, H. Keernik, F. Giardi, C. Colombi, E. Cuccia, S. Gilardoni, M. Rinaldi, M. Paglione, V. Poluzzi, D. Massabò, C. Belis, S. Grange, C. Hueglin, F. Canonaco, A. Tobler, H. Timonen, M.

[*]Equal contribution.

Aurela, M. Ehn, I. Stavroulas, A. Bougiatioti, K. Eleftheriadis, M.I. Gini, O. Zografou, M. Manousakas, G.I. Chen, C. Green, P. Pokorna, P. Vodička, R. Lhotka, J. Schwarz, A. Schemmel, S. Atabakhsh, H. Herrmann, L. Poulain, H. Flentje, L. Heikkinen, V. Kumar, H.A.D. Van Der Gon, W. Aas, S.M. Platt, K.E. Yttri, I. Salma, A. Vasanits, B. Bergmans, Y. Sosedova, J. Necki, J. Ovadnevaite, C. Lin, J. Pauraite, M. Pikridas, J. Sciare, J. Vasilescu, L. Belegante, C. Alves, J. Slowik, N. Probst-Hensch, D. Vienneau, A.S.H. Prevot**, A.A. Medbouhi**, D.T. Banos, K. De Hoogh, K. Daellenbach, E. Krymova, and I. El Haddad

In Environmental Science and Technology (ES&T) Letters, American Chemical Society (ACS) publications, 2025

Contributions by the author: Co-supervised Adrien Jouanny's Master's thesis, from which this paper originated.

Paper G **Resisting Algorithmic Desire and Manipulation in AI Recommendation Systems: An Islamic Moral Psychology of Digital Self-Mastery**

Aniss Aiman Medbouhi

Manuscript, 2026

Contributions by the author: Proposed the problem, conducted the literature review, developed the conceptual framework and normative arguments, performed the comparative ethical analysis, formulated the digital self-mastery framework and its practical and policy implications, and wrote the entire manuscript. This work originated from an essay that I independently authored as part of the KTH course Artificial Intelligence in Society (DM2585).

Poster 1 **Adaptation of ERP-based Classifiers to Novel Subjects and Tasks**

Farzaneh Taleb**, Aniss Aiman Medbouhi**, Nona Rajabi, Ali Ghadirzadeh, and Danica Kragic

Poster presented at the annual Strategic Research Area Neuroscience (StratNeuro) retreat in 2022

Contributions by the author: Designed and developed the neuroscientific experimental protocol consisting of ten electroencephalography-based rapid serial visual presentation (RSVP) tasks for brain-computer interfaces involving different cognitive functions, created the corresponding RSVP datasets, wrote part of the poster which was presented by my colleague Farzaneh Taleb.

Poster 2 **Towards Discovering the Hierarchy of the Olfactory Perceptual Space via Hyperbolic Embeddings**

Farzaneh Taleb[*]**, Aniss Aiman Medbouhi**[*], Giovanni Luca Marchetti, and Danica Kragic

Poster presented at the Computational and Systems Neuroscience (COSYNE) conference in 2025, published in Science Communications Worldwide

Contributions by the author: Proposed and formulated the problem, co-conducted the literature review, designed and implemented most

of the methods and experiments, analyzed the results, made the poster and recorded its audio presentation.

Poster 3 **Modeling the Hierarchy of the Human Olfactory Perceptual Space via Hyperbolic Embeddings**

Aniss Aiman Medbouhi[*], Farzaneh Taleb[*], Giovanni Luca Marchetti, and Danica Kragic

Poster presented at the Cognitive Computational Neuroscience (CCN) conference in 2025

Contributions by the author: Proposed and formulated the problem, co-conducted the literature review, designed and implemented most of the methods and experiments, analyzed the results, made the poster which was presented by my colleague Farzaneh Taleb, wrote most of the corresponding abstract paper. This is a later presentation of substantially the same work reported in **Poster 2**.

Contents

Abstract vii

Sammanfattning ix

Preface xi

Acknowledgment xiii

Academic Ancestors xvii

List of Included Papers and Contributions by the Author xix

Contents xxiii

List of Figures xxvii

List of Tables xxix

I Overview 1

1 Introduction 3

1.1 Motivation: From High-Dimensional Data to Structured Representations 3

1.2 Thesis Scope and Research Questions 8

1.3 Contributions and Relations Between the Papers 9

1.4 Thesis Outline 13

2 Historical and Philosophical Context: From Euclid's Fifth Postulate to Hyperbolic Data Science 15

2.1 Introduction 16

2.2 Pythagoras' Influence 16

2.3 Plato's Influence 17

2.4 Euclid's Elements 19

2.5 Euclid's Fifth Postulate and its Questioning 20

2.6 Philosophical Doubt and the Birth of Hyperbolic Geometry . . . 23

2.7 The Fifth Postulate and Gödel's Incompleteness Theorem: A Philosophical Parallel 25

2.8 Hyperbolic Geometry in Nature, Science, and Data 26

2.9 Conclusion 28

3 Structured Latent Representations: From Topology to Hyperbolic Geometry 31
3.1 Latent-Space Structure as an Inductive Bias 32
3.2 Variational Auto-Encoders and Latent-Space Structure 33
3.3 Preserving Selected Structure: **Paper A** 34
3.3.1 Preserving a Reference Representation: InvMap-VAE . . 34
3.3.2 Topological Inductive Bias: Witness Simplicial VAE 35
3.3.3 Combining Topological and Metric Constraints: Isolandmarks Witness Simplicial VAE 39
3.3.4 Experimental Observations 40
3.4 From Structural Constraints to Geometry as an Inductive Bias . 41
3.5 Hyperbolic Geometry 42
3.5.1 Negative Curvature and Hierarchical Structure 42
3.5.2 Models of Hyperbolic Space 45
3.5.3 Riemannian Operations and Optimization 49
3.6 Hyperbolic Representation Learning 52
3.7 From Hyperbolic Representation Learning to Geometric Inference 54

4 Analyzing and Interpreting Geometric Structures in Hyperbolic Representations 55
4.1 Geometric Alignment Between Hyperbolic Subsets: **Paper B** . . 56
4.1.1 Hyperbolic Delaunay Graphs 57
4.1.2 Hyperbolic Delaunay Geometric Alignment 57
4.1.3 Analysis and Interpretation with HyperDGA 58
4.2 Geometric Organization of a Hyperbolic Representation: **Paper C** 61
4.2.1 Olfactory Descriptor Data 61
4.2.2 Hyperbolic Metric Multidimensional Scaling 62
4.2.3 Analysis and Interpretation of Radial, Directional and Regional Organization 62
4.2.3.1 Radial Organization 62
4.2.3.2 Directional and Regional Organization 64
4.3 Discussion 65

5 Extracting and Interpreting Hierarchical Structures with Hyperbolic Steiner Trees 69
5.1 From Implicit Hierarchy to an Explicit Parsimonious Tree 70
5.2 HyperSteiner: A Deterministic Heuristic for Delaunay-Guided Tree Extraction 71
5.2.1 Hyperbolic Fermat Points 73
5.2.2 From Local Solutions to a Heuristic Steiner Tree 74
5.2.3 Extracting and Interpreting a Tree from Planaria Data . . . 75

5.3 Randomized HyperSteiner: Stochastic Exploration of Tree Topologies . . . 75
5.3.1 Stochastic Candidate Generation and Local Steiner Corrections . . . 77
5.3.2 Riemannian Geometric Refinement Given a Fixed Topology 78
5.3.3 Results, Limitations, and Implications . . . 78
5.4 Discussion . . . 80

6 Conclusions and Perspectives . . . 83
6.1 Answering the Research Questions . . . 83
6.2 From Structured Representations to Geometric Inference . . . 85
6.3 Limitations and Future Directions . . . 86
6.4 Potential Applications . . . 87
6.5 Societal and Sustainability Considerations . . . 88
6.6 A Final Perspective on Geometry and Inference . . . 89

7 Summary of the Included Publications . . . 91

Bibliography . . . 99

II Appended papers 113

List of Figures

.0.1 Simplified academic genealogy tree xviii

1.1.1 Hyperbolic representation of the *Planaria* dataset 4
1.1.2 Dunham adaptation of Escher Circle Limit I 7

3.3.1 Construction of a witness complex 37

4.1.1 Illustration of HyperDGA 58
4.2.2 Radial organization of olfactory descriptor profiles 63
4.2.3 Directional organization of olfactory descriptors in the Sagar dataset 65
4.2.4 Regional organization of the fruity descriptor family in GSLF . . 66

5.2.1 Three-and four-terminal full Steiner trees 72
5.2.2 Four-terminal full Steiner tree topologies 73
5.2.3 Construction of Fermat points through isoptic curves 74
5.2.4 HyperSteiner applied to a sample of the Planaria representation 76
5.3.5 Deterministic and randomized hyperbolic Steiner tree construction 77

List of Tables

1.3.1 Relations between the main operations of the thesis and the included papers. 10

Overview

1 Introduction

This chapter formulates the scientific problem addressed in the thesis and summarizes its scope, research questions, and contributions.

1.1 Motivation: From High-Dimensional Data to Structured Representations

A regenerative animal as a computational problem. Planarian flatworms are remarkable for their capacity to regenerate missing tissues and organs. Classical experiments showed that even a body fragment representing only about $1/279$ of the original animal could regenerate into a complete organism [11]. In the widely studied species *Schmidtea mediterranea*, this regenerative capacity extends to complex structures such as the central nervous system: missing head structures, including a functional brain, can be restored within about a week [12]. This ability relies on adult stem cells, called *neoblasts*, which give rise to the specialized cell types required for tissue maintenance and regeneration [13].

The regenerative phenomenon is visible at the scale of the whole animal, but the underlying process unfolds at the scale of individual cells. To study how cells progress from stem-cell states towards specialized cell types, one can examine their gene-expression profiles. DNA is the molecule in which genetic information is stored, and genes are segments of DNA containing information used by the cell. When a gene is expressed, its information is transcribed into RNA. Measuring the abundance of RNA associated with different genes therefore provides information about which genes are active and to what extent.

Single-cell RNA sequencing measures such gene expression separately in individual cells [14]. From a computational perspective, the resulting data have

a simple interpretation: each cell is an observation, each measured gene is a feature, and the corresponding value quantifies the expression of that gene in the cell. Since expression can be measured for thousands of genes, each cell is represented by a point in a space with thousands of dimensions.

The biological question, however, is not simply what value each of these thousands of coordinates takes. Of greater interest is how cells are organized and how their states change during differentiation. Stem cells, which are generic, can give rise to different specialized cell types like skin, neuron or muscle cells through developmental processes that contain both gradual transitions and branching cell-fate decisions. Using single-cell transcriptomic data from *S. mediterranea*, Plass et al. characterized the major cell types of the animal and reconstructed developmental progressions connecting stem cells to differentiated states [13]. An illustration of this differentiation process is provided in Figure 1.1.1.

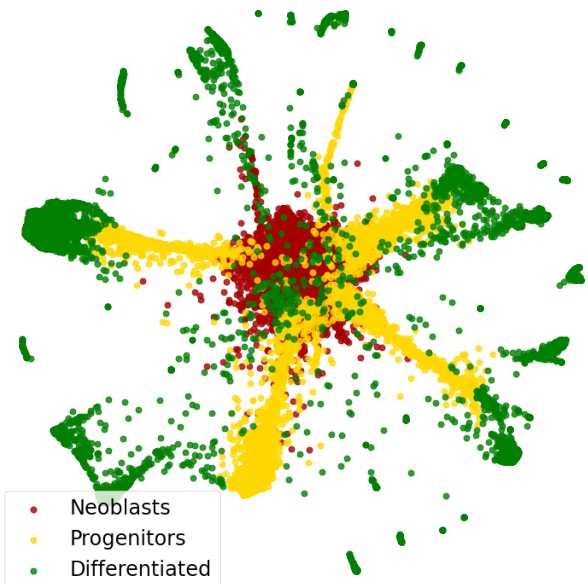


Figure 1.1.1: Hyperbolic representation from **Paper B** of the *Planaria* dataset [13] in the Poincaré disk showcasing three different types of cells within their cellular differentiation process: generic cells called neoblasts, intermediate cells called progenitors, and specialized cells called differentiated cells.

Importantly, the resulting dataset does not directly contain these developmental trajectories. Conventional single-cell RNA sequencing measures different cells rather than continuously observing the same cell throughout differentiation. The cells instead provide snapshots of different molecular states present in a population. Relationships between these states can then be used

computationally to infer transitions and branching developmental organization [15]. The measurements are observed; the developmental organization is inferred.

From high-dimensional measurements to structured representations. This example illustrates a general problem. Modern datasets can contain many measured or learned variables, while the structure relevant to a scientific or computational question may be considerably simpler. A lower-dimensional representation, as illustrated in Figure 1.1.1, can provide a more compact description of the observations, facilitate subsequent computations, and make their organization easier to analyze. When reduced to two dimensions, the representation can also make otherwise inaccessible relationships directly visible. In latent-variable generative models, a representation can serve an additional role: variational auto-encoders, for example, learn latent variables together with a generative model from which observations can be produced [16, 17].

The goal is therefore not merely to replace many coordinates by fewer ones. The usefulness of a representation depends on what information remains accessible after this transformation. In the planarian example, compressing thousands of gene-expression measurements into a few coordinates would be of limited value for studying differentiation if the relationships between developmental states were lost. Conversely, an appropriate representation may expose organization that is difficult to identify directly in the original measurement space.

A common motivation for seeking low-dimensional representations is the *manifold hypothesis*: high-dimensional observations are often modeled as concentrating near a structure of substantially lower intrinsic dimension [18]. This hypothesis should be understood as a modeling assumption rather than as a universal property of data. A smooth manifold is locally Euclidean and does not, by itself, naturally describe structures with branching points. Developmental trajectories, for example, may split into several cell fates and may therefore be better described by a more general structured object.

Constructing a useful representation therefore raises the question of *what should be preserved or exposed*. Depending on the problem, relevant information may concern metric relations, such as distances, topological properties, such as connectivity and holes, hierarchical relations, or combinations of these. Preservation is important because what can subsequently be inferred depends on what information is encoded in the representation. It is not, however, the principal focus of the thesis. The main question considered here arises *after* a structured representation has been obtained: what can be learned from its geometry?

From representation to geometric inference. This question is related to the classical problem of geometric and topological inference. In that setting, one observes a finite and possibly noisy collection of points and seeks to infer geometric or topological properties of an underlying object from which the points are sampled [19, 20]. The point cloud is therefore not treated merely as a collection of coordinates, but as evidence about an underlying structure that is not directly observed.

The setting of this thesis applies a related perspective to latent representations that have themselves been learned or constructed from data. Once such a representation is available, its geometry becomes an object of study. Relations between points or populations, the organization of properties across the representation, and possible hierarchical structures may all be encoded implicitly in the geometric configuration.

In this thesis, *geometric inference* refers broadly to the analysis and extraction of structural information encoded in geometric representations of data, together with the interpretation of the resulting quantities and structures. This perspective distinguishes several related operations. *Preservation* concerns which properties survive, or are deliberately encouraged, when a representation is constructed. *Analysis* characterizes geometric organization within a representation, whereas *extraction* constructs an explicit structure from the representation, such as a tree. Finally, *interpretation* relates the analyzed or extracted structures to meaningful properties of the data, model behavior, or reference information, or connects them to a scientific question or to theoretical properties of the chosen geometry.

The distinction is again visible in the planarian example. A low-dimensional representation may arrange cells so that developmental relations are reflected in their positions. Such an arrangement does not by itself provide an explicit hierarchy, nor does it determine which geometric quantities correspond to biologically meaningful properties. The representation makes structure accessible; further inference is required to analyze, extract, and interpret it.

Why hyperbolic geometry? The choice of geometry matters whenever the structure of the data is not well matched by Euclidean space. For example, observations that naturally lie on or near a sphere are more appropriately described using spherical rather than Euclidean geometry, while other datasets may exhibit different forms of non-Euclidean organization [21]. The geometry of a latent representation is therefore itself a modeling choice: it determines how distances, directions, neighborhoods, and other relations between represented observations are expressed. The focus of this thesis is *hyperbolic geometry*, a non-Euclidean geometry of constant negative curvature illustrated in Figure 1.1.2. In contrast, Euclidean space has zero curvature, while spherical geometry has positive curvature. Negative curvature gives hyperbolic space geometric properties that differ substantially from those of Euclidean space,

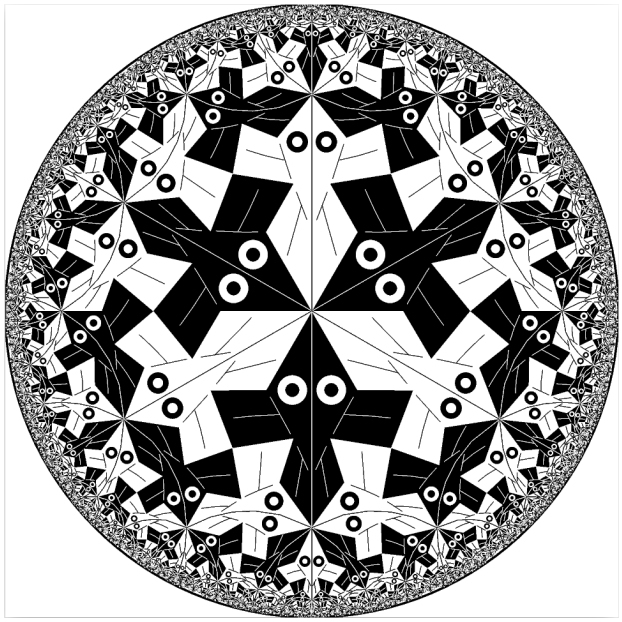


Figure 1.1.2: Dunham's adaptation of Escher *Circle Limit I*.
A tessellation is the covering of a surface with a repeating pattern without gap nor overlap. The artist Escher got particularly interested in the mathematics of tessellation after visiting Andalusia and being impressed by the geometrical symmetries present in Islamic architecture [22]. His artistic journey eventually led him to hyperbolic tessellation after meeting the geometer Coxeter and being "shocked" by a hyperbolic tessellation figure in his paper *Crystal symmetry and its generalizations*. Indeed, Escher finally found a way to "represent infinity in a finite space" and this inspired him for new artworks. He created *Circle Limit I* in 1958, the first known hyperbolic artwork. In this computerized adaptation from Dunham [23], which respects better the hyperbolic metric (on the Poincaré disk model), the backbone lines of the fish are the shortest paths on the hyperbolic disk, and all fish of the same color occupy the same hyperbolic area, i.e., they have the same (hyperbolic) size. Although an embedding of the hyperbolic space in the Euclidean one necessarily distorts distances, this illustration gives an intuition of hyperbolic geometry and its exponential expansion, as exponentially more fish can be represented toward the boundary of the hyperbolic space.

including the exponential growth of available space with distance from a point [24]. This property becomes particularly relevant when the structure of interest is hierarchical or tree-like.

A tree expands through branching: as one moves away from its root, the number of nodes can increase exponentially with depth. Hyperbolic space exhibits a compatible form of expansion, which has motivated its interpretation as a continuous geometric analogue of a tree and its use for representing hierarchical data [25, 26].

The connection between trees and hyperbolic geometry can also be made more precise. Sarkar showed that a weighted tree can be embedded in the

two-dimensional hyperbolic plane while realizing its edge lengths as hyperbolic distances, with the distortion between non-adjacent vertices made arbitrarily small [27]. Hyperbolic geometry, therefore, provides a continuous and low-dimensional space capable of accommodating branching structures. This property has motivated hyperbolic representation learning for hierarchical data and, in particular, hyperbolic representations of branching organization in single-cell data [28].

The single-cell example illustrates why this combination is appealing. Differentiation can involve gradual variation between cellular states while simultaneously branching towards distinct cell fates. A hyperbolic representation can provide a continuous geometric space in which such hierarchical organization is expressed. Yet constructing the representation is only the starting point for the questions addressed in this thesis. If different populations occupy the same hyperbolic representation, how can their geometric relation be quantified? If properties of the observations are reflected in the representation, how are they organized geometrically? If the points encode an implicit hierarchy, can this organization be extracted as an explicit parsimonious tree? And once such structures and quantities have been obtained, what can be inferred from them? These questions lead to the scope and research questions of the thesis.

1.2 Thesis Scope and Research Questions

The broader motivation of this thesis concerns structured latent representations: how relevant structure can be preserved or exposed when a representation is constructed, and what can subsequently be inferred from it. The principal methodological focus, however, is the second part of this problem. The thesis studies how structural information encoded in hyperbolic latent representations can be analyzed, extracted, and interpreted through their geometry.

Preservation plays an upstream role in this framework. If relevant information is not retained or exposed by the representation, it cannot in general be reliably recovered through subsequent analysis. This motivates the following question:

> **Motivating question — Preservation.** *How can a latent representation be constructed so that selected structure is preserved or exposed?*

The main research question instead concerns what can be inferred once a hyperbolic latent representation is available:

> **Main research question — Geometric inference.** *How can structural information encoded in hyperbolic latent representations be inferred and interpreted through their geometry?*

This question is addressed through two complementary forms of geometric inference. The first concerns the analysis and interpretation of geometric organization within an existing representation:

> **Research Question 1 (RQ1) — Analysis and interpretation.** *How can geometric organization in hyperbolic latent representations be analyzed and related to meaningful properties of the represented data or of the model that produced the representation?*

The second concerns the extraction and interpretation of an explicit hierarchical structure:

> **Research Question 2 (RQ2) — Extraction and interpretation.** *How can implicit hierarchical organization in hyperbolic latent representations be extracted as explicit parsimonious tree structures, and what can be interpreted from those structures?*

These questions distinguish between a representation and the quantities or structures inferred from it. A hyperbolic representation does not imply that the original data are intrinsically hyperbolic, and an extracted tree is not automatically identified with a ground-truth hierarchy. The conclusions depend on the information encoded in the representation and on the geometric construction used for inference. Their interpretation additionally depends on how the resulting quantities or structures relate to the problem being studied. The aim is therefore not to establish that hyperbolic geometry is universally appropriate for hierarchical data, but to investigate what forms of analysis, extraction, and interpretation become possible when data are represented in a hyperbolic latent space.

1.3 Contributions and Relations Between the Papers

The five papers address different stages of the progression from constructing a structured latent representation to inferring and interpreting information from its geometry. **Papers B–E** form the main body of the thesis on hyperbolic geometric inference. They study how information encoded in hyperbolic latent representations can be analyzed, extracted, and interpreted. **Paper A** plays a supporting upstream role by addressing the problem of preserving or encouraging selected structure when a latent representation is constructed.

There is also a distinction in geometry across the papers. **Paper A** is developed in a Euclidean latent-space setting, whereas **Papers B–E** operate on hyperbolic latent representations. The role of **Paper A** is therefore primarily conceptual with respect to the hyperbolic inference line: it establishes that the structural information available for subsequent analysis depends on how the representation is constructed. The remaining papers then investigate what can

be inferred once the latent representation is hyperbolic.

The contributions can be organized according to the four operations introduced in Section 1.1: preservation, analysis, extraction, and interpretation. These operations are not mutually exclusive, and several papers contribute to more than one of them. Table 1.3.1 summarizes these relations.

Table 1.3.1: Relations between the main operations of the thesis and the included papers.

Operation	Meaning in the thesis	Related papers
Preserve	Retain selected structure when constructing a latent representation.	Topological structure and landmark geodesic distances in **A**; pairwise distances in **C**.
Analyze	Characterize geometric or topological organization in a representation.	Topological invariants in **A**; geometric alignment between predefined subsets in **B**; radial, directional, and regional organization in **C**.
Extract	Construct explicit structures from a latent representation.	Parsimonious trees in **D** and **E**.
Interpret	Relate analyzed or extracted structures to meaningful quantities, scientific questions, model behavior, or theoretical implications.	Model evaluation and biological relations in **B**; olfactory perception in **C**; single-cell developmental progression in **D**; geometric implications of Steiner-tree parsimony in **E**.

Preserving structure. The thesis first considers what structure is retained when a latent representation is constructed. **Paper A**, *InvMap and Witness Simplicial Variational Auto-Encoders*, addresses this question explicitly from both topological and metric perspectives. InvMap-VAE starts from a prescribed low-dimensional embedding and learns a generative mapping between this representation and the original data space. Witness Simplicial VAE instead uses a witness-complex-based regularization to encourage selected topological properties of the data to be retained in the latent space. Its Isolandmarks extension additionally preserves approximate geodesic distances between landmark points, thereby combining topological and metric constraints on the latent representation. Metric preservation also provides a connection to **Paper C**.

There, a two-dimensional hyperbolic representation is constructed by minimizing the discrepancy between pairwise distances in the original data and the corresponding hyperbolic distances in the latent space. In both papers, preserving selected relations between observations constrains what information remains available after dimensionality reduction. Their emphasis is, however, different: preservation is itself a methodological objective in **Paper A**, whereas in **Paper C** it provides the basis for the subsequent geometric analysis and interpretation of the representation.

Analyzing and interpreting geometric organization. Once a structured hyperbolic representation is available, **Papers B and C** investigate what can be learned from its geometry. They address two complementary forms of organization. **Paper B**, *Hyperbolic Delaunay Geometric Alignment*, studies the geometric relation between two predefined subsets of points within a representation, whereas **Paper C**, *Geometric organization of olfactory descriptor data in the Poincaré disk*, considers the representation as a whole and examines its internal organization. **Paper B** introduces a geometry-aware score for quantifying the alignment between two subsets of points embedded in the same hyperbolic space. The score is based on adjacency relations in the Delaunay graph constructed from their union. It therefore characterizes how strongly the two subsets are geometrically intermixed and can be used to study relations between represented populations as well as changes induced by the data or by the model producing the representation. **Paper C** addresses a different question. Rather than partitioning the representation into predefined groups, it investigates how meaningful properties of the observations are organized across the representation itself. The analysis considers radial position, direction, and localized regions of the Poincaré disk, and relates these geometric patterns to properties of the olfactory descriptor data. The geometry is thus treated not merely as a means of visualization, but as a source of information about the organization and meaning of the represented data. Together, **Papers B and C** illustrate two complementary forms of geometric analysis. **Paper B** asks how predefined subsets of points are geometrically related within a representation, while **Paper C** asks how meaningful information is organized across the representation considered as a whole. Both papers construct geometric quantities and structures to characterize these relations. Their objective, however, is to analyze and interpret the organization of an existing representation, rather than to extract a new structural object such as the tree structures considered in **Papers D and E**.

Extracting and interpreting explicit hierarchies. **Papers D and E** move from characterizing geometric organization to using that geometry to extract explicit tree structures. **Paper D**, *HyperSteiner: Computing Heuristic Hyperbolic Steiner Minimal Trees*, develops an efficient heuristic for constructing Steiner trees directly in two-dimensional hyperbolic space. The inferred tree provides

a parsimonious structural hypothesis connecting the represented observations. Its application to single-cell data further demonstrates how the extracted structure can be interpreted in relation to developmental progression. **Paper E**, *Randomized HyperSteiner: A Stochastic Delaunay Triangulation Heuristic for the Hyperbolic Steiner Minimal Tree*, develops this line further by broadening the exploration of candidate tree topologies and refining their geometry. The resulting trees are not interpreted as guaranteed ground-truth hierarchies, but as parsimonious tree structures connecting the embedded observations according to their hyperbolic distances. By exploring a broader search space, the method can identify shorter Steiner trees than the deterministic heuristic of **Paper D**. The contribution of **Paper E** also has a theoretical interpretive dimension: the attainable reductions in tree length help characterize regimes in which hyperbolic geometry provides greater potential for parsimonious tree structures compared to Euclidean geometry.

Methodological connections. A recurring methodological theme across the thesis is the interaction between continuous geometry and discrete combinatorial structures. In **Paper A**, witness simplicial complexes provide a discrete representation of selected topological information, which is incorporated into representation learning through continuous regularization terms. In **Paper B**, a discrete Delaunay graph is constructed from a continuous hyperbolic representation and its connectivity is used to quantify geometric relations between predefined subsets. In **Papers D and E**, the tree topology is discrete, while its quality is evaluated through hyperbolic tree length and the locations of Steiner points are determined through continuous geometric optimization. A more specific connection between **Papers B**, **D**, and **E** is their use of hyperbolic Delaunay structures. Their role changes with the inference task. **Paper B** uses the Delaunay graph as a neighborhood structure from which geometric alignment is quantified. **Papers D and E** additionally use Delaunay triangles to identify candidate local configurations for Steiner tree construction. **Paper D** uses these configurations in a deterministic heuristic, whereas **Paper E** explores them stochastically to consider a broader set of candidate tree topologies. The same family of geometric constructions is therefore used for different purposes across the thesis: to analyze geometric organization in **Paper B** and to extract explicit tree structures in **Papers D and E**.

Connections across application domains. The papers also investigate geometric inference in different application settings. Biological data recur in **Papers B**, **D**, and **E**. In **Paper B**, biological examples are used to assess whether geometric relations between represented populations agree with domain knowledge. In **Papers D and E**, single-cell data provide an application of explicit hierarchical structure extraction, although the methodological contributions themselves concern hyperbolic computational geometry. **Paper C** considers a different domain, olfactory perception, where geometric analysis is used to

investigate how perceptual descriptor information is organized in a hyperbolic representation. Together, these applications illustrate how the same geometric-inference perspective can support different forms of interpretation across data domains.

Taken together, the papers describe a progression from the structure of a latent representation to the information that can be inferred from it. **Paper A** addresses what can be retained or encouraged during latent representation construction. **Papers B and C** analyze geometric organization within hyperbolic latent representations, while **Papers D and E** extract explicit parsimonious tree structures from those representations. Across **Papers B–E**, interpretation closes this progression by relating inferred geometric information to properties of the data or model.

1.4 Thesis Outline

The remainder of Part I develops the historical and technical context of the thesis, presents the main contributions, and concludes by synthesizing their implications.

Chapter 2 places the thesis in a broader historical and philosophical context. It traces the development from Euclidean to non-Euclidean geometry and discusses how the existence of different coherent geometries motivates treating the geometry of a latent representation as a modeling choice.

Chapter 3 introduces the technical foundations of the thesis. It first considers structured latent representations and the preservation of selected topological and metric information through **Paper A**. It then introduces hyperbolic geometry, its relation to hierarchical and tree-like structure, and the main approaches to hyperbolic representation learning used in the remainder of the thesis.

Chapter 4 addresses **RQ1** through **Papers B and C**. It considers two complementary forms of geometric analysis: quantifying the alignment between predefined subsets of a hyperbolic representation, and analyzing radial, directional, and regional organization within a representation.

Chapter 5 addresses **RQ2** through **Papers D and E**. It studies the extraction of explicit parsimonious tree structures from hyperbolic representations via heuristic hyperbolic Steiner trees, together with the interpretation of properties derived from the resulting trees.

Chapter 6 synthesizes the contributions of the thesis and answers the research questions. It discusses the relation between preservation, analysis, extraction, and interpretation, the main limitations and future directions, potential applications, and societal and sustainability considerations. The chapter concludes by returning to the philosophical perspective introduced in

Chapter 2 and reflecting on geometry as both a modeling choice and a basis for inference.

Chapter 7 summarizes the publications and manuscript included in the thesis and presents their abstracts.

Part II contains the included publications and manuscript.

2 Historical and Philosophical Context: From Euclid's Fifth Postulate to Hyperbolic Data Science

Chapter 1 introduced geometry as a modeling choice for structured latent representations. This view of geometry as a choice among mathematical structures is itself the result of a long historical development. For more than two millennia, Euclidean geometry provided the dominant framework for mathematical reasoning about space. The eventual discovery that logically coherent alternatives were possible changed not only geometry, but also the philosophical status of geometric assumptions.

This chapter retraces part of that history, from the ancient development of geometry and Euclid's *Elements* to the questioning of the fifth postulate and the emergence of non-Euclidean geometry. It then follows hyperbolic geometry into modern machine learning science and data analysis. The purpose is not to provide an exhaustive history of geometry, but to place the questions of this thesis in a broader intellectual context. The historical progression illustrates a theme that will recur throughout the thesis: the geometry employed to formulate a problem should not be taken for granted. The choice of geometry can itself be part of the model, and its consequences for representation and inference can be investigated.

Variational Auto-Encoders (VAEs) provide a concrete example. In a standard VAE, the latent variable is typically defined in a Euclidean space, together with a standard Gaussian prior [16, 17]. This choice is not geometrically neutral: the Gaussian is defined with respect to the structure of the Euclidean latent space, and notions such as distance, isotropy, and interpolation consequently inherit that geometry. If the latent space is instead assumed to be hyperbolic,

the probabilistic model must be adapted accordingly, for example by replacing the Euclidean Gaussian with a Gaussian-like distribution defined on hyperbolic space [29, 30]. The geometry of the latent space should therefore not be inherited implicitly from standard modeling conventions, but chosen deliberately as part of the model, with its consequences for representation and inference examined explicitly.

2.1 Introduction

Mathematics has long served as a means of understanding the world, evolving from practical necessity to an abstract discipline governed by logic and rigor. Euclidean geometry, with its foundations laid in antiquity, remained the dominant framework for spatial reasoning for over two millennia. However, its assumptions, particularly the contentious fifth postulate, were increasingly questioned. Mathematicians seeking to prove this postulate ultimately found themselves charting new mathematical territory, leading to the birth of hyperbolic geometry.

Hyperbolic geometry was initially seen as an abstract anomaly. Indeed, in the hyperbolic plane, through a point outside a given line there exist infinitely many lines that do not intersect the given line, in contrast with the unique parallel of Euclidean geometry. What emerged from an abstract and counter-intuitive problem concerning the foundations of geometry later became relevant in several areas of mathematics, physics, biology, neuroscience, network science, and, more recently, machine learning and data analysis, enabling more efficient representation of hierarchical data.

This history is particularly relevant to the present thesis. Modern representation learning once again makes geometry a deliberate design choice rather than a merely accepted assumption. A latent representation may be Euclidean, spherical, hyperbolic, or from other geometries with non-constant curvature. The selected geometry determines how distances, neighborhoods, directions, and other relations between represented observations are expressed. The question is therefore no longer only which geometry describes physical space, but also which geometry is useful for representing and reasoning about data.

2.2 Pythagoras' Influence

Geometry is one of the oldest branches of mathematics and can be traced back several thousand years to ancient Mesopotamia and Egypt. For example, several clay tablets from the Old Babylonian period testify to the use of the Pythagorean theorem at least a millennium before the existence of Pythagoras himself [31]. The latter, Pythagoras of Samos, contributed substantially to the development of this mathematical field that deals with properties of space such as shapes and distances, called geometry. This word is formed by the combination of Ancient Greek words "$\gamma\tilde{\eta}$" which means "earth" or "land", and

"$\mu\acute{\epsilon}\tau\rho o\nu$" which means "a measure". So the word geometry, which comes from "$\gamma\epsilon\omega\mu\epsilon\tau\rho\iota\alpha$", can be translated as "measurement of earth or land". We can see from the etymology that this field was originally about measuring and dividing the lands. Understanding the earth might be a first step towards understanding the world around us, however, geometry is not only about measuring earth.

In particular, during the 6th century BC, the Pythagoreans provided a geometric representation for numbers, relied on geometric figures to make arithmetic demonstrations, and discovered that there is a relation between the length of a vibrating string and the pitch of the sound emitted. This allowed for theorizing the origin of the musical scale. As Aristotle wrote in his first book of *Metaphysics*, the Pythagoreans understood musical harmony through arithmetic relations, associating the properties of musical scales with ratios between numbers [32].

They combined elements of philosophical schools and religious brotherhoods and were characterized by a degree of secrecy and oral transmission of doctrine. Mathematical and scientific inquiry nevertheless formed part of their activities.

Pythagorean communities were supposedly established first by Pythagoras and later developed by his followers and students in several places based on his teaching and beliefs. They combined features of philosophical schools and religious brotherhoods, with a degree of secrecy, staged initiation, and oral transmission of esoteric doctrines, while also providing an environment for mathematical and scientific inquiry [33]. They conceptualized geometry as an integral component within a broader framework of philosophical contemplation concerning the inherent perfection they attributed to the *cosmos*. It is believed that Pythagoras envisioned the study of geometry as a transformative discipline capable of harmonizing human understanding with the fundamental perfection of the universe. This led to the establishment of a secretive tradition dedicated to the exploration of philosophy and the sacred practice of geometry. Despite their predilection for esotericism, resulting in limited written documentation, the Pythagoreans actively shared their insights and convictions with their disciples and peers. However, their primary fascination resided in the philosophical implications of Pythagoras's teachings, rather than prioritizing the development of a unified mathematical system.

On the contrary, proposing a unified mathematical system was the goal of Euclid with his famous books the *Elements*, which were a pivotal moment for geometry in particular, and more generally for mathematical history.

2.3 Plato's Influence

Another important Greek influence on Euclid is undoubtedly Plato. He also reflected on the foundations of mathematics and helped shape the intellectual

tradition from which Euclid later emerged. Plato restructured underlying assumptions and provided clarifications on various definitions inherited from the Pythagoreans [33]. He underscored the notion that geometric reasoning should be done on the abstract ideas represented by geometric figures, and not on the depicted visible figures themselves. Indeed, if we reason on a square that we draw, we should think of this square as having its four sides of equal length, although it is practically impossible to draw an exact square.

However, the influence of Plato on Euclid is not necessarily Plato himself, but his legacy. While Plato did not personally produce any notable mathematical breakthroughs, he served as the focal point for the development of mathematics during his era, exerting significant influence on its advancement. It is said that the following inscription adorned the doors of his school at the Academy in Athens:

> *"Let no one ignorant of geometry enter here."*

Whether this is true or not, it at least reflects his dedication to geometry and the importance of this subject. Plato's passion for mathematics earned him the reputation not as a mathematician himself, he is indeed mostly known as a philosopher, but as "the maker of mathematicians", emphasizing his influence in nurturing the next generation of mathematicians [33]. Some content from Euclid's *Elements* is strongly influenced by contributions of persons who were at some point doing mathematics at Plato's Academy like Theaetetus of Athens or Eudoxus of Cnidus.

While Pythagoras is credited with establishing mathematics as a liberal discipline [33], Plato's influence was pivotal in integrating it into the core curriculum for the education of statesmen. The liberal arts were seven subjects taught in antiquity and medieval Western Europe traditional curriculum, consisting of the *quadrivium* of arithmetic, geometry, music and astronomy, and the *trivium* of rhetoric, grammar and logic. It dates back to Ancient Greek with their "desire for a universal understanding" [34]. Plato explicitly defends the learning of mathematics in his book *The Republic* [35]. When discussing geometry, he starts with mentioning its virtues:

> *"The knowledge at which geometry aims is knowledge of the eternal, [...] geometry will draw the mind towards truth, and create the spirit of philosophy, and raise up that which is now sadly allowed to fall down".*

Then, he argues for its universal teaching. He finally explains more direct and indirect practical advantages of learning geometry:

> *"There are the military advantages [...] and in all departments of knowledge, as experience proves, any one who has studied geometry is infinitely quicker at learning other subjects than one who has not".*

As a result, during the middle of the fourth century before Christ, the leading teachers and researchers in mathematics, as well as other topics like philosophy with the renowned Aristotle, emerged from the Platonic Academy in Athens, which had become the mathematical center of the world. Historians of mathematics generally agree to say that Euclid's mathematical contributions suggest that he likely received instruction from disciples of Plato, possibly within the academic environment of the Academy itself [33].

2.4 Euclid's Elements

History of mathematics is closely intertwined with the broader history of the world. In 323 before Christ, Alexander the Great died, and his empire was divided among his generals. It is worth noting that Aristotle, who was a student of Plato at the Academy, was actually a tutor of this emperor. Aristotle was a great philosopher, but his direct contribution to the development of mathematics was not particularly significant [33]. However, his philosophy shaped Euclid's systematic development of geometry.

Aristotle died one year after Alexander the Great, and these events occurred at the transition between two major historical periods of Antiquity: from the *Classical* period to the *Hellenistic* period, sometimes also referred to as the *Alexandrian age*. Alexander of Macedon founded many cities, among them Alexandria, which would become one of the major intellectual and mathematical centers of the world. After his death, one of his generals, Ptolemy I, established the Ptolemaic Kingdom in Egypt. His son, Ptolemy II, succeeded him, and under the early Ptolemies Alexandria flourished as an important center of scholarship. In particular, the Mouseion and its Great Library of Alexandria developed under the early Ptolemies and contributed greatly to this intellectual environment. As the art historian Germain Bazin writes in his book *Le temps des musées* [36], the Mouseion was "a college of scholars, analogous to the modern Institute for Advanced Study at Princeton or to the College de France in Paris". Ptolemaic patronage supported scholars residing and working there, as the rulers sought to attract leading scholars from the Hellenistic world in many different fields, as it was the case for the great mathematician Euclid. The Library was similarly associated with the ambitious collection of texts from across the Greek and Mediterranean worlds. This contributed to making Alexandria a center for knowledge and science through the sponsorship of the Mouseion by the royal family of the Ptolemies and subsequently by the Roman emperors. It was within this broader intellectual environment that Euclid worked in Alexandria, traditionally during the reign of Ptolemy I, and produced the work that would profoundly influence the development of mathematics [33].

The Pythagoreans and other Ancient Greeks like Plato and mathematicians from his Academy greatly influenced the work of Euclid. However, Euclid's *Elements* contrasts with earlier Pythagorean traditions, as it is a clear advance-

ment towards rigorous mathematics. The *Elements* consists of 13 books or chapters studied thoroughly for over 2000 years and that laid the foundations for *synthetic geometry*, *arithmetic* or *theory of numbers*, and *geometric algebra*. The Greeks adhered to the philosophical idea that mathematical knowledge should be grounded in logical deduction rather than empirical observation, shaping their approach to the development of geometry. They embraced the consequences of this conviction by constructing geometry as a system of mathematical propositions deduced from definitions and first principles, or axioms, in systematic treatises traditionally called *Elements*.

According to Proclus, Hippocrates compiled a set of *Elements* approximately a century prior to Euclid. Subsequent Greek mathematicians followed suit, developing analogous systems. Among the oldest preserved *Elements* is the one of Euclid, which stands as a testament to the enduring legacy of Greek mathematical thought. In Euclid's *Elements*, we witness a synthesis of mathematical notions drawn from predecessors like Theaetetus and Eudoxus. Although Euclid does not mention his specific new contributions, the organization of the Elements is commonly attributed to him, and it is presumed that he contributed also to certain proofs [33]. However, determining the precise extent of his innovative input remains a challenge. Like many ancient Greek texts, the original manuscript has been lost to time. Over the centuries, the Elements underwent numerous iterations, with later scholars commenting and introducing modifications deemed necessary. Consequently, the original composition of the Elements eludes complete reconstruction, as it exists today as a composite of various interpretations and additions.

During the Middle Ages, the dissemination of the Elements in Western Europe was facilitated by the Arabs and Moors, who recognized its foundational role in mathematical education. With over 1000 known editions, the Elements stands as one of the most widely circulated texts in Western civilization after the Bible [37].

2.5 Euclid's Fifth Postulate and its Questioning

Among the five postulates Euclid introduced in the *Elements*, the fifth stands apart in both its complexity and the historical controversy it sparked. Unlike the others, which are intuitively simple—such as the ability to draw a straight line between any two points or the equality of all right angles—the fifth postulate appears convoluted and unnatural. Euclid's formulation can be translated as follows [38]:

> *That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.*

In essence, this postulate governs the behavior of parallel lines. It asserts that

two lines that are not parallel will inevitably intersect, whereas two lines cut by a transversal that form interior angles summing to exactly 180 degrees on the same side will never meet. While this may seem obvious, its formulation is significantly more complex than the other postulates. The fifth postulate is also notable for its role in the deductive structure of the *Elements*. The first twenty-eight propositions of Book I are developed without it, and it is first invoked in Proposition 29. Its form and delayed use later attracted extensive attention from commentators and mathematicians, who repeatedly attempted to derive it from the four other postulates [38, 39]. Given Euclid's apparent reluctance to invoke it except when it became necessary, one might even wonder whether he regarded his fifth postulate as something that should follow logically from the others rather than stand as an independent assumption...

For over two millennia, mathematicians attempted to derive the fifth postulate from the first four. The idea that such an essential property of space should require an explicit assumption felt unsatisfactory. Various scholars reformulated the postulate into equivalent statements, hoping that a more intuitive expression would make it easier to prove. An important part of this history took place in the medieval Islamic world, where Greek mathematical works were not only preserved and transmitted but also critically studied and extended. The theory of parallel lines became the subject of sustained mathematical investigation [40, 41].

Among the mathematicians who worked on the problem was Ibn al-Haytham (also known as Alhazen), who proposed in the beginning of the eleventh century an attempted proof involving the behavior of equidistant lines and geometric motion. His argument did not establish the parallel postulate independently, because it relied on assumptions that are themselves closely related to the property that was to be proved. Nevertheless, his work formed part of an important sequence of investigations in which the logical dependencies underlying Euclidean geometry were progressively exposed [40, 41].

Omar Khayyam subsequently criticized earlier attempted proofs, including that of Ibn al-Haytham, and developed a different approach in his commentary *Explanations of the difficulties in the postulates of Euclid*. He considered a quadrilateral with two equal sides perpendicular to the same base, a construction now often called the *Khayyam–Saccheri quadrilateral*. The two remaining summit angles are equal, leaving three possibilities: they may be right, acute, or obtuse. Khayyam examined these alternatives systematically. The right angle hypothesis leads to the Euclidean parallel postulate, while the acute and obtuse angle hypotheses lead towards geometric situations that, in modern terminology, are associated respectively with hyperbolic and elliptic or spherical geometry [41, 40]. Khayyam did not accept the non-Euclidean alternatives as independent geometries. His purpose remained to establish Euclid's parallel postulate, and the additional principle on which his argument relied is now understood to be equivalent to the property he sought to prove.

Yet his analysis was historically important precisely because it made the three alternatives explicit. The same three possibilities that appeared as obstacles in an attempted proof would, centuries later, become associated with the three principal constant-curvature geometries: the zero-curvature Euclidean case, the negative-curvature hyperbolic case, and the positive-curvature spherical or elliptic case [41].

Nasir al-Din al-Tusi continued and developed this line of investigation in the thirteenth century. In his work on parallel lines, he critically examined the arguments of earlier mathematicians, including Ibn al-Haytham, al-Jawhari, and Khayyam, and proposed his own attempted derivations of the parallel postulate. Like Khayyam, al-Tusi considered alternatives corresponding to the different possible behaviors of the relevant angles. Although he ultimately rejected the non-Euclidean cases rather than treating them as new geometries, his analysis helped clarify the logical alternatives hidden behind Euclid's fifth postulate [41, 40].

The influence of this mathematical tradition extended beyond the medieval Islamic world. A later treatment historically associated with al-Tusi's tradition, was printed in Arabic in Rome in 1594 [42]. Ideas from the Islamic mathematical tradition concerning Euclid's fifth postulate entered European discussions of geometry. Before the publication of the 1594 Rome edition, Christopher Clavius had encountered a manuscript attributed to Nasir al-Din al-Tusi and, in the 1589 edition of his commentary on Euclid, included the attempted proof of the fifth postulate that he found there. Later, Edward Pococke translated into Latin a demonstration of the parallel postulate from the Arabic tradition associated with al-Tusi, and John Wallis subsequently published and discussed it alongside his own work on the parallel-postulate problem [39, 42].

Giordano Vitale later employed the same family of quadrilateral constructions as Khayyam, and Giovanni Girolamo Saccheri developed the analysis much further in his work *Euclid Freed from All Flaws* of 1733. Saccheri was familiar with the work associated with al-Tusi and with that of Wallis and explicitly discussed them in his treatment of the parallel problem. His method again considered the right, obtuse, and acute angle hypotheses in an attempt to eliminate the latter two and thereby prove Euclid's postulate [41].

This sequence belongs to a broader history of exchanges of mathematical and scientific knowledge between the Islamic world and Europe. The intellectual transformations that preceded and accompanied the European Renaissance drew not only on ancient Greek works, but also on traditions of mathematics, physical and life sciences, medicine, or philosophy that had been developed in the medieval Islamic world and transmitted into Latin Christendom through different translation networks [43]. European scientific development should therefore not be viewed as an isolated process, but as part of a longer history of preservation, transmission, criticism, and further development of knowledge across different intellectual traditions.

The historical path towards non-Euclidean geometry was therefore not a direct jump from Euclid to nineteenth-century Europe. Medieval Islamic mathematicians played an important role in maintaining the parallel-postulate problem as an active mathematical question, criticizing earlier arguments, developing new geometric constructions, and making explicit alternatives that would later acquire a completely different interpretation. Their goal was generally still to prove Euclid's postulate, but in attempting to do so they obtained results and distinctions that form an important part of the prehistory of hyperbolic and other non-Euclidean geometries.

Despite these efforts, all attempts to derive the fifth postulate as a theorem ultimately failed. The realization that it could not be proven led to a groundbreaking change of paradigm in mathematical thought. Instead of asking *How can the fifth postulate be proved?*, mathematicians could eventually ask *What geometry results if the fifth postulate is not assumed?*

2.6 Philosophical Doubt and the Birth of Hyperbolic Geometry

The history of the fifth postulate can also be viewed against a broader philosophical transformation concerning the status of received assumptions. René Descartes' *Discourse on the Method* laid the foundation for modern skepticism, urging scholars to question all assumptions that could not be logically deduced from self-evident truths. His famous dictum, *Cogito, ergo sum* ("I think, therefore I am"), exemplified the principle that knowledge must be rigorously justified rather than accepted based on tradition or intuition.

It would be too strong to claim that Cartesian skepticism directly caused the discovery of non-Euclidean geometry. Lobachevsky, Bolyai, and the other mathematicians involved in its development were responding to specific mathematical problems with their own historical origins. Descartes is useful here instead as a philosophical parallel. His *Philosophy of Doubt* illustrates a broader intellectual possibility: assumptions that appear obvious or deeply established can themselves become objects of examination.

Seen from this perspective, the fifth-postulate problem has an especially interesting character. For centuries, the postulate was treated as something that should follow from the remaining structure of Euclidean geometry. The eventual breakthrough came not from finally finding the desired proof, but from taking seriously the possibility that no such proof existed and examining the consequences of alternative assumptions. If the fifth postulate could not be derived from the others, was it truly a fundamental property of space?

In a letter to a fellow astronomer in 1817, Carl Friedrich Gauss wrote [44]:

> *"I am becoming more and more convinced that the necessity of our (Euclidean) geometry cannot be proved, at least not by*

> *human reason nor for human reason. Perhaps in another life we will be able to obtain insight into the nature of space, which is now unattainable."*

Around the same period, in the early nineteenth century, Nikolai Lobachevsky and János Bolyai, working independently, developed a new form of geometry, namely hyperbolic geometry, in which the Euclidean parallel postulate does not hold: instead of a unique parallel line through a given point outside a line, infinitely many non-intersecting lines exist. Gauss had developed closely related ideas earlier but did not publish them. In another letter to Friedrich Wilhelm Bessel in 1829, Gauss wrote that he feared the "howls of the Boeotians" if he expressed his views publicly, indicating his concern about the reception that non-Euclidean ideas might receive [44].

The later work of Bernhard Riemann broadened the mathematical conception of geometry considerably, while Eugenio Beltrami provided a decisive interpretation of the non-Euclidean geometry developed by Lobachevsky and Bolyai. In two works published in 1868 and 1869, Beltrami constructed several concrete models of hyperbolic geometry within familiar Euclidean coordinate domains [45, 46]. These include models now commonly known as the Klein–Beltrami disk, the Poincaré disk, and the Poincaré half-plane models. Their modern names partly conceal their historical development: the conformal disk model also builds on ideas of Riemann, while the half-plane metric had previously appeared in work by Liouville [46].

These models represent the same non-Euclidean geometry in different ways. In the projective disk model, now usually called the Klein–Beltrami model, hyperbolic geodesics are represented by Euclidean straight-line chords of the disk. In the conformal disk model, now usually called the Poincaré disk, angles are preserved, while hyperbolic geodesics appear as circular arcs orthogonal to the boundary, together with diameters. The half-plane model gives yet another conformal realization of the same geometry. Beltrami's work also related different realizations through coordinate transformations, showing that apparently different geometric pictures could describe the same intrinsic non-Euclidean space [46].

This was important not only for the acceptance of non-Euclidean geometry, but also for the concept of a mathematical model itself. Hyperbolic geometry no longer had to be understood only through an unfamiliar collection of statements about points and lines. It could be represented concretely in different coordinate systems, each making different properties of the same geometry easier to see or manipulate. Beltrami's models helped establish non-Euclidean geometry as a valid mathematical framework and showed that the same geometry can be represented in different ways, each with its own advantages.

This plurality of models has a direct connection to the present thesis. The

choice between different models of hyperbolic geometry does not change the underlying hyperbolic space, but it changes which geometric properties are represented most conveniently. **Paper C**, for example, works in the Poincaré disk, which provides an intuitive representation that has been widely used for learning and visualizing hyperbolic embeddings. In contrast, **Papers B**, **D**, and **E** make use of the Klein–Beltrami model, where hyperbolic shortest paths are represented as Euclidean straight lines, facilitating geometric constructions and computations. Transformations between hyperbolic models are therefore not merely theoretical equivalences in this thesis: they allow the same hyperbolic representation to be expressed in the model best suited to a particular analysis or computation.

This historical shift changed the status of geometry. Euclidean geometry remained a valid mathematical system, but it was no longer the only possible one. Different geometries could be developed from different assumptions. This shift provides one of the central philosophical connections to the present thesis. Choosing a non-Euclidean geometry for a latent representation does not mean that Euclidean geometry has somehow become ”wrong”. It means that the chosen geometry is part of the mathematical model. Different choices expose and encode better different relations, and their usefulness must be investigated relative to the problem being studied.

2.7 The Fifth Postulate and Gödel's Incompleteness Theorem: A Philosophical Parallel

The history of Euclid's fifth postulate offers a philosophical parallel with Gödel's incompleteness theorem. To understand this connection, it is useful first to clarify the notion of logical *undecidability*. A statement is undecidable within a given axiomatic system if neither the statement nor its negation can be proved from the axioms.

The parallel postulate provides a classical example. Euclid's original *Elements* is not a fully formal axiomatic system in the modern logical sense, since some assumptions used in its geometric arguments remain implicit and are not formally introduced as axioms [47]. Modern axiomatizations, such as those of Hilbert and Tarski, make the logical status of the parallel postulate more precise [48, 49]. If the parallel postulate is omitted, the remaining framework is commonly called *neutral* or *absolute geometry*. Within this framework, the parallel postulate cannot be proved, but neither can its negation. Euclidean geometry provides a model in which it holds, while hyperbolic geometry provides models in which it does not. The parallel postulate is therefore undecidable within neutral geometry.

This result changed the role of axioms in geometry. Rather than viewing the parallel postulate as a statement that should eventually be derived from the other axioms, mathematicians came to recognize that different additional

assumptions lead to different consistent geometries. Adding the Euclidean parallel postulate gives Euclidean geometry, whereas adopting a hyperbolic alternative gives hyperbolic geometry.

Gödel's incompleteness theorem, established more than a century later, concerns a different and stronger phenomenon. It states that any consistent and effectively axiomatized formal system sufficiently rich to express arithmetic contains statements that are undecidable within that system [50]. The parallel postulate is undecidable in neutral geometry for a different reason: neither it nor its negation follows from the remaining geometric axioms, as demonstrated by models in which the postulate holds and models in which it does not. Gödel's theorem shows instead that, for sufficiently expressive formal systems, some undecidable statements are unavoidable.

The philosophical parallel nevertheless remains useful. In both cases, what can be established depends on the assumptions of the formal framework. The history of the parallel postulate showed that a natural geometric question can remain undecidable under a given set of axioms, while Gödel later showed that, for sufficiently expressive formal systems, the existence of such undecidable statements is unavoidable. This perspective connects to a broader theme of this thesis: the geometry of a latent representation is itself a modeling assumption. Different geometric choices can support different structures and forms of inference, and the conclusions drawn from them must therefore be interpreted in relation to those choices.

2.8 Hyperbolic Geometry in Nature, Science, and Data

Although initially regarded as an abstract mathematical theory, hyperbolic geometry arises as a useful description of several natural systems [51], notably due to a physical phenomenon called "geometric frustration" that can take place for example in nanostructures of the human skin or butterfly wing [52]. It has also been discovered recently that population activity of neurons in the rat hippocampus, an important region of the brain for spatial memory and navigation, is consistent with a hyperbolic geometric organization [53]. More generally, applications from classical to quantum mechanics [54], condensed matter [55], biological soft materials [52] or neuro-mathematics [56] motivate the research in the field of triangulating hyperbolic surfaces.

The transition from physical and biological systems to data representations introduces another role for geometry. A dataset does not need literally to lie on a physical hyperbolic surface for hyperbolic geometry to be useful. Instead, one may construct a hyperbolic representation because the relations of interest are more naturally expressed in such a space.

This is particularly relevant for hierarchical and tree-like data. The advantage of hyperbolic space for representing tree-like data follows from its exponential growth. The circumference of a hyperbolic circle grows exponentially with its

radius, mirroring the exponential growth in the number of nodes with depth in a tree of constant branching factor. This geometric similarity motivates the common interpretation of hyperbolic space as a continuous analogue of a tree [24]. As mentioned in Chapter 1, Sarkar showed that only two dimensions are sufficient to embed any tree into a hyperbolic space with arbitrarily low distortion [27], enabling low-dimensional representations of tree-like data that preserve well the original pairwise distances. This provides a strong motivation for using low-dimensional hyperbolic representations when the structure of interest is hierarchical. The more detailed theory of distortion, tree metrics, and hyperbolic representations is developed in Chapter 3; here, the important point is the historical transition from a geometry created by questioning an axiom to a geometry deliberately chosen as a representation space.

This connection has motivated a growing body of work in machine learning. Works in hyperbolic dimensionality reduction show that better downstream-task performance can be achieved for hierarchical data if it is represented in a hyperbolic space [57, 58]. Hyperbolic models can be deployed to embed different examples of hierarchical data like geographic communication networks [59], internet networks [60], words [25, 61] etc. These developments form part of what can be viewed broadly as *hyperbolic data science*.

For the purposes of this thesis, *hyperbolic data science* is used as an umbrella term for methods that employ hyperbolic geometry to represent, learn, analyze, or infer structure from data. Within this terminology, *hyperbolic machine learning* concerns machine learning methods that make use of hyperbolic geometry, whereas *hyperbolic data analysis* focuses on the analysis of data once represented in a hyperbolic space.

The distinction leads directly to the focus of this thesis. A large part of the recent literature has addressed how hyperbolic representations can be constructed. Yet once such a representation has been obtained, further questions arise. How can relationships between groups of represented observations be quantified? How can meaningful variables be related to the learned geometric organization? How can a hyperbolic representation of data points that appears hierarchical be converted into an explicit parsimonious tree? What can such geometric structures tell us about the original data or about the geometry itself?

The papers in this thesis approach these questions from different directions. **Paper A** precedes the specifically hyperbolic part of the thesis and studies how selected topological and metric information can be preserved or encouraged when a latent representation is constructed. Although its latent spaces are Euclidean, it illustrates a principle that is central to the remainder of the thesis: a representation is shaped by priors and inductive biases and they determine which structures remain available for subsequent inference. If the structure of interest is the data hierarchy, then an appropriate inductive bias is to choose a hyperbolic latent geometry.

Papers B–E then operate in hyperbolic space. **Paper B** develops a method for analyzing geometric relations between predefined subsets of a hyperbolic representation. **Paper C** studies how semantic properties of olfactory data are organized within a Poincaré-disk representation. **Papers D and E** move from analyzing geometric organization to constructing explicit parsimonious tree structures connecting points according to their hyperbolic distances.

These methods contribute to what this thesis calls *hyperbolic geometric inference*: the analysis and extraction of structural information encoded in hyperbolic representations, together with the interpretation of the resulting structures. Hyperbolic geometric inference can therefore be viewed as a key element for performing hyperbolic data analysis, which itself belongs to the broader development of hyperbolic data science.

There is also a historical symmetry in this progression. The nineteenth-century development of non-Euclidean geometry established that the geometry used to reason about a problem is itself an assumption that can be changed. Modern representation learning makes this possibility computational. Geometry is not only selected abstractly; data can be mapped into the resulting space, and algorithms can for example use its distances, geodesics, density estimates, Delaunay or Steiner tree structures to perform inference.

2.9 Conclusion

The development of hyperbolic geometry and its modern use in data science illustrates the nonlinear but interconnected progression of mathematical and scientific ideas. The emergence of non-Euclidean geometry was not a single inevitable step. It followed centuries of attempts to understand the special status of Euclid's fifth postulate, together with the eventual willingness to consider that an apparently obvious property of space might instead be an assumption defining one geometry among several possible geometries.

Several pivotal moments stand out in this journey. Ancient Greek mathematics provided both the geometric problems and the deductive framework within which they were studied. Euclid organized this knowledge into the *Elements*, while the fifth postulate remained unusually resistant to derivation from the other four. Mathematicians in the medieval Islamic world, including Ibn al-Haytham, Omar Khayyam, and Nasir al-Din al-Tusi, developed important approaches to the parallel problem, finding equivalent formulations and classifying geometric cases, that became part of its long historical development. Their contributions influenced later European mathematicians, such as Clavius, Wallis, Vitale or Saccheri, and ultimately prepared the field for the radical breakthroughs of the nineteenth century. Subsequent European work continued these investigations before Lobachevsky and Bolyai finally treated an alternative to Euclidean parallelism as the basis of a geometry in its own right. Riemann broadened the conception of geometry further, and Beltrami provided

mathematical models that helped establish the legitimacy of non-Euclidean geometry.

Descartes' philosophy of doubt provides a useful philosophical parallel to this history, even if it may not be treated as a direct historical cause of non-Euclidean geometry. The broader lesson is that inherited assumptions may themselves become objects of investigation. The comparison with Gödel leads to a related reflection. The undecidability of the parallel postulate shows that neutral geometry is incomplete with respect to this particular statement: neither the Euclidean parallel postulate nor its negation can be proved from the remaining geometric assumptions. Gödel's incompleteness theorems concern a different and stronger phenomenon, showing that sufficiently expressive formal systems necessarily contain statements that cannot be decided from their axioms. Although the two results are mathematically different, both illustrate how what can be proved depends on the axioms and formal framework from which one starts.

It is interesting, although necessarily speculative, to ask how the same historical development might have unfolded under contemporary scientific conditions. Modern research provides much faster communication, computation, and international collaboration than was possible for earlier mathematicians. At the same time, contemporary research is shaped by publication incentives, funding priorities, and expectations concerning demonstrable impact. Whether these conditions would have accelerated or discouraged such a long investigation of an apparently abstract axiom is difficult to know.

This tension has accompanied many areas of mathematical science. Nicolas Rashevsky argued that "a theoretical problem may have an interest of its own" even before a definite experimental application is visible [62]. The history of hyperbolic geometry provides a particularly striking example. A mathematical structure developed through foundational questions about parallel lines eventually became useful in areas that earlier scientists may not have anticipated.

The evolution of scientific culture has also changed how mathematical ideas are produced and disseminated. The secretive traditions associated with the Pythagorean communities contrast with modern ideals of publication, collaboration, and reproducibility. Yet the central role of questioning assumptions remains. Mathematics and science in general progress not only by solving problems inside established frameworks, but sometimes by examining the frameworks themselves.

Hyperbolic data science represents one recent continuation of this history. A geometry that emerged from questioning the foundations of Euclidean geometry is now employed as a modeling choice for data whose organization may be difficult to express efficiently in Euclidean space. Just as non-Euclidean geometry later became important in areas of physics, hyperbolic geometry is re-

cently being used to model and study complex and hierarchical data structures. *Hyperbolic machine learning* provides methods for learning representations of such data in hyperbolic spaces, while *hyperbolic data analysis* asks what can subsequently be learned from these representations. While structure preservation provides an important upstream component, the work of this thesis is situated mainly within hyperbolic data analysis, through **hyperbolic geometric inference**: analyzing geometric organization, extracting explicit structures, and interpreting what those structures mean.

The historical lesson is therefore not that one geometry replaced another. Euclidean geometry remained valid after the development of hyperbolic geometry. What changed was the recognition that geometric assumptions themselves could be investigated. An analogous perspective motivates this thesis. The geometry of a latent representation should be chosen deliberately, its consequences examined, the structures it supports inferred, and the interpretation of those structures treated critically.

In this sense, the path from Euclid's fifth postulate to hyperbolic data science is not merely a historical background for the technical work presented in the remainder of the thesis. It provides the philosophical setting for a central question of this thesis:

> *What can be learned when the geometry of a representation is used as a basis for inference?*

3 Structured Latent Representations: From Topology to Hyperbolic Geometry

The previous chapter described how the development of non-Euclidean geometry eventually made geometry itself a choice among different mathematical models. This chapter considers the same idea in the context of latent representations. When high-dimensional data are mapped into a lower-dimensional space, the resulting representation is not determined by the data alone. Choices concerning the architecture, objective function, metric, topology, probability distributions, and geometry of the latent space all influence how the data are organized.

These choices can be understood through the notions of *prior knowledge* and *inductive bias*. The terminology is used here in a broad modeling sense. Following the view of Tom M. Mitchell, an inductive bias refers to assumptions or preferences that make a learning method favor some solutions or generalizations over others beyond what is determined by the observed data alone [63]. Prior knowledge refers here to information or structural assumptions available about the problem before learning the representation. Such knowledge can be incorporated into a learning method so as to induce a corresponding bias [64, 65]. Thus, *prior knowledge* concerns *what* is assumed about the relevant structure, while the *inductive bias* describes *how* the learning procedure is guided toward representations compatible with that assumption.

In this thesis, these assumptions take different forms. A prescribed embedding provides prior information about how the latent representation should be organized and is incorporated through a geometric constraint on the learned representation. Prior knowledge that selected topological structure should be

retained is incorporated through simplicial regularization. When the structure of interest is instead expected to be hierarchical or tree-like, the assumption is less specific: rather than prescribing a particular configuration of points, hyperbolic geometry is imposed as a geometric inductive bias adapted to this type of organization.

This perspective connects the two main parts of this chapter. We first consider **Paper A**, which studies how different forms of structural information can be encouraged during representation learning. We then move from constraints on a Euclidean latent representation to the choice of the latent geometry itself, introducing the hyperbolic models and Riemannian tools used throughout the remainder of the thesis.

Contributions of Paper A. **Paper A** investigates how selected geometric and topological structure can be incorporated into the latent representation of a variational auto-encoder (VAE). It contributes two complementary approaches, together with a metric extension. InvMap-VAE turns a prescribed low-dimensional embedding into a generative latent-variable model by learning mappings between the data space and that representation. Witness Simplicial VAE instead introduces a topological regularization based on a witness complex, encouraging the encoder and decoder to behave approximately as simplicial maps on the selected structure. The paper further introduces Isolandmarks Witness Simplicial VAE, which supplements this topological constraint with the preservation of approximate geodesic distances between landmark points. The methods are evaluated on a synthetic open-cylinder and the Swiss-roll datasets, where the relevant topological properties are known. InvMap-VAE produces latent representations that follow the prescribed embeddings. Witness Simplicial VAE preserves the targeted persistent Betti numbers on the open-cylinder example, but the experiments also expose a limitation of topological constraints alone: on the Swiss roll, an undesirable geometric overlap can introduce an additional one-dimensional hole. Adding the Isolandmarks metric constraint reduces this overlap and recovers the desired persistent Betti numbers although the reconstruction could still be improved. These experiments therefore illustrate both the benefits and limitations of imposing selected structural constraints, and motivate combining topological and metric information when topology alone does not sufficiently constrain the latent geometry.

3.1 Latent-Space Structure as an Inductive Bias

Let $X = \{x_i\}_{i=1}^N \subset \mathcal{X}$ denote a dataset of N points in an observation space $\mathcal{X}$, and let $Z = \{z_i\}_{i=1}^N \subset \mathcal{Z}$ denote a latent representation obtained through a map $f : \mathcal{X} \to \mathcal{Z}$.

The latent space $\mathcal{Z}$ may be Euclidean, but more generally it can be a manifold equipped with a chosen geometry. A map f can be parameterized by a set of

parameters θ, for example the weights of a neural network, and will be denoted as f_θ.

The question considered here is not only how to reduce dimensionality, but what properties the resulting representation should retain. Different learning objectives express different answers to this question. If pairwise distances are important, an objective can penalize discrepancies between distances before and after representation. If topology is important, the objective can encourage selected connectivity or higher-order topological features to remain unchanged. If an entire reference embedding is imposed, its coordinates can directly constrain the learned representation. Finally, if prior knowledge concerns the general type of organization rather than a particular configuration of points, the geometry of $\mathcal{Z}$ itself can provide the inductive bias.

These forms of prior knowledge constrain the representation in different ways. A prescribed embedding specifies a desired point configuration directly (e.g. InvMap-VAE in **Paper A**). Topological information constrains selected structural relations while allowing multiple geometric realizations (e.g. Witness Simplicial VAE in **Paper A**). Choosing the ambient geometry does not necessarily specify where individual points should be placed, but introduces a global inductive bias by determining the metric and geometric relations available throughout the latent space (e.g. hyperbolic VAEs, or hyperbolic geometry in latent representations of **Papers B–E**).

This distinction is useful for understanding the progression from **Paper A** to the hyperbolic part of the thesis **Papers B–E**.

3.2 Variational Auto-Encoders and Latent-Space Structure

Variational auto-encoders (VAEs) provide a probabilistic framework for learning latent representations together with a generative mapping back to the observation space [16, 17]. Given an observation x, the encoder defines an approximate posterior $q_\phi(z \mid x)$, while the decoder specifies a likelihood $p_\theta(x \mid z)$.

A probabilistic prior $p(z)$ is imposed on the latent variable. Training is commonly formulated as the minimization of the negative evidence lower bound, which can be written as

$$\mathcal{L}_{\text{VAE}}(x) = -\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x \mid z)] + D_{\text{KL}}\left(q_\phi(z \mid x) \,\|\, p(z)\right). \tag{3.2.1}$$

The first term encourages accurate reconstruction, whereas the second regularizes the approximate posterior relative to the latent prior.

In the standard formulation, the latent variable belongs to $\mathbb{R}^d$ and the prior is typically an isotropic Gaussian. This is a natural and convenient choice, but it also establishes a particular latent space and probability model. The

VAE objective alone does not guarantee that specific geometric or topological properties of the input data will be retained in its latent representation.

Additional structural information can be introduced in several ways. The objective can be augmented with regularization terms, the latent probability model can be modified, or the latent space itself can be replaced by a non-Euclidean manifold. These possibilities provide the main connection between **Paper A** and the later hyperbolic methods considered in this thesis.

3.3 Preserving Selected Structure: Paper A

The previous sections introduced latent representations and the VAE framework. We now consider how prior structural information can be incorporated directly during representation learning. **Paper A** investigates this question through three related approaches. InvMap-VAE uses a prescribed embedding as geometric supervision for the latent representation. Witness Simplicial VAE instead introduces a topological inductive bias through a *witness simplicial complex* constructed from the input data. Its Isolandmarks extension combines this topological constraint with the preservation of selected metric relations.

These approaches differ in the structural information they assume. InvMap-VAE specifies a complete target arrangement of the latent points. Witness Simplicial VAE specifies selected topological structure without prescribing a particular geometric realization. Isolandmarks Witness Simplicial VAE additionally constrains distances between selected *landmark* points.

3.3.1 Preserving a Reference Representation: InvMap-VAE

InvMap-VAE assumes that a low-dimensional reference embedding of the data is available. This embedding may be provided by a dimensionality-reduction method such as Isomap [66] or UMAP [67], or by another source of prior information about the desired latent organization. Rather than learning the latent representation without such guidance, InvMap-VAE encourages the VAE to reproduce the prescribed embedding while retaining its generative encoder–decoder structure.

Let $\widetilde{Z} = \{\tilde{z}_i\}_{i=1}^N \subseteq \mathbb{R}^d$ denote the reference embedding of the dataset $X = \{x_i\}_{i=1}^N$, and let $Z = \{z_i\}_{i=1}^N \subseteq \mathbb{R}^d$ denote the corresponding latent representation produced by the VAE.

Throughout this section, for two vectors or matrices A and B of the same shape containing M scalar entries, their mean squared error is denoted by

$$\mathrm{MSE}(A, B) = \frac{1}{M} \sum_{r=1}^{M} (A_r - B_r)^2, \tag{3.3.2}$$

where r indexes the scalar entries. In particular, for $Z, \widetilde{Z} \in \mathbb{R}^{N\times d}$,

$$\mathrm{MSE}(Z, \widetilde{Z}) = \frac{1}{Nd} \sum_{i=1}^{N} \sum_{j=1}^{d} \left(z_{ij} - \tilde{z}_{ij}\right)^2 . \tag{3.3.3}$$

InvMap-VAE augments the VAE objective with this discrepancy,

$$\mathcal{L}_{\mathrm{InvMap}} = \mathcal{L}_{\mathrm{VAE}} + w_{\mathrm{IM}}\, \mathrm{MSE}(Z, \widetilde{Z}), \tag{3.3.4}$$

where $w_{\mathrm{IM}} \geq 0$ controls the strength of the constraint imposed by the reference embedding.

The reference representation therefore acts as a specific form of prior structural information: the desired coordinates are supplied directly, and the additional loss encourages the learned latent representation to follow them. If the reference embedding preserves a desired geometric or topological structure, InvMap-VAE provides a way of transferring this organization into a generative latent-variable model. In addition, the learned encoder can map new observations into the representation and the decoder provides a mapping from the latent representation back to the data space.

The method is consequently dependent on the quality and availability of the reference embedding. This motivates the second approach of **Paper A**, which introduces structural information without prescribing the complete latent configuration.

3.3.2 Topological Inductive Bias: Witness Simplicial VAE

Witness Simplicial VAE starts from the assumption that selected topological properties of the data should be retained in the latent representation. The challenge is that topological invariants such as *Betti numbers* are discrete, whereas neural networks are typically trained through differentiable objectives. The method addresses this by constructing a simplicial approximation of the input data and regularizing the encoder and decoder to behave approximately as *simplicial maps* on this structure. We provide below the notions of computational topology needed to understand the methods developed in **Paper A**. For a more comprehensive introduction to computational topology, we refer the reader to the theoretical background in the author's master's thesis [68] or to an introductory textbook on the topic, such as [69].

Simplices and simplicial complexes. Given $k+1$ affinely independent points $u_0, \ldots, u_k \in \mathbb{R}^D$, the k-simplex spanned by these vertices is their convex hull,

$$[u_0, \dots, u_k] = \left\{ \sum_{j=0}^{k} \lambda_j u_j \;\middle|\; \lambda_j \geq 0,\ \sum_{j=0}^{k} \lambda_j = 1 \right\}. \tag{3.3.5}$$

A 0-simplex is a point, a 1-simplex an edge, a 2-simplex a triangle, and so forth. A face of a simplex is obtained by taking the convex hull of a subset of its vertices.

A geometric simplicial complex K is a collection of simplices satisfying two conditions: every face of a simplex in K also belongs to K, and the intersection of two simplices is either empty or a common face. Simplicial complexes therefore provide finite combinatorial structures through which the topology of more complicated spaces can be studied [69].

Betti numbers and persistent homology. The topology represented by a simplicial complex can be characterized through its homology groups. For the purposes of this thesis, it is sufficient to consider their ranks, the *Betti numbers*. Informally, the pth Betti number β_p counts independent p-dimensional holes. In particular, β_0 counts connected components, β_1 counts independent one-dimensional cycles, and β_2 counts enclosed two-dimensional voids.

For a finite point cloud, however, the topology obtained from a simplicial complex depends on the scale at which neighboring observations are connected. This motivates the use of a *filtration*, a nested family of simplicial complexes $K_{\epsilon_1} \subseteq K_{\epsilon_2} \subseteq \cdots \subseteq K_{\epsilon_m}$, with $\epsilon_1 \leq \epsilon_2 \leq \cdots \leq \epsilon_m$.

A standard construction is the Vietoris–Rips complex. Given a scale ϵ, a simplex $[x_0, \dots, x_k]$ is included whenever $d_{\mathcal{X}}(x_i, x_j) \leq \epsilon$, for all i, j.

Increasing ϵ produces the Vietoris–Rips filtration. Persistent homology tracks the appearance and disappearance of homological features throughout such a filtration. A feature born at scale b and disappearing at scale d can be represented by a point (b, d) in a *persistence diagram*. Features that persist over a larger range of scales provide candidates for the topological structure to be retained.

Witness complexes. Complexes such as the Vietoris–Rips complex can contain a large number of simplices. Witness complexes instead construct simplices using only a selected subset of the observations as vertices, providing a more compact topological approximation [70]. Let $L = \{\ell_1, \dots, \ell_M\} \subset X$ be a set of *landmarks*. Following the construction used in **Paper A**, the remaining observations act as *witnesses*. For a radius $r > 0$, define $B_r(w) = \{x \in \mathcal{X} \mid d_{\mathcal{X}}(x, w) \leq r\}$. The witness complex is then

$$W_r(X, L) = \left\{ [u_0, \dots, u_k] \;\middle|\; \{u_0, \dots, u_k\} \subseteq L,\ \exists w \in X \text{ such that } u_j \in B_r(w)\ \forall j \right\}. \tag{3.3.6}$$

Thus, a collection of landmarks forms a simplex when there exists an observation that witnesses all of them within radius r. Only the landmarks appear as vertices of the complex, while the remaining observations provide information about which landmark relations should be included. An example is shown in Figure 3.3.1: although only a subset of the data points is used as vertices, the resulting complex can still capture topological structure of the full point cloud.

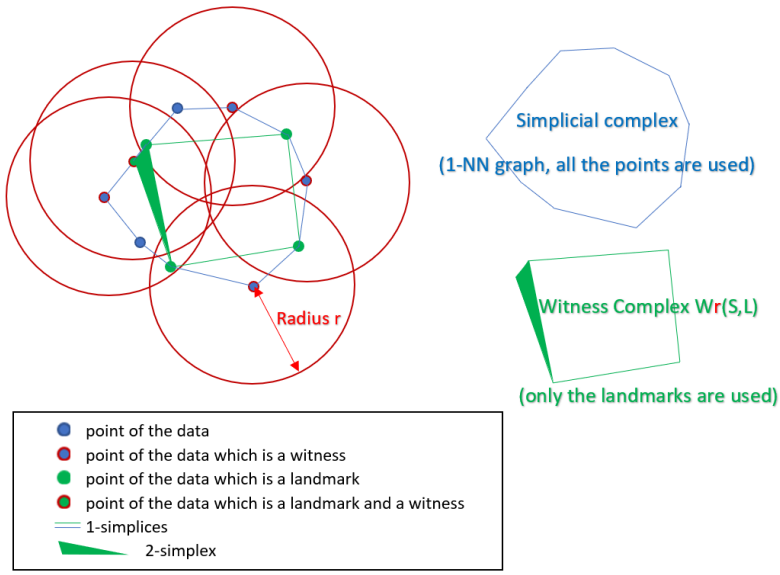


Figure 3.3.1: Example of a witness complex $W_r(S, L)$ constructed from a point cloud $S \subset \mathbb{R}^2$. A subset of the observations is selected as landmarks (green). A set of landmarks forms a simplex when they are jointly contained in a ball of radius r centered at a witness observation (red circles). The resulting witness complex, shown on the lower right, uses only the landmarks as vertices while exploiting the remaining observations to determine its simplices. Reproduced from **Paper A**.

The resulting complex depends on both the landmark set L and the scale r. Varying r produces a witness-complex filtration. In **Paper A**, this filtration is used to obtain a persistence diagram, from which a radius is selected such that the corresponding witness complex contains the topological features intended to be preserved. The number of landmarks introduces an additional trade-off: more landmarks can provide a more detailed approximation, but also increase the number of simplices entering the subsequent regularization.

Simplicial maps and simplicial regularization. Let K and K' be simplicial complexes. A map $f : \mathrm{Vert}(K) \to \mathrm{Vert}(K')$ is simplicial if the vertices of every simplex of K are mapped to the vertices of a simplex of K'. A simplicial map

induces a continuous map between the underlying topological spaces that is affine on each simplex. In particular, for $x = \sum_{j=0}^{k} \lambda_j u_j$, with $\lambda_j \geq 0$ and $\sum_{j=0}^{k} \lambda_j = 1$, the affine extension satisfies

$$f(x) = f\left(\sum_{j=0}^{k} \lambda_j u_j\right) = \sum_{j=0}^{k} \lambda_j f(u_j). \tag{3.3.7}$$

This property provides the motivation for the simplicial regularization introduced by Gallego-Posada [71] and adapted in **Paper A**. Rather than directly optimizing a discrete quantity such as a Betti number, which is not suitable as a standard differentiable learning objective, the encoder and decoder are encouraged to behave approximately as simplicial maps on the selected witness complex.

Let e_ϕ and d_θ denote the encoder and decoder mappings, respectively, and let $\sigma = [\sigma_0, \ldots, \sigma_k] \in K$ be a k-simplex of the selected witness complex. Barycentric coefficients are sampled according to the symmetric Dirichlet distribution

$$\lambda = (\lambda_0, \ldots, \lambda_k) \sim \mathrm{Dir}\left(\alpha \mathbf{1}_{k+1}\right), \tag{3.3.8}$$

where $\alpha > 0$. For $\alpha = 1$, as used in **Paper A**, this corresponds to uniform sampling over the simplex.

The encoder simplicial regularization is

$$\mathcal{L}_{\mathrm{SE}}(e_\phi, K, \alpha) = \sum_{\sigma \in K} \mathbb{E}_\lambda \left[\mathrm{MSE}\left(e_\phi\left(\sum_{j=0}^{k} \lambda_j \sigma_j\right), \sum_{j=0}^{k} \lambda_j e_\phi(\sigma_j) \right) \right]. \tag{3.3.9}$$

The two arguments of the MSE compare encoding a convex combination of the simplex vertices with taking the same convex combination after encoding the vertices. According to Equation (3.3.7), these quantities coincide for an affine map on the simplex.

Similarly, the decoder regularization is

$$\mathcal{L}_{\mathrm{SD}}(d_\theta, K, \alpha) = \sum_{\sigma \in K} \mathbb{E}_\lambda \left[\mathrm{MSE}\left(d_\theta\left(\sum_{j=0}^{k} \lambda_j e_\phi(\sigma_j)\right), \sum_{j=0}^{k} \lambda_j d_\theta(e_\phi(\sigma_j)) \right) \right]. \tag{3.3.10}$$

This term applies the same principle to the decoder: decoding a convex combination of encoded vertices is encouraged to agree with the corresponding convex combination after decoding the individual vertices.

The complete Witness Simplicial VAE objective is

$$\mathcal{L}_{\text{WSVAE}} = \mathcal{L}_{\text{VAE}} + w_{\text{SE}}\mathcal{L}_{\text{SE}} + w_{\text{SD}}\mathcal{L}_{\text{SD}}, \tag{3.3.11}$$

where $w_{\text{SE}}, w_{\text{SD}} \geq 0$ control the strengths of the encoder and decoder simplicial regularizations, respectively. The expectations in Equations (3.3.9) and (3.3.10) are approximated by Monte Carlo sampling during training.

The topological motivation follows from the relation between simplicial maps and homology. A simplicial map induces homomorphisms between the corresponding homology groups. Consequently, encouraging the encoder and decoder to behave approximately as simplicial maps provides a differentiable mechanism for favoring preservation of the homological structure represented by the selected witness complex.

This argument should be understood as a motivation rather than a guarantee of exact topological preservation. The witness complex is itself an approximation whose structure depends on the landmarks and selected filtration radius, and the simplicial conditions are imposed through finite regularization terms. The preservation of the targeted topology is therefore evaluated empirically in **Paper A** through the corresponding Betti numbers.

3.3.3 Combining Topological and Metric Constraints: Isolandmarks Witness Simplicial VAE

Preserving selected Betti numbers does not uniquely determine the geometry of a latent representation. Geometrically different configurations can share the same topology. The Isolandmarks extension of Witness Simplicial VAE therefore adds selected metric information to the topological constraint.

Let $L = \{\ell_1, \dots, \ell_M\}$ denote the landmarks used to construct the witness complex. Its 1-skeleton defines a weighted graph whose vertices are the landmarks and whose edge weights are their distances in the input space. Shortest-path distances on this graph provide approximations of the geodesic distances between landmark points, in a manner inspired by Isomap [66].

Let $D^* = \left[D^*_{ij}\right]_{i,j=1}^{M}$ denote the resulting matrix of approximate geodesic distances between landmarks in the input space, and let $D = \left[D_{ij}\right]_{i,j=1}^{M}$ denote the Euclidean distance matrix between their latent representations. The method compares the corresponding Isomap kernels

$$K(D) = -\frac{1}{2} H D^2 H, \qquad H = I - \frac{1}{M}\mathbf{1}\mathbf{1}^\top, \tag{3.3.12}$$

where the square in D^2 is applied element-wise.

The Isolandmarks Witness Simplicial VAE objective is

$$\mathcal{L}_{\text{IWSVAE}} = \mathcal{L}_{\text{WSVAE}} + w_{\text{iso}} \frac{\|K(D^*) - K(D)\|_F}{M}, \tag{3.3.13}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and $w_{\text{iso}} \geq 0$ controls the strength of the metric-preservation term.

Isolandmarks therefore combines two complementary structural constraints. The witness-complex regularization encourages preservation of selected topological structure, while the additional term encourages selected metric relations between landmarks to be retained. In the experiments of **Paper A**, this additional information was particularly useful for the Swiss-roll example, for which the simplicial regularization alone could preserve some desired topological features while still allowing an undesirable geometric overlap.

3.3.4 Experimental Observations

The experiments in **Paper A** illustrate both the benefits and the limitations of imposing selected structural constraints on the latent representation. On the synthetic open-cylinder and Swiss-roll datasets, InvMap-VAE produced latent representations that followed the prescribed Isomap or UMAP embeddings and retained their corresponding topological organization ($\beta_0 = 1$ and $\beta_1 = 1$ for the open-cylinder; $\beta_0 = 1$ and $\beta_1 = 0$ for the Swiss-roll). The method was also substantially less sensitive to neural-network initialization than the standard VAE.

Witness Simplicial VAE provided a different form of control without requiring a reference embedding. For the open-cylinder dataset, the desired Betti numbers were retained across the input and latent representations. The experiments also showed, however, that preserving selected homological features does not uniquely determine a satisfactory latent geometry. In particular, on the Swiss-roll dataset, Witness Simplicial VAE could produce an additional one-dimensional hole through an overlap in the latent representation, despite the simplicial regularization.

This observation motivated Isolandmarks Witness Simplicial VAE. By additionally preserving approximate geodesic distances between the landmark points, the method reduced this overlap and recovered a latent representation with the desired persistent Betti numbers. The experiments therefore illustrate a distinction between topological and metric preservation: topological constraints

can retain selected invariants while still allowing geometrically different representations, whereas additional metric information can further restrict the possible latent organization.

The results of **Paper A** therefore show that different forms of prior structural information constrain a representation at different levels. A reference embedding specifies a particular geometric organization, simplicial regularization targets selected topological structure, and the Isolandmarks extension supplements this with selected metric relations. This raises the broader question considered next: rather than constraining particular relationships within an otherwise Euclidean latent space, what happens when the geometry of the latent space itself is chosen according to the structure one expects to represent?

3.4 From Structural Constraints to Geometry as an Inductive Bias

The previous methods express structural assumptions through additional constraints on a Euclidean latent representation. A different possibility is to encode the assumption through the geometry of the latent space itself.

Suppose, for example, that prior knowledge does not specify a particular embedding, simplicial complex, or set of distances, but suggests that the organization of interest is hierarchical. Rather than prescribing where points should be located, one can choose a geometry whose metric properties are compatible with branching structure. Hyperbolic geometry provides such a choice.

This is a qualitatively different inductive bias. Choosing a hyperbolic latent space does not force the learned points to form a particular tree. It determines the geometric rules under which the representation is constructed: distances, geodesics, volumes, neighborhoods, probability distributions, and optimization all become hyperbolic. Hierarchical organization can then be represented within a space whose geometry is naturally compatible with the exponential growth characteristic of branching tree structures.

The distinction is particularly visible for VAEs. In a conventional VAE, the latent variable is usually modeled in Euclidean space using Gaussian distributions. If the latent space is instead hyperbolic, both the probabilistic prior and approximate posterior must be defined according to that geometry. Several analogues of Gaussian distributions have therefore been developed for hyperbolic spaces, including Riemannian normal and wrapped normal distributions [30, 29]. Thus, in the VAE setting, a hyperbolic geometric inductive bias affects not only the distance function but also the probabilistic model employed for latent variables.

For distance-based representation learning, the same choice appears in a dif-

ferent form. If a representation is constructed by preserving pairwise relations, choosing a hyperbolic latent space means that discrepancies are measured using hyperbolic rather than Euclidean distances. This is the principle used by the hyperbolic metric multidimensional scaling construction considered later in this chapter and in **Paper C**.

The common idea is therefore not a particular learning algorithm. It is that the geometry of a latent representation should be chosen deliberately according to the structure one expects the representation to express.

3.5 Hyperbolic Geometry

3.5.1 Negative Curvature and Hierarchical Structure

Exponential growth. Hyperbolic geometry is a non-Euclidean geometry of constant negative curvature. It is particularly relevant for representing data with hierarchical or tree-like structure. One reason is the way in which available space grows with distance from the origin.

For curvature normalized to -1, the circumference and area of a disk of radius r in the hyperbolic plane are respectively

$$C_{\mathbb{H}}(r) = 2\pi \sinh(r), \qquad A_{\mathbb{H}}(r) = 4\pi \sinh^2(r/2), \tag{3.5.14}$$

with sinh denoting the hyperbolic sine function defined on the real numbers by

$$\sinh(r) = \frac{e^r - e^{-r}}{2}. \tag{3.5.15}$$

Both the circumference and area of a hyperbolic disk therefore grow asymptotically proportionally to e^r. In the Euclidean plane, by contrast, the circumference and area of a Euclidean disk grow polynomially with the radius:

$$C_{\mathbb{R}^2}(r) = 2\pi r, \qquad A_{\mathbb{R}^2}(r) = \pi r^2. \tag{3.5.16}$$

This exponential growth in hyperbolic geometry is naturally compatible with branching structures. For a tree with approximately constant branching factor $b > 1$, the number of leaves or vertices at depth k grows proportionally to $b^k = e^{k \ln b}$. Hyperbolic space can therefore accommodate the exponentially increasing number of branches by providing exponentially more space at increasing radii. This geometric correspondence motivates the common interpretation of hyperbolic space as a continuous analogue of a tree [24].

Tree metrics and metric embeddings. The connection can be made more precise by considering trees as metric spaces. Let $T = (V, E, w)$ be a tree with positive edge weights $w : E \to \mathbb{R}_{>0}$. Since a tree contains a unique simple path between any two vertices $u, v \in V$, this path is also the unique shortest path when the edge weights are positive. The corresponding *tree metric* is therefore

$$d_T(u, v) = \sum_{e \in P(u,v)} w(e), \tag{3.5.17}$$

where $P(u, v) \subseteq E$ denotes the set of edges along this unique shortest path connecting u and v in the tree. Thus, embedding a tree into another metric space (Y, d_Y) means assigning each vertex $u \in V$ to a point $f(u) \in Y$ while attempting to preserve these tree distances.

An isometric embedding preserves them exactly,

$$d_Y(f(u), f(v)) = d_T(u, v) \qquad \text{for all } u, v \in V. \tag{3.5.18}$$

More generally, the deformation of the distances can be measured by the *distortion*

$$\operatorname{dist}(f) = \left(\sup_{u \neq v} \frac{d_Y(f(u), f(v))}{d_T(u, v)} \right) \left(\sup_{u \neq v} \frac{d_T(u, v)}{d_Y(f(u), f(v))} \right). \tag{3.5.19}$$

This definition is invariant under a common rescaling of the embedding. Distortion equal to one therefore corresponds to preservation of all pairwise distances up to a common scale.

From paths to general trees. Some trees admit distortion-free Euclidean embeddings. Consider, for example, the path with unit-weight edges $v_0 - v_1 - \cdots - v_{n-1}$. Its tree metric is simply $d_T(v_i, v_j) = |i - j|$, and the mapping $f(v_i) = i \in \mathbb{R}$ is an isometric embedding. Thus, a path tree can be represented without distortion in only one Euclidean dimension.

This simple example shows that some tree metrics can be represented exactly in Euclidean space. More generally, two aspects are relevant when embedding a tree metric: the distortion of its pairwise distances and the dimension of the target space. Ideally, one would preserve the tree distances with as little distortion as possible while using a low-dimensional representation. For a path, both objectives can be achieved simultaneously, since it admits an isometric embedding in one dimension. For more strongly branching trees, however, these objectives become harder to satisfy simultaneously. A standard example

is the complete binary tree, which illustrates the worst-case distortion of Euclidean tree embeddings.

Euclidean embeddings of trees. For context, classical metric-embedding results show that every n-point metric space can be embedded into an $O(\log n)$-dimensional Euclidean space with $O(\log n)$ distortion [72, 73]. Here, $O(g(n))$ denotes an asymptotic upper bound up to a constant factor, $\Omega(g(n))$ a corresponding lower bound, and $\Theta(g(n))$ indicates matching upper and lower bounds.

Trees form a more structured class of metric spaces and admit substantially better Euclidean embeddings. When the Euclidean dimension is unrestricted, every n-point tree metric can be embedded into Euclidean space with distortion $O\left(\sqrt{\log \log n}\right)$. This bound is asymptotically tight: there exist trees for which every Euclidean embedding has distortion $\Omega\left(\sqrt{\log \log n}\right)$. Hence, the optimal worst-case distortion over n-point tree metrics is $\Theta\left(\sqrt{\log \log n}\right)$ [74]. In other words, every tree with n vertices can be embedded in Euclidean space with distortion of order $\sqrt{\log \log n}$ or smaller, while for some trees the distortion must be at least of this order.

A standard example underlying this lower bound is the complete binary tree. A complete binary tree of depth h is a rooted tree in which every non-leaf vertex has two children and all leaves are at depth h. It contains $n = 2^{h+1} - 1$ vertices. Its optimal Euclidean distortion is $\Theta(\sqrt{\log h})$, which, since h grows logarithmically with n, is equivalently $\Theta\left(\sqrt{\log \log n}\right)$ when expressed in terms of the number of vertices [75].

The dimension of the target space introduces a further constraint. The bounds above allow the Euclidean dimension to be unrestricted. If instead the target dimension d is fixed, there exist n-point tree metrics whose embedding into $\mathbb{R}^d$ requires distortion $\Omega\left(n^{1/d}\right)$. This lower bound can already be obtained from a star tree, that is, a tree with one central vertex connected directly to all other vertices [74]. In particular, for the Euclidean plane, $d = 2$, the lower bound becomes $\Omega(\sqrt{n})$.

Low-distortion embeddings in the hyperbolic plane. Hyperbolic geometry provides a stronger result for tree embeddings. Sarkar showed that any weighted tree can be embedded into the two-dimensional hyperbolic plane $\mathbb{H}^2$ such that the weight of every edge is realized by the corresponding hyperbolic distance and the distortion of distances between non-adjacent vertices is smaller than $1 + \varepsilon$ for any prescribed $\varepsilon > 0$ [27]. Hence, arbitrarily small distortion can be achieved for any weighted tree while using only two hyperbolic dimensions.

The contrast is particularly relevant to low-dimensional representation learn-

ing. In unrestricted-dimensional Euclidean space, the optimal worst-case tree distortion still grows as $\Theta(\sqrt{\log \log n})$, and in a fixed Euclidean dimension it can grow considerably faster. In the hyperbolic plane, by contrast, Sarkar's construction provides distortion arbitrarily close to one independently of the number of vertices. This does not imply that every hierarchical dataset is intrinsically hyperbolic, but it provides a strong geometric motivation for choosing a hyperbolic latent space when tree-like relations are expected to be important.

Tree-likeness beyond exact trees. The relation between negative curvature and trees extends beyond exact tree metrics. In hyperbolic group theory, Gromov introduced δ-hyperbolicity as a metric notion of negative curvature [76, 77]. The definition applies to a *geodesic metric space*, that is, a metric space in which any two points can be connected by a shortest path, called a geodesic. Given three points, the three geodesics connecting them pairwise form a *geodesic triangle*.

A geodesic metric space is δ-hyperbolic if every side of every geodesic triangle lies within distance δ of the union of the other two sides. The parameter δ therefore measures how thin these triangles are. Metric trees are 0-hyperbolic: the shortest paths between three vertices overlap and form a tripod, so each side is contained in the union of the other two. In this sense, δ-hyperbolicity provides a way of describing tree-like behavior in metric spaces that are not themselves trees.

The notion can also be applied to finite graphs, where estimates of δ-hyperbolicity can be used to quantify their tree-like character [78, 79]. Together with exponential volume growth and the low-distortion embedding results discussed above, this provides further motivation for using negative curvature as an inductive bias for hierarchical and tree-like representations.

3.5.2 Models of Hyperbolic Space

A hyperbolic space can be represented using several equivalent models. These models describe the same intrinsic hyperbolic geometry but use different coordinates, making different geometric operations convenient in different settings. The Poincaré ball, Klein–Beltrami ball, and Lorentz hyperboloid are the models most relevant to this thesis.

Before introducing them, we briefly recall the Riemannian framework underlying these representations. A smooth d-dimensional manifold $\mathcal{M}$ can be viewed locally as a d-dimensional Euclidean space. At each point $x \in \mathcal{M}$, the tangent space $T_x\mathcal{M}$ is a d-dimensional vector space containing the possible infinitesimal directions at x. A tangent vector can, for example, be understood as the velocity of a smooth curve passing through x.

Formally, a *Riemannian manifold* is a pair $(\mathcal{M}, g)$, where $\mathcal{M}$ is a smooth manifold and g assigns to every point $x \in \mathcal{M}$ a positive-definite inner product $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \longrightarrow \mathbb{R}$, varying smoothly with x. The collection $g = \{g_x\}_{x\in\mathcal{M}}$ is called the *Riemannian metric*. It defines the norm of a tangent vector $v \in T_x\mathcal{M}$ as

$$\|v\|_x = \sqrt{g_x(v, v)}, \tag{3.5.20}$$

and the angle between two non-zero tangent vectors $u, v \in T_x\mathcal{M}$ by

$$\cos \angle_x(u, v) = \frac{g_x(u, v)}{\|u\|_x \|v\|_x}. \tag{3.5.21}$$

The metric also determines lengths and distances on the manifold. For a smooth curve $\gamma : [a, b] \to \mathcal{M}$, its Riemannian length is

$$L_g(\gamma) = \int_a^b \sqrt{g_{\gamma(t)}\left(\dot\gamma(t), \dot\gamma(t)\right)}\, \mathrm{d}t. \tag{3.5.22}$$

The distance between two points is then defined as the infimum of the lengths of all curves connecting them,

$$d_{\mathcal{M}}(x, y) = \inf_{\substack{\gamma(a)=x \\ \gamma(b)=y}} L_g(\gamma). \tag{3.5.23}$$

Geodesics can be understood as the counterparts of straight lines on a curved space: locally, they follow length-minimizing paths determined by the Riemannian metric. In hyperbolic space, any two points are connected by a unique minimizing geodesic, whose length therefore realizes the distance in Equation (3.5.23). The different models below represent these same intrinsic geodesics and distances using different coordinates.

Poincaré ball. For curvature normalized to -1, the d-dimensional Poincaré ball is

$$\mathbb{P}^d = \left\{x \in \mathbb{R}^d \ : \ \|x\| < 1\right\}, \tag{3.5.24}$$

where $\|\cdot\|$ denotes the Euclidean norm. Since $\mathbb{P}^d$ is an open subset of $\mathbb{R}^d$, its tangent space can be identified with $\mathbb{R}^d$ at every point.

Let

$$\lambda_x = \frac{2}{1 - \|x\|^2} \tag{3.5.25}$$

denote the conformal factor. The Poincaré Riemannian metric is

$$g_x^{\mathbb{P}} = \lambda_x^2 g^{\mathrm{E}} = \left(\frac{2}{1 - \|x\|^2} \right)^2 g^{\mathrm{E}}, \tag{3.5.26}$$

where g^{E} denotes the Euclidean metric. At every point, it is given by the ordinary Euclidean inner product $\langle \cdot, \cdot \rangle_{\mathrm{E}}$, so that for $u, v \in \mathbb{R}^d$,

$$g^{\mathrm{E}}(u, v) = \langle u, v \rangle_{\mathrm{E}} = \sum_{i=1}^{d} u_i v_i. \tag{3.5.27}$$

Since $T_x\mathbb{P}^d$ can be identified with $\mathbb{R}^d$, for tangent vectors $u, v \in T_x\mathbb{P}^d$ we equivalently have

$$g_x^{\mathbb{P}}(u, v) = \lambda_x^2 \langle u, v \rangle_{\mathrm{E}}. \tag{3.5.28}$$

The metric is therefore *conformal*: at any point x, all tangent directions are scaled by the same positive factor λ_x. Consequently, the hyperbolic angle between two curves at their point of intersection is equal to the Euclidean angle between their tangent vectors in the Poincaré ball, while lengths are locally scaled by λ_x.

The corresponding geodesic distance between $x, y \in \mathbb{P}^d$ is

$$d_{\mathbb{P}}(x, y) = \operatorname{arcosh}\left(1 + \frac{2\|x - y\|^2}{(1 - \|x\|^2)(1 - \|y\|^2)} \right), \tag{3.5.29}$$

where the inverse hyperbolic cosine is defined for $t \geq 1$ by

$$\operatorname{arcosh}(t) = \log\left(t + \sqrt{t^2 - 1} \right). \tag{3.5.30}$$

Although the model occupies a bounded Euclidean ball, its boundary is at infinite hyperbolic distance. Hyperbolic distances therefore increase rapidly as points approach the boundary. Geodesics are represented by Euclidean diameters or by circular arcs that intersect the boundary orthogonally. Together with conformality, this bounded representation makes the Poincaré model particularly intuitive for visualizing low-dimensional hyperbolic representations [25].

The two-dimensional Poincaré disk is used in **Paper C** both for constructing the olfactory representations and for their subsequent radial, directional, and regional analysis.

Klein–Beltrami ball. The Klein–Beltrami model is defined on the same open Euclidean unit ball,

$$\mathbb{K}^d = \{x \in \mathbb{R}^d \ : \ \|x\| < 1\}, \tag{3.5.31}$$

but uses different coordinates and a different Riemannian metric. Unlike the Poincaré model, the Klein model is not conformal. Its main advantage for the computational geometry developed later in this thesis is that hyperbolic geodesics are represented by Euclidean straight-line chords.

For $x, y \in \mathbb{K}^d$, the hyperbolic distance can be written as

$$d_{\mathbb{K}}(x, y) = \operatorname{arcosh}\left(\frac{1 - \langle x, y\rangle_{\mathrm{E}}}{\sqrt{(1 - \|x\|^2)(1 - \|y\|^2)}}\right). \tag{3.5.32}$$

The Poincaré and Klein coordinates of the same hyperbolic point are related by

$$k = \frac{2p}{1 + \|p\|^2}, \qquad p \in \mathbb{P}^d, \quad k \in \mathbb{K}^d. \tag{3.5.33}$$

This allows a representation obtained in Poincaré coordinates to be converted to Klein coordinates when straight geodesics or projective constructions are computationally advantageous.

The Klein–Beltrami model plays an important role in **Papers B, D, and E**. In particular, it supports efficient constructions of hyperbolic Voronoi and Delaunay structures through their relation to Euclidean power diagrams. These constructions are introduced where they are used in Chapters 4 and 5.

Lorentz hyperboloid. The Lorentz model realizes hyperbolic space as a hyperboloid in $\mathbb{R}^{d+1}$. Its ambient space is equipped with the Lorentzian bilinear form

$$\langle x, y\rangle_{\mathrm{L}} = -x_0 y_0 + \sum_{j=1}^{d} x_j y_j. \tag{3.5.34}$$

Unlike the Euclidean inner product, this bilinear form is not positive definite. The d-dimensional Lorentz hyperboloid is

$$\mathbb{L}^d = \{x \in \mathbb{R}^{d+1} : \langle x, x\rangle_{\mathrm{L}} = -1,\ x_0 > 0\}. \tag{3.5.35}$$

At a point $x \in \mathbb{L}^d$, the tangent space is

$$T_x\mathbb{L}^d = \{v \in \mathbb{R}^{d+1} : \langle x, v\rangle_{\mathrm{L}} = 0\}. \tag{3.5.36}$$

Although the ambient Lorentzian form is indefinite, its restriction to the tangent space is positive definite and defines the Riemannian metric of the hyperboloid,

$$g^{\mathbb{L}}_x(u, v) = \langle u, v\rangle_{\mathrm{L}}, \qquad u, v \in T_x\mathbb{L}^d. \tag{3.5.37}$$

The resulting geodesic distance is

$$d_{\mathbb{L}}(x, y) = \operatorname{arcosh}\left(-\langle x, y\rangle_{\mathrm{L}}\right). \tag{3.5.38}$$

The Lorentz representation is convenient for learning and Riemannian computations and has been used for large-scale hierarchical embeddings [26]. In particular, its distance function avoids the fraction appearing in Equation (3.5.29), whose denominator approaches zero near the boundary of the Poincaré ball. It therefore avoids this particular source of numerical instability [26].

Conversion from Lorentz to Klein coordinates is particularly simple:

$$(x_0, x_1, \dots, x_d) \longmapsto \left(\frac{x_1}{x_0}, \dots, \frac{x_d}{x_0}\right). \tag{3.5.39}$$

The three representations above describe the same intrinsic hyperbolic space. A point can therefore be expressed in the model most convenient for a particular operation without changing its underlying hyperbolic relations. In this thesis, this equivalence is particularly useful for moving between models suited to representation learning, visualization, and computational geometry.

3.5.3 Riemannian Operations and Optimization

The Riemannian structure introduced above also determines how optimization is performed in hyperbolic space. In Euclidean optimization, an update can be added directly to the current point. On a manifold, however, optimization

directions belong to the tangent space at the current point, and an update must be mapped back onto the manifold.

Let $F : \mathcal{M} \to \mathbb{R}$ be a differentiable function on a Riemannian manifold $(\mathcal{M}, g)$. At a point $z \in \mathcal{M}$, the change of F along a tangent direction $v \in T_z\mathcal{M}$ is described by the *directional derivative*. More precisely, if $\gamma(t)$ is any smooth curve satisfying $\gamma(0) = z$ and $\dot{\gamma}(0) = v$, then

$$\mathrm{d}F_z[v] = \frac{\mathrm{d}}{\mathrm{d}t} F\left(\gamma(t)\right)\Big|_{t=0}. \tag{3.5.40}$$

The quantity $\mathrm{d}F_z[v]$ therefore describes how rapidly F changes at z when moving in the direction v.

In Euclidean space, the directional derivative can be written in terms of the Euclidean gradient as

$$\mathrm{d}F_z[v] = \langle \nabla_{\mathbb{E}} F(z), v \rangle_{\mathbf{E}}. \tag{3.5.41}$$

On a Riemannian manifold, the Euclidean inner product is replaced by the local Riemannian metric g_z. The *Riemannian gradient* $\nabla_{\mathcal{M}} F(z) \in T_z\mathcal{M}$ is defined as the unique tangent vector satisfying

$$g_z\left(\nabla_{\mathcal{M}} F(z), v\right) = \mathrm{d}F_z[v], \qquad \forall v \in T_z\mathcal{M}. \tag{3.5.42}$$

The existence and uniqueness of this vector follow from the Riesz representation theorem. The differential $\mathrm{d}F_z$ is a linear functional on the finite-dimensional tangent space $T_z\mathcal{M}$, while g_z defines an inner product on this space. The Riesz representation theorem therefore guarantees the existence of a unique tangent vector whose inner product with any $v \in T_z\mathcal{M}$ is equal to $\mathrm{d}F_z[v]$.

The definition also explains why $\nabla_{\mathcal{M}} F(z)$ determines the direction of steepest increase. To compare directions independently of their magnitude, we restrict to tangent vectors of unit Riemannian norm; otherwise, the directional derivative could be increased arbitrarily by simply scaling the vector. Among all tangent vectors v of unit Riemannian norm, $\|v\|_z = 1$, the Cauchy–Schwarz inequality gives

$$\mathrm{d}F_z[v] = g_z\left(\nabla_{\mathcal{M}} F(z), v\right) \leq \left\|\nabla_{\mathcal{M}} F(z)\right\|_z. \tag{3.5.43}$$

Equality is attained when v points in the same direction as $\nabla_{\mathcal{M}} F(z)$. Hence, the Riemannian gradient determines the local direction of steepest increase of

F with respect to the metric g, while $-\nabla_{\mathcal{M}} F(z)$ determines the direction of steepest decrease.

In local coordinates, the Riemannian metric can be represented by a positive-definite matrix $G(z)$ such that

$$g_z(u, v) = u^\top G(z) v. \tag{3.5.44}$$

Combining this expression with the definition of the Riemannian gradient gives

$$\nabla_{\mathcal{M}} F(z) = G(z)^{-1} \nabla_{\mathbb{E}} F(z). \tag{3.5.45}$$

Thus, the Riemannian gradient is not generally the ordinary Euclidean gradient: the metric modifies the gradient according to the local geometry of the manifold.

For the Poincaré ball, the metric tensor is $G(x) = \lambda_x^2 I$. Hence,

$$\nabla_{\mathbb{P}} F(x) = \lambda_x^{-2} \nabla_{\mathbb{E}} F(x) = \left(\frac{1 - \|x\|^2}{2} \right)^2 \nabla_{\mathbb{E}} F(x). \tag{3.5.46}$$

For $v \in T_z \mathcal{M}$, the initial conditions

$$\gamma_v(0) = z, \qquad \dot{\gamma}_v(0) = v \tag{3.5.47}$$

determine a unique geodesic γ_v locally. In hyperbolic space, geodesic completeness ensures that this geodesic is defined for all $t \in \mathbb{R}$. The *exponential map* at z is then defined by

$$\exp_z(v) = \gamma_v(1). \tag{3.5.48}$$

It maps the tangent vector v back to the manifold by following the geodesic determined by the initial position z and velocity v. In hyperbolic space, the norm $\|v\|_z$ is equal to the geodesic distance from z to $\exp_z(v)$.

Conversely, the *logarithmic map* maps a point on the manifold back to a tangent vector at z. It is locally the inverse of the exponential map,

$$\log_z(y) = v \quad \iff \quad \exp_z(v) = y. \tag{3.5.49}$$

The Riemannian gradient and the exponential map together provide the basic ingredients for extending gradient-based optimization to a manifold. The gradient $\nabla_{\mathcal{M}} F(z_t)$ gives a descent direction in the tangent space $T_{z_t}\mathcal{M}$, and the exponential map follows this direction on the manifold. A *Riemannian gradient-descent* step can therefore be written as

$$z_{t+1} = \exp_{z_t}\left(-\eta_t \nabla_{\mathcal{M}} F(z_t)\right), \tag{3.5.50}$$

where $\eta_t > 0$ is the step size [80]. In practice, the exponential map may also be replaced by a simpler *retraction* [81]. A retraction maps the tangent-space update back to the manifold while locally approximating the exponential map. It is useful when explicit evaluation of the exponential map is unnecessary or computationally more expensive. This is the approach employed during the geometric refinement in **Paper E**.

Optimization methods that maintain additional tangent-space quantities require one further operation. Tangent vectors at different points belong to different tangent spaces, so a vector in $T_x\mathcal{M}$ cannot in general be directly reused in $T_y\mathcal{M}$. *Parallel transport* provides a way to move such vectors between tangent spaces while preserving the Riemannian inner product. Along the geodesic from x to y, it defines a map

$$\mathcal{P}_{x\to y} : T_x\mathcal{M} \to T_y\mathcal{M}, \tag{3.5.51}$$

satisfying

$$g_y\left(\mathcal{P}_{x\to y}(u), \mathcal{P}_{x\to y}(v)\right) = g_x(u, v). \tag{3.5.52}$$

This is useful, for example, for transferring momentum between successive tangent spaces in adaptive optimization methods. Riemannian variants of methods such as Adam combine these geometric operations with adaptive first-order optimization [82]. In **Paper C**, optimization on the Poincaré disk employs Riemannian Adam.

3.6 Hyperbolic Representation Learning

Hyperbolic geometry has been used in different ways to construct and study low-dimensional representations of hierarchical data. Early work considered the visualization of large hierarchies in biology [83, 84]. A related line of work considered geometric descriptions of complex networks. Serrano et al. proposed that complex networks may be organized by an underlying hidden metric space, in which distances between nodes influence their probability of connection [85]. Krioukov et al. subsequently developed an explicit hyperbolic

framework for complex networks such as the Internet [24, 60]. Hyperbolic mapping methods were also developed for wireless ad hoc networks and sensor nets [59].

Hyperbolic multidimensional scaling provides another approach, constructing a low-dimensional hyperbolic representation from pairwise distances [86, 87]. **Paper C** follows this approach to obtain a two-dimensional representation of olfactory perceptual data in the Poincaré disk. Other dimensionality-reduction methods for hyperbolic data include variants of PCA and t-SNE [57, 58].

Poincaré embeddings provided an influential approach for learning hyperbolic representations of symbolic data whose relations may exhibit a latent hierarchical structure [25]. Nickel and Kiela extended this framework to the Lorentz model, which can provide favorable optimization properties [26].

Poincaré Maps later adapted hyperbolic embedding to single-cell data by constructing pairwise relations from a nearest-neighbor graph and embedding the cells in the Poincaré disk [28]. Representations obtained with Poincaré Maps are used in **Papers B**, **D**, and **E**: **Paper B** analyzes relations within such representations, while **Papers D** and **E** use them as input for extracting parsimonious tree structures.

Hyperbolic geometry has also been incorporated into neural-network architectures. Hyperbolic neural networks extend standard neural operations to hyperbolic space [88], allowing hyperbolic-valued representations to be processed directly within neural networks.

Related developments introduced hyperbolic Variational Auto-Encoders [30, 29]. In this case, the latent variable lies on a hyperbolic manifold, such as the Poincaré ball or the Lorentz hyperboloid, and the prior and approximate posterior must be defined accordingly. Different constructions have been proposed, including Riemannian or wrapped normal distributions. Mixed-curvature VAEs generalize this setting by defining the latent space as a product of constant-curvature spaces, with different components chosen from Euclidean, spherical, and hyperbolic geometries [89]. **Paper B** also considers hyperbolic latent representations obtained from a hyperbolic VAE and analyzes their geometric organization after training.

More recently, hyperbolic operations and representations have been explored in attention and Transformer architectures [90, 91, 92, 93], including recent work on large language and foundation models [94, 95]. These developments further broaden the range of learned hyperbolic representations on which geometric analysis may be performed.

3.7 From Hyperbolic Representation Learning to Geometric Inference

The methods considered in this chapter introduce structure into a representation in different ways. In **Paper A**, InvMap-VAE uses a prescribed reference embedding, Witness Simplicial VAE targets selected topological structure, and its Isolandmarks extension additionally preserves selected metric relations. Hyperbolic representation learning instead specifies a geometry of the representation space adapted to hierarchical data, without prescribing a particular configuration of the embedded points.

These choices concern the construction of the representation. The main focus of the included papers (**Papers B–E**) is what can be inferred once a structured hyperbolic representation is available. **Papers B** and **C** analyze geometric organization and relate it to properties of the represented data or model, while **Papers D** and **E** aim to extract parsimonious tree structures that can subsequently be interpreted.

Given a hyperbolic representation, Chapter 4 concerns its geometric analysis and interpretation, while Chapter 5 considers the extraction and interpretation of hierarchical structures via heuristic hyperbolic Steiner trees. Together, these two directions constitute the hyperbolic geometric inference developed in this thesis and form its main focus.

4 Analyzing and Interpreting Geometric Structures in Hyperbolic Representations

The previous chapter considered how structures can be preserved in a latent representation, either through explicit constraints or through the choice of its geometry. Once a hyperbolic latent representation is available, the focus shifts to what structural information can be inferred and interpreted from its geometry.

This chapter addresses **RQ1**:

> *How can geometric organization in hyperbolic latent representations be analyzed and related to meaningful properties of the represented data or of the model that produced the representation?*

Papers B and C address this question from complementary perspectives. **Paper B** considers geometric relations between predefined subsets of a hyperbolic representation, whereas **Paper C** studies how properties of the represented observations are organized across the representation as a whole.

Contributions of Papers B and C. **Paper B** contributes Hyperbolic Delaunay Geometric Alignment (HyperDGA), a geometry-aware score for comparing two predefined subsets of a hyperbolic representation. The score is defined from the connectivity of the hyperbolic Delaunay graph constructed on their union, through the proportion of edges connecting points from different subsets. The Delaunay graph is computed in the Klein–Beltrami model using its relation to Euclidean power diagrams. HyperDGA is experimentally validated on

both synthetic and real-world biological data, and compared with hyperbolic versions of geometric distances between sets. The experiments show that HyperDGA responds to controlled perturbations of the represented data, can be used to follow changes in the latent representation of a hyperbolic VAE during training, and captures relations between represented single-cell biological populations. **Paper C** contributes a complementary framework for analyzing the geometric organization of both continuous and binary descriptor data of human olfactory perception in hyperbolic representations. Descriptor profiles are embedded in the Poincaré disk using hyperbolic metric multidimensional scaling, with distance preservation evaluated before the geometry is interpreted. The resulting representation is decomposed into radial and angular organization. Hyperbolic radius is related to dataset-specific entropy measures that summarize global properties of descriptor profiles. Angular organization is then treated according to the data type: continuous descriptors are analyzed through dominant directions obtained by tangent-space regression, whereas binary descriptors are characterized through localized regions estimated by hyperbolic kernel density estimation. Restricted permutation tests and robustness analyses are employed to assess these geometric patterns. In the continuous Sagar dataset, the results indicate that diffuse descriptor profiles lie closer to the center and more concentrated profiles closer to the boundary, while some specific descriptors exhibit strong directional trends. In the binary GoodScents–Leffingwell dataset, the results suggest that molecules assigned a larger number of odor descriptors occupy more peripheral regions of the disk, while molecules closer to the center exhibit more diffuse profiles across orthogonal modes of variation. In addition, related binary descriptors occupy coherent high-density regions, with broad odor families and finer subcategories appearing in distinct angular sectors of the disk.

The next chapter will address **RQ2**, where geometric inference instead takes the form of extracting an explicit parsimonious tree from the latent representation and interpreting properties derived from that tree.

4.1 Geometric Alignment Between Hyperbolic Subsets: Paper B

A common problem in representation analysis is to compare two groups of points embedded in the same space. Several methods have approached this problem through the geometric or topological organization of the representations. Geometry Score [96] uses a topological approximation of the datasets, while Improved Precision and Recall [97] approximates their support through unions of local neighborhoods. Geometric Component Analysis [98] instead constructs an ϵ-proximity graph and derives local and global measures from its connected components. The choice of a fixed neighborhood scale can be problematic when the representation contains regions of different densities. Delaunay Component Analysis (DCA) [99] addresses this issue by replacing the fixed-scale proximity graph with the Delaunay graph, whose connectivity

is determined by the spatial configuration of the points.

These methods are formulated for Euclidean representation spaces and cannot be applied directly when distances and neighborhoods are defined by hyperbolic geometry. **Paper B** addresses this gap by introducing Hyperbolic Delaunay Geometric Alignment (HyperDGA), a geometric alignment score based on the hyperbolic Delaunay graph inspired by DCA.

4.1.1 Hyperbolic Delaunay Graphs

Consider two finite subsets of a hyperbolic latent representation, $A, B \subseteq Z \subseteq \mathbb{H}^d$, and let $P = A \cup B$. For each point $p \in P$, its hyperbolic Voronoi cell is

$$V(p) = \{x \in \mathbb{H}^d \mid d_{\mathbb{H}}(x,p) \leq d_{\mathbb{H}}(x,q) \text{ for all } q \in P\}. \tag{4.1.1}$$

The Delaunay complex is the nerve of this Voronoi decomposition, and its 1-skeleton defines the Delaunay graph. Two points are therefore connected when their Voronoi cells intersect. Unlike a fixed-radius graph, the resulting neighborhood structure is determined by the spatial configuration of the complete point set P.

Computing this construction directly in hyperbolic space is less straightforward than in the Euclidean case. **Paper B** therefore uses the Klein–Beltrami model $\mathbb{K}^d$. As introduced in the previous chapter, the Klein and Poincaré models represent the same hyperbolic geometry and points can be transformed between them. In the Klein model, hyperbolic Voronoi cells can be obtained from a Euclidean power diagram [100]. This allows standard computational geometry algorithms to be used, after removing intersections that lie outside the Klein ball and therefore do not correspond to valid hyperbolic Delaunay relations.

For the analysis that follows, only the resulting graph connectivity is required. The Delaunay graph therefore acts as a geometry-dependent neighborhood graph through which the relative organization of A and B can be examined.

4.1.2 Hyperbolic Delaunay Geometric Alignment

The construction and interpretation of the HyperDGA score are illustrated in Figure 4.1.1 below.

Let $\mathcal{E}$ denote the edges of the Delaunay graph constructed on $A \cup B$. An edge is called *heterogeneous* when its two endpoints belong to different subsets. Denoting the set of such edges by $\tilde{\mathcal{E}} \subseteq \mathcal{E}$, **Paper B** defines

$$\text{HyperDGA}(A,B) = 1 - \frac{|\tilde{\mathcal{E}}|}{|\mathcal{E}|}. \tag{4.1.2}$$

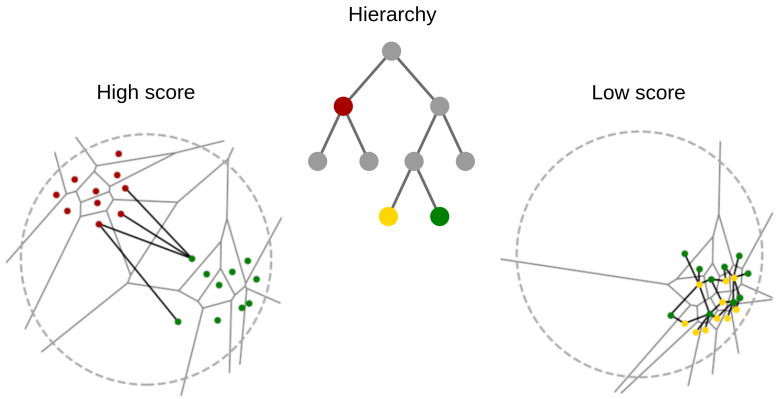


Figure 4.1.1: Illustration of HyperDGA for two subsets of a hyperbolic representation, adapted from **Paper B**. The Delaunay graph is constructed on the union of the subsets, and edges joining points from different subsets are heterogeneous (shown in black). Strong intermixing produces more heterogeneous edges and therefore a lower HyperDGA score, while separated subsets produce a higher score. The hyperbolic Voronoi diagrams are shown in gray on top of the Poincaré disks.

The score has a direct geometric interpretation. When A and B occupy similar regions of the representation, many local Delaunay neighborhoods contain points from both sets and the number of heterogeneous edges is large. HyperDGA is therefore low. Conversely, when the two sets occupy more separated regions, heterogeneous connections become less frequent and the score increases.

Figure 4.1.1 illustrates this idea for hierarchical data. The Delaunay graph is constructed on the union of the two subsets and their labels are subsequently used to identify heterogeneous edges. In the illustrated hierarchy, subsets that are closer are more strongly intermixed and share more Delaunay neighborhoods, whereas subsets lying further apart exhibit fewer connections across the two groups.

4.1.3 Analysis and Interpretation with HyperDGA

Paper B considers three settings that illustrate how the geometric alignment score can be employed and interpreted in hyperbolic data analysis. Across these experiments, HyperDGA is compared with hyperbolic versions of the Chamfer and Wasserstein distances. These baselines compare the two subsets using hyperbolic point-to-point distances, through nearest-neighbor relations for the Chamfer distance and an optimal transport formulation for the Wasserstein distance. HyperDGA instead relies on the connectivity of the hyperbolic Delaunay graph.

Noise inference and evolutionary biology. The first uses synthetic data derived from a binary tree inspired by [29] to simulate protein evolution. Each tree node is associated with a binary sequence, and progressively larger perturbations are introduced by randomly changing the sequence entries in order to simulate mutations. A hyperbolic variational auto-encoder [29] is employed to represent these data in a two-dimensional hyperbolic latent space. As the perturbation increases, the encoded perturbed population becomes progressively less aligned with the reference population. HyperDGA follows this change consistently. In this controlled setting, the monotonic relation between HyperDGA and the perturbation level shows that geometric alignment in the latent representation retains information about the change introduced in the input data. The baseline measures are also strongly correlated with the perturbation level, however, HyperDGA yields a stronger correlation than the Chamfer distance and a correlation comparable to the Wasserstein distance. The perturbation level can therefore be inferred from the representation. A potential application could be to compare the genetic divergence of populations that have evolved from a common ancestral population through their geometric alignment in a hyperbolic representation. The controlled simulated mutation experiment suggests that HyperDGA can retain information about the amount of accumulated genetic change, potentially providing insights into processes such as *speciation* in evolutionary biology.

Model behavior and latent representation evaluation. The same synthetic dataset is then used to examine the representation learned during training of the hyperbolic VAE. The encoding of the training data is compared with the re-encoding of its reconstructions. As training progresses, HyperDGA is strongly associated with both the VAE loss and a score assessing how well distances in the original binary tree are preserved in the hyperbolic latent representation. Among the three measures considered, HyperDGA exhibits the strongest correlations with both quantities. Changes in HyperDGA are also associated with the formation of new clusters in the latent representation. In this case, the score is not used to characterize a relation in the data, but to obtain information about the behavior of the model that produced the representation. Such information could also be employed to compare the latent representations produced by different hyperbolic auto-encoder-based models.

Single-cell analysis and potential biomedical applications. Finally, **Paper B** considers Poincaré representations [28] of three real-world single-cell RNA sequencing datasets: *Olsson* [101], *Paul* [102], and *Planaria* [13]. Each contains predefined cell populations related through cellular differentiation. HyperDGA is computed between selected populations and compared with known developmental relations. HyperDGA and the two baseline measures are consistent with the expected biological relations across the three datasets.

The *Planaria* comparison also illustrates a difference between the measures: for neoblast and progenitor populations, which occupy visibly distinct regions of the representation, HyperDGA reflects their separation more strongly than the Chamfer and Wasserstein distances. More broadly, this type of analysis suggests potential applications in oncology. For example, single-cell profiles from a tumor and from healthy reference tissue could be represented jointly in a hyperbolic latent space. Rather than comparing malignant and healthy cells as two broad groups, HyperDGA could be computed between a malignant cell population and several healthy populations corresponding to different differentiation states. The resulting alignment profile would indicate which healthy states are most strongly intermixed with the malignant population in the representation. If samples from different disease stages or treatment time points were available, the comparison could be repeated to analyze how these geometric relations change during disease progression or treatment. Such changes could support inference about shifts in malignant cell-state composition. If such shifts were also related to differences in treatment response, they could provide information relevant to treatment selection for precision medicine. These inferences would, however, remain conditional on the chosen representation and would not by themselves establish cellular lineage, biological mechanism, or clinical outcome. This potential biomedical application is supported by single-cell studies of acute myeloid leukemia showing that malignant cells can be related to different healthy hematopoietic cell types and differentiation states [103, 104], and that the composition of malignant cell states can change following chemotherapy treatment [105]. Hematopoietic differentiation involves branching transitions between cell states, making it a relevant setting in which to investigate whether a hyperbolic representation, with the geometric properties discussed in the previous chapter, provides a useful basis for such comparisons.

The three settings illustrate two distinct forms of interpretation. In the mutation and biological experiments, geometric alignment is related to properties or relations associated with the represented data. In the VAE experiment, it is instead related to changes in the model and in the representation produced during training. This distinction corresponds directly to the two forms of interpretation considered in **RQ1**: properties of the represented data and properties of the model that produced the representation.

The Delaunay graph is important here as an analytical instrument. Although it is an explicit geometric object, the graph itself is not interpreted as a recovered hierarchy or biological structure. Its role is to define neighborhood relations from which the geometric alignment between already specified groups can be quantified.

4.2 Geometric Organization of a Hyperbolic Representation: Paper C

The previous section starts from semantic groups that are known in advance and uses geometry to characterize their relation. A complementary problem is to study how semantic properties are distributed across a representation without first reducing the analysis to a pair of predefined subsets. **Paper C** considers this question for representations of olfactory descriptor data.

Several studies have attempted to organize odor perception using similarity judgments, sorting tasks, descriptor ratings, factor analysis, and multidimensional scaling [106, 107, 108, 109, 110]. These studies identify meaningful perceptual dimensions, but no single low-dimensional organization of odor quality has emerged, and the resulting representations depend on the stimuli, descriptors, participants, and analysis method [111]. More recent work has learned representations from molecular and perceptual information, including the Principal Odor Map [112], while taxonomy-based approaches have explicitly considered hierarchical relations among odor descriptors [113].

Most low-dimensional representations of odor perception have been Euclidean. Previous work by Zhou et al. [114] showed that statistics of natural odor mixtures and perceptual descriptions are compatible with hyperbolic spaces of dimensions 3 and more. **Paper C** addresses a different question: given a two-dimensional hyperbolic representation of olfactory descriptor profiles, what interpretable organization can be identified from its geometry?

4.2.1 Olfactory Descriptor Data

Paper C considers two complementary datasets whose different forms of annotation support different geometric analyses.

The Sagar dataset [115] contains continuous perceptual ratings of monomolecular odorants from three participants. Each participant rated 160 odorants, and the analysis uses 15 perceptual descriptors shared across participants, including intensity, pleasantness, fruity, floral, sweet, and musky. Each observation is therefore represented by a continuous descriptor profile.

The GoodScents–Leffingwell (GSLF) dataset distributed with OpenPOM [116] contains 4983 molecules annotated with 138 expert-defined odor descriptors. In contrast to the Sagar ratings, these annotations are binary: a descriptor is either present or absent for a given molecule. Each molecule is therefore represented by a multi-label binary descriptor vector.

This distinction is important for the subsequent analysis. Continuous Sagar ratings allow gradual changes in descriptor values to be associated with directions in the representation. Binary GSLF annotations do not define such gradients

and are instead analyzed through the regions occupied by molecules carrying a given label.

4.2.2 Hyperbolic Metric Multidimensional Scaling

For each dataset, **Paper C** constructs a two-dimensional representation in the Poincaré disk. Let $X = \{x_i\}_{i=1}^{N} \subseteq \mathbb{R}^D$ denote the descriptor vectors, and let $D_{ij}^{\text{in}} = \|x_i - x_j\|$ be their pairwise Euclidean distances. The corresponding representation $Z = \{z_i\}_{i=1}^{N} \subseteq \mathbb{P}^2$ is obtained by minimizing

$$\mathcal{L}_{\text{hMDS}} = \frac{1}{N^2} \sum_{i,j=1}^{N} \left(D_{ij}^{\text{in}} - d_{\mathbb{P}}(z_i, z_j) \right)^2 . \tag{4.2.3}$$

This is the metric multidimensional scaling principle expressed using hyperbolic distances [86, 87]. The points are optimized directly on the Poincaré disk using Riemannian Adam optimization [82].

This construction also provides a direct connection with the preservation perspective introduced in the previous chapter. Before interpreting the embedding, **Paper C** evaluates how well hyperbolic pairwise distances retain the pairwise distance structure of the input descriptor space, using Pearson and Spearman correlations between the input and embedding distances. The results show substantial distance preservation in both datasets. The subsequent geometric analyses are therefore interpreted with respect to representations that retain a substantial part of the input metric structure.

The use of two dimensions allows center-to-boundary variation to be distinguished from angular organization and makes the resulting patterns directly visualizable. Principal radial and directional associations were statistically assessed using permutation tests adapted to the structure of the datasets and to the analysis setting.

4.2.3 Analysis and Interpretation of Radial, Directional and Regional Organization

4.2.3.1 Radial Organization

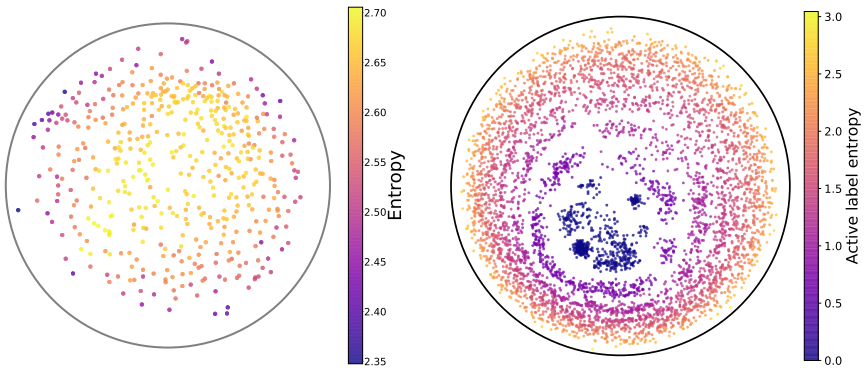
The first analysis considers the hyperbolic radius $r_i = d_{\mathbb{P}}(z_i, 0)$, that is, the hyperbolic distance of an embedded observation from the origin of the Poincaré disk. The radial analysis asks whether meaningful scalar properties of the descriptor profiles vary systematically from the origin toward the boundary.

A central quantity in **Paper C** is descriptor entropy. Given a probability vector $p_i = (p_{i,1}, \ldots, p_{i,D})$, the entropy of observation i is

$$H_i = -\sum_{k=1}^{D} p_{i,k} \log p_{i,k}. \tag{4.2.4}$$

The construction of p_i, and therefore the interpretation of H_i, depends on the type of descriptor data.

Figure 4.2.2 illustrates the radial organization observed for the two datasets. Although the corresponding entropy measures have different interpretations, both examples relate a global property of the descriptor profile to center-to-boundary position in the hyperbolic representation.



(a) Sagar rating profile entropy

(b) GSLF active label entropy

Figure 4.2.2: Radial organization of descriptor-profile entropy in representative hyperbolic embeddings from **Paper C**. **(a)** In the Sagar representation, rating profile entropy decreases toward the boundary: diffuse descriptor profiles tend to occur closer to the origin, while more concentrated profiles occur further outward. **(b)** In the GSLF representation, active label entropy increases toward the boundary, with molecules carrying more active odor descriptors tending to occur further from the origin.

For the continuous Sagar ratings, p_i is obtained by applying a softmax to the descriptor values. The resulting *rating profile entropy* is high when the rating strength is distributed across many descriptors and low when the profile is dominated by a smaller number of descriptors. A strong negative association is observed between this quantity and hyperbolic radius, as illustrated in Figure 4.2.2a. Diffuse descriptor profiles tend to be located closer to the origin, whereas more concentrated profiles tend to occur closer to the boundary. The radial organization is stronger for this global property of the descriptor profile than for any individual descriptor. The same qualitative relationship persists when intensity is removed, when correlated descriptors are pruned, after transforming the descriptors to orthogonal modes of variation, and when the participants are considered separately or through averaged ratings. These analyses support the interpretation that the radial organization reflects a global property of the descriptor profile rather than being driven by a single descriptor or by the original descriptor basis.

The binary annotations in GSLF give the entropy a different interpretation. Here, p_i is uniform over the descriptors active for molecule i, so that the

active label entropy is $\log K_i$, where K_i is the number of active labels. It therefore measures descriptor multiplicity rather than the spread of continuous descriptor strengths. This quantity increases toward the boundary of the disk, as illustrated in Figure 4.2.2b. When the descriptor vectors are instead expressed through orthogonal modes of variation, the corresponding entropy decreases toward the boundary. These patterns are not contradictory, because the quantities summarize different properties of the descriptor profile. They illustrate that radial position can encode global organization, but its interpretation depends on the property being related to it.

4.2.3.2 Directional and Regional Organization

Radius only describes center-to-boundary position. **Paper C** therefore also examines the organization around the disk.

For the continuous Sagar data, embedded observations are first mapped to the tangent space at the origin,

$$u_i = \log_0(z_i) \in T_0\mathbb{P}^2. \tag{4.2.5}$$

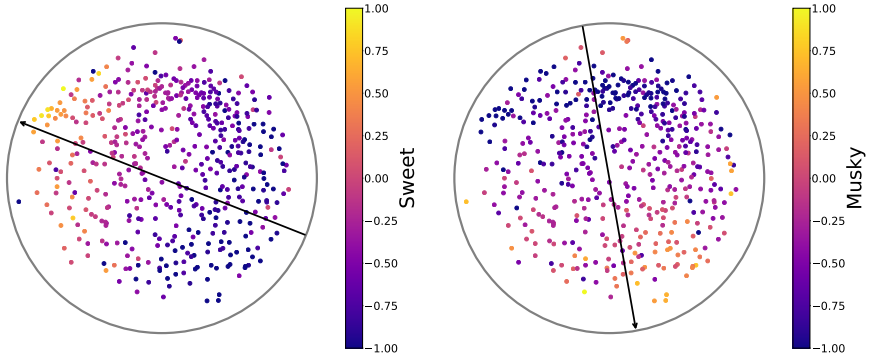
For each continuous descriptor k, a linear model

$$x_{i,k} \approx \beta_k^\top u_i + b_k \tag{4.2.6}$$

is fitted in the tangent space. The vector β_k defines the direction along which the fitted descriptor value increases most strongly. The purpose is not to assign an absolute angle to an odor quality, but to determine whether a descriptor exhibits a dominant directional trend across the representation. The strength of this directional organization is quantified by the coefficient of determination R^2 of the fitted model. A larger R^2 indicates that the descriptor is better summarized by a global directional trend in the tangent space. Figure 4.2.3 illustrates this directional organization for sweet and musky, which exhibit distinct trends across the same Sagar representation.

Several descriptors show clear directional organization, with sweet, musky, fruity, and pleasantness among the strongest examples. Pleasantness is particularly relevant because it has repeatedly been identified as an important organizing dimension of olfactory perception [117, 118, 108, 119, 120]. The emergence of a strong pleasantness direction is therefore consistent with previous work, while the same representation also exhibits the distinct radial descriptor-profile structure described above.

Different optimizations can produce representations related by rotations or reflections without changing their pairwise geometry. What is relevant is therefore the presence and relative organization of directional trends rather than their absolute angles.



(a) Sweet (b) Musky

Figure 4.2.3: Directional organization of the descriptors sweet and musky in a representative Sagar embedding from **Paper C**. Both descriptors exhibit graded variation across the Poincaré disk, but along distinct directions. These examples illustrate how individual perceptual descriptors can be associated with different directional trends in the same hyperbolic representation.

The binary GSLF annotations require a different treatment. Since a binary descriptor does not provide a graded value, it does not define a direction of increase. Instead, **Paper C** uses hyperbolic kernel density estimation to identify regions of the Poincaré disk in which molecules carrying a particular descriptor are concentrated. This analysis is related to descriptor-region visualizations used for the Principal Odor Map [112], but the density estimation is adapted to the hyperbolic geometry of the Poincaré disk. Figure 4.2.4 illustrates this regional organization for the fruity descriptor family and related odor descriptors.

The broader fruity region in Figure 4.2.4 overlaps with several related descriptors, including different fruit qualities, while some descriptors occupy more distinct local regions. These regional patterns complement the radial analysis of the same representation. Radius is associated with global properties of the descriptor profile, whereas the descriptor-specific density regions exhibit an angular organization around the disk.

4.3 Discussion

Papers B and C illustrate two complementary ways in which the geometry of a hyperbolic representation can be used for inference.

In **Paper B**, group labels are supplied in advance through the choice of subsets. The question is then relational: how are these groups positioned with respect to one another? The hyperbolic Delaunay graph provides neighborhood relations from which their geometric alignment can be quantified. Depending on the

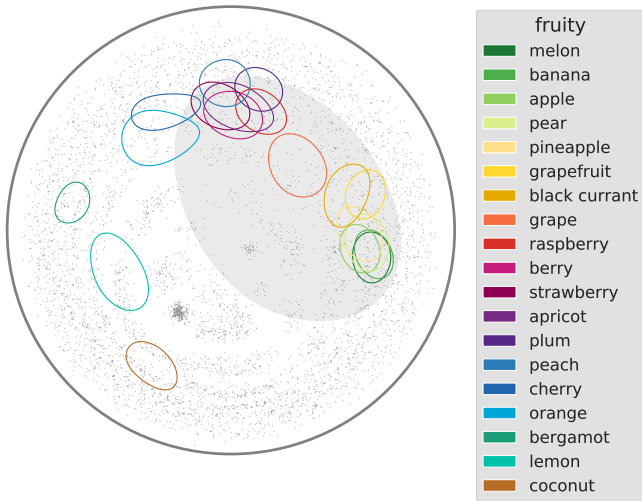


Figure 4.2.4: Regional organization of the fruity descriptor family in a representative GoodScents–Leffingwell representation from **Paper C**. High-density regions show where molecules annotated with fruity and related descriptors are concentrated in the Poincaré disk.

setting, this alignment can then be interpreted in terms of a distribution shift in the input data, the behavior of the model that produced the representation, or semantics and developmental biological relations between represented populations.

In **Paper C**, the analysis instead begins with the complete representation and asks how semantic variables are distributed through it. Different geometric quantities characterize different forms of organization. Hyperbolic radius is associated with global properties of descriptor profiles, tangent-space directions characterize graded trends in continuous descriptors, and localized density regions describe the organization of binary annotations. The appropriate geometric construction therefore depends both on the type of organization being investigated and on the form of the available data, such as continuous ratings or binary annotations.

The approaches of **Papers B and C** could also be combined when observations from several participants are represented in a common hyperbolic space, as in the Sagar union representation of **Paper C**. The observations associated with each participant would define predefined subsets, allowing HyperDGA to quantify inter-participant geometric alignment and compare it with differences in perceptual ratings or other behavioral measures. The same approach could potentially be employed to compare healthy and olfactory-impaired participants, or to follow changes in alignment with a healthy reference during olfactory rehabilitation. Such geometric changes would need to be interpreted

together with behavioral or clinical measures.

The two papers also illustrate how preservation and geometric inference can interact. In **Paper C**, preservation is part of the construction of the representation: hyperbolic metric MDS seeks to retain the pairwise-distance structure of the descriptor data, after which the resulting geometry is analyzed and interpreted. In **Paper B**, preservation is examined through information that remains reflected in the latent geometry. In the controlled perturbation experiment, geometric alignment retains information about the amount of change introduced in the input data. In the VAE experiment, HyperDGA is related to how well the latent organization is preserved through reconstruction and to how well distances from the original binary tree are retained in the representation. In the single-cell experiments, known developmental relations between cell populations are reflected in their geometric alignment. Together, these examples show how structural information retained or exposed by a representation can provide the basis for subsequent geometric analysis and interpretation.

These interpretations remain conditional on the representation and on the geometric quantity being analyzed. In **Paper B**, HyperDGA quantifies relations encoded in a given representation; it does not by itself establish an underlying biological hierarchy or causal explanation. For example, a change in alignment between represented cellular populations does not by itself establish evolutionary ancestry, biological mechanism, or clinical prognosis. Likewise, the organization identified in **Paper C** describes a constructed two-dimensional Poincaré representation of olfactory descriptor data. It does not imply that olfactory perception, or its neural representation, is intrinsically two-dimensional or hyperbolic. Given that the Sagar dataset contains ratings from only three participants, an important direction for future work is to investigate whether the radial entropy relation and the principal directional trends are reproduced across larger and more diverse groups of participants.

The VAE experiment in **Paper B** illustrates a different limitation of the evaluation setting. Poor alignment between the encoding of the input data and the re-encoding of its reconstructions indicates that the model fails to preserve the organization of its latent representation through the reconstruction process. Strong alignment, however, establishes only that the two representations are consistent according to the chosen comparison; it does not by itself guarantee that the learned representation preserves meaningful structure or is suitable for a particular downstream task. This asymmetry concerns the evaluation setting rather than HyperDGA itself and applies similarly to other set-based comparison measures, such as the baselines considered in **Paper B**.

There are also limitations to the geometric constructions themselves. HyperDGA depends on the discrete connectivity of the Delaunay graph and is therefore discontinuous as the positions of the points change. This makes it useful as an analysis and evaluation quantity, but not directly as a differentiable

objective for gradient-based representation learning. Similarly, the directional analysis in **Paper C** provides a global first-order summary and may not capture nonlinear or multi-region variation in a descriptor.

Together, **Papers B and C** address **RQ1** by showing how geometric organization in hyperbolic latent representations can be analyzed at different scales and related to meaningful properties of the represented data or of the model producing the representation. Both papers infer geometric organization from an existing representation, through neighborhood relations, radial and directional patterns, or localized regions. The next chapter considers a different form of geometric inference, in which the goal is to extract an explicit parsimonious tree whose structure is itself the object to be analyzed and interpreted.

When the represented data are known to possess an underlying hierarchy, one may ask whether the representation contains sufficient geometric organization for an explicit tree to be inferred from it. If a reference hierarchy is available, an inferred tree could in principle be compared with that reference, complementing evaluations based on metric distortion or relations between predefined subsets. The representation could then be assessed not only by how well it preserves pairwise relations, but also by how well it preserves higher-level hierarchical organization. When no trusted reference hierarchy is available, the extracted tree instead provides a candidate structural hypothesis whose properties can be interpreted in relation to the application. This distinction motivates the next chapter. When a reference hierarchy is available, an inferred tree can be compared with it to evaluate the representation. Without such a reference, the inferred tree is instead interpreted as a candidate parsimonious structure.

This transition also gives the hyperbolic Delaunay construction a second role in the thesis. In **Paper B**, Delaunay geometry is used as an analytical tool: the graph provides neighborhood relations from which the geometric alignment of predefined subsets is measured. Chapter 5 moves from analysis to extraction. There, **Papers D and E** use hyperbolic Delaunay geometry in the search for parsimonious hyperbolic Steiner trees, where the resulting tree is itself the structural object to be extracted and interpreted for geometric inference.

5 Extracting and Interpreting Hierarchical Structures with Hyperbolic Steiner Trees

The previous chapter considered how geometric organization in a hyperbolic latent representation can be analyzed and related to properties of the represented data or of the model that produced the representation. When the organization of interest is hierarchical, a further problem is to extract this implicit organization as an explicit tree.

This chapter addresses **RQ2**:

> *How can implicit hierarchical organization in hyperbolic latent representations be extracted as explicit parsimonious tree structures, and what can be interpreted from those structures?*

Contributions of Papers D and E. These papers provide complementary deterministic and stochastic approaches to address this question through Steiner minimal trees (SMT). **Paper D** introduces HyperSteiner, a deterministic Delaunay-based heuristic for constructing an approximate SMT in the two-dimensional hyperbolic plane. The method adapts the Smith–Lee–Liebman strategy to hyperbolic geometry by combining a hyperbolic Minimum Spanning Tree (MST) and Delaunay triangulation with local three- and four-terminal Steiner constructions. To make these local constructions computationally tractable, the paper derives formulations of hyperbolic isoptic curves and Fermat points in the Klein–Beltrami model that can be solved numerically. The paper further validates the method on synthetic data and on a hyperbolic representation of Planaria single-cell

data, where tree-derived distances from the inferred hierarchy are shown to correlate with developmental progression. **Paper E** contributes Randomized HyperSteiner (RHS), which addresses the restricted topology exploration of the deterministic construction through stochastic Delaunay sampling. RHS generates candidate Steiner points through randomized exploration, applies local Steiner corrections, and refines Steiner-point locations via Riemannian optimization. The paper further establishes that, for fixed terminals and a fixed tree topology, the Steiner-point configuration minimizing total hyperbolic tree length is unique. Its empirical evaluation shows that the broader topology search can produce shorter trees than HyperSteiner and baselines across the considered synthetic and real-data settings. Finally, RHS is shown empirically to achieve reductions in tree length relative to the MST of up to 43% in near-boundary settings, approaching the 50% theoretical upper bound.

In this setting, geometric inference concerns both the extraction of a parsimonious tree from the latent representation and the interpretation of properties derived from the inferred tree.

5.1 From Implicit Hierarchy to an Explicit Parsimonious Tree

Hyperbolic latent representations may encode hierarchical relations without explicitly representing a tree. Points that occupy different parts of the space may reflect branching or multiscale organization, while the representation itself specifies neither edges nor internal branching points. Extracting a tree therefore requires an additional structural criterion.

Tree reconstruction from data has been studied extensively in areas such as phylogenetics and cell lineage inference [121, 122]. Classical distance-based methods, including UPGMA [123] and Neighbor Joining [124], construct an abstract tree from pairwise dissimilarities. Character-based approaches instead operate directly on observed characters and include Maximum Parsimony [125], Maximum Likelihood [126], and Bayesian methods [127]. Connections between phylogenetic reconstruction and Steiner trees have also been studied by formulating evolutionary reconstruction as the minimization of a cost over tree structures [128].

The setting considered in this chapter is different. **Papers D and E** take an existing hyperbolic latent representation as input and construct a tree directly within the same geometric space. The extraction criterion is therefore geometric: candidate trees are compared through their total hyperbolic length. Recent approaches have also combined hyperbolic geometry with probabilistic phylogenetic inference [129, 130, 131], but these methods incorporate explicit evolutionary models into the inference procedure. Here, the objective is instead to extract a parsimonious embedded tree from the geometry already encoded in the latent representation.

This leads naturally to the *Steiner minimal tree problem*. Let $Z = \{z_i\}_{i=1}^{N} \subseteq \mathbb{H}^2$

denote a hyperbolic latent representation, and let $P \subseteq Z$ be the points selected as terminals, with $P = Z$ when all represented observations are used. A tree $T = (V, E)$ may contain additional vertices $S = V \setminus P$, called *Steiner points*. Its total length is

$$L(T) = \sum_{(u,v)\in E} d_{\mathbb{H}}(u, v). \tag{5.1.1}$$

The Steiner minimal tree (SMT) seeks the choice of tree topology and Steiner-point locations that minimizes its total length while connecting all terminals.

If no additional Steiner points are allowed, the problem reduces to the Minimum Spanning Tree (MST). Allowing Steiner points can produce a shorter network by introducing branching locations that are not themselves observed data points. This motivates a notion of geometric parsimony based on total tree length in the hyperbolic representation: among trees connecting the same terminals, a shorter tree is considered more parsimonious, and may be preferred in some applications.

The reduction relative to the MST can be summarized through the local Steiner ratio

$$\rho(P) = \frac{L(\mathrm{SMT}(P))}{L(\mathrm{MST}(P))} \leq 1. \tag{5.1.2}$$

The global Steiner ratio of the hyperbolic plane is known to be $1/2$ [132]. Thus, particular hyperbolic configurations can in principle admit Steiner trees approaching half the length of their corresponding MSTs. This does not imply that hyperbolic Steiner trees are always substantially shorter: the achievable reduction depends strongly on the geometric configuration of the terminals.

Finding an exact SMT is computationally difficult, and the problem is NP-hard in standard settings [133]. **Papers D and E** therefore consider heuristic constructions that exploit the geometry of the representation to restrict and explore candidate tree structures.

5.2 HyperSteiner: A Deterministic Heuristic for Delaunay-Guided Tree Extraction

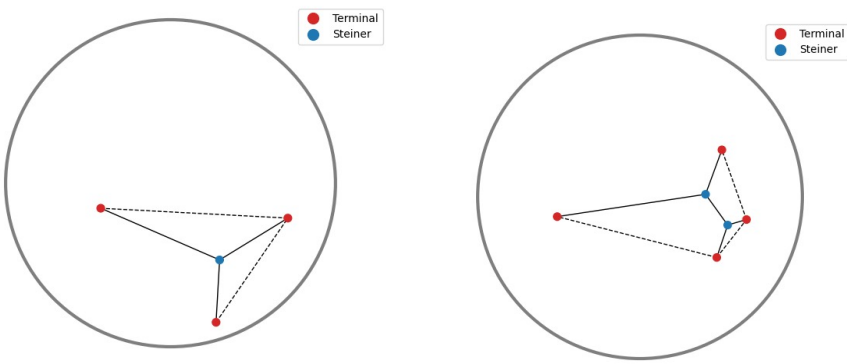
Steiner tree problems are defined more generally in metric spaces [134] and have been studied under several non-Euclidean metrics [135, 136, 137, 138]. In the Euclidean plane, efficient heuristic methods often exploit Delaunay triangulations to identify promising local Steiner configurations [139]. The SLL algorithm [140] is a prominent example of this strategy. Related constructions have been extended to spherical geometry [141], while explicit full Steiner trees in the hyperbolic plane have been studied through hyperbolic geometric constructions [142].

Given this background, **Paper D** adapts the Smith–Lee–Liebman (SLL) heuristic [140] to the hyperbolic plane. The main idea is to use the Delaunay

triangulation to identify small local configurations in which adding Steiner points can reduce the length of an initial MST.

The Delaunay construction already appeared in Chapter 4. As discussed in Section 4.1.1, hyperbolic Voronoi cells in the Klein–Beltrami model can be computed through Euclidean power diagrams [100]. Whereas **Paper B** uses the resulting Delaunay graph to quantify alignment between predefined subsets, **Papers D and E** use the Delaunay triangulation to restrict and explore the local configurations considered during tree construction.

HyperSteiner first constructs the hyperbolic MST and the corresponding Delaunay triangulation. It then considers local Steiner problems involving three or four terminals. A *full Steiner tree* (FST) for m terminals contains the maximal number $m - 2$ of Steiner points. Thus, a three-terminal FST contains one Steiner point, while a four-terminal FST contains two. Figure 5.2.1 illustrates these two local problems, together with their corresponding MSTs.



(a) Three-terminal FST (b) Four-terminal FST

Figure 5.2.1: Examples of local full Steiner trees considered by HyperSteiner for three and four terminals represented in the Klein-Beltrami disk. Solid lines correspond to the edges of HyperSteiner while the dashed ones correspond to the Minimum Spanning Tree.

For the three-terminal case, HyperSteiner selects Delaunay triangles containing two edges of the MST and for which an FST exists. For the four-terminal case, it considers a selected triangle together with an adjacent Delaunay triangle when their union contains three edges of the MST. The union contains four terminals, for which the two possible FST topologies illustrated in Figure 5.2.2 are evaluated and the shorter one is retained.

The resulting local FSTs are ordered according to their local Steiner ratio, with smaller ratios considered first, and combined greedily while preventing cycles in order to obtain a heuristic approximation of the global SMT. A central

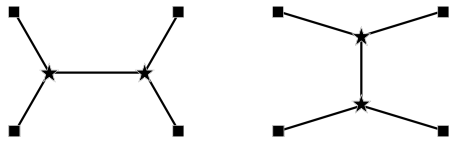


Figure 5.2.2: The two possible full Steiner tree topologies for a four-terminal local problem, reproduced from **Paper D**. Squares denote terminals and stars denote Steiner points.

geometric difficulty is therefore the efficient computation of these three- and four-terminal Steiner problems in hyperbolic space.

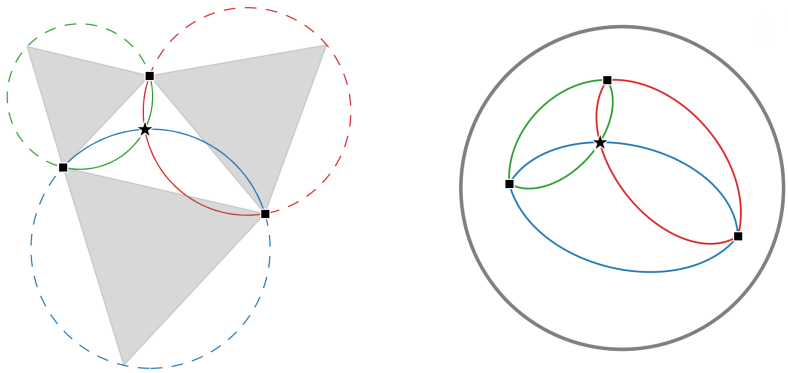
5.2.1 Hyperbolic Fermat Points

For three terminals, an FST contains a single Steiner point as illustrated in Figure 5.2.1a. When such a tree exists, this Steiner point is the *Fermat point*: it minimizes the sum of the distances to the three terminals, and the three incident edges meet at angles of 120°.

In the Euclidean plane, the Fermat point can be obtained through the classical Torricelli construction. An equivalent characterization uses *isoptic curves*. Given two terminals x and y, an isoptic curve for an angle α is the locus, that is, the set of all points s from which the segments joining s to x and y meet at the prescribed angle α. As s moves along the curve, this angle remains constant. In the Euclidean plane, these loci are circular arcs. The Fermat point of three terminals is therefore the common intersection of the three 120° isoptic curves associated with the three pairs of terminals, as illustrated in Figure 5.2.3a. This is equivalent to the Torricelli construction.

The classical Euclidean Torricelli construction cannot be transferred unchanged to hyperbolic space. Its construction relies on Euclidean geodesics and angles simultaneously. In the Poincaré disk for example, angles are preserved but hyperbolic geodesics are represented by circular arcs rather than straight segments. Conversely, in the Klein–Beltrami model used in **Paper D**, geodesics are represented by straight segments but Euclidean angles are not equal to hyperbolic angles. The Euclidean circular construction therefore does not directly characterize the hyperbolic Fermat point.

What remains intrinsic to the problem is the 120° condition at a Steiner point. More generally, the incident edges at a Steiner point of an SMT on a Riemannian surface meet at angles of 120° [143]. **Paper D** therefore characterizes the Fermat point through hyperbolic isoptic curves: for each pair of terminals, the corresponding curve consists of the points from which the two hyperbolic geodesics toward the terminals meet at 120°. Their common intersection gives the hyperbolic Fermat point, as shown in Figure 5.2.3b.



(a) Euclidean construction (b) Klein–Beltrami construction

Figure 5.2.3: Fermat-point construction through 120° isoptic curves, reproduced from **Paper D**. For each pair of terminals, the corresponding curve contains the points from which the two terminals form an angle of 120°. In the Euclidean plane, these curves are circular arcs corresponding to the Torricelli construction; in the Klein–Beltrami model, they are defined according to the hyperbolic metric. Their common intersection determines the Fermat point.

A central contribution of **Paper D** is to turn this geometric characterization into a computational procedure. Using the hyperbolic cosine rule, the isoptic curves can be characterized by equations in the Klein–Beltrami coordinates. Although the three pairwise curves intersect at the Fermat point, two are sufficient to determine a candidate solution. Computing the hyperbolic Fermat point can therefore be reduced to solving a system of two nonlinear equations using numerical root-finding methods, such as Powell's hybrid algorithm [144]. **Paper D** also derives an algebraic reformulation, enabling the use of numerical solvers for systems of polynomial equations [145].

The three-terminal construction then provides the computational building block for the four-terminal FSTs used by HyperSteiner. For each of the two possible four-terminal topologies, the two Steiner points are obtained iteratively through successive three-terminal Fermat-point constructions following [142].

5.2.2 From Local Solutions to a Heuristic Steiner Tree

The local solutions are incorporated into the SLL construction. Starting from the MST, HyperSteiner identifies candidate Delaunay configurations, computes their local Steiner solutions, and orders them according to their local Steiner ratio. Compatible local trees are then combined greedily while avoiding cycles.

The Delaunay triangulation serves two purposes. First, it restricts the number

of local configurations that need to be considered. Second, it ties the tree construction directly to the geometry of the hyperbolic representation. Since the Delaunay triangulation restricts the number of local problems that need to be considered, HyperSteiner retains the $\mathcal{O}(|P| \log |P|)$ asymptotic complexity of Delaunay-based SLL heuristics.

This efficiency comes with a limitation: the construction is deterministic and explores only the local configurations selected from the MST and Delaunay triangulation. It may therefore miss alternative tree structures with shorter total length. **Paper E** addresses this limitation by introducing stochastic exploration of candidate tree structures.

5.2.3 Extracting and Interpreting a Tree from Planaria Data

Paper D illustrates how the extracted tree can be interpreted using the *Planaria* single-cell RNA sequencing dataset [13]. The cells are first represented in a two-dimensional hyperbolic space using the Poincaré map [28]. HyperSteiner is then applied directly to this latent representation. Figure 5.2.4 illustrates the extraction on a 1% sample of the data for visual clarity, showing the resulting HyperSteiner tree together with the corresponding MST.

The Planaria dataset provides an auxiliary notion of developmental progression through pseudotime. In **Paper D**, the tree-geodesic distance from a chosen root cell is used as an estimate of cell age and compared with pseudotime estimates based on PAGA [15]. HyperSteiner improves this prediction relative to the MST, while Neighbor Joining obtains a slightly lower error at substantially higher computational cost.

The purpose of this experiment is not to claim that the inferred Steiner tree recovers the true cellular lineage. Rather, it illustrates how an explicit tree extracted from the latent geometry can support the interpretation of properties of the represented cells. In particular, quantities derived from the tree, such as root-to-cell distance, can be related to biologically meaningful properties such as cellular age or developmental progression. The inserted Steiner points should therefore be interpreted here as geometric branching locations rather than as identified biological entities. A natural direction for future work is to examine whether, when combined with additional biological information, these branching locations and the paths connecting them can be related to common progenitor states or to developmental lineage trajectories.

5.3 Randomized HyperSteiner: Stochastic Exploration of Tree Topologies

The deterministic construction of HyperSteiner restricts the search to local configurations selected from the MST and Delaunay triangulation. This makes the method efficient, but the resulting tree depends strongly on these initial local

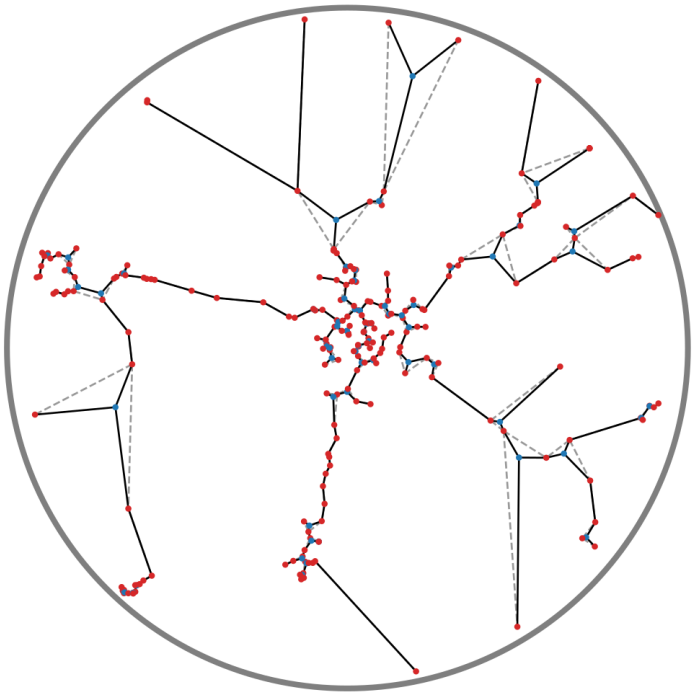
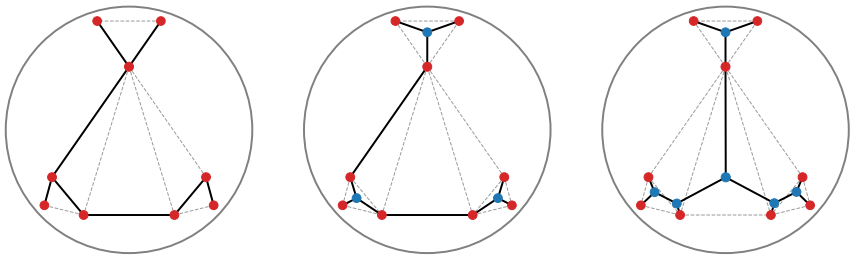


Figure 5.2.4: HyperSteiner applied to a sample of the hyperbolic Planaria representation. Terminals are shown in red while Steiner points are in blue. The solid black edges correspond to the heuristic hyperbolic Steiner tree, while the dashed grey edges show the hyperbolic Minimum Spanning Tree on the same terminals. The added Steiner points allow the inferred tree to introduce branching locations that are not restricted to observed cells.

choices and may miss shorter alternative topologies. **Paper E** addresses this limitation with Randomized HyperSteiner (RHS), which introduces stochasticity into the Delaunay-based construction.

RHS builds on earlier randomized Delaunay-based heuristics for the Euclidean SMT problem. Beasley [146] introduced an iterative construction in which the Delaunay triangulation is recomputed after each Steiner point insertion, while stochastic exploration is used to search for alternative configurations. Laarhoven and de Jong [147] later used randomized sampling of Delaunay configurations to broaden the search without exhaustively considering all local candidates. These methods motivate the use of randomization in **Paper E**: rather than restricting the search to the local configurations selected from the MST, RHS samples Delaunay triangles to generate candidate Steiner points and thereby explores a broader set of tree topologies. The resulting candidates can lead to tree structures that are not considered by deterministic HyperSteiner.

Figure 5.3.5 illustrates this difference. For the same set of terminals, HyperSteiner improves on the MST but remains restricted to a particular branching structure, whereas RHS obtains a different tree with shorter total length.



(a) MST (b) HyperSteiner (c) Randomized HyperSteiner

Figure 5.3.5: Comparison of the MST, HyperSteiner, and Randomized HyperSteiner for the same set of terminals, reproduced from **Paper E**. Red points denote terminals, blue points denote Steiner points, and dashed edges show the auxiliary hyperbolic Delaunay triangulation. In this example, the randomized construction produces a different branching structure with shorter total tree length.

5.3.1 Stochastic Candidate Generation and Local Steiner Corrections

RHS starts from the best tree found so far and expands its set of Steiner-point candidates through randomized Delaunay sampling. At each expansion step, it computes the hyperbolic Delaunay triangulation of the terminals together with the current candidates. Delaunay triangles are then selected probabilistically, and the hyperbolic barycenter of each selected triangle is added as a new candidate Steiner point. These barycenters provide inexpensive proposals for new branching locations rather than final Steiner-point positions. As new candidates are inserted, the Delaunay triangulation is recomputed, allowing different local configurations to be explored.

The expanded candidate set is then processed according to structural properties of an SMT [146]. RHS constructs the MST on the terminals and candidate Steiner points and examines the degree of each candidate. A degree-three candidate is replaced, when possible, by the Fermat point of its three neighbors. A degree-four candidate is replaced by the two Steiner points of the corresponding four-terminal FST. Candidates that do not satisfy the required degree conditions are removed. These steps reuse the three- and four-terminal constructions introduced for HyperSteiner in Section 5.2.1. RHS also enforces the angle condition of an SMT. Angles between adjacent edges must be at least 120°, while the three edges incident to a Steiner point meet at exactly 120°. If two adjacent edges form an angle smaller than 120°, RHS considers

the corresponding three-terminal problem and, when a Fermat point exists, replaces the local configuration by the associated three-terminal FST.

These stages therefore play complementary roles. Randomly sampled barycenters propose new branching locations and broaden the set of tree structures explored. The degree and angle checks then correct these proposals using local Steiner constructions and can modify the resulting tree topology. Once a candidate topology is obtained, the Steiner-point locations are further refined through Riemannian optimization.

5.3.2 Riemannian Geometric Refinement Given a Fixed Topology

The local Steiner constructions determine candidate branching structures and Steiner-point locations, but these locations do not necessarily jointly minimize the total length of the tree. **Paper E** therefore further refines the Steiner points by minimizing the total hyperbolic tree length defined in Equation (5.1.1) using Riemannian gradient descent [80].

As introduced in Chapter 3, Riemannian optimization performs gradient-based optimization while respecting the geometry of the underlying manifold. Here, the terminals and the tree topology are kept fixed, and only the locations of the Steiner points are optimized. The refinement can therefore shorten the tree geometrically, but it does not add or remove Steiner points or change which vertices are connected. Changes to the topology are instead introduced by the stochastic expansion and the local degree and angle corrections described above.

A theoretical contribution of **Paper E** is to show that, for fixed terminals and a fixed tree topology, the placement of the Steiner points that minimizes the total tree length is unique, extending the corresponding Euclidean result [148]. This result applies to the continuous optimization of Steiner-point locations for a given topology; it does not determine which topology gives the globally shortest tree.

Within each exploration trial, Riemannian refinement is combined with the degree and angle corrections described above. The resulting tree is then compared with the best tree found so far. If it is shorter, it is retained; otherwise, RHS returns to the previous best solution and starts a new trial.

5.3.3 Results, Limitations, and Implications

Paper E compares RHS with the MST, HyperSteiner from **Paper D**, and Neighbor Joining on synthetic data distributions and the Planaria representation. Across the considered experiments, RHS produces shorter trees than HyperSteiner. For distributions concentrated near the center of the hyperbolic disk, the improvement is relatively modest, whereas it becomes substantially larger for distributions concentrated near the boundary. This indicates that

the broader exploration of candidate tree topologies performed by RHS is particularly beneficial for some terminal configurations.

The near-boundary experiments also provide a connection with the Steiner ratio discussed in Section 5.1. The local Steiner ratio $\rho(P)$ defined in Equation 5.1.2 quantifies how much the connection can be shortened from the MST by allowing additional Steiner points. In the Euclidean plane, the classical Gilbert–Pollak conjecture predicts $\rho(P) \geq \sqrt{3}/2 \approx 0.866$ for arbitrary finite point sets [148], which would correspond to a maximal reduction of approximately 13.4% relative to the MST. A recent computer-assisted result, developed using a system based on a large language model, established the lower bound $\rho \geq 0.8559$ [149], improving substantially on the previous bound of 0.824 [150]. In the hyperbolic plane, by contrast, the global Steiner ratio is known to be $1/2$ [132]. Hence, there exist hyperbolic configurations for which the SMT approaches half the length of the corresponding MST. The potential benefit of introducing Steiner branching can therefore be substantially larger in hyperbolic geometry than suggested by the classical Euclidean benchmark.

This difference is reflected in the experiments of **Paper E**. In the configuration designed to approach the hyperbolic limit, RHS reaches a maximum average reduction of approximately 43% relative to the MST, whereas HyperSteiner achieves considerably smaller reductions. This approaches the limiting 50% reduction associated with the hyperbolic Steiner ratio and is substantially larger than the approximately 13.4% reduction associated with the conjectured Euclidean $\sqrt{3}/2$ benchmark. The experiments on the Planaria representation likewise show a consistent improvement over HyperSteiner.

The ratio also provides information when comparing different configurations within the same hyperbolic space. A value close to one indicates that introducing Steiner points provides little shortening relative to the MST, whereas a smaller value indicates that the terminal configuration admits more efficient branching under the hyperbolic metric. The experiments therefore suggest that some configurations expose the potential advantage of hyperbolic Steiner branching more strongly than others.

These observations also clarify the practical relation between HyperSteiner and RHS. HyperSteiner provides a deterministic and comparatively inexpensive construction and can be preferable when computational efficiency or reproducibility is the main concern. RHS spends additional computation exploring a broader set of candidate topologies and refining their geometry, and is therefore more appropriate when reducing total tree length is important and the additional computational cost can be afforded. The difference is particularly relevant for configurations in which the local deterministic construction of HyperSteiner leaves substantial room for further shortening.

Finally, RHS remains a heuristic method: although the Steiner-point locations have a unique minimizing configuration once a topology is fixed, there is

no guarantee that the stochastic topology search finds the globally optimal topology.

5.4 Discussion

Papers D and E address the extraction part of **RQ2** by constructing an explicit tree directly in the hyperbolic latent space. The SMT objective provides the criterion for this extraction: among trees connecting the same terminals, configurations with smaller total hyperbolic length are preferred. The resulting tree may introduce Steiner points, allowing branching locations that are not restricted to the observed data points.

A useful distinction emerging from the two papers is between the *combinatorial* problem of choosing a tree topology and the *continuous* problem of locating its Steiner points. **Paper D** addresses both through a deterministic Delaunay-based heuristic: the Delaunay triangulation and MST restrict the candidate topologies, while three- and four-terminal Steiner problems determine the corresponding local branching locations. **Paper E** separates these aspects more explicitly. Randomized Delaunay expansion explores alternative topologies, whereas, once a topology is fixed, Riemannian optimization adjusts the Steiner-point locations to reduce the total tree length. The uniqueness result of **Paper E** shows that the fixed-topology continuous problem has a unique minimizing Steiner-point configuration. The topology itself remains a combinatorial problem addressed heuristically.

This use of the Delaunay construction also differs from Chapter 4. In **Paper B**, the Delaunay graph provides neighborhood relations from which the geometric alignment of predefined subsets is measured. In **Papers D and E**, Delaunay geometry instead guides the construction of a new structural object: an embedded tree connecting the represented points.

The interpretation of this tree remains conditional on the latent representation and on the SMT objective. A shorter tree is more parsimonious according to total hyperbolic length, but this alone does not imply that its topology corresponds to the hierarchy that generated the original data. The tree is instead interpreted as a parsimonious structural hypothesis derived from the geometry of the representation. The Planaria experiment in **Paper D** provides a concrete example. Since the ground-truth cellular hierarchy is unknown, the inferred tree is not evaluated as a reconstruction of the true lineage. Instead, root-to-cell distances along the tree are compared with developmental progression estimated using PAGA. The experiment therefore evaluates whether a quantity derived from the tree is related to a biologically meaningful property of the cells. Similarly, the inserted Steiner points represent geometric branching locations; identifying them with progenitor states or lineage branching events would require additional biological evidence.

An important limitation of both methods is that they are developed for two-

dimensional hyperbolic representations. HyperSteiner achieves $\mathcal{O}(|P|\log|P|)$ complexity by restricting the search to local Delaunay configurations, which can exclude shorter topologies. RHS reduces this restriction through stochastic exploration, but does not remove the underlying combinatorial difficulty. Extending the constructions beyond two dimensions and developing more systematic topology-search procedures therefore remain natural directions for future work.

The relation between heuristic and approximation methods also provides a useful perspective on the development from **Papers D and E**. In **Paper D**, polynomial-time approximation schemes are discussed as an important difference between the Euclidean and hyperbolic SMT problems. While Euclidean methods provide a controllable approximation error at increased computational cost, no corresponding approximation scheme for the hyperbolic SMT problem was available at the time. HyperSteiner instead provided an efficient Delaunay-based deterministic heuristic, but without an explicit approximation guarantee. **Paper E** improves the heuristic approach by exploring a broader set of candidate topologies, but the topology search remains stochastic and has no guarantee of finding the globally optimal SMT. Based on the remaining difference between the approximately 43% reduction obtained by RHS and the limiting 50% reduction associated with the hyperbolic Steiner ratio, the paper identifies quadtree-based approximation schemes as a direction for future work.

Since then, Kisfaludi-Bak et al. have derived a randomized $(1+\varepsilon)$-approximation scheme for the Steiner tree problem in fixed-dimensional hyperbolic space [151]. This addresses the theoretical gap identified in **Paper D** and the direction proposed in **Paper E**: the quality of a hyperbolic Steiner tree can now, in principle, be approximated with a prescribed error relative to the optimal solution. An interesting next step is therefore to implement their algorithm and evaluate this approximation scheme on the configurations considered in **Paper E**. A sufficiently accurate approximation would provide a controlled reference for the optimal tree length and help determine whether the remaining difference between RHS and the 50% limit is mainly due to the heuristic topology search or to the optimum of the finite configurations themselves. The distinction is important: decreasing ε approaches the optimal SMT for a given instance, but does not imply that this optimum approaches a 50% reduction, since the Steiner ratio of $1/2$ is a limiting result rather than a property of every hyperbolic point configuration. From a practical perspective, this also suggests comparing the approximation–computation trade-off of the new scheme with the heuristic constructions of HyperSteiner and RHS.

Overall, **Papers D and E** address **RQ2** by providing methods for extracting explicit parsimonious trees from hyperbolic latent representations and by showing how quantities derived from these trees can be interpreted in relation to the represented data.

6 Conclusions and Perspectives

This thesis has considered how structural information can be preserved, analyzed, extracted, and interpreted through latent data representations. Its principal focus has been on the latter stages of this progression: once data have been represented in a hyperbolic latent space, what can be inferred from the resulting geometry? The contributions address this question through the analysis of geometric organization and the extraction of explicit parsimonious tree structures. This chapter summarizes the answers to the research questions at the level of the thesis as a whole, discusses the main remaining limitations and directions for future work, and returns to the broader perspective on geometry introduced in Chapter 2.

6.1 Answering the Research Questions

Motivating question — Preservation. The motivating question of the thesis concerns how a latent representation can be constructed so that selected structure is preserved or exposed. The methodologies considered in Chapter 3 show that this can be approached at different levels. As developed in **Paper A**, prescribed geometric organization can be imposed directly, selected topological properties can be encouraged through simplicial regularization, and metric relations can be incorporated explicitly into the learning objective. The latter was also employed in **Paper C**. When prior knowledge concerns the general type of organization expected in the data, the geometry of the latent space itself can provide an inductive bias. In particular, when hierarchical or tree-like relations are expected to be important, hyperbolic geometry provides a space whose metric properties are adapted to branching structure. Choosing a hyperbolic latent space does not impose or guarantee a particular hierarchy, but it can facilitate the preservation or exposure of hierarchical organization in the representation.

The main implication for the remainder of the thesis is that preservation and inference cannot be considered independently. What can subsequently be analyzed or extracted depends on what information remains accessible in the representation. Preservation is therefore not the principal problem studied in this thesis, but it establishes the conditions under which later geometric inference can be meaningful.

RQ1 — Analysis and interpretation. Chapter 4 considered how geometric organization in hyperbolic latent representations can be analyzed and related to meaningful properties of the represented data or of the model producing the representation. The main conclusion is that there is no single geometric quantity that characterizes all relevant forms of organization. Different questions and different datasets require different constructions. Relations between predefined subsets can be characterized through geometry-dependent neighborhood relations as in **Paper B**, while organization across a complete representation may instead be described through radial position, directions, or localized regions, as performed in **Paper C**.

Geometric analysis therefore requires a choice of quantity that is appropriate both to the geometry of the representation and to the type of information being investigated. Interpretation then requires an additional step: the resulting geometric quantity must be related to properties of the represented data, to the behavior of the model, or to available reference information such as labels, measurements, or known structures. The geometric quantities describe how the representation is organized, while their interpretation depends on how this organization relates to the problem being studied.

RQ2 — Extraction and interpretation. Chapter 5 considered the stronger problem of converting implicit hierarchical organization into an explicit parsimonious tree structure. A point configuration does not by itself specify a unique hierarchy, and an additional criterion is therefore required to extract one. In **Papers D and E**, this criterion is to minimize total hyperbolic tree length. Steiner points allow branching locations to be introduced beyond the observed data points, and shorter trees are preferred according to this geometric notion of parsimony.

A useful distinction emerging from these contributions is between two parts of the extraction problem. Choosing which vertices are connected and where branching occurs is a combinatorial problem over tree topologies. Once a topology has been chosen, determining the locations of its Steiner points is a continuous geometric optimization problem. HyperSteiner and Randomized HyperSteiner address these two aspects in different ways.

The resulting structure must consequently be interpreted conditionally. If a trusted reference hierarchy is available, the extracted tree can be compared with it to evaluate how well hierarchical organization is represented. When no

such reference exists, the tree instead provides a parsimonious structural hypothesis. Quantities or structures derived from the tree, such as tree distances or hierarchies, can then be related to meaningful properties of the represented data.

Main research question — Geometric inference. Taken together, these results provide an answer to the main research question of the thesis. Structural information encoded in a hyperbolic latent representation can be inferred by constructing geometric quantities or explicit structures that make aspects of its organization accessible. These results become interpretable when they are related to properties of the data, to the model that produced the representation, or to reference information such as labels, measurements, or known structures. The resulting conclusions are therefore conditional on three elements: the information encoded in the representation, the geometric construction used for inference, and the information used for interpretation.

6.2 From Structured Representations to Geometric Inference

A broader connection across the thesis concerns the changing role of geometry. During representation construction, geometry acts as a modeling assumption or inductive bias. Choosing a hyperbolic latent space determines how distances, geodesics, neighborhoods, and volumes are expressed, and thereby influences which organizations can be represented efficiently. Once the representation has been constructed, the same geometry becomes a basis for inference. Distances can be measured, neighborhoods constructed, directions analyzed, triangulations computed, and embedded trees extracted. Geometry therefore plays two related roles: it first constrains how information is represented and then determines the mathematical relations through which that information can be examined.

A second recurring theme is the interaction between continuous and discrete structures. Several contributions in this thesis move between these two views on data. In **Paper A**, simplicial complexes provide discrete representations of topology, from which algebraic invariants such as Betti numbers are obtained, while the corresponding topological information is incorporated into representation learning through a continuous regularization objective. In **Paper B**, a discrete Delaunay graph is constructed from the continuous hyperbolic representation and its connectivity is used to quantify geometric changes between predefined subsets. In **Papers D and E**, the topology of a tree is discrete, whereas its parsimony is evaluated through continuous hyperbolic lengths and the locations of its Steiner points are determined through geometric optimization. In particular, **Paper D** formulates the continuous hyperbolic Fermat point problem through a system of algebraic equations, while **Paper E** uses Riemannian optimization to refine Steiner point locations for a fixed topology. Across these settings, discrete structures make aspects of the latent

geometry explicit, while continuous geometric information determines how these structures are constructed, evaluated, or refined.

This perspective also clarifies why geometric inference cannot be completely separated from representation construction. Two representations of the same observations may preserve different relations and can consequently lead to different inferred structures.

6.3 Limitations and Future Directions

One general limitation of the approaches considered in this thesis is that most of the inference is conditional on a particular representation. An important future direction is therefore to study further the robustness of geometric conclusions across alternative representations of the same data. Changes in the representation method, latent dimensionality, curvature, initialization, sampling, or preprocessing may affect the inferred alignment, directions, regions, or tree structures. Rather than treating a single inferred structure as fixed, future methods could quantify which conclusions remain stable across such variations. This would also provide a natural basis for representing uncertainty in geometric inference.

A related direction is to close the loop between representation construction and subsequent inference. The main progression studied in this thesis goes from a representation to the structures that can be inferred from it. The reverse direction is also possible. If geometric analysis identifies information that should be retained, that information could be incorporated into the construction of future representations. An inferred or known hierarchy could, for example, provide information for evaluating or constraining a latent representation. More generally, quantities currently used only for downstream analysis could motivate differentiable objectives or other forms of feedback during representation learning.

Another direction concerns the relation between topology and hyperbolic geometry. **Paper A** uses computational topology to introduce structural constraints during representation learning, whereas the remaining papers primarily use metric and computational-geometric constructions in hyperbolic space. Topological data analysis is not restricted to Euclidean spaces: persistent homology and related constructions can be defined from general metric data. An interesting continuation of this thesis would therefore be to develop or adapt topological constructions that make more explicit use of hyperbolic geometry and its associated neighborhood structures. Hyperbolic witness complexes, Delaunay-based complexes, or other geometry-aware filtrations could provide tools for studying topological organization directly in hyperbolic latent representations.

Topology may also contribute before the choice of representation geometry has been fixed. The olfactory study by Zhou et al. [114] discussed in Chapter 4

and **Paper C**, for example, compares Betti curves derived from odor data with those generated from candidate geometric models and finds low-dimensional hyperbolic models consistent with both natural odor statistics and perceptual data. This suggests a broader role for topological summaries: rather than only evaluating whether selected topology has been preserved, they may also help assess which geometric models are compatible with observed relations. Such an approach would provide a natural connection between the topological perspective of **Paper A** and the hyperbolic geometric inference developed in the remainder of the thesis in **Papers B–E**.

Several questions also remain open for explicit hierarchy extraction. HyperSteiner and RHS are developed for two-dimensional hyperbolic representations, and extending their geometric constructions to higher dimensions would broaden the representations to which explicit tree extraction can be applied. The combinatorial search over tree topologies remains another central difficulty. HyperSteiner obtains computational efficiency by restricting this search, while RHS explores a broader set of possibilities at additional computational cost. The recent approximation scheme from Kisfaludi-Bak et al. [151] discussed in Chapter 5 provides a complementary direction by allowing the hyperbolic Steiner tree problem to be approximated with a prescribed error relative to its optimum. Applying this algorithm to the configurations studied in **Paper E** would provide a controlled approximation of the optimum. This would help separate two sources of the remaining gap to the limiting Steiner ratio experimented with RHS: limitations of the heuristic topology search and geometric properties of the finite configurations themselves.

6.4 Potential Applications

The methods developed in the thesis were evaluated in several domains, but their combination also suggests applications that were not investigated directly.

Olfactory perception. A first direction is to extend the analysis of **Paper C**. Larger and more diverse participant cohorts would make it possible to examine whether the radial organization and principal directional trends observed in the current data are reproduced across more individuals. Beyond increasing the number of participants, richer olfactory annotation could provide more interesting data for hyperbolic representation and inference. In particular, trained olfactory panels or domain experts could provide more fine-grained descriptor data. Such data would make it possible to investigate organization with greater hierarchical depth.

This is particularly relevant for hyperbolic geometry, whose representational advantages become more pronounced when the structure of interest contains substantial branching. With a sufficiently hierarchical olfactory representation, HyperSteiner or RHS could be employed to convert an implicit geometric

hierarchy into an explicit parsimonious tree. If an expert taxonomy were available, it could provide a reference against which the extracted structure could be compared. In the absence of such a reference, the tree would instead provide a candidate organization whose branching structure could be examined together with perceptual and semantic information.

Biomedical and single-cell data. Biomedical data provide another possible application of the methods developed in this thesis. As discussed in Chapter 4, HyperDGA could be used to compare predefined populations in a hyperbolic representation, for example to study how a population of interest is geometrically related to different reference populations.

When the underlying process is expected to have a branching organization, HyperSteiner or RHS could be employed to extract an explicit parsimonious tree from the representation. In single-cell data, such a tree could be compared with pseudotime, known developmental stages, or lineage information to study whether its structure is consistent with a possible developmental trajectory. The extracted tree should not be interpreted directly as a true cell lineage, but it could provide a geometric hypothesis to be evaluated against independent biological information.

6.5 Societal and Sustainability Considerations

Geometric inference can support scientific analysis by making organization in complex representations more accessible. The biological applications considered in this thesis illustrate possible positive uses. Methods for comparing cellular populations or generating candidate developmental structures may contribute to the analysis of differentiation, disease progression, or treatment-related changes when combined with appropriate biological evidence. Such applications are broadly aligned with the objective of Sustainable Development Goal 3, Good Health and Well-Being, through their potential contribution to biomedical research.

The same ability to infer structure also creates risks of misuse or overinterpretation. A geometric relation in a latent representation does not necessarily establish a biological mechanism, and an extracted Steiner tree does not by itself establish a lineage. Similar considerations would apply if geometric inference were used on representations of people, behavior, or social data: groups, directions, or hierarchies inferred from a model could be assigned a meaning that is not justified by the underlying evidence. Responsible use therefore requires maintaining the distinction made throughout this thesis between a structure inferred under a particular representation and criterion and a structure independently established in the phenomenon being studied. For sensitive applications, this also requires attention to the quality, provenance, and possible biases of the data from which the representation is constructed.

Sustainability also concerns the computational resources required by the methods themselves. The work in this thesis primarily considers low-dimensional representations. In addition to supporting visualization and interpretation, low dimensionality limits the storage and computational requirements of many subsequent geometric operations. The two-dimensional setting of the hyperbolic analyses and tree constructions was motivated primarily by interpretability and computational tractability rather than by an explicit environmental objective, but these choices also avoid the cost of performing the corresponding operations in unnecessarily high-dimensional spaces.

The development from HyperSteiner to RHS provides a more direct example of a computational trade-off. Both papers report computational time together with the relative lengths of the resulting tree constructions. HyperSteiner restricts the topology search and provides an efficient deterministic heuristic. RHS explores a broader set of candidate structures and obtains shorter trees, but requires additional computation. The preferable method therefore depends on the application: when the additional reduction in tree length is not critical, the less expensive construction may be sufficient. Computational resources can in this way be treated as part of the methodological trade-off rather than assuming that the most computationally expensive solution should always be preferred.

More broadly, the representational efficiency of hyperbolic geometry suggests an interesting question for large-scale learned representations. Hierarchical relations can be represented with low distortion in low-dimensional hyperbolic spaces. If similar information in large models could be encoded using fewer latent dimensions than in a comparable Euclidean representation, this could reduce memory requirements and some subsequent computations. Recent work on hyperbolic Transformer and foundation-model representations makes this question more relevant.

Such a sustainability benefit should not, however, be assumed from dimensionality alone. Hyperbolic operations may themselves be more expensive than their Euclidean counterparts, and representation-related computation may constitute only part of the overall cost of a large model. A relevant future evaluation would therefore compare not only representation quality and dimensionality, but also memory use, computation time, and, where possible, energy consumption. The sustainability question is consequently empirical: whether a more structurally appropriate geometry can produce sufficient representational efficiency to compensate for the additional cost of operating in that geometry.

6.6 A Final Perspective on Geometry and Inference

Chapter 2 described how the development of non-Euclidean geometry changed the status of geometric assumptions. Euclidean geometry remained mathematically valid, but it was no longer the only coherent framework within

which geometric questions could be posed. Different assumptions produced different geometries, and the choice between them became itself an object of mathematical investigation.

Representation learning gives this observation a computational counterpart. The geometry of a latent space is not simply a neutral coordinate system. Choosing a Euclidean, hyperbolic, spherical, or another geometry determines how distances, geodesics, neighborhoods, and other relations are expressed. The technical chapters of this thesis illustrate a further point: even after a geometry has been chosen, inference introduces another level of assumptions. Delaunay adjacency defines one form of neighborhood relation, radial distance defines one form of organization, density estimation identifies particular regions, and total Steiner-tree length defines a particular notion of parsimony. The structure eventually interpreted therefore depends not only on the observed data, but on the representation through which the data are expressed and on the criterion through which that representation is analyzed.

This does not imply that geometry is arbitrary. A geometric model can be evaluated according to whether it preserves relevant relations, provides useful and stable descriptions of the data, supports efficient computation, and produces conclusions that agree with independent evidence. The relevant question is therefore not whether a dataset is ”truly” hyperbolic in an absolute sense, but whether a hyperbolic representation provides an appropriate model for the structure of interest and whether the inferences drawn from it are supported by the available evidence.

The progression developed in this thesis follows from this view. Selected structure can first be preserved or exposed when a representation is constructed. Its geometry can then be analyzed to characterize organization, or used to extract an explicit structure of interest such as a parsimonious tree. Finally, the resulting quantities and structures can be interpreted in relation to the problem from which the data originated. Preservation, analysis, extraction, and interpretation are therefore not independent objectives, but successive parts of a geometric-inference process whose conclusions remain tied to the assumptions under which the representation and the inference were constructed.

7 Summary of the Included Publications

The thesis is based on the papers and manuscript listed in this chapter. The symbol * denotes equal contribution.

Paper A [1]

InvMap and Witness Simplicial Variational Auto-Encoders

Aniss Aiman Medbouhi, Vladislav Polianskii, Anastasia Varava, and Danica Kragic

Machine Learning and Knowledge Extraction, 2023.

Abstract: Variational auto-encoders (VAEs) are deep generative models used for unsupervised learning, however their standard version is not topology-aware in practice since the data topology may not be taken into consideration. In this paper, we propose two different approaches with the aim to preserve the topological structure between the input space and the latent representation of a VAE. Firstly, we introduce InvMap-VAE as a way to turn any dimensionality reduction technique, given an embedding it produces, into a generative model within a VAE framework providing an inverse mapping into original space. Secondly, we propose the Witness Simplicial VAE as an extension of the simplicial auto-encoder to the variational setup using a witness complex for computing the simplicial regularization, and we motivate this method theoretically using tools from algebraic topology. The Witness Simplicial VAE is independent of any dimensionality reduction technique and together with its extension, Isolandmarks Witness Simplicial VAE, preserves the persistent Betti numbers of a dataset better than a standard VAE.

Paper B [2]

Hyperbolic Delaunay Geometric Alignment

Aniss Aiman Medbouhi, Giovanni Luca Marchetti, Vladislav Polianskii, Alexander Kravberg, Petra Poklukar, Anastasia Varava, and Danica Kragic

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), 2024.

Abstract: Hyperbolic machine learning is an emerging field aimed at representing data with a hierarchical structure. However, there is a lack of tools for evaluation and analysis of the resulting hyperbolic data representations. To this end, we propose Hyperbolic Delaunay Geometric Alignment (HyperDGA) – a similarity score for comparing datasets in a hyperbolic space. The core idea is counting the edges of the hyperbolic Delaunay graph connecting datapoints across the given sets. We provide an empirical investigation on synthetic and real-life biological data and demonstrate that HyperDGA outperforms the hyperbolic version of classical distances between sets. Furthermore, we showcase the potential of HyperDGA for evaluating latent representations inferred by a Hyperbolic Variational Auto-Encoder.

Paper C [3]

Geometric Organization of Olfactory Descriptor Data in the Poincaré Disk

Aniss Aiman Medbouhi, Farzaneh Taleb, Giovanni Luca Marchetti, and Danica Kragic

Preprint, arXiv:2609.09573, 2026. Submitted to *Chemical Senses*.

Abstract: Odor quality is commonly represented using high dimensional descriptor profiles, yet their low dimensional organization remains unclear. We investigated whether a two-dimensional hyperbolic embedding can provide an interpretable representation of this structure. We applied hyperbolic metric multidimensional scaling to two complementary datasets: 480 Sagar rating profiles from three participants rating 160 odorants on 15 continuous descriptors, and 4983 GoodScents–Leffingwell molecules annotated with 138 binary descriptors. The embeddings substantially preserved pairwise descriptor distances, supporting subsequent analyses of radial and angular organization. In Sagar, *rating profile entropy* was strongly and negatively associated with hyperbolic radius (Pearson $r = -0.77 \pm 0.05$), with diffuse profiles closer to the center and concentrated profiles closer to the boundary. This radial organization emerged primarily at the level of the full descriptor profile, rather than any individual descriptor, and remained robust across alternative descriptor representations, participant specific analyses, and averaged ratings. *Sweet*, *musky*, *fruity*, and *pleasantness* showed the strongest directional trends (mean $R^2 = 0.40$ to 0.51). In GoodScents–Leffingwell, *active label entropy*, reflecting descriptor multiplicity, increased with radius ($r = 0.87 \pm 0.03$), whereas *orthogonalized descriptor entropy*, reflecting spread across orthogonal modes, decreased with radius ($r = -0.88 \pm 0.01$). Related binary descriptors occupied coherent localized high-density regions. These findings reveal complementary radial and angular organization in the hyperbolic representation of olfactory descriptor data. They support hyperbolic mapping as an interpretable descriptive framework in which radius summarizes global profile properties, while the angular component captures continuous descriptor gradients and categorical organization.

Paper D [4]

HyperSteiner: Computing Heuristic Hyperbolic Steiner Minimal Trees

Alejandro García-Castellanos*, **Aniss Aiman Medbouhi***, Giovanni Luca Marchetti, Erik J. Bekkers, and Danica Kragic

Proceedings of the 27th Society for Industrial and Applied Mathematics (SIAM) Symposium on Algorithm Engineering and Experiments (ALENEX), 2025.

Abstract: We propose HyperSteiner – an efficient heuristic algorithm for computing Steiner minimal trees in the hyperbolic space. HyperSteiner extends the Euclidean Smith-Lee-Liebman algorithm, which is grounded in a divide-and-conquer approach involving the Delaunay triangulation. The central idea is rephrasing Steiner tree problems with three terminals as a system of equations in the Klein-Beltrami model. Motivated by the fact that hyperbolic geometry is well-suited for representing hierarchies, we explore applications to hierarchy discovery in data. Results show that HyperSteiner infers more realistic hierarchies than the Minimum Spanning Tree and is more scalable to large datasets than Neighbor Joining.

Paper E [5]

Randomized HyperSteiner: A Stochastic Delaunay Triangulation Heuristic for the Hyperbolic Steiner Minimal Tree

Aniss Aiman Medbouhi*, Alejandro García-Castellanos*, Giovanni Luca Marchetti, Daniël M. Pelt, Erik J. Bekkers, and Danica Kragic

The 29th International Conference on Artificial Intelligence and Statistics (AISTATS), 2026.

Abstract: We study the problem of constructing Steiner Minimal Trees (SMTs) in hyperbolic space. Exact SMT computation is NP-hard, and existing hyperbolic heuristics such as HyperSteiner are deterministic and often get trapped in locally suboptimal configurations. We introduce **Randomized HyperSteiner** (RHS), a stochastic Delaunay triangulation heuristic that incorporates randomness into the expansion process and refines candidate trees via Riemannian gradient descent optimization. Experiments on synthetic data sets and a real-world single-cell transcriptomic data show that RHS outperforms Minimum Spanning Tree (MST), Neighbour Joining, and vanilla HyperSteiner (HS). In near-boundary configurations, RHS can achieve a 32% reduction in total length over HS, demonstrating its effectiveness and robustness in diverse data regimes.

Declaration on the use of Generative AI

In line with the requirement from the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology and the doctoral programmes, the author acknowledges the use of Artificial Intelligence (AI) tools, in particular ChatGPT serie 5 from Open AI, to get feedback and assist with language editing and text structure. The author has critically reviewed and revised all generated material and bears full responsibility for the content of the thesis.

Bibliography

[1] Aniss Aiman Medbouhi, Vladislav Polianskii, Anastasia Varava, et al. “InvMap and Witness Simplicial Variational Auto-Encoders”. In: *Machine Learning and Knowledge Extraction* 5.1 (2023), pp. 199–236. ISSN: 2504-4990. DOI: 10.3390/make5010014. URL: https://www.mdpi.com/2504-4990/5/1/14.

[2] Aniss Aiman Medbouhi, Giovanni Luca Marchetti, Vladislav Polianskii, et al. “Hyperbolic Delaunay Geometric Alignment”. In: *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*. Ed. by Albert Bifet, Jesse Davis, Tomas Krilavičius, et al. Cham: Springer Nature Switzerland, 2024, pp. 111–126. ISBN: 978-3-031-70352-2.

[3] Aniss Aiman Medbouhi, Farzaneh Taleb, Giovanni Luca Marchetti, et al. “Geometric organization of olfactory descriptor data in the Poincaré disk”. arXiv preprint, 2026. arXiv: 2609.09573 [cs.CE]. URL: https://arxiv.org/abs/2609.09573.

[4] Alejandro García-Castellanos, Aniss Aiman Medbouhi, Giovanni Luca Marchetti, et al. “HyperSteiner: Computing Heuristic Hyperbolic Steiner Minimal Trees”. In: *Proceedings of the 27th Society for Industrial and Applied Mathematics (SIAM) Symposium on Algorithm Engineering and Experiments (ALENEX)*. 2025, pp. 194–208. DOI: 10.1137/1.9781611978339.16. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611978339.16. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611978339.16.

[5] Aniss Aiman Medbouhi, Alejandro García-Castellanos, Giovanni Luca Marchetti, et al. “Randomized HyperSteiner: A Stochastic Delaunay Triangulation Heuristic for the Hyperbolic Steiner Minimal Tree”. In: *The 29th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2026. URL: https://openreview.net/forum?id=cq2lc89TqM.

[6] A. Jouanny, A. Upadhyay, J. Jiang, et al. “Machine-Learning-Driven Reconstruction of Organic Aerosol Sources across Dense Monitoring Networks in Europe”. In: *Environmental Science and Technology (ES&T) Letters, American Chemical Society (ACS) publications* 12.11 (2025), pp. 1523–1531. ISSN: 2328-8930. DOI: 10.1021/acs.estlett.5c00771. URL: https://pubs.acs.org/doi/10.1021/acs.estlett.5c00771.

[7] Aniss Aiman Medbouhi. “Resisting Algorithmic Desire and Manipulation in AI Recommendation Systems: An Islamic Moral Psychology of Digital Self-Mastery”. Manuscript, 2026.

[8] Farzaneh Taleb, Aniss Aiman Medbouhi, Nona Rajabi, et al. *Adaptation of ERP-based Classifiers to Novel Subjects and Tasks*. Poster presented at the annual Strategic Research Area Neuroscience (StratNeuro) retreat in 2022.

[9] Farzaneh Taleb, Aniss Aiman Medbouhi, Giovanni Luca Marchetti, et al. *Towards Discovering the Hierarchy of the Olfactory Perceptual Space via Hyperbolic Embeddings*. Poster presented at the Computational and Systems Neuroscience (COSYNE) conference in 2025, published in Science Communications Worldwide. DOI: 10.57736/0d78-6155. URL: https://www.world-wide.org/cosyne-25/towards-discovering-hierarchy-olfactory-037f6ccb.

[10] Aniss Aiman Medbouhi, Farzaneh Taleb, Giovanni Luca Marchetti, et al. *Modeling the Hierarchy of the Human Olfactory Perceptual Space via Hyperbolic Embeddings*. Extended abstract, Poster A143. Amsterdam, The Netherlands, Poster presented at the Cognitive Computational Neuroscience (CCN) conference in 2025. URL: https://2025.ccneuro.org/poster/?id=FijOA6b5lY.

[11] Thomas Hunt Morgan. “Experimental Studies of the Regeneration of Planaria maculata”. In: *Archiv für Entwicklungsmechanik der Organismen* 7 (1898), pp. 364–397. DOI: 10.1007/BF02161491.

[12] Thomas Sandmann, Matthias C. Vogg, Suthira Owlarn, et al. “The Head-Regeneration Transcriptome of the Planarian Schmidtea mediterranea”. In: *Genome Biology* 12 (2011), R76. DOI: 10.1186/gb-2011-12-8-r76.

[13] Mireya Plass, Jordi Solana, F. Wolf, et al. “Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics”. In: *Science* 360 (Apr. 2018), eaaq1723.

[14] Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, et al. “mRNA-Seq whole-transcriptome analysis of a single cell”. In: *Nature Methods* 6.5 (2009), pp. 377–382. DOI: 10.1038/nmeth.1315.

[15] F. Wolf, Fiona Hamey, Mireya Plass, et al. “PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells”. In: *Genome biology* 20 (Mar. 2019). DOI: 10.1186/s13059-019-1663-x.

[16] Diederik P. Kingma and Max Welling. “Auto-Encoding Variational Bayes.” In: *ICLR*. Ed. by Yoshua Bengio and Yann LeCun. 2014.

[17] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. “Stochastic Backpropagation and Approximate Inference in Deep Generative Models”. In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research 2. Bejing, China: PMLR, 22–24 Jun 2014, pp. 1278–1286.

[18] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. “Testing the Manifold Hypothesis”. In: *Journal of the American Mathematical Society* 29.4 (2016), pp. 983–1049. DOI: 10.1090/jams/852.

[19] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. “Geometric Inference for Probability Measures”. In: *Foundations of Computational Mathematics* 11.6 (2011), pp. 733–751. DOI: 10.1007/s10208-011-9098-0.

[20] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. “Finding the Homology of Submanifolds with High Confidence from Random Samples”. In: *Discrete & Computational Geometry* 39 (2008), pp. 419–441. DOI: 10.1007/s00454-008-9053-2.

[21] Xavier Pennec. “Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements”. In: *Journal of Mathematical Imaging and Vision* 25.1 (2006). A preliminary appeared as INRIA RR-5093, January 2004., pp. 127–154.

[22] Doris Schattschneider. “The mathematical side of M.C. Escher”. In: *Notices of the American Mathematical Society* 57 (June 2010).

[23] Douglas Dunham. “Hyperbolic symmetry”. In: *Computers & Mathematics with Applications* 12.1, Part B (1986), pp. 139–153. ISSN: 0898-1221. DOI: https://doi.org/10.1016/0898-1221(86)90147-1.

[24] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, et al. “Hyperbolic geometry of complex networks”. In: *Phys. Rev. E* 82 (3 Sept. 2010), p. 036106.

[25] Maximillian Nickel and Douwe Kiela. “Poincaré Embeddings for Learning Hierarchical Representations”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. Von Luxburg, S. Bengio, et al. Vol. 30. Curran Associates, Inc., 2017.

[26] Maximillian Nickel and Douwe Kiela. “Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 3779–3788.

[27] Rik Sarkar. “Low Distortion Delaunay Embedding of Trees in Hyperbolic Plane”. In: *Graph Drawing*. Ed. by Marc van Kreveld and Bettina Speckmann. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 355–366. ISBN: 978-3-642-25878-7.

[28] Anna Klimovskaia, David Lopez-Paz, Léon Bottou, et al. “Poincaré Maps for Analyzing Complex Hierarchies in Single-Cell Data”. In: *Nature Communications* (2020). DOI: 10.1038/s41467-020-16822-4.

[29] Yoshihiro Nagano, Shoichiro Yamaguchi, Yasuhiro Fujita, et al. “A Wrapped Normal Distribution on Hyperbolic Space for Gradient-Based Learning”. In: *International Conference on Machine Learning*. 2019.

[30] Emile Mathieu, Charline Le Lan, Chris J. Maddison, et al. “Continuous Hierarchical Representations with Poincaré Variational Auto-Encoders”. In: *Advances in Neural Information Processing Systems*. 2019.

[31] Jens Høyrup. “Pythagorean ’rule’ and ’theorem’: mirror of the relation between Babylonian and Greek Mathematics”. English. In: *Babylon : Focus mesopotamischer Geschichte, Wiege früher Gelehrsamkeit, Mythos in der*

Moderne. Ed. by Johannes Renger. SDV Saarländische Druckerei und Verlag GmbH, 1999.

[32] Aristotle. *Metaphysics, Volume I: Books 1–9*. Vol. 271. Loeb Classical Library. Translated by Hugh Tredennick. Cambridge, MA: Harvard University Press, 1933. ISBN: 9780674992993.

[33] Carl B. Boyer and Uta C. Merzbach. "A history of mathematics". English. In: ed. by Johannes Renger. Wiley, 1989.

[34] Nigel Tubbs. *Philosophy and Modern Liberal Arts Education: Freedom is to Learn*. English. 2015. ISBN: 9781137358912.

[35] Plato. *The Republic of Plato*. 3rd ed. Translated by Benjamin Jowett. Oxford: Clarendon Press, 1908.

[36] Germain Bazin. *Le temps des musées*. French. Ed. by Desoer. Desoer, 1967.

[37] Phillip S. Jones Lucas N. H. Bunt and Jack D. Bedient. *The Historical Roots of Elementary Mathematics*. English. Ed. by Dover Publications. Dover Publications, 1988.

[38] Euclid. *The Thirteen Books of Euclid's Elements*. Vol. 1. Translated from the text of Heiberg, with introduction and commentary, by Thomas L. Heath. Cambridge: Cambridge University Press, 1908.

[39] Vincenzo De Risi. "The Development of Euclidean Axiomatics: The Systems of Principles and the Foundations of Mathematics in Editions of the Elements in the Early Modern Age". In: *Archive for History of Exact Sciences* 70 (2016), pp. 591–676. DOI: 10.1007/s00407-015-0173-9.

[40] Rashed Roshdi and Régis Morelon. *Encyclopedia of the History of Arabic Science*. Vol. 2. Routledge, 1996. ISBN: 0-415-02063-8. DOI: 10.1017/S0007087497283202.

[41] Boris A. Rosenfeld. *A History of Non-Euclidean Geometry: Evolution of the Concept of a Geometric Space*. Vol. 12. Studies in the History of Mathematics and Physical Sciences. New York, NY: Springer, 1988. ISBN: 9780387964584. DOI: 10.1007/978-1-4419-8680-1.

[42] Gregg De Young. "Further Adventures of the Rome 1594 Arabic Redaction of Euclid's Elements". In: *Archive for History of Exact Sciences* 66 (2012), pp. 265–294. DOI: 10.1007/s00407-012-0094-9.

[43] Charles Burnett, David C. Lindberg, and Michael H. Shank. "Translation and Transmission of Greek and Islamic Science to Latin Christendom". In: *The Cambridge History of Science*. The Cambridge History of Science. Cambridge University Press, 2013, pp. 341–364.

[44] Jeremy J. Gray. *Worlds out of nothing: a course in the history of geometry in the 19th Century*. Springer Undergraduate Mathematics Series. London: Springer, 2007. DOI: 10.1007/978-1-84628-633-9.

[45] Eugenio Beltrami. *Saggio di interpretazione della geometria Non-Euclidea*. ita. Napoli: s.n., 1868.

[46] Nicola Arcozzi. "Beltrami's Models of Non-Euclidean Geometry". In: (Dec. 2012). DOI: 10.1007/978-3-0348-0227-7_1.

[47] Torkel Franzén. *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*. A K Peters, 2005.

[48] David Hilbert. *The Foundations of Geometry*. 2nd ed. Authorized English translation by Edgar Jerome Townsend. Open Court Publishing Company, 1902.

[49] Alfred Tarski. “What is Elementary Geometry?” In: *The Axiomatic Method*. Ed. by Leon Henkin, Patrick Suppes, and Alfred Tarski. Vol. 27. Studies in Logic and the Foundations of Mathematics. Elsevier, 1959, pp. 16–29. DOI: https://doi.org/10.1016/S0049-237X(09)70017-5. URL: https://www.sciencedirect.com/science/article/pii/S0049237X09700175.

[50] Kurt Gödel. “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I”. In: *Monatshefte für Mathematik und Physik* 38 (1931), pp. 173–198. DOI: 10.1007/BF01700692.

[51] Stephen Hyde, Barry W. Ninham, Sten Andersson, et al. “The Role of Curvature in Condensed Matter: Physics, Chemistry and Biology”. In: *The Language of Shape*. Amsterdam: Elsevier Science B.V., 1997. ISBN: 978-0-444-81538-5.

[52] M.E. Evans and G.E. Schröder-Turk. “In a material world: Hyperbolic geometry in biological materials”. eng. In: *Asia Pacific Mathematics Newsletter* 5.2 (2015), pp. 21–30. ISSN: 2010-3484.

[53] Huanqiu Zhang, P. Rich, Albert Lee, et al. “Hippocampal spatial representations exhibit a hyperbolic geometry that expands with experience”. In: *Nature Neuroscience* 26 (Dec. 2022), pp. 1–9.

[54] N.L. Balazs and A. Voros. “Chaos on the pseudosphere”. In: *Physics Reports* 143.3 (1986), pp. 109–240. ISSN: 0370-1573.

[55] Fran çois Sausset, Gilles Tarjus, and Pascal Viot. “Tuning the Fragility of a Glass-Forming Liquid by Curving Space”. In: *Phys. Rev. Lett.* 101 (15 Oct. 2008), p. 155701.

[56] Pascal Chossat, Grégory Faye, and Olivier Faugeras. “Bifurcation of hyperbolic planforms”. In: *Journal of Nonlinear Science* 21.4 (2011), pp. 465–498.

[57] Ines Chami, Albert Gu, Dat P Nguyen, et al. “HoroPCA: Hyperbolic Dimensionality Reduction via Horospherical Projections”. In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, July 2021, pp. 1419–1429.

[58] Yunhui Guo, Haoran Guo, and Stella X. Yu. “CO-SNE: Dimensionality Reduction and Visualization for Hyperbolic Data”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2022, pp. 21–30.

[59] R. Kleinberg. “Geographic Routing Using Hyperbolic Space”. In: *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*. 2007, pp. 1902–1909.

[60] Marián Boguñá, Fragkiskos Papadopoulos, and Dmitri Krioukov. “Sustaining the Internet with Hyperbolic Mapping”. In: *Nature communications* 1 (Sept. 2010), p. 62.

[61] A. Tifrea, G. Becigneul, and O.-E. Ganea. “Poincaré GloVe: Hyperbolic Word Embeddings”. In: *7th International Conference on Learning Representations (ICLR)*. May 2019.

[62] Gualtiero Piccinini. “The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts’s “Logical Calculus of Ideas Immanent in Nervous Activity””. In: *Synthese* 141.2 (2004), pp. 175–215. DOI: 10.1023/B:SYNT.0000043018.52445.3e.

[63] Tom M. Mitchell. *The Need for Biases in Learning Generalizations*. Tech. rep. CBM-TR-117. New Brunswick, NJ: Computer Science Department, Rutgers University, May 1980.

[64] William W. Cohen. "Compiling Prior Knowledge Into an Explicit Bias". In: *Machine Learning Proceedings 1992*. Ed. by Derek Sleeman and Peter Edwards. San Francisco (CA): Morgan Kaufmann, 1992, pp. 102–110. ISBN: 978-1-55860-247-2. DOI: https://doi.org/10.1016/B978-1-55860-247-2.50018-8. URL: https://www.sciencedirect.com/science/article/pii/B9781558602472500188.

[65] Christian W. Omlin and Sean Snyders. "Inductive bias strength in knowledge-based neural networks: application to magnetic resonance spectroscopy of breast tissues". In: *Artificial Intelligence in Medicine* 28.2 (2003). Knowledge-based Neurocomputing in Medicine, pp. 121–140. ISSN: 0933-3657. DOI: https://doi.org/10.1016/S0933-3657(03)00062-9. URL: https://www.sciencedirect.com/science/article/pii/S0933365703000629.

[66] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. "A Global Geometric Framework for Nonlinear Dimensionality Reduction". In: *Science* 290.5500 (2000), p. 2319.

[67] Leland McInnes, John Healy, and James Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. [Reference implementation available at http://github.com/lmcinnes/umap.] 2018. DOI: 10.48550/ARXIV.1802.03426. URL: https://arxiv.org/abs/1802.03426.

[68] Aniss Aiman Medbouhi. *Towards topology-aware Variational Auto-Encoders: from InvMap-VAE to Witness Simplicial VAE. Master thesis, KTH Royal Institute of Technology, Sweden*. 2022. URL: https://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akth%3Adiva-309487.

[69] Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction*. American Mathematical Society, 2010, pp. I–XII, 1–241. ISBN: 978-0-8218-4925-5.

[70] Vin de Silva and Gunnar Carlsson. "Topological estimation using witness complexes". In: *IEEE Symposium on Point-based Graphic* (2004), pp. 157–166.

[71] J.D. Gallego-Posada. *Simplicial AutoEncoders: A connection between Algebraic Topology and Probabilistic Modelling. Master thesis, University of Amsterdam, Netherlands*. 2018.

[72] Jean Bourgain. "On lipschitz embedding of finite metric spaces in Hilbert space". In: *Israel Journal of Mathematics* 52 (1985), pp. 46–52.

[73] N. Linial, E. London, and Y. Rabinovich. "The geometry of graphs and some of its algorithmic applications". In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. 1994, pp. 577–591.

[74] Jiří Matoušek. "On embedding trees into uniformly convex Banach spaces". In: *Israel Journal of Mathematics* 114 (1999), pp. 221–237. DOI: 10.1007/BF02785579.

[75] Nathan Linial and Michael E. Saks. "The Euclidean Distortion of Complete Binary Trees". In: *Discrete & Computational Geometry* 29 (2002), pp. 19–21. URL: https://api.semanticscholar.org/CorpusID:33364229.

[76] M. Gromov. “Hyperbolic Groups”. In: *Essays in Group Theory*. Ed. by S. M. Gersten. New York, NY: Springer New York, 1987, pp. 75–263. ISBN: 978-1-4613-9586-7.

[77] B.H. Bowditch. *A Course on Geometric Group Theory*. MSJ memoirs. Mathematical Society of Japan, 2006. ISBN: 9784931469358.

[78] Aaron B. Adcock, Blair D. Sullivan, and Michael W. Mahoney. “Tree-Like Structure in Large Social and Information Networks”. In: *2013 IEEE 13th International Conference on Data Mining*. 2013, pp. 1–10.

[79] Hervé Fournier, Anas Ismail, and Antoine Vigneron. “Computing the Gromov hyperbolicity of a discrete metric space”. In: *Information Processing Letters* 115.6 (2015), pp. 576–579. ISSN: 0020-0190.

[80] Silvere Bonnabel. “Stochastic Gradient Descent on Riemannian Manifolds”. In: *IEEE Transactions on Automatic Control* 58.9 (Sept. 2013), pp. 2217–2229. ISSN: 1558-2523. DOI: 10.1109/tac.2013.2254619. URL: http://dx.doi.org/10.1109/TAC.2013.2254619.

[81] Maximillian Nickel and Douwe Kiela. “Poincaré embeddings for learning hierarchical representations”. In: *Advances in neural information processing systems* 30 (2017).

[82] Gary Becigneul and Octavian-Eugen Ganea. “Riemannian Adaptive Optimization Methods”. In: *International Conference on Learning Representations*. 2019. URL: https://openreview.net/forum?id=r1eiqi09K7.

[83] John Lamping, Ramana Rao, and Peter Pirolli. “A focus+context technique based on hyperbolic geometry for visualizing large hierarchies”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’95. Denver, Colorado, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 401–408. ISBN: 0201847051. DOI: 10.1145/223904.223956. URL: https://doi.org/10.1145/223904.223956.

[84] Jonathan Bingham and Sucha Sudarsanam. “Visualizing large hierarchical clusters in hyperbolic space”. In: *Bioinformatics* 16.7 (July 2000), pp. 660–661. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/16.7.660. eprint: https://academic.oup.com/bioinformatics/article-pdf/16/7/660/48836038/bioinformatics_16_7_660.pdf. URL: https://doi.org/10.1093/bioinformatics/16.7.660.

[85] M. Ángeles Serrano, Dmitri Krioukov, and Marián Boguñá. “Self-Similarity of Complex Networks and Hidden Metric Spaces”. In: *Phys. Rev. Lett.* 100 (7 Feb. 2008), p. 078701. DOI: 10.1103/PhysRevLett.100.078701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.078701.

[86] Jörg A Walter. “H-MDS: a new approach for interactive visualization with multidimensional scaling in the hyperbolic space”. In: *Information Systems* 29.4 (2004). Knowledge Discovery and Data Mining (KDD 2002), pp. 273–292. ISSN: 0306-4379. DOI: https://doi.org/10.1016/j.is.2003.10.002. URL: https://www.sciencedirect.com/science/article/pii/S0306437903001005.

[87] Frederic Sala, Chris De Sa, Albert Gu, et al. “Representation Tradeoffs for Hyperbolic Embeddings”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 4460–4469.

[88] Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. “Hyperbolic neural networks”. In: *Advances in neural information processing systems* 31 (2018).

[89] Ondrej Skopek, Octavian-Eugen Ganea, and Gary Bécigneul. “Mixed-curvature Variational Autoencoders”. In: *International Conference on Learning Representations*. 2020.

[90] Caglar Gulcehre, Misha Denil, Mateusz Malinowski, et al. “Hyperbolic Attention Networks”. In: *International Conference on Learning Representations*. 2019. URL: https://openreview.net/forum?id=rJxHsjRqFQ.

[91] Weize Chen, Xu Han, Yankai Lin, et al. “Fully Hyperbolic Neural Networks”. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Smaranda Muresan, Preslav Nakov, and Aline Villavicencio. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 5672–5686. DOI: 10.18653/v1/2022.acl-long.389. URL: https://aclanthology.org/2022.acl-long.389/.

[92] Ryohei Shimizu, YUSUKE Mukuta, and Tatsuya Harada. “Hyperbolic Neural Networks++”. In: *International Conference on Learning Representations*. 2021. URL: https://openreview.net/forum?id=Ec85b0tUwbA.

[93] Menglin Yang, Harshit Verma, Delvin Ce Zhang, et al. “Hypformer: Exploring Efficient Transformer Fully in Hyperbolic Space”. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. KDD ’24. Barcelona, Spain: Association for Computing Machinery, 2024, pp. 3770–3781. ISBN: 9798400704901. DOI: 10.1145/3637528.3672039. URL: https://doi.org/10.1145/3637528.3672039.

[94] Neil He, Rishabh Anand, Hiren Madhu, et al. *HELM: Hyperbolic Large Language Models via Mixture-of-Curvature Experts*. 2025. arXiv: 2505.24722 [cs.LG]. URL: https://arxiv.org/abs/2505.24722.

[95] Neil He, Hiren Madhu, Ngoc Bui, et al. “Hyperbolic Deep Learning for Foundation Models: A Survey”. In: *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*. KDD ’25. ACM, Aug. 2025, pp. 6021–6031. DOI: 10.1145/3711896.3736564. URL: http://dx.doi.org/10.1145/3711896.3736564.

[96] Valentin Khrulkov and Ivan Oseledets. “Geometry Score: A Method For Comparing Generative Adversarial Networks”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 2621–2629.

[97] Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, et al. “Improved Precision and Recall Metric for Assessing Generative Models”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, et al. Vol. 32. Curran Associates, Inc., 2019.

[98] Petra Poklukar, Anastasiia Varava, and Danica Kragic. “GeomCA: Geometric Evaluation of Data Representations”. In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 18–24 Jul 2021, pp. 8588–8598.

[99] Petra Poklukar, Vladislav Polianskii, Anastasiia Varava, et al. “Delaunay Component Analysis for Evaluation of Data Representations”. In: *International Conference on Learning Representations*. 2022.

[100] Frank Nielsen and Richard Nock. “Hyperbolic Voronoi Diagrams Made Easy”. In: *2010 International Conference on Computational Science and Its Applications*. 2010, pp. 74–80.

[101] Andre Olsson, Meenakshi Venkat, Viren Chaudhri, et al. “Single-cell analysis of mixed-lineage states leading to a binary cell fate choice”. In: *Nature* 537 (Aug. 2016).

[102] Franziska Paul, Ya’ara Arkin, Amir Giladi, et al. “Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors”. In: *Cell* 163 (Nov. 2015).

[103] Sergio Triana, Dominik Vonficht, Lea Jopp-Saile, et al. “Single-cell proteogenomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states”. In: *Nature Immunology* 22 (Dec. 2021). DOI: 10.1038/s41590-021-01059-0.

[104] Peter van Galen, Volker Hovestadt, Marc H. Wadsworth II, et al. “Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity”. In: *Cell* 176.6 (2019), 1265–1281.e24. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2019.01.031. URL: https://www.sciencedirect.com/science/article/pii/S0092867419300947.

[105] Matteo Naldini, Gabriele Casirati, Matteo Barcella, et al. “Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia”. In: *Nature Communications* 14 (Mar. 2023). DOI: 10.1038/s41467-023-36969-0.

[106] Susan S Schiffman. “Contributions to the physicochemical dimensions of odor: a psychophysical approach”. In: *Annals of the New York Academy of Sciences* 237.1 (1974), pp. 164–183.

[107] Amir Madany Mamlouk and Thomas Martinetz. “On the dimensions of the olfactory perception space”. In: *Neurocomputing* 58-60 (2004). Computational Neuroscience: Trends in Research 2004, pp. 1019–1025. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2004.01.161. URL: https://www.sciencedirect.com/science/article/pii/S0925231204001663.

[108] Alexei Koulakov, Brian E Kolterman, Armen Enikolopov, et al. “In search of the structure of human olfactory space”. In: *Frontiers in systems neuroscience* 5 (2011), p. 9271.

[109] Marcelo O. Magnasco, Andreas Keller, and Leslie B. Vosshall. “On the dimensionality of olfactory space”. In: *bioRxiv* (2015). DOI: 10.1101/022103. eprint: https://www.biorxiv.org/content/early/2015/07/06/022103.full.pdf. URL: https://www.biorxiv.org/content/early/2015/07/06/022103.

[110] Markus Meister. “On the dimensionality of odor space”. In: *eLife* 4 (July 2015). Ed. by Alexander Borst, e07865. ISSN: 2050-084X. DOI: 10.7554/eLife.07865. URL: https://doi.org/10.7554/eLife.07865.

[111] Kathrin Kaeppler and Friedrich Mueller. “Odor classification: a review of factors influencing perception-based odor arrangements”. In: *Chemical senses* 38.3 (2013), pp. 189–209.

[112] Brian K. Lee, Emily J. Mayhew, Benjamin Sanchez-Lengeling, et al. “A principal odor map unifies diverse tasks in olfactory perception”. In: *Science* 381.6661 (2023), pp. 999–1006. DOI: 10.1126/science.ade4401. eprint: https://www.science.org/doi/pdf/10.1126/science.ade4401. URL: https://www.science.org/doi/abs/10.1126/science.ade4401.

[113] Akshay Sajan, Stijn Sluis, Reza Haydarlou, et al. “Hierarchies of Smell: Structuring the molecular odor space using Semantic Taxonomies and machine learning”. In: *Chemical Senses* (July 2026), bjag020. ISSN: 1464-3553. DOI: 10.1093/chemse/bjag020. eprint: https://academic.oup.com/chemse/advance-article-pdf/doi/10.1093/chemse/bjag020/68718409/bjag020.pdf. URL: https://doi.org/10.1093/chemse/bjag020.

[114] Yuansheng Zhou, Brian H. Smith, and Tatyana O. Sharpee. “Hyperbolic geometry of the olfactory space”. In: *Science Advances* 4.8 (2018), eaaq1458. DOI: 10.1126/sciadv.aaq1458. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aaq1458. URL: https://www.science.org/doi/abs/10.1126/sciadv.aaq1458.

[115] Vivek Sagar, Laura K Shanahan, Christina M Zelano, et al. “High-precision mapping reveals the structure of odor coding in the human brain”. In: *Nature neuroscience* (2023), pp. 1–8.

[116] Aryan Amit Barsainyan, Ritesh Kumar, Pinaki Saha, et al. *OpenPOM*. 2023. URL: https://github.com/ARY2260/openpom.

[117] Ernest C Crocker and LF Henderson. *Analysis and classification of odors: an effort to develop a workable method*. Robbins Perfumer Company, 1927.

[118] Rehan M. Khan, Chung-Hay Luk, Adeen Flinker, et al. “Predicting Odor Pleasantness from Odorant Structure: Pleasantness as a Reflection of the Physical World”. In: *The Journal of Neuroscience* 27 (2007), pp. 10015–10023. URL: https://api.semanticscholar.org/CorpusID:13710261.

[119] Kobi Snitz, Adi Yablonka, Tali Weiss, et al. “Predicting odor perceptual similarity from odor structure”. In: *PLoS computational biology* 9.9 (2013), e1003184.

[120] Clara C. Licon, Céline Manesse, Morgane Dantec, et al. “Pleasantness and trigeminal sensations as salient dimensions in organizing the semantic and physiological spaces of odors”. In: *Scientific Reports* 8 (2018), p. 8444. DOI: 10.1038/s41598-018-26510-5.

[121] D. Cieslik. *Shortest Connectivity: An Introduction with Applications in Phylogeny*. Combinatorial Optimization. Springer US, 2006. ISBN: 9780387235394. URL: https://books.google.se/books?id=4E0r3oWkn6AC.

[122] Wuming Gong, Alejandro A. Granados, Jingyuan Hu, et al. “Benchmarked approaches for reconstruction of in vitro cell lineages and in silico models of C. elegans and M. musculus developmental trees”. In: *Cell Systems* 12.8 (2021), 810–826.e4. ISSN: 2405-4712. DOI: https://doi.org/10.1016/j.

cels.2021.05.008. URL: https://www.sciencedirect.com/science/article/pii/S2405471221001940.

[123] Robert R. Sokal and Charles Duncan Michener. "A statistical method for evaluating systematic relationships". In: *University of Kansas science bulletin* 38 (1958), pp. 1409–1438. URL: https://api.semanticscholar.org/CorpusID:61950873.

[124] N Saitou and M Nei. "The neighbor-joining method: a new method for reconstructing phylogenetic trees." In: *Molecular Biology and Evolution* 4.4 (July 1987), pp. 406–425. ISSN: 0737-4038. DOI: 10.1093/oxfordjournals.molbev.a040454. eprint: https://academic.oup.com/mbe/article-pdf/4/4/406/11167444/7sait.pdf. URL: https://doi.org/10.1093/oxfordjournals.molbev.a040454.

[125] Walter M. Fitch. "Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology". In: *Systematic Biology* 20.4 (Dec. 1971), pp. 406–416. ISSN: 1063-5157. DOI: 10.1093/sysbio/20.4.406. eprint: https://academic.oup.com/sysbio/article-pdf/20/4/406/4697394/20-4-406.pdf. URL: https://doi.org/10.1093/sysbio/20.4.406.

[126] J Felsenstein. "Evolutionary trees from DNA sequences: a maximum likelihood approach". In: *Journal of molecular evolution* 17.6 (1981), pp. 368–376. ISSN: 0022-2844. DOI: 10.1007/bf01734359. URL: https://doi.org/10.1007/BF01734359.

[127] John P. Huelsenbeck, Fredrik Ronquist, Rasmus Nielsen, et al. "Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology". In: *Science* 294.5550 (2001), pp. 2310–2314. ISSN: 00368075, 10959203. URL: http://www.jstor.org/stable/3085235 (visited on 12/18/2023).

[128] Les Foulds, Michael Hendy, and David Penny. "A graph theoretic approach to the development of minimal phylogenetic trees". In: *Journal of molecular evolution* 13 (Aug. 1979), pp. 127–49. DOI: 10.1007/BF01732868.

[129] Takahiro Mimori and Michiaki Hamada. "GeoPhy: Differentiable Phylogenetic Inference via Geometric Gradients of Tree Topologies". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh, T. Naumann, A. Globerson, et al. Vol. 36. Curran Associates, Inc., 2023, pp. 36591–36619. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/732c5757aa5577de9b103332cf7ac0bf-Paper-Conference.pdf.

[130] Matthew Macaulay and Mathieu Fourment. "Differentiable phylogenetics via hyperbolic embeddings with Dodonaphy". In: *Bioinformatics Advances* 4.1 (June 2024), vbae082. ISSN: 2635-0041. DOI: 10.1093/bioadv/vbae082. eprint: https://academic.oup.com/bioinformaticsadvances/article-pdf/4/1/vbae082/58778658/vbae082.pdf. URL: https://doi.org/10.1093/bioadv/vbae082.

[131] Alex Chen, Philippe Chlenski, Kenneth Munyuza, et al. "Variational Combinatorial Sequential Monte Carlo for Bayesian Phylogenetics in Hyperbolic Space". In: *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*. Ed. by Yingzhen Li, Stephan Mandt, Shipra Agrawal, et al. Vol. 258. Proceedings of Machine Learning Research. PMLR, Mar. 2025,

pp. 2962–2970. URL: https://proceedings.mlr.press/v258/chen25f.html.

[132] Nobuhiro Innami and Byung Hak Kim. “Steiner ratio for hyperbolic surfaces”. In: *Proceedings of the Japan Academy, Series A, Mathematical Sciences* 82.6 (June 2006). Publisher: The Japan Academy, pp. 77–79. ISSN: 0386-2194. DOI: 10.3792/pjaa.82.77. (Visited on 08/02/2023).

[133] Michael R Garey, Ronald L Graham, and David S Johnson. “The complexity of computing Steiner minimal trees”. In: *SIAM journal on applied mathematics* 32.4 (1977), pp. 835–859.

[134] Dietmar Cieslik. “The Steiner ratio of metric spaces”. In: *The Steiner Ratio* (2001), pp. 79–98.

[135] Maurice Hanan. “On Steiner’s problem with rectilinear distance”. In: *SIAM Journal on Applied mathematics* 14.2 (1966), pp. 255–265.

[136] GY Yan, A Albrecht, GHF Young, et al. “The Steiner tree problem in orientation metrics”. In: *journal of computer and system sciences* 55.3 (1997), pp. 529–546.

[137] Alexandr Olegovich Ivanov, Alexey Avgustinovich Tuzhilin, and Dietmar Cieslik. “Steiner ratio for Riemannian manifolds”. In: *Russian Mathematical Surveys* 55.6 (2000), p. 1150.

[138] Andrew E Logan. *The Steiner Problem on Closed Surfaces of Constant Curvature*. Brigham Young University, 2015.

[139] M. Zachariasen and P. Winter. “Concatenation-Based Greedy Heuristics for the Euclidean Steiner Tree Problem”. en. In: *Algorithmica* 25.4 (Aug. 1999), pp. 418–437. ISSN: 1432-0541. DOI: 10.1007/PL00009287. URL: https://doi.org/10.1007/PL00009287 (visited on 08/10/2023).

[140] J MacGregor Smith, DT Lee, and Judith S Liebman. “An O (n log n) heuristic for Steiner minimal tree problems on the Euclidean metric”. In: *Networks* 11.1 (1981), pp. 23–39.

[141] John Dolan, Richard Weiss, and J MacGregor Smith. “Minimal length tree networks on the unit sphere”. In: *Annals of Operations Research* 33 (1991), pp. 501–535.

[142] Denise Halverson and D March. “Steiner tree constructions in hyperbolic space”. In: *preprint* (2005).

[143] J. F. Weng. “Steiner Trees on Curved Surfaces”. In: *Graphs and Combinatorics* 17 (2001), pp. 353–363.

[144] M. J. D. Powell. “A Hybrid Method for Nonlinear Equations”. In: *Numerical Methods for Nonlinear Algebraic Equations*. Ed. by P. Rabinowitz. Gordon and Breach, 1970.

[145] Tianran Chen and Tien-Yien Li. “Homotopy continuation method for solving systems of nonlinear and polynomial equations”. In: *Communications in Information and Systems* 15.2 (2015), pp. 119–307.

[146] J. E. Beasley and F. Goffinet. “A delaunay triangulation-based heuristic for the euclidean steiner problem”. In: *Networks* 24.4 (1994), pp. 215–224. DOI: https://doi.org/10.1002/net.3230240404. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/net.3230240404. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/net.3230240404.

[147] Jon Laarhoven and Jeffrey Ohlmann. “A randomized Delaunay triangulation heuristic for the Euclidean Steiner tree problem in $\mathbb{R}^d$”. In: *J. Heuristics* 17 (Aug. 2011), pp. 353–372. DOI: 10.1007/s10732-010-9137-z.

[148] E. N. Gilbert and H. O. Pollak. “Steiner Minimal Trees”. In: *SIAM Journal on Applied Mathematics* 16.1 (1968), pp. 1–29. DOI: 10.1137/0116001. eprint: https://doi.org/10.1137/0116001. URL: https://doi.org/10.1137/0116001.

[149] Yisi Ke, Tianyu Huang, Yankai Shu, et al. “Towards Solving the Gilbert-Pollak Conjecture via Large Language Models”. In: *Forty-third International Conference on Machine Learning*. 2026. URL: https://openreview.net/forum?id=Cqq3HGRIuM.

[150] F. Chung and R. Graham. “A new bound for Euclidean Steiner minimal trees”. In: *Annals of the New York Academy of Sciences* 440 (Dec. 2006), pp. 328–346. DOI: 10.1111/j.1749-6632.1985.tb14564.x.

[151] Sándor Kisfaludi-Bak, Saeed Odak, Satyam Singh, et al. “Gap-ETH-Tight Algorithms for Hyperbolic TSP and Steiner Tree”. In: *42nd International Symposium on Computational Geometry (SoCG 2026)*. Ed. by Hee-Kap Ahn, Michael Hoffmann, and Amir Nayyeri. Vol. 367. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026, 64:1–64:17. ISBN: 978-3-95977-418-5. DOI: 10.4230/LIPIcs.SoCG.2026.64. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.SoCG.2026.64.

Appended papers