Persistence Theory: From Quiver Representations to Data Analysis

Steve Y. Oudot
Persistence Theory: From Quiver Representations to Data Analysis
Persistence Theory: From Quiver Representations to Data Analysis

Steve Y. Oudot
# Contents

Preface vii

Introduction 1

**Part 1. Theoretical Foundations** 11

Chapter 1. Algebraic Persistence 13

1. A quick walk through the theory of quiver representations 14
2. Persistence modules and interval decompositions 19
3. Persistence barcodes and diagrams 21
4. Extension to interval-indecomposable persistence modules 25
5. Discussion 26

Chapter 2. Topological Persistence 29

1. Topological constructions 29
2. Calculations 39

Chapter 3. Stability 49

1. Metrics 50
2. Proof of the stability part of the Isometry Theorem 54
3. Proof of the converse stability part of the Isometry Theorem 60
4. Discussion 61

**Part 2. Applications** 65

Chapter 4. Topological Inference 67

1. Inference using distance functions 71
2. From offsets to filtrations 78
3. From filtrations to simplicial filtrations 81

Chapter 5. Topological Inference 2.0 85

1. Simple geometric predicates 88
2. Linear size 92
3. Scaling up with the intrinsic dimensionality of the data 95
4. Side-by-side comparison 104
5. Natural images 106
6. Dealing with outliers 110

Chapter 6. Clustering 115

1. Contributions of persistence 117
2. ToMATo 118
3. Theoretical guarantees 122
4. Experimental results 126
5. Higher-dimensional structure 129

Chapter 7. Signatures for Metric Spaces 133
1. Simplicial filtrations for arbitrary metric spaces 138
2. Stability for finite metric spaces 140
3. Stability for totally bounded metric spaces 142
4. Signatures for metric spaces equipped with functions 146
5. Computations 146

Part 3. Perspectives 153

Chapter 8. New Trends in Topological Data Analysis 155
1. Optimized inference pipeline 156
2. Statistical topological data analysis 158
3. Topological data analysis and machine learning 160

Chapter 9. Further prospects on the theory 163
1. Persistence for other types of quivers 163
2. Stability for zigzags 165
3. Simplification and reconstruction 165

Appendix A. Introduction to Quiver Theory with a View Toward Persistence 167
1. Quivers 168
2. The category of quiver representations 168
3. Classification of quiver representations 170
4. Reflections 175
5. Proof of Gabriel’s theorem: the general case 186
6. Beyond Gabriel’s theorem 191

Bibliography 197

List of Figures 213

Index 217
Preface

It is in the early 2000’s that persistence emerged as a new theory in the field of applied and computational topology. This happened mostly under the impulsion of two schools: the one led by H. Edelsbrunner and J. Harer at Duke University, the other led by G. Carlsson at Stanford University. After more than a decade of a steady development, the theory has now reached a somewhat stable state, and the community of researchers and practitioners gathered around it has grown in size from a handful of people to a couple hundreds. In other words, persistence has become a mature research topic.

The existing books and surveys on the subject \[48, 114, 115, 119, 141, 245\] are largely built around the topological aspects of the theory, and for particular instances such as the persistent homology of the family of sublevel sets of a Morse function on a compact manifold. While this can be useful for developing intuition, it does create bias in how the subject is understood. A recent monograph [72] tries to correct this bias by focusing almost exclusively on the algebraic aspects of the theory, and in particular on the mathematical properties of persistence modules and of their diagrams.

The goal pursued in the present book is to put the algebraic part back into context, to give a broad view of the theory including also its topological and algorithmic aspects, and to elaborate on its connections to quiver theory on the one hand, to data analysis on the other hand. While the subject cannot be treated with the same level of detail as in [72], the book still describes and motivates the main concepts and ideas, and provides sufficient insights into the proofs so the reader can understand the mechanisms at work.

Throughout the exposition I will be focusing on the currently most stable instance of the theory: 1-dimensional persistence. Other instances, such as multidimensional persistence or persistence indexed over general partially ordered sets, are comparatively less well understood and will be mentioned in the last part of the book as directions for future research. The background material on quiver theory provided in Chapter 1 and Appendix A should help the reader understand the challenges associated with them.

Reading guidelines. There are three parts in the book. The first part (Chapters 1 through 3 and Appendix A) focuses on the theoretical foundations of persistence. The second part (Chapters 4 through 7) deals with a selected set of
applications. The third part (Chapters 8 and 9) talks about future prospects for both the theory and its applications. The document has been designed in the hope that it can provide something to everyone among our community, as well as to newcomers with potentially different backgrounds:

- Readers with a bias towards mathematical foundations and structure theorems will find the current state of knowledge about the decomposability of persistence modules in Chapter 1 and about the stability of their diagrams in Chapter 3. To those who are curious about the connections between persistence and quiver theory, I recommend reading Appendix A.
- Readers with a bias towards algorithms will find a survey of the methods used to compute persistence in Chapter 2 and a thorough treatment of the algorithmic aspects of the applications considered in Part 2.
- Practitioners in applied fields who want to learn about persistence in general will find a comprehensive yet still accessible exposition spanning all aspects of the theory, including its connections to some applications. To those I recommend the following walk through Part 1 of the document:
  a) The general introduction,
  b) Sections 1.1 through 3.3 of Chapter 1
  c) Sections 1.1 and 2.1 of Chapter 2
  d) Sections 1, 2.1 and 4 of Chapter 3.

Then, they can safely read Parts 2 and 3.

For the reader’s convenience, the introduction of each chapter in Parts 1 and 2 mentions the prerequisites for reading the chapter and provides references to the relevant literature. As a general rule, I would recommend reading [115] or [142] prior to this book, as these references give quite accessible introductions to the field of applied and computational topology.

Acknowledgements. First of all, I want to express my gratitude towards the people who have contributed to shape persistence theory as we know it today. Among them, let me thank my co-authors, with whom I had an exciting time developing some of the ideas presented in this book: Jean-Daniel Boissonnat, Mickaël Buchet, Mathieu Carrière, Frédéric Chazal, David Cohen-Steiner, Vin de Silva, Jie Gao, Marc Glisse, Leonidas Guibas, Benoît Hudson, Clément Maria, Facundo Mémoli, Gary Miller, Maksim Ovsjanikov, Donald Sheehy, Primoz Skraba, and Yue Wang.

Second, I want to thank the people who have helped me design the book and improve its content. Among them, my gratitude goes primarily to Michael Lesnick, for his careful reading of early versions of the manuscript and for his insightful comments that greatly helped improve Part 1 and Appendix A. I am also grateful to the anonymous referees, who provided me with valuable feedback on the flow of the book and on its readability. I also want to thank the people who have proofread excerpts from the manuscript and helped me improve the content and exposition locally: Eddie Aamari, Jean-Daniel Boissonnat, Frédéric Chazal, Jérémy Cochoy, Pawel Dlotko, Marc Glisse, Bertrand Michel. Let me apologize in advance to those whose names I may have forgotten in this list.

Finally, I want to thank Sergei Gelfand, Christine Thivierge, and the American Mathematical Society for their interest in the book and for their support to finalize it.

Palaiseau, June 2015
Bibliography

198 BIBLIOGRAPHY


[90] É. Colin de Verdière, G. Ginot, and X. Goaoc. “Helly numbers of acyclic families”. In: Advances in Mathematics 253 (2014), pp. 163–193 (cit. on pp. [161] [162]).


[95] W. Crawley-Boevey. Personal notes, published as part of [63]. 2013 (cit. on pp. [27] [28]).


[139] I. M. Gelfand and V. A. Ponomarev. “Problems of linear algebra and classification of quadruples of subspaces in a finite-dimensional vector space”. In:


## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>A planar point set with several underlying geometric structures</td>
<td>1</td>
</tr>
<tr>
<td>0.2</td>
<td>The dendrogram produced by single-linkage on the data set of Figure 0.1</td>
<td>2</td>
</tr>
<tr>
<td>0.3</td>
<td>The barcode produced by persistence on the data set of Figure 0.1</td>
<td>2</td>
</tr>
<tr>
<td>0.4</td>
<td>A collection of images from the Columbia Object Image Library</td>
<td>3</td>
</tr>
<tr>
<td>0.5</td>
<td>A smooth function and a piecewise linear approximation</td>
<td>4</td>
</tr>
<tr>
<td>1.1</td>
<td>The four decorated points corresponding to intervals $[b_j^\pm, d_j^\pm]$</td>
<td>22</td>
</tr>
<tr>
<td>1.2</td>
<td>A classical example in persistence theory</td>
<td>22</td>
</tr>
<tr>
<td>1.3</td>
<td>A decorated point $(b^\pm, d^\pm)$ belonging to a rectangle $R$</td>
<td>23</td>
</tr>
<tr>
<td>1.4</td>
<td>Additivity of $\mu_V$ under vertical and horizontal splitting</td>
<td>24</td>
</tr>
<tr>
<td>1.5</td>
<td>A quadrant, horizontal strip, vertical strip, and finite rectangle</td>
<td>24</td>
</tr>
<tr>
<td>2.1</td>
<td>A simplicial filtration of the octahedron</td>
<td>31</td>
</tr>
<tr>
<td>2.2</td>
<td>The full persistence diagram of the height function from Example 1.11</td>
<td>33</td>
</tr>
<tr>
<td>2.3</td>
<td>The extended persistence diagram of the height function from Example 1.11</td>
<td>35</td>
</tr>
<tr>
<td>2.4</td>
<td>The pyramid for $n = 3$</td>
<td>38</td>
</tr>
<tr>
<td>2.5</td>
<td>Run of the matrix reduction algorithm on a filtration of a solid triangle</td>
<td>42</td>
</tr>
<tr>
<td>2.6</td>
<td>A filtered planar graph and the spanning tree of negative edges</td>
<td>47</td>
</tr>
<tr>
<td>3.1</td>
<td>A smooth function, a piecewise linear approximation, and their diagrams</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>The snapping rule</td>
<td>54</td>
</tr>
<tr>
<td>3.3</td>
<td>Tracking down point $p \in \text{dgm}(V)$ through its successive matchings</td>
<td>56</td>
</tr>
<tr>
<td>3.4</td>
<td>Identifying $\phi, \psi$ with the families of vertical and horizontal morphisms</td>
<td>58</td>
</tr>
<tr>
<td>3.5</td>
<td>Effects of various interpolation parameters on the persistence diagrams</td>
<td>60</td>
</tr>
<tr>
<td>4.1</td>
<td>Distribution of Matter in the Universe</td>
<td>68</td>
</tr>
<tr>
<td>4.2</td>
<td>Sampling of a curve winding around a torus in $\mathbb{R}^3$</td>
<td>69</td>
</tr>
<tr>
<td>4.3</td>
<td>Examples of medial axes</td>
<td>71</td>
</tr>
<tr>
<td>4.4</td>
<td>Some compact sets with positive reach</td>
<td>73</td>
</tr>
<tr>
<td>4.5</td>
<td>The generalized gradient of the distance to $K$</td>
<td>74</td>
</tr>
<tr>
<td>4.6</td>
<td>Generalized gradient of $d_K$ and its associated flow</td>
<td>74</td>
</tr>
<tr>
<td>4.7</td>
<td>A compact set that is not homotopy equivalent to its small offsets</td>
<td>76</td>
</tr>
<tr>
<td>4.8</td>
<td>Example with no homologically valid offset of the point cloud</td>
<td>76</td>
</tr>
</tbody>
</table>
4.9 Barcode of the offsets filtration of the point cloud from Figure 4.2 79
4.10 The corresponding persistence diagrams 80
4.11 Union of balls and corresponding Čech complex 81

5.1 Truncated barcode of the Čech filtration on the Clifford data set 86
5.2 A point cloud $P$ and its Rips complex $R_{2i}(P)$ 89
5.3 Landmarks set, witnesses set, and corresponding witness complex 90
5.4 Relationship between the offsets and mesh filtrations 93
5.5 Logscale barcode of the mesh filtration on the Clifford data set 95
5.6 Scale-rank plot obtained from the data set of Figure 4.2 96
5.7 Logscale barcode of the oscillating Rips zigzag on the Clifford data set 102
5.8 Logscale barcode of the Morozov zigzag on the Clifford data set 103
5.9 Projecting the natural images data down using dimensionality reduction 107
5.10 Experimental results obtained using the witness filtration 108
5.11 Logscale barcode of the Morozov zigzag on $Q_{k,x}, k = 1, 200, x = 30\%$ 108
5.12 Logscale barcode of the Morozov zigzag on $Q_{k,x}, k = 24,000, x = 30\%$ 109
5.13 Parametrization of the space of high-contrast $3 \times 3$ patches 109
5.14 Sublevel sets of the distance-like functions $d_P$, $d_{\mu_P,m}$, and $d_{\mu_P,m}^P$ 111

6.1 Point cloud, dendrogram, and hierarchy produced by persistence 116
6.2 ToMATo in a nutshell 119
6.3 Separation of $d_{\text{gm}}(f)$ and partial matching with $d_{\text{gm}}(\hat{G})$ 122
6.4 Influence of the partial matching on the regions $D_1$ and $D_2$ 124
6.5 A function $f$ with unstable basins of attraction 125
6.6 Outputs of the algorithm on a uniform sampling of the domain of $f$ 125
6.7 The twin spirals data set processed using a smaller neighborhood radius 126
6.8 Persistence diagrams obtained on the twin spirals data set 127
6.9 Result of spectral clustering on the twin spirals data set 127
6.10 Biological data set 128
6.11 Quantitative evaluation of the quality of the output of ToMATo 129
6.12 The generalized ToMATo in a nutshell 130

7.1 A database of 3d shapes with six classes 134
7.2 Some toy examples 137
7.3 Bottleneck distance matrix, confusion matrix, and MDS embedding 139
7.4 Overview of the proof of Theorem 7.2 140
7.5 Effect of truncating a filtration as per (7.14) on its persistence diagram 147
7.6 An example where no ball of the offset is ever covered 149
7.7 The additive weight function $t \mapsto s_t(x_k)$ associated to point $x_k \in X$ 149

8.1 A cover of the unit circle, and its nerve before and after refinement 161
A.1 Example of a quiver and its underlying undirected graph \hfill 168
A.2 The Dynkin diagrams \hfill 169
A.3 The Euclidean diagrams \hfill 173
A.4 The pyramid travelled down when turning an $A_5$-type quiver into $L_5$ \hfill 186
Index

μ-reach, 77
q(uadrant)-tame, 230 350 354

Artin algebra, 174 193

complex
α-, 83
i-Delaunay, 83
(Vietoris-)Rips, 89 139
Čech, 81 139
nerve, 81
sparse Rips, 150
weighted Rips, 112
witness, 66 139
correspondence, 143
Coxeter functors, 179

decorated
multiset, 25
number, 21
persistence diagram, 22
point, 23
diamond, 129 132
Mayer-Vietoris, 58
principle, 179 189
dimension
doubling, 87
distance
bottleneck, 50
distance-like function, 78
Gromov-Hausdorff, 133 143
Hausdorff, 69
interleaving, 52
power, 111
to compact set, 69
to measure, 110
witnessed k-, 112
duality
EP, 36
elder rule, 120
elder rule, 6 31
endomorphism ring, 171
filter, 81

filtered space, 30
filtration, 29
relative, 31
right-, 184
sublevel-sets, 31
superlevel-sets, 51
truncated, 147

interleaving, 52
multiplicative, 87
sublevel-sets, 92
weak, 55
interpolation lemma, 57
interval, 20
(in)decomposable, 21
decomposition, 20
module, 20
representation, 174 175
rule, 21

local endomorphism ring, 171

medial axis, 71

morphism
of degree ε, 63
of representations, 18
Morse-type function, 37
multivalued map, 142

natural images, 109
nerve, 51
lemma, 51
persistent, 52

observable category, 28 63
offset, 63
outliers, 110

partial matching, 50
path, 192
algebra, 171 192

persistence
barcode, 22 26
diagram, 22 26
(d1, d2)-separated, 120

217
extended, 34
hierarchy, 2, 44, 116
kernels, images, cokernels, 39
measure, 23
module, 20
streamlined, 183
zigzag, 39
persistent
(co-)homology, 30, 52, 53
(co-)homology group, 30
poset representation, 19, 194
pyramid, 37, 185
theorem, 39
quiver, 14, 168
A_n-type, 14, 169
A_n-type, 173
\mathbb{N}, \mathbb{Z}, 13, 194
algebra, 19, 192
category, 19, 194
Dynkin, \mathcal{L}_n, 16, 169
Euclidean, 173
linear \mathcal{L}_n, 13, 180
poset, 19, 194
tame, 17, 174, 191
wild, 17, 174, 191
with relations, 19, 195
quiver representation, 14, 168
(in-)decomposable, 16, 170
category, 16, 170
classification, 16, 170
dimension, 170
dimension vector, 170
direct sum, 15
interval, 14, 156
morphism, 15, 169
kern, image, cokernel, 15
shift, 53
pointwise finite-dimensional, 18
trivial, 13
reach, 72
reflection, 175
functor, 176
functors theorem, 178
root, 174
sampling
\varepsilon-sample, 136
furthest-point, 97
subsampling, 95
signatures
persistence-based, 138
snapping
lemma, 53
principle, 54
Sparse Voronoi Refinement, 93
subrepresentation, 16, 168
symmetry
theorem
converse stability, 49
Crawley-Boevey, 19, 195
Gabriel, 17
Gabriel I, 171
Gabriel II, 174
interval decomposition, 20
isometry, 49
Kac, 171
Krull-Remak-Schmidt, 16, 171
pyramid, 39
reflection functors, 178
stability, 19, 61, 63
Webb, 118, 194
time of appearance, 34, 39
Tits form, 172
ToMATo, 118
topological feature, 39
totally bounded space, 180
undecorated
persistence diagram, 22
weak feature size, 75
zigzag, 39
image Rips (iR-ZZ), 98
level-sets, 57
manipulations, 99
module, 20
Morozov (M-ZZ), 98
oscillating Rips (oR-ZZ), 98
Selected Published Titles in This Series

209 Steve Y. Oudot, Persistence Theory: From Quiver Representations to Data Analysis, 2015
205 Pavel Etingof, Shlomo Gelaki, Dmitri Nikshych, and Victor Ostrik, Tensor Categories, 2015
204 Victor M. Buchstaber and Taras E. Panov, Toric Topology, 2015
203 Donald Yau and Mark W. Johnson, A Foundation for PROPs, Algebras, and Modules, 2015
201 Christopher L. Douglas, John Francis, André G. Henriques, and Michael A. Hill, Editors, Topological Modular Forms, 2014
200 Nikolai Nadirashvili, Vladimir Tkachev, and Serge Vladuț, Nonlinear Elliptic Equations and Nonassociative Algebras, 2014
198 Jörg Jahnel, Brauer Groups, Tamagawa Measures, and Rational Points on Algebraic Varieties, 2014
197 Richard Evan Schwartz, The Octagonal PETs, 2014
196 Silouanos Brazitikos, Apostolos Giannopoulos, Petros Valettas, and Beatrice-Helen Vritsiou, Geometry of Isotropic Convex Bodies, 2014
195 Ching-Li Chai, Brian Conrad, and Frans Oort, Complex Multiplication and Lifting Problems, 2014
194 Samuel Herrmann, Peter Imkeller, Ilya Pavlyukevich, and Dierk Peithmann, Stochastic Resonance, 2014
193 Robert Rumely, Capacity Theory with Local Rationality, 2013
192 Messoud Efendiev, Attractors for Degenerate Parabolic Type Equations, 2013
191 Grégory Berhuy and Frédérique Oggier, An Introduction to Central Simple Algebras and Their Applications to Wireless Communication, 2013
190 Aleksandr Pukhlikov, Birationally Rigid Varieties, 2013
189 Alberto Elduque and Mikhail Kochetov, Gradings on Simple Lie Algebras, 2013
188 David Lannes, The Water Waves Problem, 2013
186 Gregory Berkolaiko and Peter Kuchment, Introduction to Quantum Graphs, 2013
185 Patrick Iglesias-Zemmour, Diffeology, 2013
184 Frederick W. Gehring and Kari Hag, The Ubiquitous Quasidisk, 2012
180 Martin W. Liebeck and Gary M. Seitz, Unipotent and Nilpotent Classes in Simple Algebraic Groups and Lie Algebras, 2012
179 Stephen D. Smith, Subgroup Complexes, 2011
178 Helmut Brass and Knut Petras, Quadrature Theory, 2011
176 Peter E. Kloeden and Martin Rasmussen, Nonautonomous Dynamical Systems, 2011

For a complete list of titles in this series, visit the AMS Bookstore at www.ams.org/bookstore/survseries/.
Persistence theory emerged in the early 2000s as a new theory in the area of applied and computational topology. This book provides a broad and modern view of the subject, including its algebraic, topological, and algorithmic aspects. It also elaborates on applications in data analysis. The level of detail of the exposition has been set so as to keep a survey style, while providing sufficient insights into the proofs so the reader can understand the mechanisms at work.

The book is organized into three parts. The first part is dedicated to the foundations of persistence and emphasizes its connection to quiver representation theory. The second part focuses on its connection to applications through a few selected topics. The third part provides perspectives for both the theory and its applications. The book can be used as a text for a course on applied topology or data analysis.

For additional information and updates on this book, visit www.ams.org/bookpages/surv-209