

Discrete Calculus

Leo J. Grady • Jonathan R. Polimeni

Discrete Calculus

Applied Analysis on Graphs
for Computational Science

 Springer

Dr. Leo J. Grady
Siemens Corporate Research
755 College Road East
Princeton, NJ 08540-6632
USA
leo.grady@siemens.com

Dr. Jonathan R. Polimeni
Athinoula A. Martinos Center
for Biomedical Imaging
Department of Radiology
Massachusetts General Hospital
Harvard Medical School
Charlestown, MA 02129
USA
jonp@nmr.mgh.harvard.edu

ISBN 978-1-84996-289-6
DOI 10.1007/978-1-84996-290-2
Springer London Dordrecht Heidelberg New York

e-ISBN 978-1-84996-290-2

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2010930985

© Springer-Verlag London Limited 2010

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licenses issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Cover design: KünkelLopka GmbH

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Leo (John) Grady dedicates the book to his very patient wife, Amy Grady, daughter Alexandra Grady, son Leo James Grady (who arrived during the course of writing) and the memory of his late father, Leo Joseph Grady.

Preface

The goal of this book is to present the topic of discrete calculus to scientists and engineers and to show how the theory can be applied to solving a wide variety of real-world problems. We feel that discrete calculus allows us to unify many approaches to data analysis and content extraction while being accessible enough to be widely applied in many fields and disciplines. This project initially began as a tutorial on discrete calculus and its applications, and we hope that this work can provide an introduction to discrete calculus and demonstrate its effectiveness in many problem domains.

This book benefited enormously from the friends and colleagues who provided software, data, and their time in reading the manuscript. In particular, for software we would like to thank Sébastien Bogleux for help with total variation code for an arbitrary graph and Camille Couprie for power watershed code. For use of their data, we wish to thank Jason Bohland, James Fowler, David Gleich, and Robert Sumner. For comments on the manuscript we would like to thank Christopher Alvino, Mukund Balasubramanian, Jason Bohland, Sébastien Bogleux, Gaëlle Desbordes, Mathieu Desbrun, Noha El-Zehiry, Gareth Funka-Lea, David Gleich, Oliver Hinds, Anil Hirani, Hiroshi Ishikawa, Robert Kotiuga, Elliot Saltzman, Dheeraj Singaraju, Ganesh Sundaramoorthi and Enzo Tonti. We thank Eric Schwartz for all his help and support over the years. We would like to thank Wayne Wheeler, Simon Rees, and Catherine Brett at Springer for their enthusiasm and for initially proposing this project. Leo Grady would also like to thank Chenyang Xu for convincing him to take on this project and to Gareth Funka-Lea for his patience and support throughout the writing of this book.

Princeton, New Jersey
Charlestown, Massachusetts

Leo Grady
Jonathan Polimeni

Contents

1	Discrete Calculus: History and Future	1
1.1	Discrete Calculus	1
1.1.1	Origins of Vector Calculus	2
1.1.2	Origins of Discrete Calculus	3
1.1.3	<i>Discrete vs. Discretized</i>	4
1.2	Complex Networks	6
1.3	Content Extraction	7
1.4	Organization of the Book	8
1.5	Intended Audience	8

Part I A Brief Review of Discrete Calculus

2	Introduction to Discrete Calculus	13
2.1	Topology and the Fundamental Theorem of Calculus	14
2.2	Differential Forms	16
2.2.1	Exterior Algebra and Antisymmetric Tensors	17
2.2.2	Differentiation and Integration of Forms	26
2.2.3	The Hodge Star Operator	32
2.2.4	Differential Forms and Linear Pairings	37
2.3	Discrete Calculus	38
2.3.1	Discrete Domains	39
2.3.2	Discrete Forms and the Coboundary Operator	48
2.3.3	Primal and Dual Complexes	50
2.3.4	The Role of a Metric: the Metric Tensor, the Discrete Hodge Star Operator, and Weighted Complexes	54
2.3.5	The Dual Coboundary Operator	59
2.3.6	The Discrete Laplace–de Rham Operator	60
2.4	Structure of Discrete Physical Laws	63
2.5	Examples of Discrete Calculus	63
2.5.1	Fundamental Theorem of Calculus and the Generalized Stokes’ Theorem	65
2.5.2	The Helmholtz Decomposition	73

- 2.5.3 Matrix Representation of Discrete Calculus Identities . . . 77
- 2.5.4 Elliptic Equations 79
- 2.5.5 Diffusion 82
- 2.5.6 Advection 86
- 2.6 Concluding Remarks 88
- 3 Circuit Theory and Other Discrete Physical Models 91**
 - 3.1 Circuit Laws 93
 - 3.2 Steady-State Solutions 94
 - 3.2.1 Dependent Sources 97
 - 3.2.2 Energy Minimization 99
 - 3.3 AC Circuits 101
 - 3.4 Connections Between Circuit Theory and Other Discrete Domains 104
 - 3.4.1 Spring Networks 104
 - 3.4.2 Random Walks 106
 - 3.4.3 Gaussian Markov Random Fields 110
 - 3.4.4 Tree Counting 112
 - 3.4.5 Linear Algebra Applied to Circuit Analysis 117
 - 3.5 Conclusion 121

Part II Applications of Discrete Calculus

- 4 Building a Weighted Complex from Data 125**
 - 4.1 Determining Edges and Cycles 126
 - 4.1.1 Defining an Edge Set 126
 - 4.1.2 Defining a Cycle Set 129
 - 4.2 Deriving Edge Weights 135
 - 4.2.1 Edge Weights to Reflect Geometry 135
 - 4.2.2 Edge Weights to Penalize Data Outliers 136
 - 4.2.3 Edge Weights to Cause Repulsion 143
 - 4.2.4 Edge Weights to Represent Joint Statistics 144
 - 4.2.5 Deducing Edge Weights from Observations 144
 - 4.3 Obtaining Higher-Order Weights to Penalize Outliers 147
 - 4.3.1 Weights Beyond Flows 149
 - 4.4 Metrics Defined on a Complex 150
 - 4.5 Conclusion 153
- 5 Filtering on Graphs 155**
 - 5.1 Fourier and Spectral Filtering on a Graph 156
 - 5.1.1 Graphs that Are Not Shift-Invariant 159
 - 5.1.2 The Origins of High Frequency Noise 162
 - 5.2 Energy Minimization Methods for Filtering 163
 - 5.2.1 The Basic Energy Minimization Model 163
 - 5.2.2 Extended Basic Energy Model 167
 - 5.2.3 The Total Variation Model 168

- 5.3 Filtering with Implicit Discontinuities 170
- 5.4 Filtering with Explicit, but Unknown, Discontinuities 172
- 5.5 Filtering by Gradient Manipulation 174
- 5.6 Nonlocal Filtering 174
- 5.7 Filtering Vectors and Flows 175
 - 5.7.1 Translating Scalar Filtering to Flow Filtering 176
- 5.8 Filtering Higher-Order Cochains 179
- 5.9 Applications 180
 - 5.9.1 Image Processing 180
 - 5.9.2 Three-Dimensional Mesh Filtering 185
 - 5.9.3 Filtering Data on a Surface 187
 - 5.9.4 Geospatial Data 192
 - 5.9.5 Filtering Flow Data—Brain Connectivity 193
- 5.10 Conclusion 197

- 6 Clustering and Segmentation 199**
 - 6.1 Targeted Clustering 200
 - 6.1.1 Primal Targeted Clustering 201
 - 6.1.2 Dual Targeted Clustering 210
 - 6.2 Untargeted Clustering 215
 - 6.2.1 Primal Untargeted Clustering 216
 - 6.2.2 Dual Untargeted Clustering 220
 - 6.3 Semi-targeted Clustering 221
 - 6.3.1 The *k*-Means Model 221
 - 6.4 Clustering Higher-Order Cells 227
 - 6.4.1 Clustering Edges 227
 - 6.5 Applications 229
 - 6.5.1 Image Segmentation 229
 - 6.5.2 Social Networks 235
 - 6.5.3 Machine Learning and Classification 236
 - 6.5.4 Gene Expression 240
 - 6.6 Conclusion 242

- 7 Manifold Learning and Ranking 243**
 - 7.1 Manifold Learning 244
 - 7.1.1 Multidimensional Scaling and *Isomap* 245
 - 7.1.2 Laplacian Eigenmaps and Spectral Coordinates 247
 - 7.1.3 Locality Preserving Projections 249
 - 7.1.4 Relationship to Clustering 251
 - 7.1.5 Manifold Learning on Edge Data 251
 - 7.2 Ranking 253
 - 7.2.1 PageRank 253
 - 7.2.2 HITS 256
 - 7.3 Applications 257

- 7.3.1 Shape Characterization 257
- 7.3.2 Point Correspondence 260
- 7.3.3 Web Search 262
- 7.3.4 Judicial Citation 264
- 7.4 Conclusion 266
- 8 Measuring Networks 267**
 - 8.1 Measures of Graph Connectedness 268
 - 8.1.1 Graph Distance 268
 - 8.1.2 Node Centrality 269
 - 8.1.3 Distance-Based Properties of a Graph 270
 - 8.2 Measures of Graph Separability 274
 - 8.2.1 Clustering Measures 274
 - 8.2.2 Small-World Graphs 277
 - 8.3 Topological Measures 279
 - 8.4 Geometric Measures 281
 - 8.4.1 Discrete Gaussian Curvature 282
 - 8.4.2 Discrete Mean Curvature 283
 - 8.5 Applications 285
 - 8.5.1 Social Networks 285
 - 8.5.2 Chemical Graph Theory 286
 - 8.6 Conclusion 288
- Appendix A Representation and Storage of a Graph and Complex 291**
 - A.1 General Representations for Complexes 291
 - A.1.1 Cells List Representation 291
 - A.1.2 Operator Representation 292
 - A.2 Representation of 1-Complexes 293
 - A.2.1 Neighbor List Representation 293
- Appendix B Optimization 295**
 - B.1 Real-Valued Optimization 296
 - B.1.1 Unconstrained Direct Solutions 297
 - B.1.2 Constrained Direct Solutions 298
 - B.1.3 Descent Methods 308
 - B.1.4 Nonconvex Energy Optimization over Real Variables 312
 - B.2 Integer-Valued Optimization 312
 - B.2.1 Linear Objective Functions 312
 - B.2.2 Quadratic Objective Functions 314
 - B.2.3 General Integer Programming Problems 318
- Appendix C The Hodge Theorem: A Generalization of the Helmholtz
Decomposition 319**
 - C.1 The Helmholtz Theorem 319
 - C.2 The Hodge Decomposition 326

Contents	xiii
Summary of Notation	331
References	335
Index	353
Color Plates	361

Acronyms

AC	Alternating Current
BEM	Basic Energy Model
CDP	Constitutive Determination Problem
CFL	Courant–Friedrichs–Lewy
CRT	Chromatographic Retention Time
CT	Computed Tomography
DC	Direct Current
DFT	Discrete Fourier Transform
DNA	Deoxyribonucleic Acid
fMRI	functional Magnetic Resonance Imaging
GMRF	Gaussian Markov Random Field
GPU	Graphics Processing Unit
HITS	Hyperlink-Induced Topic Search
IRLS	Iteratively Reweighted Least Squares
KCL	Kirchhoff’s Current Law
KM	<i>K</i> -Means
KVL	Kirchhoff’s Voltage Law
LE	Laplacian Eigenmaps
LII	Legal Information Institute
LP	Linear Programming
LPP	Locality Preserving Projections
LTI	Linear Time Invariant
MAP	Maximum <i>A Posteriori</i>
MDS	Multidimensional Scaling
ML	Maximum Likelihood
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
MPCV	MultiPhase Chan–Vese
mRNA	messenger Ribonucleic Acid
MS	Mumford–Shah
MSF	Maximum Spanning Forest

NP-Hard	Non-deterministic Polynomial-Time Hard
PCA	Principle Components Analysis
PDE	Partial Differential Equation
QP	Quadratic Programming
QSAR	Quantitative Structure–Activity Relationships
QSPR	Quantitative Structure–Property Relationships
RGB	Red Green Blue
RLC	Resistor–Inductor–Capacitor
SDP	Semidefinite Programming
SNF	Smith Normal Form
SPD	Symmetric Positive Definite
SPECT	Single Photon Emission Computed Tomography
TV	Total Variation