

DE GRUYTER

Simon Blatt, Philipp Reiter, Armin Schikorra (Eds.)

NEW DIRECTIONS IN GEOMETRIC AND APPLIED KNOT THEORY



PARTIAL DIFFERENTIAL EQUATIONS
AND MEASURE THEORY

Simon Blatt, Philipp Reiter, Armin Schikorra (Eds.)

New Directions in Geometric and Applied Knot Theory

Simon Blatt, Philipp Reiter, Armin Schikorra
(Eds.)

New Directions in Geometric and Applied Knot Theory

Edited by

Series Editor: Gianluca Crippa

Managing Editor: Agnieszka Bednarczyk-Draż

Language Editor: Adam Tod Levertton

DE GRUYTER

ISBN 978-3-11-057148-6

e-ISBN (PDF) 978-3-11-057149-3

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2018 © Simon Blatt, Philipp Reiter, Armin Schikorra and Chapters' Contributors, published by de
Gruyter, Walter de Gruyter GmbH, Berlin/Boston

Cover image: © Henrik Schumacher

☼ Printed on acid-free paper

Printed in Germany

www.degruyter.com

Contents

1	Introduction — 1
---	------------------

Paweł Strzelecki and Heiko von der Mosel

Geometric curvature energies: facts, trends, and open problems — 8

2.1	Facts — 8
2.2	Trends and open problems — 15
	Bibliography — 27

Takeyuki Nagasawa

On Möbius invariant decomposition of the Möbius energy — 36

3.1	O'Hara's knot energies — 36
3.2	Freedman-He-Wang's procedure and the Kusner-Sullivan conjecture — 38
3.3	Basic properties of the Möbius energy — 39
3.4	The Möbius invariant decomposition — 41
3.4.1	The decomposition — 42
3.4.2	Variational formulae — 48
3.4.3	The Möbius invariance — 64
	Bibliography — 75

Henrik Schumacher

Pseudogradient Flows of Geometric Energies — 77

4.1	Introduction — 77
4.2	Banach Bundles — 80
4.2.1	General Fiber Bundles — 80
4.2.2	Banach Bundles and Hilbert Bundles — 83
4.3	Riesz Structures — 84
4.3.1	Riesz Structures — 85
4.3.2	Riesz Bundle Structures — 87
4.3.3	Riesz Manifolds — 89
4.4	Pseudogradient Flow — 90
4.5	Applications — 92
4.5.1	Minimal Surfaces — 92
4.5.2	Elasticae — 97
4.5.3	Euler-Bernoulli Energy and Euler Elastica — 100
4.5.4	Willmore Energy — 101
4.6	Final Remarks — 104
	Bibliography — 107

Sebastian Scholtes

Discrete knot energies — 109

5.1 Introduction — 109

5.1.1 Notation — 110

5.2 Möbius Energy — 111

5.3 Integral Menger Curvature — 114

5.4 Thickness — 116

A.1 Appendix: Postlude in Γ -convergence — 118

Bibliography — 120

Radmila Sazdanovic

Khovanov homology and torsion — 125

6.1 Introduction — 125

6.2 Definition and structure of Khovanov link homology — 126

6.3 Torsion of Khovanov link homology — 129

6.4 Homological invariants of alternating and quasi-alternating cobordisms — 133

Bibliography — 134

Elizabeth Denne

Quadriseccants and essential secants of knots — 138

7.1 Introduction — 138

7.2 Quadriseccants — 139

7.2.1 Essential secants — 140

7.2.2 Results about quadriseccants — 142

7.2.3 Counting quadriseccants and quadriseccant approximations. — 143

7.3 Key ideas in showing quadriseccants exist — 145

7.3.1 Triseccants and quadriseccants. — 146

7.3.2 Structure of the set of triseccants. — 147

7.4 Applications of essential secants and quadriseccants — 148

7.4.1 Total curvature — 148

7.4.2 Second Hull — 149

7.4.3 Ropelength — 150

7.4.4 Distortion — 154

7.4.5 Final Remarks — 154

Bibliography — 155

Gyo Taek Jin

Polygonal approximation of unknots by quadriseccants — 159

8.1 Introduction — 159

8.2 Quadriseccant approximation of knots — 160

8.3 Quadriseccants of Polygonal Unknots — 162

8.4	Quadriseccants of Smooth Unknots —	166
8.5	Finding Quadriseccants —	170
8.6	Test for Good Approximations —	171
	Bibliography —	174

Erin Brine-Doyle, Madeline Shogren, Emily Vecchia, and Eric J. Rawdon

Open knotting — 176

9.1	Introduction —	176
9.2	Defining open knotting —	178
9.2.1	Single closure techniques —	180
9.2.2	Stochastic techniques —	184
9.2.3	Other closure techniques —	187
9.2.4	Topology of knotted arcs —	189
9.3	Visualizing knotting in open chains using the knotting fingerprint —	190
9.4	Features of knotting fingerprints, knotted cores, and crossing changes —	194
9.5	Conclusions —	201
	Bibliography —	202

Yuanan Diao, Claus Ernst, Uta Ziegler, and Eric J. Rawdon

The Knot Spectrum of Random Knot Spaces — 205

10.1	Introduction —	206
10.2	Basic mathematical background in knot theory —	208
10.3	Spaces of random knots, knot sampling and knot identification —	209
10.4	An analysis of the behavior of $P_{\mathcal{K}}$ with respect to length and radius —	212
10.4.1	$P_{\mathcal{K}}(L, R)$ as a function of length L for fixed R —	212
10.4.2	$P_{\mathcal{K}}(L, R)$ as a function of confinement radius R for fixed L —	213
10.4.3	Modeling $P_{\mathcal{K}}$ as a function of length and radius. —	213
10.5	Numerical results —	214
10.5.1	The numerical analysis of $P_{\mathcal{K}}(L, R)$ based on the old data —	214
10.5.2	The numerical analysis of $P_{\mathcal{K}}(L, R)$ based on the new data —	216
10.5.3	The location of local maxima of $P_{\mathcal{K}}(L, R)$ —	216
10.6	The influence of the confinement radius on the distributions of knot types —	217
10.6.1	3-, 4-, and 5-crossing knots —	218
10.6.2	6-crossing knots —	218
10.6.3	7-crossing knots —	219
10.6.4	8-crossing knots —	220
10.6.5	9-crossing knots —	223
10.6.6	10-crossing knots —	225

10.7	The influence of polygon length on the distributions of knot types in the presence of confinement —	227
10.7.1	3-, 4-, and 5-crossing knots —	227
10.7.2	6-crossing knots —	228
10.7.3	7-crossing knots —	229
10.7.4	8-crossing knots —	229
10.7.5	9-crossing knots —	232
10.7.6	10-crossing knots —	233
10.8	Conclusions —	234
	Bibliography —	235

Kyle Chapman, Kenneth C. Millett, and Laura Plunkett

Sampling Spaces of Thick Polygons — 238

11.1	Introduction —	238
11.2	Classical Perspectives —	239
11.2.1	Thickness of polygons —	239
11.2.2	Self-avoiding random walks —	240
11.2.3	Closed polygons: fold algorithm —	240
11.2.4	Closed polygons: crankshaft algorithm —	241
11.2.5	Quaternionic Perspective —	242
11.3	Sampling Thick Polygons —	243
11.3.1	Primer on Probability Theory —	243
11.3.2	Open polygons: Plunkett algorithm [18] —	244
11.3.3	Closed polygons: Chapman algorithm —	250
11.4	Discussion and Conclusions —	253
	Bibliography —	255

Eugene L. Starostin and Gert H.M. van der Heijden

Equilibria of elastic cable knots and links — 258

12.1	Introduction —	258
12.2	Theory of elastic braids made of two equidistant strands —	259
12.2.1	Equidistant curves, reference frames and strains —	259
12.2.2	Equations for the standard 2-braid —	262
12.2.3	Kinematics equations —	264
12.3	Numerical solution —	265
12.3.1	Torus knots —	266
12.3.2	Torus links —	271
12.4	Concluding remarks —	273
	Bibliography —	275

Renzo L. Ricca and Francesca Maggioni

Groundstate energy spectra of knots and links: magnetic versus bending energy — 276

- 13.1 Introduction — **276**
 - 13.2 Magnetic knots and links in ideal conditions — **277**
 - 13.3 The prototype problem — **278**
 - 13.4 Relaxation of magnetic knots and constrained minima — **280**
 - 13.5 Groundstate magnetic energy spectra — **281**
 - 13.6 Bending energy spectra — **283**
 - 13.7 Magnetic energy versus bending energy — **284**
 - 13.8 Conclusions — **286**
- Bibliography — 286**

1 Introduction

*Is the “cable spaghetti” on the floor really knotted
or is it enough to pull on both ends of the wire
to completely unfold it?*

Since its invention, knot theory has always been a place of interdisciplinarity. The first knot tables composed by Tait have been motivated by Lord Kelvin’s theory of atoms. But it was not until a century ago that tools for rigorously distinguishing the trefoil and its mirror image as members of two *different* knot classes were derived. In the first half of the twentieth century, knot theory seemed to be a place mainly driven by algebraic and combinatorial arguments, mentioning the contributions of Alexander, Reidemeister, Seifert, Schubert, and many others. Besides the development of higher dimensional knot theory, the search for new knot invariants has been a major endeavour since about 1960.

At the same time, connections to applications in DNA biology and statistical physics have surfaced. DNA biology had made a huge progress and it was well understood that the topology of DNA strands matters: modeling of the interplay between molecules and enzymes such as topoisomerases necessarily involves notions of ‘knottedness’.

Any configuration involving long strands or flexible ropes with a relatively small diameter leads to a mathematical model in terms of curves. Therefore knots appear almost naturally in this context and require techniques from algebra, (differential) geometry, analysis, combinatorics, and computational mathematics.

The discovery of the Jones polynomial in 1984 has led to the great popularity of knot theory not only amongst mathematicians and stimulated many activities in this direction. Many deep connections between the Jones polynomial and quantum physics were found. In the 1990s, monographs of Adams, Kauffman, and Livingston addressed a wide audience of mathematicians, from undergraduate students to research professionals, disseminated new results and ideas and thereby added to the perception that knot theory is a vivid discipline having impact far beyond its roots.

Today we find that knots appear in almost all mathematical disciplines, having important applications in the sciences and, most importantly, that concepts from one field have impact on problems in other areas. Even more, we will see that applications do not only make use of existing theories developed in entirely theoretical frameworks, but that questions from the sciences also stimulate theoretical developments in turn.

In this edition we focus on four aspects that thematically contour its basis and background and these will be outlined below. Of course, this choice is not meant to be exhaustive, e.g., we do not cover recent developments in the theory of low-dimensional

topology and restrict ourselves to one of the most thrilling and active parts of algebraic knot theory: Khovanov homology.

Applications in the sciences

As already pointed out, the initial development of knot theory has been driven by physical motivation. In the second half of the last century, many additional fields of application have emerged. We will illustrate this variety of them by commenting on some examples.

At the beginning of the 1960s, it was realized that many molecules consist of long polymeric chains. The general aim is to investigate their structure and the relationship of the observed structures to biological features, their formation, and function. In this context, the question about the likeliness of knots in these structures arose almost naturally.

In this time, Frisch and Wasserman as well as Delbrück stated that ring polymers would be knotted with probability increasing to one as their length tends to infinity. This conjecture stimulated the study of polygonal knots and could be settled only about thirty years later.

An important point in this context is to identify and characterize “knotting” using a topological concept that is applicable to open protein knots. Several approaches have been developed to achieve this goal. As polymers are often modeled by open polygons, there is a need to generalize quantities used for the characterization of closed curves, rods, and ropes (such as writhe, self-linking number, and linking number). In order to consider a large variety of possible states, random walks regarding polygons have successfully been investigated and are still a subject of recent activity. Certain knot invariants such as the Jones polynomial have been employed in algorithms to automatically determine the knot class of a randomly chosen polygon.

In certain situations one can in fact work with closed curves. Since the 1960s it has been known that there exist some specific circular DNA molecules. In other cases the two sticky ends of the DNA are very close to each other, so modeling by a closed curve seems reasonable.

According to experimental observations, molecules forming more complex knots migrate faster than those with less complex knots. The speed seems to be proportional to the average crossing number of the corresponding ideal knots. The latter ones minimize the ropelength, i.e., the quotient of length over thickness of a given knot. While this definition seems to be conceptionally elementary, it becomes quite involved from an analytical point of view due to its non-smoothness. This is a major issue in geometric knot theory which will be addressed below. The average crossing number is a frequently used and studied functional in polymer topology and turned out to be a good measure of polymer compaction.

Long DNA chains are often packed compactly under extreme confining conditions. Although the general DNA packing mechanism still lacks a complete understanding, there have been recent attempts to develop a corresponding DNA packing model and test whether the kind of knot spectrum observed in the experiments can be computationally reproduced by studying equilateral random polygons confined in a sphere.

Knots and the concept of ropelength also appear in the context of mathematical physics, for instance in ideal magnetohydrodynamics, with potential applications in the study of astrophysical flows. Magnetic knots are tubular embeddings of the magnetic field in some torus centered on a smooth oriented loop that is knotted in the fluid domain. They are driven by the Lorentz force to inflexion-free spiral knots and braids. The topological crossing number provides a lower bound on the minimum magnetic energy of knots. Combining results on magnetic energy relaxation and data on numerical knot tightening reveals relationships between ropelength and groundstate energy minima.

Models of closed rods in elasticity theory exhibit knotted configurations when applying a suitable amount of twist in the case where self-avoidance is not present. Imposing repulsive forces leads to a variety of different shapes and bifurcation phenomena which has not been fully understood yet.

Geometric knot theory

In the 1990s, the notion of knot energies led to the definition of many functionals on a suitable space of knots, so calculus of variations came into play. Originating from an idea of Fukuhara, the general aim is to investigate geometric properties of a given knotted curve in order to gain information on its knot type. This new subfield emerged from the search of particularly “nice” shapes of knots. In a broader sense, it is part of geometric curvature energies which include geometric integrals measuring smoothness and bending for objects that a priori do not have to be smooth, also covering the higher-dimensional case.

Moffatt speculated that the DNA molecule seeks to attain some minimum energy state. Up to now, the energy involved here has not been identified, but this idea is one of the main motivations to study repulsive functionals in geometric knot theory. There have been several attempts to employ self-avoiding energies for mathematical models in microbiology.

In fact, knot energies basically model self-avoidance, i.e., the functionals blow up on embedded curves converging to a curve with a self-intersection. Following the negative gradient flow of this functional should simultaneously “untangle” the curve and prevent it from leaving the ambient knot class. As long as there are no saddle points of the energy within the unknot class, any such flow is a candidate for a retraction of the unknot class onto the circles which exists due to the Smale conjecture.

In this context, questions of existence and regularity of minimizers are of great interest. The first knot energies defined on smooth curves by O'Hara (which contain as a special case the famous “Möbius energy”) provide several interesting features. The definition is quite straight-forward: one penalizes distant strands of the curve with a small euclidean distance. It was the seminal discovery due to Freedman, He, and Wang that the Möbius energy is invariant under the full group of Möbius transformations in \mathbb{R}^3 which paved the way to proving deep theorems on existence and regularity of minimizers within certain knot classes essentially by geometric means.

The conjecture that the Möbius energy of links is minimized by the stereographic projection of the standard Hopf link has been settled using the min-max theory of minimal surfaces which had been developed for proving the Willmore conjecture by Marques and Neves. There are still other open hypotheses concerning the Möbius energy, for instance the non-existence of minimizers in composite knot classes as stated by Kusner and Sullivan.

Ropelength, mentioned above, albeit conceptually being the most elementary knot energy one can think of, turns out to be analytically quite involved due to taking suprema which is a non-smooth operation. The existence of *ideal knots* (ropelength minimizers within a given knot class) has been established independently by several research groups. The problem of ropelength criticality is quite hard—due to the lack of an explicit analytical characterization of the shape of a (non-trivial) ideal knot, many contributions focus on discretization and numerical visualization.

Analytically, the situation becomes more accessible when suprema in the definition of ropelength are replaced by integration. However, the regularity properties of these (families of) functionals still lack complete understanding. Additionally, several generalizations to higher dimensions have been studied. Especially the investigation of the integral Menger curvature reveals interesting connections to deep questions in complex analysis such as Painlevé's problem.

Up to now, gradient flows have been analyzed mainly for a sub-family of O'Hara's energies. Albeit many numerical schemes rely on suitable gradient-based discretizations, there is still a lack of a rigorous numerical analysis.

There is a close connection between the geometric setup in knot theory and recently developed methods from analysis of non-local critical partial differential equations. More precisely, techniques in the setup of so-called fractional harmonic maps could be extended to the Möbius energy to prove regularity for critical points. This was quite surprising, and it opens up a new link from knot theory to classical geometric analysis, making available the well-studied techniques from harmonic maps and harmonic analysis. This is an active area of research, with the hope of eventually being able to solve open problems with the well-developed tools from classical geometric and harmonic analysis, where purely knot-related geometric arguments fail.

Connections to differential geometry

A classical example for the impact of knottedness to the geometry of curves is the Fáry–Milnor theorem: The total curvature of a curve is bounded from below in terms of the bridge number. Consequently, any curve having total curvature less than 4π must be unknotted.

A more subtle consequence of knottedness is the existence of quadrisecants, i.e., straight lines intersecting the curve at least four times, which has been established by Pannwitz (1933). Surprisingly, there are many of them. Denne sharpened this result a couple of years ago to stating that a (tame) knotted curve even possesses an *alternating* quadrisecant. In fact, quadrisecants greatly improve the known lower bounds on ropelength.

In the same spirit, configuration spaces come into play when proving that Jordan curves of finite total curvature (without cusps) have inscribed squares.

Combinatorial knot theory

Classically, a knot invariant is a functional f mapping a given knot k to some object $f(k)$ with the property that $f(k_1) = f(k_2)$ if the knots k_1, k_2 belong to the same knot class. For instance, we have $f(k) \in \{\text{yes}, \text{no}\}$ in case of tricolorability, and the fundamental group of the complement $\pi_1(\mathbb{S}^3 \setminus k)$ is another knot invariant. Up to now, there is still no easily computable knot invariant that allows to distinguish any two given knot classes. Therefore, in order to capture features of different knot types, it is of great interest to define and investigate new knot invariants.

They all contribute to a sort of big ‘tool box’ which can be applied to further characterize knot classes and add valuable pieces to the ‘puzzle’ that knot theory sometimes consists of.

Although the definition may only involve quite elementary geometric means, the corresponding theory turns out to be rather challenging. This applies, e.g., to the stick index which is the least number of line segments required to build the knot in space. A related quantity is the cubic lattice stick index. Another example is the arc index, the minimum number of pages required to present a given knot type, where an arc presentation is an embedding k of a knot or link in finitely many pages of an “open-book” so that each of its pages meets k in a single simple arc. This particular invariant has a long history dating back to Brunn who looked for a projection of a given link with a single singular point of higher multiplicity.

Sometimes even the definition of a knot invariant may demand deep algebraic techniques. In the early 1980s, Jones established a connection to statistical mechanics. A system of particles where only neighbors interact can be modeled by a lattice or, more generally, a graph. If the partition function, the sum over all possible states of this system, can be chosen to respect Reidemeister moves, it becomes a knot invariant.

An example is the Potts model which is intimately related to knot theory, more precisely to the Jones polynomial and to the Khovanov homology, providing a link between algebraic topology and theoretical physics. The Khovanov homology is a rather new knot invariant introduced by Khovanov in the 1990s and can be obtained from categorification of the Jones polynomial.

Exciting physical interpretation and connections of Khovanov homology using avant-garde ideas from for example quantum fields and strings have already been found. And many new exciting developments in this directions are still to be expected. Or, as Edward Witten put it in a talk on knot and quantum theory at the Institute of Advances Study in October 2011: “Probably the full story involves physics ideas that we do not understand yet.”

The aim of this book is to present short surveys and open questions on different active research directions related to knot theory in the broadest sense that are thought of as invitation and challenge to junior and senior researchers. In this sense, it complements the existing introducing (e.g., the above-mentioned monographs) and more advanced textbooks (like Crowell–Fox, Rolfsen, Burde–Zieschang, and many others).

On the other hand, this collection emphasizes connections between the different fields and provides a prototypical example for the unity of mathematics. In this light, distinguishing between “pure” and “applied” fields does not seem to be a meaningful category.

Joining certain aspects of knot theory, this presentation shall provide an overview on the state of the art, popularize open problems, and reveal synergies and future directions.

The fields of modern knot theory use quite different tools. However, each such area is deeply grounded on results from the others, so new developments will certainly foster prospective research. To bring together those various aspects seems to us like a great opportunity for all sides.

This edition is intended as a platform to exchange and update on ideas, open problems, and interesting directions on which future collaborations over the borders of areas can be based. In the past, many of the contributing authors have successfully collaborated in different formations which adds to the flexibility and vitality of this area.

We would be glad if this collection would further stimulate research activities and help disseminate ideas between different areas related to knots and knotted structures.

We are very much indebted to all authors who accepted the challenge of presenting the state of the art of their field of expertise to a broader readership as well as to our referees for their thorough work.

This edition has been initiated on the occasion of the workshop *Geometric Energies with Links to Applications, Topology, and Open Problems* which has been held in

September 2015 at the University of Basel. We would like to express our gratitude to the Swiss National Science Foundation, grant no IZ32Z0 160298/1, as well as to the Mathematical Institute in Basel for their generous support.

We are very grateful to Professor Dr. Gianluca Crippa for his interest in our project and for adopting it in the newly launched series *Partial Differential Equations and Measure Theory*. Furthermore we would like to thank the staff at De Gruyter, especially Dr. Agnieszka Bednarczyk-Drag, Dr. Konrad Kieling, and Dr. Agata Morka.

Salzburg | Essen | Freiburg, March 2017

S. B. | Φ . R. | A. S.

Paweł Strzelecki and Heiko von der Mosel

Geometric curvature energies: facts, trends, and open problems

Abstract: This survey focuses on geometric curvature functionals, that is, geometrically defined self-avoidance energies for curves, surfaces, or more general k -dimensional sets in \mathbb{R}^d . Previous investigations of the authors and collaborators concentrated on the regularising effects of such energies, with a priori estimates in the regime above scale-invariance that allowed for compactness and variational applications for knotted curves and surfaces under topological restrictions. We briefly describe the impact of geometric curvature energies on geometric knot theory. Currently, various attempts are being made to obtain a deeper understanding of the energy landscape of these highly singular and nonlinear nonlocal interaction energies. Moreover, a regularity theory for critical points is being developed in the setting of fractional Sobolev spaces. We describe some of these current trends and present a list of open problems.

Keywords: geometric curvature energies, singular integrals, critical points, regularity theory, geometric knot theory, elastic knots, rectifiability

2.1 Facts

Energies. Geometric curvature functionals are characterised as geometrically defined energies on a priorily non-smooth k -dimensional subsets Σ^k of \mathbb{R}^d , and these functionals are designed to penalise self-intersections. In addition, there are regularising effects: finite energy implies some higher degree of smoothness of Σ . One of the first examples is that of the *Möbius energy* on rectifiable curves $\gamma \subset \mathbb{R}^d$, introduced by J. O'Hara [82] and investigated analytically by M. H. Freedman et al. [41, 50],


$$\mathcal{E}_{\text{Möb}}(\gamma) := \int_{\gamma} \int_{\gamma} \left[\frac{1}{|x-y|^2} - \frac{1}{d_{\gamma}(x,y)^2} \right] d\mathcal{H}^1(x) d\mathcal{H}^1(y), \quad (2.1.1)$$

where $d_{\gamma}(x, y)$ denotes the intrinsic distance between the points x, y on the curve γ , and \mathcal{H}^1 stands for the one-dimensional Hausdorff-measure. The first summand in the integrand resembles a Coulomb-type *repulsive potential* suitably regularised by the second term so as to obtain finite energy for smooth embedded curves. Another exam-

Paweł Strzelecki, Institute of Mathematics, University of Warsaw, Banacha 2, PL-02-097 Warsaw, Poland, E-mail: p.strzelecki@mimuw.edu.pl

Heiko von der Mosel, RWTH Aachen University, Templergraben 55, D-52062 Aachen, Germany, heiko@instmath.rwth-aachen.de

<https://doi.org/10.1515/9783110571493-001>

Open Access.  © 2018 Paweł Strzelecki and Heiko von der Mosel, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

ple is that of *rope length*^{2.1} defined as the quotient of length $\mathcal{L}(\gamma)$ and thickness $\Delta[\gamma]$ of curves γ , where the latter is a non-smooth functional introduced by O. Gonzalez and J. H. Maddocks [48], defined as

$$\Delta[\gamma] := \inf_{\substack{x, y, z \in \gamma \\ x \neq y \neq z \neq x}} R(x, y, z). \quad (2.1.2)$$

Here, $R(x, y, z)$ denotes the circumcircle radius of the three curve points x, y , and z . Thickness can be regarded as a hard core potential or steric constraint, in contrast to repulsive potentials.

Both functionals have had their impact on the modeling of macromolecules such as DNA and proteins [25, 76, 77], and on *geometric knot theory*, where one studies relations between the geometry of space curves and the knot types they represent. Several geometric curvature energies can be minimised within given knot classes to obtain particularly nice representatives of that knot, for example *ideal knots* as rope length minimisers [29, 49]. Bounds on the energy sometimes imply bounds on knot invariants like stick number or crossing number, hence bounds on the number of knot classes that possess representatives below these energy thresholds; see, e.g., [24, 73].

However, both these extreme forms of energies have serious drawbacks. The highly singular integrals involved in the definition of any kind of repulsive potential like (2.1.1) need some sort of regularisation, and – besides the ambiguity in the choice of such a regularisation – it is by no means clear how to generalise this concept to higher dimensional objects. The steric constraint of given thickness (2.1.2) or the rope-length functional, on the other hand, is a non-smooth quantity imposing challenging technical problems, e.g., for the derivation and analysis of variational equations. This led to our systematic research between 2007 and 2012, devoted to a whole range of intermediate energies on curves and surfaces interpolating in some sense between hard steric constraints and “soft” repulsive potentials. Examples of such energies on one-dimensional sets include [119]

$$\mathcal{U}_p(\gamma) := \int_{\gamma} \sup_{\substack{y, z \in \gamma \\ z \neq y \neq x \neq z}} \frac{1}{R^p(x, y, z)} d\mathcal{H}^1(x), \quad p \geq 1, \quad (2.1.3)$$

or the double integral [114]

$$\mathcal{I}_p(\gamma) := \int_{\gamma} \int_{\gamma} \sup_{\substack{z \in \gamma \\ z \neq x \neq y \neq z}} \frac{1}{R^p(x, y, z)} d\mathcal{H}^1(x) d\mathcal{H}^1(y), \quad p \geq 2, \quad (2.1.4)$$

2.1 The name of that functional is coined after the following geometric variational problem: given a rope of fixed constant thickness, what is the minimum length of this rope required to tie a given knot?

and also *integral Menger*^{2.2} *curvature* [115]

$$\mathcal{M}_p(\gamma) := \int_{\gamma} \int_{\gamma} \int_{\gamma} \frac{1}{R^p(x, y, z)} d\mathcal{H}^1(x) d\mathcal{H}^1(y) d\mathcal{H}^1(z), \quad p \geq 3. \quad (2.1.5)$$

On a fixed loop γ of unit length, these energies are ordered as

$$\mathcal{M}_p^{1/p}(\gamma) \leq \mathcal{J}_p^{1/p}(\gamma) \leq \mathcal{U}_p^{1/p}(\gamma) \leq \frac{1}{\Delta[\gamma]}, \quad (2.1.6)$$

where the last term is the ropelength of γ . Moreover, the p -th root of \mathcal{M}_p , \mathcal{J}_p , and of \mathcal{U}_p tends to ropelength as $p \rightarrow \infty$ both on fixed conformations of knots and in the sense of Γ -convergence.

Besides averaging and maximising over the *multi-point interactions* in the circumradius we investigated *tangent-point interactions* such as [121, 57]

$$\mathcal{E}_p(\gamma) := \int_{\gamma} \int_{\gamma} \frac{1}{r_{\text{tp}}^p(x, y)} d\mathcal{H}^1(x) d\mathcal{H}^1(y), \quad p \geq 2, \quad (2.1.7)$$

as well, where $r_{\text{tp}}(x, y)$ is the radius of the unique circle through two given curve points x and y that is additionally tangent to γ in x .

We also introduced and studied geometric curvature energies on higher-dimensional sets such as *thickness for surfaces* $\Sigma \subset \mathbb{R}^d$ [118, 117], where one minimises over all pairs of points $x, y \in \Sigma$ the *tangent-point radius* $R_{\text{tp}}(x, y)$ of the smallest sphere through x and y that is tangent to Σ in x . Later, we investigated *integral Menger curvature for surfaces* $\Sigma \subset \mathbb{R}^3$ [120],

$$\mathcal{M}_p(\Sigma) := \int_{\Sigma} \int_{\Sigma} \int_{\Sigma} \int_{\Sigma} \mathcal{K}^p(x, y, z, \xi) d\mathcal{H}^2(x) d\mathcal{H}^2(y) d\mathcal{H}^2(z) d\mathcal{H}^2(\xi), \quad p \geq 8, \quad (2.1.8)$$

where the integrand \mathcal{K} is defined^{2.3} on tetrahedra $T = (x, y, z, \xi)$ with vertices $x, y, z, \xi \in \Sigma$, as

$$\mathcal{K}(T) = \mathcal{K}(x, y, z, \xi) := \frac{\mathcal{H}^3(T)}{\text{area}(T) \text{diam}^2(T)}. \quad (2.1.9)$$

2.2 Karl Menger considered in the 1930's the circumradius $R(x, y, z)$ of three curve points $x, y, z \in \gamma$ knowing that the coalescent limit of $R(x, y, z)$ as x and y tend to z coincides with the local radius of curvature if the curve γ is sufficiently smooth. Menger was also aware of the fact that there is an elementary formula for the circumradius solely in terms of the mutual distances of the points x, y , and z . By means of *multipoint functions* such as $R(\cdot, \cdot, \cdot)$ Menger indeed intended to develop a purely metric geometry in contrast to classic differential geometry. The idea of using Menger curvature as a tool – both in harmonic analysis and in modeling – has been re-discovered in the last 20 years.

2.3 The most obvious choice to take as integrand in (2.1.8) a negative power of the *circumsphere radius* of a tetrahedron does *not* serve our purposes since there are smooth embedded surfaces for which such an integrand would not be bounded; see our detailed discussion on various integrands in [120, Appendix B].

Here $\mathcal{H}^3(T)$ is the volume of the tetrahedron T and $\text{area}(T)$ the sum of its facet areas. For k -dimensional submanifolds $\Sigma^k \subset \mathbb{R}^d$ we looked at the *tangent-point energy* [123]

$$\mathcal{E}_p(\Sigma) := \int_{\Sigma} \int_{\Sigma} \frac{1}{R_{\text{tp}}^p(x, y)} d\mathcal{H}^k(x) d\mathcal{H}^k(y), \quad (2.1.10)$$

and S. Kolasiński investigated in his Ph.D. thesis *integral Menger curvature for $\Sigma^k \subset \mathbb{R}^d$* [61, 62]

$$\mathcal{M}_p(\Sigma) := \underbrace{\int_{\Sigma} \cdots \int_{\Sigma}}_{k+2 \text{ integrals}} \mathcal{K}^p(x_0, \dots, x_{k+1}) d\mathcal{H}^k(x_0) \cdots d\mathcal{H}^k(x_{k+1}) \quad (2.1.11)$$

for $p > k(k+2)$, where the integrand generalises (and simplifies) the one given in (2.1.9) to

$$\mathcal{K}(T) = \mathcal{K}(x_0, \dots, x_{k+1}) := \frac{\mathcal{H}^{k+1}(T)}{(\text{diam}(T))^{k+2}}, \quad (2.1.12)$$

for $(k+1)$ -dimensional simplices $T = (x_0, \dots, x_{k+1})$ with each vertex x_i on Σ .

Regularising effects. Summarising the essential results of this systematic research (which is well documented in a number of publications [119, 114, 115, 121, 117, 118, 120, 123, 61, 62], we can say the following: we have a pretty clear understanding of the topological and regularising effects of each of these energies, with sharp regularity statements and uniform a priori estimates. For example, a rectifiable curve γ with finite integral Menger curvature $\mathcal{M}_p(\gamma)$ for some $p > 3$ (i.e., above the scale-invariant case $p = 3$) is homeomorphic to the unit-circle or unit-interval, and the arclength parametrisation of that curve satisfies the uniform a priori estimate

$$|\gamma'(s) - \gamma'(t)| \lesssim \mathcal{M}_p(\gamma[s, t])^{1/p} |s - t|^{1-3/p} \quad \text{for all } s, t. \quad (2.1.13)$$

In other words, $\gamma \in C^{1,1-(3/p)}$, so although \mathcal{M}_p does not capture the pointwise value of local curvature (which may be simply undefined even if $\mathcal{M}_p(\gamma)$ is finite), it does capture the average oscillation of the unit tangent vector; see [115, Theorem 1.2]. We may interpret this result as a *geometric Morrey-Sobolev embedding*: the integrand corresponds to a very weak form of curvature integrated to some power $p > 3$, and the total domain of integration is three-dimensional; the classic Morrey-Sobolev theorem applied to second derivatives (instead of curvature) would give exactly the optimal Hölder exponent $1 - (3/p)$ for the first derivatives.

Likewise for higher-dimensional subsets of \mathbb{R}^d , exemplified by integral Menger curvature $\mathcal{M}_p(\Sigma)$ for two-dimensional surfaces in Euclidean 3-space, as defined in (2.1.8); see [120, Theorem 1.4]: if an admissible two-dimensional set $\Sigma \subset \mathbb{R}^3$ satisfies $\mathcal{M}_p(\Sigma) < \infty$ for some $p > 8$ (again above the scale-invariant case $p = 8$), then Σ is actually an orientable $C^{1,1-(8/p)}$ -submanifold with a controlled local graph representation: There is a uniform radius $R > 0$ depending only on p and the energy

value $\mathcal{M}_p(\Sigma)$, such that for each $x \in \Sigma$ the intersection $B_R(x) \cap \Sigma$ equals the graph of a $C^{1,1-(8/p)}$ -function with uniform estimates on this function solely depending on p and $\mathcal{M}_p(\Sigma)$. Again, this is a geometric variant of the Morrey-Sobolev embedding theorem with optimal Hölder exponent $1 - (8/p)$ for the oscillation of tangent planes. Similar results hold for k -dimensional admissible sets $\Sigma^k \subset \mathbb{R}^d$ with finite integral Menger curvature $\mathcal{M}_p(\Sigma)$ for $p > k(k+2)$, or finite tangent point energy $\mathcal{E}_p(\Sigma)$ for $p > 2k$, both in the regime above scale-invariance; see [61, 62, 123]. A thorough discussion of the respective admissibility class of sets can be found, e.g., in [64]. At this point, we may roughly describe our mild requirements on the set Σ^k as a certain degree of local flatness around many (but not all) points, together with an amount of connectivity to allow for some degree-theoretic arguments.

That these regularity estimates are indeed sharp, can be seen either by explicit examples constructed in [66, 57], or by the complete *characterisation of energy spaces* for all these energies in the work of S. Blatt and Kolasiński [17, 13, 11]: based on our results that finite energy implies that the admissible sets are already C^1 -submanifolds in \mathbb{R}^d , they use the explicit structure of the energies to estimate locally the seminorms of fractional Sobolev spaces to find that Σ has finite energy *if and only if* Σ is embedded and has local graph representations of exactly that Sobolev regularity. Recall, e.g., from [128, Section 2.2.2], that a function $u \in L^p(\mathbb{R}^k)$ belongs to the *Sobolev-Slobodeckii space* $W^{m+s,p}(\mathbb{R}^k)$ for some $m \in \mathbb{N}$, $s \in (0, 1)$, and $p \in [1, \infty)$ if u belongs to the classic Sobolev space $W^{m,p}$ and satisfies, in addition,

$$\|u\|_{m+s,p}^p := \|u\|_{W^{m,p}(\mathbb{R}^k)}^p + \sum_{|\alpha|=m} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \frac{|D^\alpha u(x) - D^\alpha u(y)|^p}{|x - y|^{k+sp}} dy dx < \infty. \quad (2.1.14)$$

As an example, let us mention Blatt's and Kolasiński's characterisation of the energy space for integral Menger curvature as defined in (2.1.11); see [17, Corollary 1.2]: If $p > k(k+2)$ and $\Sigma^k \subset \mathbb{R}^d$ is an admissible set, then its integral Menger curvature $\mathcal{M}_p(\Sigma)$ is finite if and only if Σ is a submanifold with local graph representation of class $W^{1+s,p}(\mathbb{R}^k, \mathbb{R}^{d-k})$, where $s = 1 - (k(k+1)/p) \in (0, 1)$. Blatt and Kolasiński treated also all intermediate energies where up to all but two integrations in (2.1.11) are replaced by maximisations, obtaining corresponding fractional Sobolev spaces as the exact energy spaces with suitably adapted differentiability and integrability. The missing cases where only one integration is left, i.e.,

$$\int_{\Sigma} \sup_{x_1, \dots, x_{k+1} \in \Sigma} \mathcal{K}^p(x_0, x_1, \dots, x_{k+1}) d\mathcal{H}^k(x_0), \quad (2.1.15)$$

generalising (2.1.3) to k -dimensional sets $\Sigma^k \subset \mathbb{R}^d$, and the *global tangent-point energy* (cf. (2.1.10)),

$$\int_{\Sigma} \sup_{y \in \Sigma} \frac{1}{R_{\text{tp}}^p(x, y)} d\mathcal{H}^k(x), \quad (2.1.16)$$

were treated in cooperation with Kolasiński in [64] leading to the theorem that finite energy *characterises* embedded submanifolds of classic Sobolev regularity $W^{2,p}$ if $p > k$; see [64, Theorem 1.4]. This result may be compared to Allard's famous $C^{1,\alpha}$ -regularity theorem [3, 37] for k -dimensional varifolds whose generalised mean curvature is p -integrable, where, again, $p > k$.

Connections to geometric knot theory. The uniform estimates obtained in [119, 114, 115] for finite energy curves γ (like the one in (2.1.13)) together with a uniform geometric rigidity of these curves (replacing the excluded volume constraint of thickness) was used to connect the respective energies to *geometric knot theory*, as described in detail in [116, Section 4]; see also the recent surveys [122, 124]. This geometric rigidity^{2.4} means, roughly speaking, that the curve may be equipped with a necklace of consecutive double-cones whose size and opening angle are determined purely in terms of the respective energy [116, Proposition 4.7]. The circular cross-sections of each piece of this necklace, i.e., of each such double cone (with its two tips located on the curve γ), are intersected by γ transversally and exactly in one point; see Figure 2.1. Once this necklace is established one can fairly easily construct an ambient isotopy from γ to the inscribed polygon made of the consecutive double cones' axes.

Thus, any one of the geometric curvature energies for curves (2.1.3)–(2.1.5), (2.1.7), bounds the *stick number*, which is the minimal number of straight segments you need to build a polygonal representative of the same knot type. Since stick number is a knot invariant, any such energy bounds the number of knot types: given any constant $E \geq 0$ there is a nonnegative integer $N(E)$ depending only on E , such that at most $N(E)$ knot types can be represented by curves of geometric curvature energy below the energy threshold E .

On the other hand, the double-cone property described above also serves as a substitute of the excluded volume constraint given by finite ropelength. This allows us to control the *average crossing number* $\text{acn}(\gamma)$, where you count the number of self-intersections of every planar projection of the given curve γ and then average over all directions of projections. Indeed, Freedman et al. derived in [41, Section 3] a double integral formula for $\text{acn}(\gamma)$,

$$\text{acn}(\gamma) := \frac{1}{4\pi} \int_0^1 \int_0^1 \frac{|\gamma'(s) \times \gamma'(t) \cdot (\gamma(t) - \gamma(s))|}{|\gamma(t) - \gamma(s)|^3} ds dt, \quad (2.1.17)$$

where \times denotes the usual cross-product in \mathbb{R}^3 . While the local interaction terms in that formula may be estimated by the local smoothness properties of a finite energy curve γ , one can follow the strategy of G. Buck and J. Simon [24] for curves of finite ropelength, to estimate the global interaction terms by estimating the volume of a spatial region necessary to fit in a maximally compactified curve γ . Only here, one

2.4 Referred to as *diamond property* in [116].

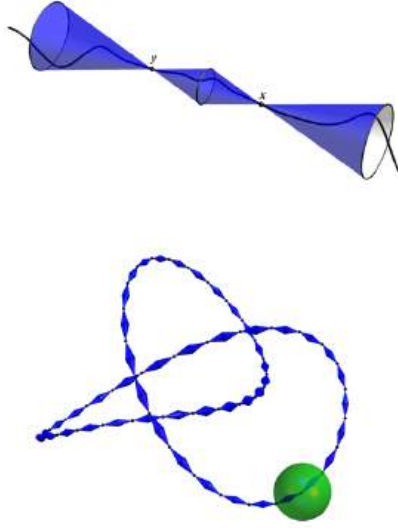


Fig. 2.1: Top: The curve γ is trapped in a conical region with two tips at $x, y \in \gamma$ and does not meander back and forth: each cross section of the double-cones contain exactly one point of the curve. Bottom: Necklace made of such small double cones with vertices along the curve γ that have pairwise disjoint interiors. The polygonal curve joining the consecutive vertices of the cones is ambient isotopic to γ .

has to replace the excluded volume constraint by the double-cone condition, so that our constants are far from being optimal; see [116, Proposition 4.13]. Since the average crossing number bounds the classic knot invariant *crossing number*, we thus have established another means to control the number of knot types below given energy thresholds.

In addition, we could show that all these energies are *charge* and *tight*, which means that they blow up along sequences that converge to curves with self-intersections and also along sequences where one small knotted subarc pulls tight, i.e., vanishes in the limit. Being tight distinguishes these geometric curvature energies from the Möbius energy (2.1.1): O'Hara showed in [83, Theorem 3.1] that the Möbius energy does *not* prevent the pull-tight phenomenon. Moreover, \mathcal{U}_p and \mathcal{I}_p could be shown to distinguish the knot from the unknot: there is a gap between the infimum over unknots and the infimum over non-trivially knotted curves. The infima of all these energies (in contrast, e.g., to the Möbius^{2.5} energy) are attained on each given knot

2.5 The Möbius energy can, however, be minimised within prescribed *prime* (or *irreducible*) knot classes; see [41, Theorem 4.3].

class. Freedman et al. [41] showed that the Möbius energy $\mathcal{E}_{\text{Möb}}$ is uniquely minimised by the round circle, a knot energy with that property is called *basic*. A. Abrams et al. [1] extended that uniqueness result to a larger family of energies. Also ropelength is basic, and more generally, \mathcal{U}_p as well; see [119, Lemma 7]. A more recent monotonicity formula^{2.6} for compact free boundary surfaces by A. Volkmann [129, Section 5] implies the same for the tangent-point energy \mathcal{E}_p (see (2.1.7)), and hence also \mathcal{J}_p by the arguments in [116, Proof of Cor. 3.7]. All these results resolve some of the open problems formulated in geometric knot theory, e.g. in [125, Section 2], or in [86, Chapter 8], and give some first insights into the presumably complicated energy landscape of these energies on knot space. Almost nothing is known about the actual shape of knotted energy minimisers, apart from the explicit continuous family of ideal links (minimising ropelength in fixed link classes) presented by J. Cantarella, R. B. Kusner, and J. M. Sullivan in [29], and necessary criticality conditions for ropelength minimisers [110, 26, 27]. Moreover, studying ideal knots in \mathbb{R}^4 lead to the discovery of unique explicit solution families of longest (thick) ropes on the two-sphere by H. Gerlach and the second author; see [45] and the popular account in [44].

2.2 Trends and open problems

Regularity. Higher regularity of local minimisers or critical points is only known in a few cases. Freedman et al. [41] used the Möbius-invariance of the Möbius energy $\mathcal{E}_{\text{Möb}}$ defined in (2.1.1) to apply reflection arguments to show that local minimisers are of class $C^{1,1}$, and they derived the Euler-Lagrange equation

$$\begin{aligned} \delta \mathcal{E}_{\text{Möb}}(\gamma, h) &:= \lim_{\tau \rightarrow 0} \frac{\mathcal{E}_{\text{Möb}}(\gamma + \tau h) - \mathcal{E}_{\text{Möb}}(\gamma)}{\tau} \\ &= 2 \lim_{\varepsilon \searrow 0} \iint_{|u-v| \geq \varepsilon} \left(\frac{\gamma'(u) \cdot h'(u)}{|\gamma'(u)|^2} - \frac{(\gamma(u) - \gamma(v)) \cdot (h(u) - h(v))}{|\gamma(u) - \gamma(v)|^2} \right) \\ &\quad \times \frac{|\gamma'(v)| |\gamma'(u)|}{|\gamma(u) - \gamma(v)|^2} du dv \end{aligned} \quad (2.2.1)$$

for injective and regular curves γ and perturbations h both of class $C^{1,1}$. Later Zh.-X. He [50] used this Euler-Lagrange equation to improve the regularity of local $\mathcal{E}_{\text{Möb}}$ -minimisers to C^∞ -smoothness; see also [98]. Quite recently, there has been considerable progress through the work of Blatt, P. Reiter, and A. Schikorra, who established the following deep regularity result [22, Theorem I] in the correct fractional Sobolev space $W^{3/2,2}$ corresponding – according to Blatt’s earlier work [11] – to finite Möbius energy:

2.6 Or alternatively, a secant map approach of Blatt reminiscent of an argument to prove the classic Fenchel inequality; see [129, Section 5].

Theorem 2.2.1 ($\mathcal{E}_{\text{Möb}}$ -critical points are smooth). *Any arclength parametrised critical point $\gamma \in W^{3/2,2}$ of the Möbius energy is C^∞ -smooth.*

It is remarkable that no use at all is made of the Möbius-invariance of $\mathcal{E}_{\text{Möb}}$ to prove Theorem 2.2.1 in contrast to the previous work of Freedman et al. Here, one uses the Euler-Lagrange equation (suitably extended to the correct fractional Sobolev spaces) to first gain some additional regularity, i.e., a slightly higher integrability of the tangent [22, Theorem III], before a bootstrapping process can be started. One should point out that, similarly to other geometric equations like the variational equation for the Willmore functional, the Euler-Lagrange equation for $\mathcal{E}_{\text{Möb}}$ is in a sense critical, which requires some very intricate techniques that were developed in the context of fractional harmonic mappings [34, 33, 102, 101].

Somewhat less involved is the regularity proof for other members of O'Hara's families of repulsive potentials [82, 84, 85], namely for the energies \mathcal{E}^α , where a power $\alpha \in (2, 3)$ replaces the quadratic power in the denominators of (2.1.1). Blatt and Reiter established the Fréchet-differentiability of \mathcal{E}^α on the space of regular curves of finite energy, and proved C^∞ -smoothness for arclength parametrised critical points [19, Theorems 1.1 & 1.2].

To carry over this regularity program to critical points of geometric curvature energies such as the tangent-point energy (2.1.7) or integral Menger curvature (2.1.5), Blatt and Reiter embedded those energies into larger two-parameter families of energies by decoupling the integrability exponent p into different powers for numerator and denominator of the integrands. In this way they obtain, for instance, *modified tangent-point energies*

$$\mathcal{TP}^{(p,q)} := \int_{\gamma} \int_{\gamma} \frac{1}{r^{(p,q)}(x,y)} d\mathcal{H}^1(x) d\mathcal{H}^1(y), \quad (2.2.2)$$

by replacing the p -th power of the *inverse tangent-point radius*

$$\frac{1}{r_{\text{tp}}(x,y)} = \frac{2 \text{dist}(T_x, y)}{|x-y|^2} \text{ by the less geometric expression } \frac{1}{r^{(p,q)}(x,y)} := \frac{\text{dist}(T_x, y)^q}{|x-y|^p}.$$

(Here, T_x denotes the tangent-line to γ through the point $x \in \gamma$.) In the parameter regime $q > 1$ and $p \in (q+2, 2q+1)$ the modified tangent-point energies turn out to be well-behaved knot energies that are minimisable in every knot class. The fractional Sobolev regularity $W^{(p-1)/q,q}$ characterises finite energy (see [20, Theorems 1.1 & 1.3]), and allows a first variation formula even without Cauchy principal values [20, Theorem 1.4] in contrast to the variational equations of O'Hara's repulsive energies. Blatt and Reiter then identify a *non-degenerate* parameter range $q = 2$, $p \in (4, 5)$ that permits a regularity result [20, Theorem 1.5]. Unfortunately this range excludes the original tangent-point energy $\mathcal{E}_p = \mathcal{TP}^{2p,p}$ (cf. (2.1.7)).

Theorem 2.2.2 ($\mathcal{TP}^{(p,2)}$ -critical points are smooth). *For $p \in (4, 5)$ any $\mathcal{TP}^{(p,2)}$ -critical arclength parametrised injective curve of class $W^{(p-1)/2,2}$ is C^∞ -smooth.*

A similar result holds also for a subfamily of *modified integral Menger energies*; see [21], but apart from that non-degenerate parameter regime, in particular for (the original) integral Menger curvature (2.1.5) and tangent point energies (2.1.7) the regularity of critical points, even that of local minimisers remains open.

Open questions 2.2.3. *Despite degeneracies in the non-local Euler-Lagrange equations, is there any chance to prove additional regularity (beyond the fractional Sobolev regularity characterising energy spaces) for local minimisers or critical points of tangent-point energies (2.1.7), or of integral Menger curvature (2.1.5)? Does the optimal regularity depend on p , and if yes, what happens with that p -dependent regularity in the limit $p \rightarrow \infty$? Does that lead to new insights into the still open optimal regularity of ideal knots?*

We know that the geometric curvature energies in higher dimensions such as integral Menger curvature (2.1.8), (2.1.11), or the tangent-point energy (2.1.10) can be minimised within given isotopy classes of submanifolds [117, Theorem 7.1], [120, Theorem 1.7], [65, Corollary 1], but nothing is known about higher regularity of these minimisers, not to speak of a regularity statement about possible critical points. Not even a variational equation has been derived so far in higher dimensions. In case of the non-smooth energies (2.1.15), (2.1.16), and also for the one-dimensional prototype (2.1.3), non-smooth analysis tools such as Clarke gradients would have to be applied to derive the variational differential inclusion, similar to the analysis performed for the ropelength functional for curves involving the non-smooth expression (2.1.2) for thickness; see [109, 110, 26, 27].

Open questions 2.2.4. *What are the Euler-Lagrange-equations for higher-dimensional geometric curvature energies like integral Menger curvature (2.1.5) or tangent-point energies (2.1.10)? Is there any chance to prove higher regularity of local minimisers or critical points of these energies? What form do the expected variational inclusions have for the non-smooth geometric energies (2.1.3), (2.1.15), (2.1.16)? What can be said about the regularity of thick knotted surfaces minimising area?*

Below or in the scale-invariant regime. Integral Menger curvature \mathcal{M}_p on one-dimensional sets $E \subset \mathbb{C}$ with integrability exponent $p = 2$ (well below the scale-invariant exponent $p = 3$) has played a fundamental rôle in harmonic analysis, e.g., in the proof of the famous Vitushkin conjecture on the removability of compact subsets of the complex plane for complex analytic functions; see, for instance, X. Tolsa's quite recent excellent monograph [126]. Motivated by some of G. David's methods [35] for his final proof of this conjecture, J.-C. Léger [70] proved the following remarkable rectifiability result:

Theorem 2.2.5 (Rectifiability for sets of finite integral Menger curvature \mathcal{M}_2). *Any Borel set $E \subset \mathbb{R}^n$ with $0 < \mathcal{H}^1(E) < \infty$ satisfying*

$$\mathcal{M}_2(E) = \int_E \int_E \int_E \frac{1}{R^2(x, y, z)} d\mathcal{H}^1(x) d\mathcal{H}^1(y) d\mathcal{H}^1(z) < \infty \quad (2.2.3)$$

is 1-rectifiable, i.e., there exists a countable family of Lipschitz functions $f_i : \mathbb{R} \rightarrow \mathbb{R}^n$ such that $\mathcal{H}^1(E \setminus \bigcup_i f_i(\mathbb{R})) = 0$.

So, even below scale-invariance, integral Menger curvature has regularising effects on sets. In our context of geometric curvature energies one is naturally lead to the question if one can generalise Léger's deep result to sets of higher dimensions? What are suitable generalisations of the integrand in (2.2.3) which is defined on point triples forming two-dimensional simplices. Already for our generalisations to surfaces and submanifolds as given in (2.1.9) and (2.1.12) we had discussed several variants of integrands defined on general $(k+1)$ -dimensional simplices; see, e.g., the introduction and appendix of [120]. Recently, M. Meurer [75] presented a collection of integrands for which Léger's result could indeed be extended to arbitrary dimensions and co-dimensions, including, e.g., the integrand, defined on $(k+1)$ -dimensional simplices $T = (x_0, x_1, \dots, x_{k+1})$; see [75, Section 3.2],

$$\mathcal{K}_M(T) = \mathcal{K}_M(x_0, \dots, x_{k+1}) := \frac{\mathcal{H}^{k+1}(T)}{\text{diam}(T)^{(k+1)(k+2)/2}}, \quad (2.2.4)$$

which is one out of several possible generalisations of Léger's integrand $1/R(x, y, z)$ in (2.2.3). Meurer could prove the following rectifiability theorem [75, Theorem 1.1].

Theorem 2.2.6 (Rectifiability in arbitrary dimensions). *Any Borel set $E \subset \mathbb{R}^n$ with $0 < \mathcal{H}^k(E) < \infty$ satisfying*

$$\mathcal{M}_2(E) := \underbrace{\int_E \dots \int_E}_{k+2 \text{ integrals}} \mathcal{K}_M^2(x_0, \dots, x_{k+1}) d\mathcal{H}^k(x_0) \dots d\mathcal{H}^k(x_{k+1}) < \infty \quad (2.2.5)$$

is k -rectifiable, i.e., can be covered (up to sets of \mathcal{H}^k -measure zero) by a countable union of Lipschitz images of \mathbb{R}^k .

Meurer's class of admissible integrands includes also the discrete curvatures used by G. Lerman and J. T. Whitehouse in [71, 72] to give a characterisation of David's and S. Semmes' concept of uniform rectifiability; cf [36, Theorem 1.57]. J. Azzam and Tolsa [7] recently established a new rectifiability criterion in terms of P. Jones's β -numbers [56] which are fundamentally related to integral Menger curvature as shown in [70, 75]. Interestingly, however, and somewhat surprising is the fact, that the integrands (2.1.9) and (2.1.12) we studied in the integrability regime above scale-invariance, are

not included in Meurer's class of integrands; they do scale differently. At this point it remains open, if one can replace \mathcal{K}_M in (2.2.5) by the expression \mathcal{K} defined in (2.1.12).

In the definition of rectifiability one covers the set (up to a set of measure zero) by Lipschitz images, and one might think about improving the regularity of the covering images. The step from Lipschitz to C^1 -images is immediate by Whitney's extension theorem; see, e.g. [112, Section 3, Lemma 11.1], but improving that to $C^{1,\alpha}$ (as in the regime above scale-invariance) is highly non-trivial. This was recently accomplished by Kolasinski [63] also for a large class of discrete curvatures with a certain overlap with Meurer's class including (2.2.4), so that, e.g., the following *higher order rectifiability result* holds true and can be deduced from [63, Theorem 1.1].

Theorem 2.2.7 ($C^{1,\alpha}$ -rectifiability). *Any Borel set $E \subset \mathbb{R}^n$ with $0 < \mathcal{H}^k(E) < \infty$ and a.e. positive lower density, satisfying*

$$\mathcal{M}_p(E) := \underbrace{\int_E \cdots \int_E}_{k+2 \text{ integrals}} \mathcal{K}_M^p(x_0, \dots, x_{k+1}) d\mathcal{H}^k(x_0) \cdots d\mathcal{H}^k(x_{k+1}) < \infty \quad (2.2.6)$$

for some $p > 2$ is k -rectifiable of class $C^{1,\alpha}$ for some positive Hölder exponent $\alpha = \alpha(p)$, i.e., the set E can be covered (up to sets of \mathcal{H}^k -measure zero) by a countable union of k -dimensional $C^{1,\alpha}$ -submanifolds of \mathbb{R}^n .

Open questions 2.2.8. *Can one extend Meurer's rectifiability result to the integrands (2.1.9) or (2.1.12) of integral Menger curvature or to the tangent-point energies (2.1.10) defined on a suitable wide class of non-smooth sets? How does Meurer's result relate to other recent rectifiability results like [127, 7]?*

Not much is known about geometric curvature energies in the scale-invariant regime, but simple scaling arguments reveal the fact that cone-type singularities do lead to infinite geometric curvature energies; see Figure 2.2. S. Scholtes could indeed demonstrate that embedded polygons have finite integral Menger curvature \mathcal{M}_p if and only if $p \in (0, 3)$; see [103]. Recall that $p = 3$ is the scale-invariant exponent for integral Menger curvature for curves. In addition, Scholtes established certain weak tangential properties of arbitrary (a priori fairly wild) sets at every point if the one-dimensional set $E \subset \mathbb{R}^n$ has finite integral Menger curvature $\mathcal{M}_3(E)$ [104].

So, one can indeed hope for mild regularising effects, like for the energy \mathcal{U}_p for curves for $p = 1$, where we proved in [119, Theorem 1] that finite \mathcal{U}_1 -energy implies that the curve is embedded and in the Sobolev class $W^{2,1}$. However, not every embedded $W^{2,1}$ -curve has finite U_1 -energy; see [119, Example pp. 120–121]. Finiteness of the tangent-point energy \mathcal{E}_p in the scale-invariant case $p = 2$ (see definition (2.1.7)) yields at least a topological one-dimensional manifold – possibly with boundary; see [121, Theorem 1.1]. Only for the Möbius energy (2.1.1), whose Möbius-invariance implies scale-invariance, one has Blatt's [11] characterisation of the appropriate energy spaces

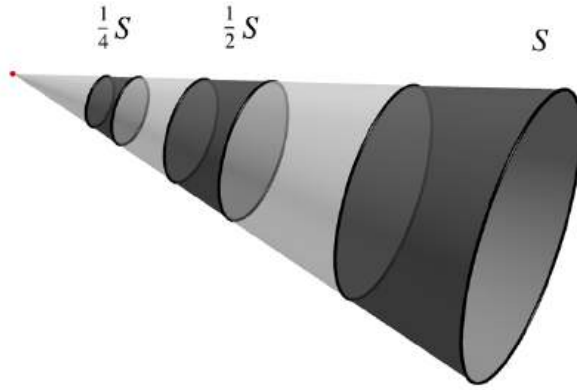


Fig. 2.2: A cone has infinite scale-invariant geometric curvature \mathcal{E} , since scaling of a fixed portion S of the cone leads to the same quantum of energy $\mathcal{E}(S) = \mathcal{E}(S/2) = \mathcal{E}(S/4) = \dots$. Adding up these infinitely many contributions $\mathcal{E}(S/2^i)$, $i \in \mathbb{N}$ leads to a divergent series of positive real numbers as a lower bound for the cone's energy if $(S/2^i) \cap (S/2^j) = \emptyset$ for $i \neq j$.

as the fractional Sobolev space $W^{3/2,2}$ (assuming injective arclength parametrised curves), and already earlier Blatt and Reiter used an idea of He to construct a closed bi-Lipschitz curve with finite Möbius energy that is not differentiable [18, Corollary 4.2]. But very recently, Blatt has established a nice approximation result on convolutions of curves whose tangents have vanishing mean oscillations which in particular implies that arclength parametrised curves of finite Möbius energy can be approximated in the $W^{3/2,2}$ -norm and in energy^{2.7} by smooth curves; see [14, Theorem 1.3]. At present there are a few suggestions how to generalise the Möbius energy to higher-dimensional submanifolds – we are aware of Kusner and Sullivan [67, 68] and D. Auckly and L. Sadun [5] (see also the very recent contribution by O'Hara and G. Solanes [88], [87]) – but no satisfactory analysis regarding regularity or variational issues has been performed yet.

2.7 This has various consequences in geometric knot theory, for instance, it completes Scholtes' recent investigations on a discrete version of the Möbius energy for polygons with n edges, that can now be shown to Γ -converge to the Möbius energy (2.1.1) as $n \rightarrow \infty$; see [105, Theorem 1.1] and [14, Theorem 3.8]. We do not address the very interesting questions regarding suitable discretisations and merely refer to the work of Rawdon et al. [92, 93, 95, 78, 94, 96, 97, 106] on discretised versions of ropelength, and to [108, 105, 107] for discretisations of a few other geometric curvature energies.

The scale-invariant exponent for integral Menger curvature $\mathcal{M}_p(\Sigma)$ on two-dimensional surfaces $\Sigma \subset \mathbb{R}^3$ is $p = 8$, and we proved a Fenchel-type theorem [120, Theorem 1.2].

Theorem 2.2.9 (Fenchel-type theorem). *There is an absolute constant $\gamma_0 > 0$ such that $\mathcal{M}_8(\Sigma) \geq \gamma_0$ for any closed compact connected two-dimensional Lipschitz surface $\Sigma \subset \mathbb{R}^3$.*

Due to our rather rough estimates the constant γ_0 is far from being optimal.

Open questions 2.2.10. *What is the optimal regularity of finite energy Lipschitz curves or submanifolds for any of the geometric curvature energies in the scale-invariant case? How regular are submanifolds of finite energy for a suitable generalisation of the Möbius energy to higher dimensional objects? How much geometric curvature energy does one really need to close a curve or a surface, in other words what are the optimal constants in Fenchel-type theorems like Theorem 2.2.9? Is every k -dimensional manifold \mathcal{M}^k immersed in \mathbb{R}^n automatically embedded if its image has finite scale-invariant tangent-point energy \mathcal{E}_{2k} (see (2.1.10)), or finite scale-invariant integral Menger curvature $\mathcal{M}_{k(k+2)}$ (see (2.1.11))?*

Existence of critical points. For all geometric curvature energies above scale-invariance one can find (at least one) minimising knot in a given isotopy class. This even works for higher-dimensional geometric curvature energies such as integral Menger curvature or tangent-point energies for submanifolds as described above. But are there other critical points, and how can one prove their existence? One of the first attempts in that direction is the work of D. Kim and Kusner [59] on the Möbius energy. They applied R. S. Palais' *principle of symmetric criticality* [89] to obtain $\mathcal{E}_{\text{Möb}}$ -critical torus knots by minimising the Möbius energy within the appropriate subclass of torus knots enjoying particular symmetries. In addition, together with G. Stengle [59, p. 4] they used classic residue calculus from complex analysis to calculate their energy values. Further numerical experiments lead them to conjecture that most of these $\mathcal{E}_{\text{Möb}}$ -critical torus knots are not local minimisers. For the non-smooth ropelength functional Cantarella et al. [28] successfully modified Palais' symmetric criticality principle to find new critical points in several symmetry classes of knots and links, e.g. in the non-trivial (a, b) -torus knots. They used their numerical ropelength minimising algorithm RIDGE RUNNER to compute their respective values for ropelength; see Figure 2.3.

In ongoing cooperative work with A. Gilsbach we apply Palais' principle to O'Hara's repulsive energies, integral Menger curvature, and tangent-point energies to produce symmetric critical configurations in every prescribed knot class. Specifically, in non-trivial (a, b) -torus knot classes we even obtain two distinct symmetric critical knots with this method [46], [47]. Very helpful in that context is the knowledge of the respec-

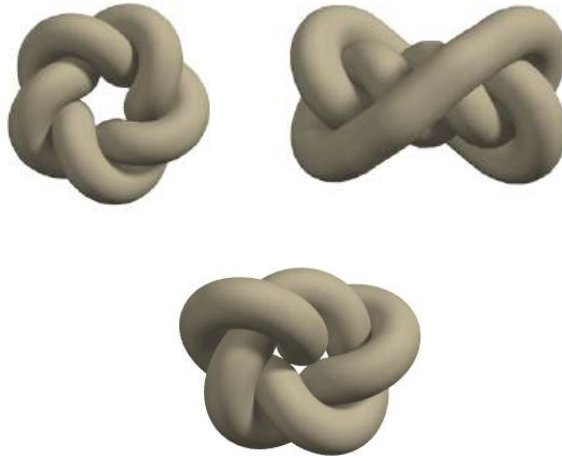
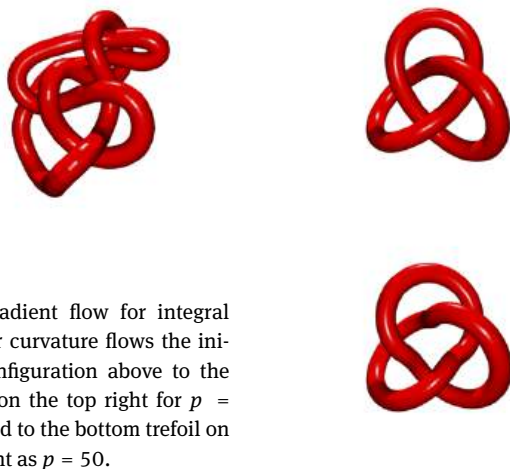


Fig. 2.3: A 5-fold and a 2-fold symmetric ropelength-critical $(2, 5)$ -torus knot, both with ropelength values distinctly larger than the unsymmetric global ropelength minimiser on the right according to the computations with RIDGE RUNNER of Cantarella et al. in [28]. (Images by courtesy of J. Cantarella.)

tive correct energy spaces described in Section 2.1. Gilsbach also uses Γ -convergence arguments to show that her symmetric critical points of integral Menger curvature \mathcal{M}_p do converge to ropelength-critical points as $p \rightarrow \infty$. Recently, Gilsbach has modified T. Hermes' numerical code [52] to actually compute the energy values of the symmetric critical points of integral Menger curvature. Hermes had rigorously derived the first variation formula for integral Menger curvature in the suitable fractional Sobolev space, and could prove that the round circle is a critical point. He created a numerical tool to explore the presumably quite complicated energy landscape of integral Menger curvature. His numerical experiments exhibit among other things the ability of the Menger gradient flow to untangle complicated unknots to the round circle after fairly short time, as well as varying features as p approaches infinity. For p only slightly above the scale-invariant exponent one finds smoothing as the predominant feature (while keeping the curves embedded in contrast to, e.g., the classic mean curvature flow on space curves), whereas for large p , say $p \geq 50$, the similarity to Cantarella's RIDGE RUNNER (corresponding to the case $p = \infty$) is striking [4]: both flows try to embed the curves as nicely as possible.

A second variation formula has been derived and analysed in detail for the Möbius energy by A. Ishizeki and T. Nagasawa [54]. They used a very interesting decomposition theorem for the Möbius energy itself [53], and studied recently the Möbius-invariance of the various parts of that decomposition [55]. Also quite recently J. Knapp-



The gradient flow for integral Menger curvature flows the initial configuration above to the trefoil on the top right for $p = 3.5$, and to the bottom trefoil on the right as $p = 50$.

Fig. 2.4: Different parameters $p > 3$ lead to different final configurations for the gradient flow of the rescaled integral Menger curvature but the knot type is preserved. (Images by courtesy of T. Hermes.)

mann [60] succeeded in deriving rigorously a second variation formula for integral Menger curvature \mathcal{M}_p on curves in the appropriate fractional Sobolev spaces.

The only approach to deal with higher-dimensional critical points for geometric curvature energies is the ingenious paper by A. Nabutovsky [80] who combined complexity theory with real algebraic geometry to prove the existence of infinitely many critical unknotted hyperspheres in \mathbb{R}^n for $n \geq 6$ for a higher-dimensional variant of ropelength.

Open questions 2.2.11. *Are there critical points for geometric curvature energies in every prescribed knot class other than the known global minimisers? In particular, are there critical unknots different^{2.8} from the round circle? And if so, how many are there? Can we relate critical points of different geometric curvature energies with each other, e.g., via Γ -convergence? What can be said about the stability of such critical points? Is it possible to find critical configurations in higher dimensions?*

2.8 Energies that allow such critical unknots would therefore not be suitable to give an alternative proof of the Smale conjecture by means of a gradient flow as, e.g., suggested by Freedman et al. in [41, p. 41] for the Möbius energy. We do not address here the very challenging topic of gradient flows for geometric curvature energies and just refer to the pioneering work of Blatt on the gradient flow for the Möbius energy and other O'Hara energies [12, 15, 16]



Fig. 2.5: Springy knots: figure-eight knot, mathematician's loop, and Chinese button knot. Wire models manufactured by WHY KNOTS, Aptos, in 1980; photographs by B. Bollwerk, Aachen.

Further implications on geometric knot theory. The round circle minimises many of the known geometric knot energies [41, Corollary 2.2], [1], [119, Lemma 7], [129, Corollary 5.12], and we expect the same for integral Menger curvature due to strong numerical evidence based on Hermes' numerical experiments with his gradient flow algorithm [52, Section 4.3]. In addition, we mentioned the explicit continuous families of ropelength-minimising links constructed by Cantarella et al. [29]. More recently, I. Agol, F. C. Marques, and A. Nèves applied their ingenious min-max-theory for minimal currents to resolve not only the famous Willmore conjecture [74] but also a conjecture by Freedman et al. by proving that the stereographic projection of the standard Hopf-link minimises the Möbius energy; see [2]. Apart from these results nothing is known analytically about the shape of non-trivially knotted minimising curves. For the ropelength-minimising trefoil, the so-called *ideal trefoil* one has presumably fairly accurate numerical solutions [8, 30, 9, 4, 91] and some local analytic information on the possible shape of general ideal knots [110, 26, 39, 40, 27] extracted from the complicated necessary conditions.

If one combines geometric curvature energies with (higher order) local energies like the classic bending energy $\mathcal{E}_{\text{bend}}(\gamma) := \int_{\gamma} \kappa^2 ds$, one can study minimal configurations for such energies under topological constraints on the competing curves or surfaces. This leads to the concept of *elastic knots* that can be obtained as limits of minimisers γ_{ϑ} of the total energy

$$\mathcal{E}_{\vartheta}(\gamma) := \mathcal{E}_{\text{bend}}(\gamma) + \vartheta \frac{1}{\Delta[\gamma]} \quad (2.2.7)$$

as $\vartheta \rightarrow 0$. Recall that $\Delta[\gamma]$ denotes the thickness as defined in (2.1.2), so that the particular geometric curvature energy chosen in (2.2.7) is the ropelength functional if one restricts to curves of length one. Indeed, it can be shown [43, Theorem 2.2] that in every given knot class one finds such a limiting curve γ_0 , which has smaller bending energy than any knotted competitor. However, as one would expect from the simple toy models of knotted wires designed by J. C. Langer, see Figure 2.5, γ_0 has self-intersections unless the given knot class is the unknot in which case γ_0 is the round circle [43, Proposition 3.1].

Is there anything one can say about the actual shape of the elastic knot γ_0 for non-trivial knot classes? This is indeed the case for any $(2, b)$ -torus knot as shown in [43, Corollary 6.5]:

Theorem 2.2.12 (Elastic $(2, b)$ -torus knots). *For any odd integer $|b| \geq 3$ the unique elastic $(2, b)$ -torus knot is the doubly covered circle. In particular, the elastic trefoil is the doubly covered circle.*

This result confirms mechanical and numerical experiments (see Figure 2.6), as well as the heuristics and Metropolis Monte Carlo simulations of R. Gallotti and O. Pierre-Louis [42, 90], and the numerical gradient-descent results by S. Avvakumov and A. Sossinsky [6]. However, adding twist changes the geometry of the springy wire drastically; see bottom right of Figure 2.6. And there is no theory yet, describing these twist effects for knotted elastic wires.

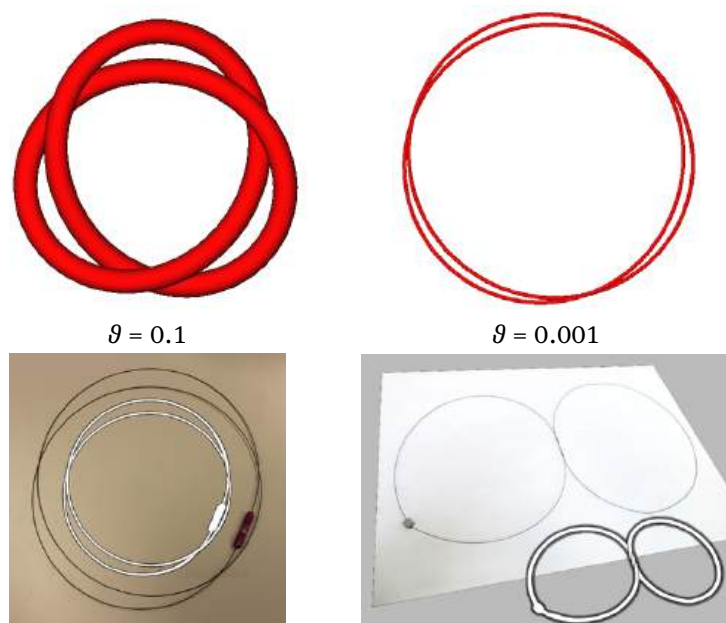


Fig. 2.6: Top: Numerically computed minimisers of the total energy \mathcal{E}_g in the class of trefoils approaching the doubly covered circle as g tends to zero. Bottom left: Mechanical experiments: The springy trefoil knot is close to the doubly covered circle. Bottom right: Adding twist leads to a stable flat trefoil configuration close to a planar figure-eight. (Wire models by courtesy of J. H. Maddocks, Lausanne.)

Open questions 2.2.13. *Does the round circle minimise integral Menger curvature (2.1.5)? Can one prove more about the actual shape of ideal knots? What is the shape of links with more than two components minimising the Möbius energy? Can one identify the shapes of global minimisers of other geometric curvature energies for curves? What can be said about elastic knots for non-trivial knot classes different from $(2, b)$ -torus knots? What is the shape of twisted elastic knots?*

A higher-dimensional branch of geometric knot theory is naturally much less developed yet. The shape of possible minimising configurations for higher-dimensional geometric curvature energies is wide open and seems currently out of reach. There is one exception, however: We proved in [64, Theorem 1.5] with the isoperimetric inequality and a simple measure-theoretic argument the following uniqueness result for the global tangent-point-energy (2.1.16):

Theorem 2.2.14 (Spheres are unique minimisers). *The round sphere uniquely minimises the global tangent-point energy (2.1.16) among all compact embedded C^1 -hypersurfaces in \mathbb{R}^n .*

Recently, we proved in [65] that many higher-dimensional geometric curvature energies including integral Menger curvature (2.1.8), (2.1.11) or tangent-point energies (2.1.10), or their more singular variants (2.1.15), (2.1.16), are valuable knot energies. All these energies are self-repulsive (on the scale above scale-invariance), lower-semicontinuous on sublevel sets with respect to Hausdorff-convergence, they enjoy nice compactness properties and can thus be minimised in given isotopy classes; see [65, Theorem 2, Corollary 1]. They also bound the number of isotopy types with explicit constants only depending on the energy level and the integrability exponent, on a diameter bound, and on the dimensions [65, Theorem 1 & Remark 1.1]. In particular, one has the following boundedness result.

Theorem 2.2.15 (Isotopy finiteness). *Let $E, d > 0$ be given. Then there are at most $K = K(E, d, k, n, p)$ different ambient isotopy types among all k -dimensional Lipschitz submanifolds $\Sigma \subset \mathbb{R}^n$ with integral Menger curvature $\mathcal{M}_p(\Sigma) \leq E$ and $\text{diam } \Sigma \leq d$. This constant may be estimated as*

$$\log \log K \leq c(k, n, p) \left(|\log d| + \log(E^{1/p} + 1) + 1 \right). \quad (2.2.8)$$

This result can be compared to a whole series of finiteness theorems of diffeomorphism types under given bounds on classic curvatures, beginning with the work of J. Cheeger [31], and extended by many others, see, e.g., Cheeger's exhaustive survey [32] and the references therein. The notable difference is here that we deal with embedded submanifolds of lower regularity, whose Riemannian metrics are just Hölder continuous so that the classic notion of curvature does not make sense. The geometric curvature energies in the regime above scale-invariance turn out to be valuable substitutes. The

only comparable result with this emphasis is the work of O. Durumeric [38] who, however, works in the context of $C^{1,1}$ -submanifolds with positive thickness.

Higher-dimensional variants of elastic knots have not been discussed explicitly yet, but L. Simon's pioneering work [113] solves the problem of minimising the Willmore energy

$$\int_{\Sigma} H^2(x) d\mathcal{H}^2(x) \quad (2.2.9)$$

in the class of two-dimensional embedded surfaces with prescribed genus or under alternative constraints; see also [10, 100, 111, 69, 58, 99, 79]. But minimising the Willmore energy or related functionals such as the Helfrich functional [51, 81] on given isotopy classes has to the best of our knowledge not been investigated yet — with the exception of recent work of P. Breuning, J. Hirsch, and E. Mäder-Baumdicker [23] on Willmore minimising Klein bottles.

Open questions 2.2.16. *Is it possible to identify the shape of global minimisers of higher-dimensional geometric curvature energies? The explicit constant in Theorem 2.2.15 is far from being optimal, what is the best constant bounding the number of isotopy types under given energy values? For curves the energies could often be related to knot invariants or to quantities like the average crossing number (2.1.17) controlling knot invariants. Are there meaningful topological invariants or geometric quantities for higher-dimensional knots that could be controlled by means of higher-dimensional geometric curvature energies? Are there higher-dimensional elastic knots, for instance minimisers of the Willmore functional in arbitrary prescribed isotopy classes? And can one say anything about their shapes?*

Acknowledgment: The first author was partially supported by NCN Grant no. 2013/10/M/ST1/00416 *Geometric curvature energies for subsets of the Euclidean space*. The second author's work is partially funded by DFG Grant no. Mo 966/7-1 *Geometric curvature functionals: energy landscape and discrete methods* and by the Excellence Initiative of the German federal and state governments.

Bibliography

- [1] Aaron Abrams, Jason Cantarella, Joseph H. G. Fu, Mohammad Ghomi, and Ralph Howard. Circles minimize most knot energies. *Topology*, 42(2):381–394, 2003.
- [2] Ian Agol, Fernando C. Marques, and André Neves. Min-max theory and the energy of links. *J. Amer. Math. Soc.*, 29(2):561–578, 2016.
- [3] William K. Allard. On the first variation of a varifold. *Ann. of Math. (2)*, 95:417–491, 1972.

- [4] Ted Ashton, Jason Cantarella, Michael Piatek, and Eric J. Rawdon. Knot tightening by constrained gradient descent. *Exp. Math.*, 20(1):57–90, 2011.
- [5] David Auckly and Lorenzo Sadun. A family of Möbius invariant 2-knot energies. In *Geometric topology (Athens, GA, 1993)*, volume 2 of *AMS/IP Stud. Adv. Math.*, pages 235–258. Amer. Math. Soc., Providence, RI, 1997.
- [6] Sergey Avvakumov and Alexey Sossinsky. On the normal form of knots. *Russ. J. Math. Phys.*, 21(4):421–429, 2014.
- [7] Jonas Azzam and Xavier Tolsa. Characterization of n -rectifiability in terms of Jones’ square function: Part II. *Geom. Funct. Anal.*, 25(5):1371–1412, 2015.
- [8] Justyna Baranska, Piotr Pieranski, and Eric J. Rawdon. Ropelength of tight polygonal knots. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 293–321. World Sci. Publ., Singapore, 2005.
- [9] Justyna Baranska, Sylwester Przybyl, and Piotr Pieranski. Curvature and torsion of the tight closed trefoil knot. *Eur. Phys. J. B*, 66(4):547–556, 2008.
- [10] Matthias Bauer and Ernst Kuwert. Existence of minimizing Willmore surfaces of prescribed genus. *Int. Math. Res. Not.*, (10):553–576, 2003.
- [11] Simon Blatt. Boundedness and regularizing effects of O’Hara’s knot energies. *J. Knot Theory Ramifications*, 21(1):1250010, 9, 2012.
- [12] Simon Blatt. The gradient flow of the Möbius energy near local minimizers. *Calc. Var. Partial Differential Equations*, 43(3-4):403–439, 2012.
- [13] Simon Blatt. The energy spaces of the tangent-point energies. *J. Topol. Anal.*, 5(261):261–270, 2013.
- [14] Simon Blatt. Curves between Lipschitz and C^1 and their relation to geometric knot theory. *ArXiv e-prints*, February 2016.
- [15] Simon Blatt. The Gradient Flow of O’Hara’s Knot Energies. *Math. Ann.*, 370(3-4):993–1061, 2018.
- [16] Simon Blatt. The Gradient Flow of the Möbius energy: ε -regularity and consequences. *ArXiv e-prints*, January 2016.
- [17] Simon Blatt and Sławomir Kolasiński. Sharp boundedness and regularizing effects of the integral Menger curvature for submanifolds. *Adv. Math.*, 230(3):839–852, 2012.
- [18] Simon Blatt and Philipp Reiter. Does finite knot energy lead to differentiability? *J. Knot Theory Ramifications*, 17(10):1281–1310, 2008.
- [19] Simon Blatt and Philipp Reiter. Stationary points of O’Hara’s knot energies. *Manuscripta Math.*, 140(1-2):29–50, 2013.
- [20] Simon Blatt and Philipp Reiter. Regularity theory for tangent-point energies: the non-degenerate sub-critical case. *Adv. Calc. Var.*, 8(2):93–116, 2015.
- [21] Simon Blatt and Philipp Reiter. Towards a regularity theory for integral Menger curvature. *Ann. Acad. Sci. Fenn. Math.*, 40(1):149–181, 2015.
- [22] Simon Blatt, Philipp Reiter, and Armin Schikorra. Harmonic analysis meets critical knots. Critical points of the Möbius energy are smooth. *Trans. Amer. Math. Soc.*, 368(9):6391–6438, 2016.

- [23] Patrick Breuning, Jonas Hirsch, and Elena Mäder-Baumdicker. Existence of minimizing Willmore Klein bottles in Euclidean four-space. *Geom. Topol.*, 21(4):2485–2526, 2017.
- [24] Gregory Buck and Jonathan Simon. Energy and length of knots. In *Lectures at KNOTS '96 (Tokyo)*, volume 15 of *Ser. Knots Everything*, pages 219–234. World Sci. Publ., River Edge, NJ, 1997.
- [25] Jorge A. Calvo, Kenneth C. Millett, Eric J. Rawdon, and Andrzej Stasiak, editors. *Physical and numerical models in knot theory*, volume 36 of *Series on Knots and Everything*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005. Including applications to the life sciences.
- [26] Jason Cantarella, Joseph H. G. Fu, Rob Kusner, John M. Sullivan, and Nancy C. Wrinkle. Criticality for the Gehring link problem. *Geom. Topol.*, 10:2055–2116 (electronic), 2006.
- [27] Jason Cantarella, Joseph H. G. Fu, Robert B. Kusner, and John M. Sullivan. Ropelength criticality. *Geom. Topol.*, 18(4):1973–2043, 2014.
- [28] Jason Cantarella, Joseph H. G. Fu, Matt Mastin, and Jennifer Ellis Royal. Symmetric criticality for tight knots. *J. Knot Theory Ramifications*, 23(2):1450008, 17, 2014.
- [29] Jason Cantarella, Robert B. Kusner, and John M. Sullivan. On the minimum ropelength of knots and links. *Invent. Math.*, 150(2):257–286, 2002.
- [30] Matthias Carlen, Ben Laurie, John H. Maddocks, and Jana Smutny. Biarcs, global radius of curvature, and the computation of ideal knot shapes. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 75–108. World Sci. Publ., Singapore, 2005.
- [31] Jeff Cheeger. Finiteness theorems for Riemannian manifolds. *Amer. J. Math.*, 92:61–74, 1970.
- [32] Jeff Cheeger. Structure theory and convergence in Riemannian geometry. *Milan J. Math.*, 78(1):221–264, 2010.
- [33] Francesca Da Lio. Fractional harmonic maps into manifolds in odd dimension $n > 1$. *Calc. Var. Partial Differential Equations*, 48(3-4):421–445, 2013.
- [34] Francesca Da Lio and Tristan Rivière. Sub-criticality of non-local Schrödinger systems with antisymmetric potentials and applications to half-harmonic maps. *Adv. Math.*, 227(3):1300–1348, 2011.
- [35] Guy David. Unrectifiable 1-sets have vanishing analytic capacity. *Rev. Mat. Iberoamericana*, 14(2):369–479, 1998.
- [36] Guy David and Stephen Semmes. *Analysis of and on uniformly rectifiable sets*, volume 38 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1993.
- [37] John P. Duggan. $W^{2,p}$ regularity for varifolds with mean curvature. *Comm. Partial Differential Equations*, 11(9):903–926, 1986.
- [38] Oguz C. Durumeric. Thickness formula and C^1 -compactness for $C^{1,1}$ Riemannian submanifolds, 2002.

- [39] Oguz C. Durumeric. Local structure of ideal shapes of knots. *Topology Appl.*, 154(17):3070–3089, 2007.
- [40] Oguz C. Durumeric. Local structure of ideal knots. II. Constant curvature case. *J. Knot Theory Ramifications*, 18(11):1525–1537, 2009.
- [41] Michael H. Freedman, Zheng-Xu He, and Zhenghan Wang. Möbius energy of knots and unknots. *Ann. of Math. (2)*, 139(1):1–50, 1994.
- [42] Riccardo Gallotti and Olivier Pierre-Louis. Stiff knots. *Phys. Rev. E* (3), 75(3):031801, 14, 2007.
- [43] Henryk Gerlach, Philipp Reiter, and Heiko von der Mosel. The elastic trefoil is the doubly covered circle. *Arch. Ration. Mech. Anal.*, 225(1):89–139, 2017.
- [44] Henryk Gerlach and Heiko von der Mosel. On sphere-filling ropes. *Amer. Math. Monthly*, 118(10):863–876, 2011.
- [45] Henryk Gerlach and Heiko von der Mosel. What are the longest ropes on the unit sphere? *Arch. Ration. Mech. Anal.*, 201(1):303–342, 2011.
- [46] A. Gilsbach and H. von der Mosel. Symmetric critical knots for O’Hara’s energies. *Topol. Appl.*, 242, (15): 73–102, 2018.
- [47] Alexandra Gilsbach. *On symmetric critical points of knot energies*. PhD thesis, RWTH Aachen University, 2018.
- [48] Oscar Gonzalez and John H. Maddocks. Global curvature, thickness, and the ideal shapes of knots. *Proc. Natl. Acad. Sci. USA*, 96(9):4769–4773 (electronic), 1999.
- [49] Oscar Gonzalez, John H. Maddocks, Friedemann Schuricht, and Heiko von der Mosel. Global curvature and self-contact of nonlinearly elastic curves and rods. *Calc. Var. Partial Differential Equations*, 14(1):29–68, 2002.
- [50] Zheng-Xu He. The Euler-Lagrange equation and heat flow for the Möbius energy. *Comm. Pure Appl. Math.*, 53(4):399–431, 2000.
- [51] Wolfgang Helfrich. Elastic Properties of Lipid Bilayers: Theory and Possible Experiments. *Z. Naturforsch.*, 28c:693–703, 1973.
- [52] Tobias Hermes. Analysis of the first variation and a numerical gradient flow for integral Menger curvature. *ArXiv e-prints*, August 2014.
- [53] Aya Ishizeki and Takeyuki Nagasawa. A decomposition theorem of the Möbius energy I: Decomposition and Möbius invariance. *Kodai Math. J.*, 37(3):737–754, 2014.
- [54] Aya Ishizeki and Takeyuki Nagasawa. A decomposition theorem of the Möbius energy II: variational formulae and estimates. *Math. Ann.*, 363(1-2):617–635, 2015.
- [55] Aya Ishizeki and Takeyuki Nagasawa. The invariance of decomposed Möbius energies under inversions with center on curves. *J. Knot Theory Ramifications*, 25(2):1650009, 22, 2016.
- [56] Peter W. Jones. Rectifiable sets and the traveling salesman problem. *Invent. Math.*, 102(1):1–15, 1990.
- [57] Malte Laurens Kampschulte. The symmetrized tangent-point energy. Master’s thesis, RWTH Aachen University, 2012.

- [58] Laura Gioia Andrea Keller, Andrea Mondino, and Tristan Rivière. Embedded surfaces of arbitrary genus minimizing the Willmore energy under isoperimetric constraint. *Arch. Ration. Mech. Anal.*, 212(2):645–682, 2014.
- [59] Denise Kim and Rob Kusner. Torus knots extremizing the Möbius energy. *Experiment. Math.*, 2(1):1–9, 1993.
- [60] Jan Knappmann. *On the second variation of integral Menger curvature*. PhD thesis, RWTH Aachen University, 2018. In progress.
- [61] Sławomir Kolasiński. *Integral Menger curvature for sets of arbitrary dimension and codimension*. PhD thesis, Institute of Mathematics, University of Warsaw, 2011.
- [62] Sławomir Kolasiński. Geometric Sobolev-like embedding using high-dimensional Menger-like curvature. *Trans. Amer. Math. Soc.*, 367(2):775–811, 2015.
- [63] Sławomir Kolasiński. Higher order rectifiability of rectifiable sets via averaged discrete curvatures. *Rev. Mat. Iberoam.*, 33(3):861–884, 2017.
- [64] Sławomir Kolasiński, Paweł Strzelecki, and Heiko von der Mosel. Characterizing $W^{2,p}$ submanifolds by p -integrability of global curvatures. *Geom. Funct. Anal.*, 23(3):937–984, 2013.
- [65] Sławomir Kolasiński, Paweł Strzelecki, and Heiko von der Mosel. Compactness and isotopy finiteness for submanifolds with uniformly bounded geometric curvature energies. *ArXiv e-prints*, April 2015, To appear in *Communications in Analysis and Geometry*.
- [66] Sławomir Kolasiński and Marta Szumańska. Minimal Hölder regularity implying finiteness of integral Menger curvature. *Manuscripta Math.*, 141(1-2):125–147, 2013.
- [67] Robert B. Kusner and John M. Sullivan. Möbius energies for knots and links, surfaces and submanifolds. In *Geometric topology (Athens, GA, 1993)*, volume 2 of *AMS/IP Stud. Adv. Math.*, pages 570–604. Amer. Math. Soc., Providence, RI, 1997.
- [68] Robert B. Kusner and John M. Sullivan. Möbius-invariant knot energies. In *Ideal knots*, volume 19 of *Ser. Knots Everything*, pages 315–352. World Sci. Publ., River Edge, NJ, 1998.
- [69] Ernst Kuwert and Reiner Schätzle. Minimizers of the Willmore functional under fixed conformal class. *J. Differential Geom.*, 93(3):471–530, 2013.
- [70] Jean-Christophe L  ger. Menger curvature and rectifiability. *Ann. of Math. (2)*, 149(3):831–869, 1999.
- [71] Gilad Lerman and J. Tyler Whitehouse. High-dimensional Menger-type curvatures. Part II: d -separation and a menagerie of curvatures. *Constr. Approx.*, 30(3):325–360, 2009.
- [72] Gilad Lerman and J. Tyler Whitehouse. High-dimensional Menger-type curvatures. Part I: Geometric multipoles and multiscale inequalities. *Rev. Mat. Iberoam.*, 27(2):493–555, 2011.

- [73] Richard A. Litherland, Jonathan K. Simon, Oguz C. Durumeric, and Eric J. Rawdon. Thickness of knots. *Topology Appl.*, 91(3):233–244, 1999.
- [74] Fernando C. Marques and André Neves. Min-max theory and the Willmore conjecture. *Ann. of Math. (2)*, 179(2):683–782, 2014.
- [75] Martin Meurer. Integral Menger Curvature and Rectifiability of n -dimensional Borel sets in Euclidean N -space. *Trans. Amer. Math. Soc.*, 370(2):1185–1250, 2018.
- [76] Cristian Micheletti and Davide Marenduzzo. Modelling DNA as a flexible thick polymer: DNA elasticity and packaging thermodynamics. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 127–147. World Sci. Publ., Singapore, 2005.
- [77] Cristian Micheletti, Davide Marenduzzo, and Enzo Orlandini. Polymers with spatial or topological constraints: theoretical and computational results. *Phys. Rep.*, 504(1):1–73, 2011.
- [78] Kenneth C. Millett and Eric J. Rawdon. Energy, ropelength, and other physical aspects of equilateral knots. *J. Comput. Phys.*, 186(2):426–456, 2003.
- [79] Stefan Müller and Matthias Röger. Confined structures of least bending energy. *J. Differential Geom.*, 97(1):109–139, 2014.
- [80] Alexander Nabutovsky. Non-recursive functions, knots “with thick ropes”, and self-clenching “thick” hyperspheres. *Comm. Pure Appl. Math.*, 48(4):381–428, 1995.
- [81] Johannes C. C. Nitsche. Boundary value problems for variational integrals involving surface curvatures. *Quart. Appl. Math.*, 51(2):363–387, 1993.
- [82] Jun O’Hara. Energy of a knot. *Topology*, 30(2):241–247, 1991.
- [83] Jun O’Hara. Energy functionals of knots. In *Topology Hawaii (Honolulu, HI, 1990)*, pages 201–214. World Sci. Publishing, River Edge, NJ, 1992.
- [84] Jun O’Hara. Family of energy functionals of knots. *Topology Appl.*, 48(2):147–161, 1992.
- [85] Jun O’Hara. Energy functionals of knots. II. *Topology Appl.*, 56(1):45–61, 1994.
- [86] Jun O’Hara. *Energy of knots and conformal geometry*, volume 33 of *Series on Knots and Everything*. World Scientific Publishing Co. Inc., River Edge, NJ, 2003.
- [87] Jun O’Hara and Gil Solanes. Möbius invariant energies and average linking with circles. *Tohoku Math. J. (2)*, 67(1):51–82, 2015.
- [88] Jun O’Hara and Gil Solanes. Regularized Riesz energies of submanifolds. *ArXiv e-prints*, December 2015, Mathematische Nachrichten, available online <https://onlinelibrary.wiley.com/doi/epdf/10.1002/mana.201600083>.
- [89] Richard S. Palais. The principle of symmetric criticality. *Comm. Math. Phys.*, 69(1):19–30, 1979.
- [90] Olivier Pierre-Louis. On the geometry of stiff knots. *Eur. Phys. J. B*, 71(2):281–288, 2009.
- [91] Sylwester Przybył and Piotr Pieranski. High resolution portrait of the ideal trefoil knot. *J. Phys. A*, 47(28):285201, 20, 2014.

- [92] Eric J. Rawdon. Approximating the thickness of a knot. In *Ideal knots*, volume 19 of *Ser. Knots Everything*, pages 143–150. World Sci. Publ., River Edge, NJ, 1998.
- [93] Eric J. Rawdon. Approximating smooth thickness. *J. Knot Theory Ramifications*, 9(1):113–145, 2000.
- [94] Eric J. Rawdon. Can computers discover ideal knots? *Experiment. Math.*, 12(3):287–302, 2003.
- [95] Eric J. Rawdon and Jonathan Simon. Möbius energy of thick knots. *Topology Appl.*, 125(1):97–109, 2002.
- [96] Eric J. Rawdon and Jonathan K. Simon. Polygonal approximation and energy of smooth knots. *J. Knot Theory Ramifications*, 15(4):429–451, 2006.
- [97] Eric J. Rawdon and Joseph Worthington. Error analysis of the minimum distance energy of a polygonal knot and the Möbius energy of an approximating curve. *J. Knot Theory Ramifications*, 19(8):975–1000, 2010.
- [98] Philipp Reiter. Regularity theory for the Möbius energy. *Commun. Pure Appl. Anal.*, 9(5):1463–1471, 2010.
- [99] Tristan Rivière. Variational principles for immersed surfaces with L^2 -bounded second fundamental form. *J. Reine Angew. Math.*, 695:41–98, 2014.
- [100] Reiner Schätzle. The Willmore boundary problem. *Calc. Var. Partial Differential Equations*, 37(3-4):275–302, 2010.
- [101] Armin Schikorra. Interior and Boundary-Regularity for Fractional Harmonic Maps on Domains. *ArXiv e-prints*, March 2011.
- [102] Armin Schikorra. Regularity of $n/2$ -harmonic maps into spheres. *J. Differential Equations*, 252(2):1862–1911, 2012.
- [103] Sebastian Scholtes. For which positive p is the integral Menger curvature \mathcal{M}_p finite for all simple polygons?, 2012. Preprint no. 50 (2011), RWTH Aachen.
- [104] Sebastian Scholtes. Tangency properties of sets with finite geometric curvature energies. *Fund. Math.*, 218(2):165–191, 2012.
- [105] Sebastian Scholtes. Discrete Möbius energy. *J. Knot Theory Ramifications*, 23(9):1450045, 16, 2014.
- [106] Sebastian Scholtes. Discrete thickness. *Mol. Based Math. Biol.*, 2:73–85, 2014.
- [107] Sebastian Scholtes. Discrete knot energies. In *New Directions in Geometric and Applied Knot Theory*, De Gruyter, Berlin/Boston, 2018.
- [108] Sebastian Scholtes, Henrik Schumacher, and Max Wardetzky. Convergence of discrete elastica. In *Oberwolfach Reports Vol. 9, no. 3*. European Mathematical Society (EMS), Zürich, 2012.
- [109] Friedemann Schuricht and Heiko von der Mosel. Euler-Lagrange equations for nonlinearly elastic rods with self-contact. *Arch. Ration. Mech. Anal.*, 168(1):35–82, 2003.
- [110] Friedemann Schuricht and Heiko von der Mosel. Characterization of ideal knots. *Calc. Var. Partial Differential Equations*, 19(3):281–305, 2004.
- [111] Johannes Schygulla. Willmore minimizers with prescribed isoperimetric ratio. *Arch. Ration. Mech. Anal.*, 203(3):901–941, 2012.

- [112] Leon Simon. *Lectures on geometric measure theory*, volume 3 of *Proceedings of the Centre for Mathematical Analysis, Australian National University*. Australian National University, Centre for Mathematical Analysis, Canberra, 1983.
- [113] Leon Simon. Existence of surfaces minimizing the Willmore functional. *Comm. Anal. Geom.*, 1(2):281–326, 1993.
- [114] Paweł Strzelecki, Marta Szumańska, and Heiko von der Mosel. A geometric curvature double integral of Menger type for space curves. *Ann. Acad. Sci. Fenn. Math.*, 34(1):195–214, 2009.
- [115] Paweł Strzelecki, Marta Szumańska, and Heiko von der Mosel. Regularizing and self-avoidance effects of integral Menger curvature. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 9(1):145–187, 2010.
- [116] Paweł Strzelecki, Marta Szumańska, and Heiko von der Mosel. On some knot energies involving Menger curvature. *Topology Appl.*, 160(13):1507–1529, 2013.
- [117] Paweł Strzelecki and Heiko von der Mosel. On a mathematical model for thick surfaces. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 547–564. World Sci. Publ., Singapore, 2005.
- [118] Paweł Strzelecki and Heiko von der Mosel. Global curvature for surfaces and area minimization under a thickness constraint. *Calc. Var. Partial Differential Equations*, 25(4):431–467, 2006.
- [119] Paweł Strzelecki and Heiko von der Mosel. On rectifiable curves with L^p -bounds on global curvature: self-avoidance, regularity, and minimizing knots. *Math. Z.*, 257(1):107–130, 2007.
- [120] Paweł Strzelecki and Heiko von der Mosel. Integral Menger curvature for surfaces. *Adv. Math.*, 226(3):2233–2304, 2011.
- [121] Paweł Strzelecki and Heiko von der Mosel. Tangent-point self-avoidance energies for curves. *J. Knot Theory Ramifications*, 21(5):1250044, 28, 2012.
- [122] Paweł Strzelecki and Heiko von der Mosel. Menger curvature as a knot energy. *Physics Reports*, 530:257–290, 2013.
- [123] Paweł Strzelecki and Heiko von der Mosel. Tangent-point repulsive potentials for a class of non-smooth m -dimensional sets in \mathbb{R}^n . Part I: Smoothing and self-avoidance effects. *J. Geom. Anal.*, 23(3):1085–1139, 2013.
- [124] Paweł Strzelecki and Heiko von der Mosel. How averaged menger curvatures control regularity and topology of curves and surfaces. *Journal of Physics: Conference Series*, 544(1):012018, 2014.
- [125] John M. Sullivan. Approximating ropelength by energy functions. In *Physical knots: knotting, linking, and folding geometric objects in \mathbb{R}^3* (Las Vegas, NV, 2001), volume 304 of *Contemp. Math.*, pages 181–186. Amer. Math. Soc., Providence, RI, 2002.
- [126] Xavier Tolsa. *Analytic capacity, the Cauchy transform, and non-homogeneous Calderón-Zygmund theory*, volume 307 of *Progress in Mathematics*. Birkhäuser/Springer, Cham, 2014.

- [127] Xavier Tolsa. Characterization of n -rectifiability in terms of Jones' square function: part I. *Calc. Var. Partial Differential Equations*, 54(4):3643–3665, 2015.
- [128] Hans Triebel. *Theory of function spaces*, volume 38 of *Mathematik und ihre Anwendungen in Physik und Technik [Mathematics and its Applications in Physics and Technology]*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1983.
- [129] Alexander Volkmann. A monotonicity formula for free boundary surfaces with respect to the unit ball. *Comm. Anal. Geom.*, 24(1):195–221, 2016.

Takeyuki Nagasawa

On Möbius invariant decomposition of the Möbius energy

3.1 O'Hara's knot energies

In his early paper [17], O'Hara took inspiration from Fukuhara's paper [6] to propose the use of the knot energy

$$\mathcal{E}_{(\alpha,p)}(\mathbf{f}) = \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \left(\frac{1}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^3}^\alpha} \frac{1}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^\alpha} \right)^p ds_1 ds_2$$

to determine the “canonical configuration” of knots of a given knot type. Here $\mathbf{f} : \mathbb{R}/\mathcal{L}\mathbb{Z} \rightarrow \mathbb{R}^3$ represents a knot, *i.e.*, a closed curve without self-intersections that is parametrized by arc-length s and has total length \mathcal{L} . The notation \mathcal{D} represents the intrinsic distance (*i.e.*, the distance along the curve) between two points $\mathbf{f}(s_1)$ and $\mathbf{f}(s_2)$, and α and p are positive constants. Because the canonical configuration should be of a shape that makes it as easy as possible to determine how it is knotted, O'Hara first considered the negative power of the Euclidean distance of two points as the energy density. However, this may result in the divergence of energy for all knots. Thus, to avoid such a situation, he subtracted the intrinsic distance raised to the same power. This means that the energy calculates the proximity of every two points, but excludes the singularity at the diagonal set $\{(s_1, s_2) \in (\mathbb{R}/\mathcal{L}\mathbb{Z})^2 \mid s_1 = s_2\}$. This procedure extracts the convergence part from the divergence integral, which O'Hara calls renormalization, and is mathematically equivalent to Hadamard's regularization.


Let us consider the problem of minimizing O'Hara's energy in a given knot type $[K]$, *i.e.*, the ambient class of given knot K .

Remark 3.1.1. When $\alpha p \neq 2$, the energy does not exhibit scaling invariance, and therefore we consider the problem under a length constraint.

Because the energy \mathcal{E} is non-negative, there exists a minimizing sequence in the set of “all knots”. The first problem is the existence of minimizers in the class $[K]$. When $\alpha p \geq 2$, the energy induces self-repulsiveness. That is, the knots cannot be continuously deformed into a different class while preserving the finiteness of the energy.

Takeyuki Nagasawa, Department of Mathematics, Graduate School of Science and Engineering, Saitama University, Japan, E-mail: tnagasaw@rimath.saitama-u.ac.jp

<https://doi.org/10.1515/9783110571493-002>

Open Access.  © 2018 Takeyuki Nagasawa, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

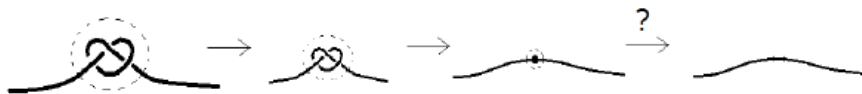


Fig. 3.1: pull-tight (The image was firstly appeared in [16].)

Theorem 3.1.1 ([19]). $\mathcal{E}_{(\alpha,p)}$ is finite for smooth embedded knots if and only if α and p satisfy

$$\alpha \leq 2, \text{ or } p < \frac{1}{\alpha - 2} \text{ and } 2 < \alpha < 4.$$

Furthermore, if $\alpha p \geq 2$, then $\mathcal{E}_{(\alpha,p)}$ induces self-repulsiveness.

This means that there exists an *energy wall* between two classes of knots when $\alpha p \geq 2$. In other words, when $\alpha p < 2$, there is no *energy wall*, and therefore the minimizing sequence can freely move into other knot classes. This suggests that minimizers are not guaranteed to exist for any given knot type except for the trivial type.

Furthermore there is an *energy hole* when $\alpha p = 2$. This is the case when the energy is scaling-invariant. Scaling-invariance may allow the pull-tight phenomena to occur along the sequence. If such a phenomena occurs, then the limit knot, if it exists, is not of the same knot type as the sequence. The energy behavior along the pull-tight for $\mathcal{E}_{(\alpha,p)}$ has been studied by O'Hara [18].

Theorem 3.1.2 ([18]). Let a knot K_ε be a connected sum of K and a small tangle T_ε . The difference of the energy $D(\varepsilon) = \mathcal{E}_{(\alpha,p)}(K_\varepsilon) - \mathcal{E}_{(\alpha,p)}(K)$ behaves as follows in a pull-tight process $T_\varepsilon \rightarrow \{\text{a point}\}$:

- $D(\varepsilon)$ blows up when $\alpha p > 2$.
- $D(\varepsilon)$ converges to a positive constant when $\alpha p = 2$.
- $D(\varepsilon)$ vanishes when $\alpha p < 2$.

Consequently, when $\alpha p = 2$, a knot can be moved into another class while preserving the finiteness of the energy, through a pull-tight process that we call the *energy hole*. This effect is related to the concentration of energy density at scaling-invariant energies.

Hence, when $\alpha p > 2$, the energy wall is present but without energy holes, and we can expect that the direct method of calculus of variations can be applied. Indeed, the following result was demonstrated by O'Hara.

Theorem 3.1.3 ([19]). Let $n = 3$. Then, there exists a minimizer (under rescaling) for any knot types if $\alpha p > 2$.

We call the cases $\alpha p > 2$, $= 2$, and < 2 *subcritical*, *critical*, and *supercritical*, respectively.

3.2 Freedman-He-Wang's procedure and the Kusner-Sullivan conjecture

A pull-tight entails the disappearance of a tangle. The above results implies that a pull-tight can occur in critical and supercritical cases. This shows that the argument of minimizing sequence fails in such cases. Nevertheless, the argument applies for the case $(\alpha, p) = (2, 1)$ in *prime* knot classes. This remarkable result was proven by Freedman-He-Wang [5].

Definiton 3.2.1. Let $n = 3$. A knot is a *composite knot* if it is a connected sum of two non-trivial knots. A *prime knot* is a knot that is neither composite nor trivial.

The authors showed that $\mathcal{E}_{(2,1)}$ is not only invariant under scaling, but also under Möbius transformations. Since then, this has been called the *Möbius energy*.

Theorem 3.2.1 ([5]). There exists a minimizer of $\mathcal{E}_{(2,1)}$ on each *prime* or *trivial* knot type.

The key point is avoiding the occurrence of a pull-tight along the minimizing sequence. If the knot is prime, then we can enlarge the tangle by inversion with respect to a sphere near the shrinking tangle, without changing the energy level. The authors passed to the limit of the minimizing sequence together with, if necessary, the enlarged tangle, and showed the limit knot is a minimizer in the given knot class. If the knot is composite, then two tangles may shrink simultaneously. Hence, such a method does not apply.

For a knot class $[K]$, we denote

$$\tilde{\mathcal{E}}_{(2,1)}([K]) = \inf_{f \in [K]} \mathcal{E}_{(2,1)}(f) - 4.$$

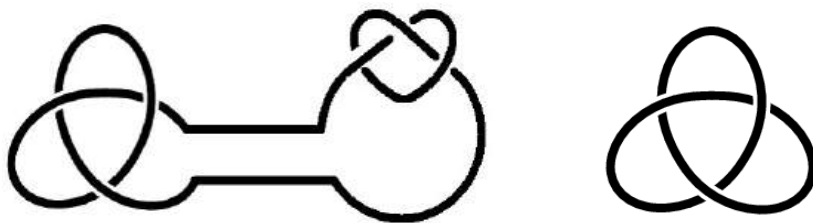


Fig. 3.2: A composite knot
(The images were firstly appeared in [16].)

A prime knot

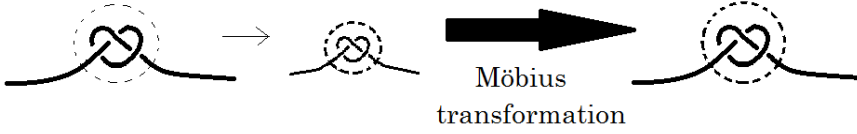


Fig. 3.3: Freedman-He-Wang's procedure (The image was firstly appeared in [16].)

Note that $\tilde{\mathcal{E}}_{(2,1)}([K])$ exists, because the energy density is non-negative. Kusner and Sullivan [14] investigated the energy $\mathcal{E}_{(2,1)}$ for various knots numerically, and proposed the following conjecture.

Conjecture 3.2.1 (The Kusner-Sullivan conjecture [14]).

1. *There do not exist minimizers of composite knot type.*
2. Assume that $\mathbf{f} \in [K]$ is composite, and it is a connected sum $\mathbf{f}_1 \# \mathbf{f}_2$, for $\mathbf{f}_i \in [K_i]$. Then, it holds that

$$\tilde{\mathcal{E}}_{(2,1)}([K]) = \tilde{\mathcal{E}}_{(2,1)}([K_1]) + \tilde{\mathcal{E}}_{(2,1)}([K_2]).$$

As far as the author knows, this remains an open conjecture.

3.3 Basic properties of the Möbius energy

In what follows, we consider the Möbius energy in \mathbb{R}^n , which is denoted simply by \mathcal{E} :

$$\mathcal{E}(\mathbf{f}) = \iint_{(\mathbb{R}/\mathbb{Z})^2} \left(\frac{1}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} - \frac{1}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2} \right) ds_1 ds_2.$$

Möbius invariance still holds for the n -dimensional case. As stated in the previous section, the minimizing problem for the Möbius energy was partially solved and remains partially unsolved. Hence, it seems important to study further properties of \mathcal{E} . To begin with it, we summarize some basic properties that are given without proofs.

Because our energy was introduced to determine the “canonical” configuration of knots, we expect that the finiteness of energy suggests some regularity of curves. Indeed, we have the bi-Lipschitz estimate for a curve with finite energy, which means that the curve cannot bend rapidly.

Because we use the arc-length parameter, the fact that

$$\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n} \leq \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))$$

is trivial. That is, \mathbf{f} is Lipschitz continuous, with the Lipschitz constant 1. The finiteness of \mathcal{E} implies the converse estimate.

Theorem 3.3.1 ([19, 2]). If $\mathcal{E}(\mathbf{f}) < M$, then there exists $\lambda = \lambda(M) > 0$ such that

$$\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n} \geq \lambda \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2)).$$

We call these estimates *bi-Lipschitz continuity*. In the following, we assume that \mathbf{f} is bi-Lipschitz.

The finiteness of energy not only implies bi-Lipschitz continuity but also the integrability of (fractional) derivatives.

Definition 3.3.1 (Sobolev-Slobodeckij space). For a non-negative integer j , and $\alpha \in (0, 1)$, $W^{j+\alpha,p}(\mathbb{R}/\mathcal{L}\mathbb{Z}, \mathbb{R}^n)$ is defined as

$$W^{j+\alpha,p}(\mathbb{R}/\mathcal{L}\mathbb{Z}, \mathbb{R}^n) = \{\mathbf{f} \in W^{j,p}(\mathbb{R}/\mathcal{L}\mathbb{Z}, \mathbb{R}^n) \mid [\mathbf{f}^{(j)}]_{\alpha,p} < \infty\},$$

$$[\mathbf{f}^{(j)}]_{\alpha,p} = \left(\int_{\mathbb{R}/\mathcal{L}\mathbb{Z} - \frac{\mathcal{L}}{2}}^{\frac{\mathcal{L}}{2}} \int_{\mathbb{R}/\mathcal{L}\mathbb{Z} - \frac{\mathcal{L}}{2}}^{\frac{\mathcal{L}}{2}} \frac{\|\mathbf{f}^{(j)}(s_1 + s_2) - \mathbf{f}^{(j)}(s_1)\|_{\mathbb{R}^n}^p}{|s_2|^{ap+1}} ds_2 ds_1 \right)^{\frac{1}{p}}$$

with the norm

$$\|\mathbf{f}\|_{W^{j+\alpha,p}} = \|\mathbf{f}\|_{W^{j,p}} + [\mathbf{f}^{(j)}]_{\alpha,p}.$$

When $p = 2$, we denote $W^{j+\alpha,2}(\mathbb{R}/\mathcal{L}\mathbb{Z})$ by $H^{j+\alpha}(\mathbb{R}/\mathcal{L}\mathbb{Z})$.

The following result was given by Blatt [2].

Theorem 3.3.2 ([2]). The finiteness result $\mathcal{E}(\mathbf{f}) < \infty$ implies the bi-Lipschitz continuity of \mathbf{f} and $\mathbf{f} \in H^{3/2}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathcal{L}\mathbb{Z})$. The converse is also true, i.e., if \mathbf{f} is bi-Lipschitz and belongs to $H^{3/2}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathcal{L}\mathbb{Z})$, then $\mathcal{E}(\mathbf{f})$ is finite.

This indicates the proper domain of \mathcal{E} .

Remark 3.3.1. O'Hara and Blatt demonstrated that the above results not only hold for the Möbius energy, but also for the general O'Hara's energy. For details, we refer the reader to the papers of those authors.

The criticality of energy can provide further information regarding regularity. Several results have been established. Freedman, He, and Wang [5] and He [7] demonstrated the regularity of local minimizers.

Theorem 3.3.3 ([5, 7]). Local minimizers of \mathcal{E} with respect to $L^\infty(\mathbb{R}/\mathcal{L}\mathbb{Z})$ topology are smooth.

Reiter [21] proved the regularity for not only local minimizers, but also critical points.

Theorem 3.3.4 ([21]). Any critical point of \mathcal{E} in $W^{2,2}(\mathbb{R}/\mathcal{L}\mathbb{Z})$ is smooth.

Recently, Blatt, Reiter, and Schikorra [4] improved on the assumptions of the previous result.

Theorem 3.3.5 ([4]). Any critical point of \mathcal{E} with finite energy is smooth.

3.4 The Möbius invariant decomposition

In this section, we present a decomposition of the Möbius energy, where each part retains the Möbius invariance. This is based on a series of papers [10]–[13] written with Ishizeki. See also the dissertation of that author [9]. According to Theorems 3.3.1 and 3.3.2, we may assume the existence of the unit tangent vector $\boldsymbol{\tau}(s) = \boldsymbol{f}'(s)$ almost everywhere, which satisfies some integrability condition. Then, the energy may be decomposed into three parts:

$$\mathcal{E}(\boldsymbol{f}) = \mathcal{E}_1(\boldsymbol{f}) + \mathcal{E}_2(\boldsymbol{f}) + 4.$$

See Theorem 3.4.1 below for details. This was recently demonstrated by our research group [10]. The first decomposed energy \mathcal{E}_1 is an analogue of the Gagliardo semi-norm of $\boldsymbol{\tau}$ in the fractional Sobolev space $H^{1/2}(\mathbb{R}/\mathbb{L}\mathbb{Z})$. Indeed, if we replace the Euclidean distance in the denominator of the energy density of \mathcal{E}_1 (see \mathcal{M}_1 in Theorem 3.4.1) with the intrinsic distance, we find that

$$\mathcal{E}_1(\boldsymbol{f}) \approx \frac{1}{2} \iint \frac{\|\boldsymbol{\tau}(s_1) - \boldsymbol{\tau}(s_2)\|^2}{\mathcal{D}(\boldsymbol{f}(s_1), \boldsymbol{f}(s_2))^2} ds_1 ds_2 = \frac{1}{2} [\boldsymbol{f}']_{H^{1/2}}^2.$$

This implies that the domain of \mathcal{E} is $H^{3/2}(\mathbb{R}/\mathbb{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathbb{L}\mathbb{Z})$, as was shown by Blatt. The energy density of the second energy \mathcal{E}_2 has a determinant structure, which implies a cancellation of the integrand. O’Hara helpfully informed the current author that the second energy is a constant multiple of the O’Hara-Solanes energy ([20]). As a by-product, we can provide an alternative proof of the known fact that the right circle, *i.e.*, the circle with constant curvature, is the only global minimizer of the Möbius energy. In Section 3.4.1, we give the precise statement of the decomposition and a sketch of the proof.

Using this decomposition, we can derive explicit expressions for variational formulae and reasonable estimates in several function spaces, including absolute integrability ([11, 13]). Indeed, because the last part, “4”, is an absolute constant, we can ignore this when considering variational problems. This fact shortens the derivation of variational formulae, and enables us to find their “good” estimates for them in several functional spaces [11]. Furthermore, we can determine the L^2 -gradient of each decomposed energy that contains the fractional Laplacian $(-\Delta_s)^{3/2}$ as the principal term [13]. See Section 3.4.2 for further details.

Surprisingly, each of the components of the decomposition is Möbius invariant. In Section 3.4.3, we focus on this fact, which has been studied in [10, 12]. As a conse-

quence, we can demonstrate that right circles are the only global minimizers of \mathcal{E}_1 with the minimum value $2\pi^2$. This seems to be related to the fact that the first eigenvalue of $(-\Delta_s)^{3/2}$ is $\left(\frac{2\pi}{\mathcal{L}}\right)^3$.

3.4.1 The decomposition

The precise statement of the decomposition is as follows.

Theorem 3.4.1 ([10]). Assume that \mathbf{f} satisfies the bi-Lipshitz estimate, and that it belongs to $H^{3/2}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathcal{L}\mathbb{Z})$. Let $\boldsymbol{\tau}(s) = \mathbf{f}'(s)$ be the unit tangent vector. Then, it holds that

$$\mathcal{E}(\mathbf{f}) = \mathcal{E}_1(\mathbf{f}) + \mathcal{E}_2(\mathbf{f}) + 4,$$

where

$$\begin{aligned} \mathcal{E}_i(\mathbf{f}) &= \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \mathcal{M}_i(\mathbf{f}) \, ds_1 ds_2, \\ \mathcal{M}_1(\mathbf{f}) &= \frac{\|\boldsymbol{\tau}(s_1) - \boldsymbol{\tau}(s_2)\|_{\mathbb{R}^n}^2}{2\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2}, \\ \mathcal{M}_2(\mathbf{f}) &= \frac{2}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^4} \det \begin{pmatrix} \boldsymbol{\tau}(s_1) \cdot \boldsymbol{\tau}(s_2) & (\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_1) \\ (\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_2) & \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 \end{pmatrix}. \end{aligned}$$

Proof. In order to deform the energy density, we first consider following integration in the sense of Cauchy's principal value,

$$\lim_{\varepsilon \rightarrow +0} \iint_{|s_1 - s_2| \geq \varepsilon} \cdots \, ds_1 ds_2,$$

and derive absolute integrability later. We differentiate $\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2$ with respect to s_2 . In the sense of distributions, we have that

$$\frac{d}{dx} \log |x| = \text{p.v.} \frac{1}{x},$$

where

$$\langle \text{p.v.} \frac{1}{x}, \varphi \rangle = \lim_{\varepsilon \rightarrow +0} \int_{|x| \geq \varepsilon} \frac{\varphi(x)}{x} \, dx$$

for $\varphi \in C_0^\infty(\mathbb{R})$ (see [15]). Using the periodicity of $\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))$, the distributional derivative is given by

$$\frac{\partial}{\partial s_2} \log \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2)) = \begin{cases} -\text{p.v.} \frac{1}{s_1 - s_2 - \mathcal{L}} & (s_2 + \frac{\mathcal{L}}{2} < s_1 \leq s_2 + \mathcal{L} \pmod{\mathcal{L}}), \\ -\text{p.v.} \frac{1}{s_1 - s_2} & (s_2 \leq s_1 < s_2 + \frac{\mathcal{L}}{2} \pmod{\mathcal{L}}). \end{cases}$$

Regarding this as a distribution of s_1 , we have that it is differentiable for $s_1 \neq s_2 + \mathcal{L}/2$ in the weak sense, and we obtain that

$$\frac{\partial^2}{\partial s_1 \partial s_2} \log \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2)) = \text{p.v.} \frac{1}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2}.$$

As a function of s_2 , $\frac{\partial}{\partial s_2} \log \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))$ has a jump discontinuity at $s_2 = s_1 + \mathcal{L}/2$ with gap $-4/\mathcal{L}$. As functions of s_2 , $\frac{\partial^2}{\partial s_1 \partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}$ is bounded at $s_2 = s_1 \pm \frac{\mathcal{L}}{2}$, and $\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}$ is continuous at the same points. Therefore, we obtain that

$$\begin{aligned} & \int_{\varepsilon \leq |s_1 - s_2| \leq \frac{\mathcal{L}}{2}} \frac{1}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2} ds_2 \\ &= \lim_{\delta \rightarrow +0} \int_{\varepsilon \leq |s_1 - s_2| \leq \frac{\mathcal{L}}{2} - \delta} \frac{\partial^2}{\partial s_1 \partial s_2} \left(\log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n} - \log \frac{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right) ds_2 \\ &= \int_{\varepsilon \leq |s_1 - s_2| \leq \frac{\mathcal{L}}{2}} \frac{\partial^2}{\partial s_1 \partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n} ds_2 + \left[\frac{\partial}{\partial s_1} \log \frac{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right]_{s_2=s_1-\varepsilon}^{s_2=s_1+\varepsilon} - \frac{4}{\mathcal{L}}. \end{aligned}$$

We integrate this with respect to s_1 . Using periodicity, we note that

$$\begin{aligned} & \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left[\frac{\partial}{\partial s_1} \log \frac{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right]_{s_2=s_1-\varepsilon}^{s_2=s_1+\varepsilon} ds_1 \\ &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left[\frac{(\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_1)}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} - \frac{1}{s_1 - s_2} \right]_{s_2=s_1-\varepsilon}^{s_2=s_1+\varepsilon} ds_1 \\ &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left\{ \frac{(\mathbf{f}(s_1) - \mathbf{f}(s_1 + \varepsilon)) \cdot \boldsymbol{\tau}(s_1)}{\|\mathbf{f}(s_1) - \mathbf{f}(s_1 + \varepsilon)\|_{\mathbb{R}^n}^2} - \frac{(\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)) \cdot \boldsymbol{\tau}(s_1 + \varepsilon)}{\|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} + \frac{2}{\varepsilon} \right\} ds_1 \\ &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left\{ -\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_1 + \varepsilon)\|_{\mathbb{R}^n} + \frac{2(\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)) \cdot \boldsymbol{\tau}(s_1)}{\|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} + \frac{2}{\varepsilon} \right\} ds_1 \\ &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left[-2\varepsilon \left(\frac{1}{\|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} - \frac{1}{\varepsilon^2} \right) \right. \\ & \quad \left. + \frac{2}{\|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} \int_{s_1}^{s_1+\varepsilon} (1 - \boldsymbol{\tau}(s_3) \cdot \boldsymbol{\tau}(s_1)) ds_3 \right] ds_1. \end{aligned}$$

We have that

$$\begin{aligned}
& \left| 2\varepsilon \left(\frac{1}{\|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} - \frac{1}{\varepsilon^2} \right) \right| \\
&= \frac{2\varepsilon}{\varepsilon^2 \|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} \int_{s_1}^{s_1+\varepsilon} \int_{s_1}^{s_1+\varepsilon} (1 - \boldsymbol{\tau}(s_3) \cdot \boldsymbol{\tau}(s_4)) ds_3 ds_4 \\
&= \frac{1}{\varepsilon \|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} \int_{s_1}^{s_1+\varepsilon} \int_{s_1}^{s_1+\varepsilon} \|\boldsymbol{\tau}(s_3) - \boldsymbol{\tau}(s_4)\|_{\mathbb{R}^n}^2 ds_3 ds_4
\end{aligned}$$

and by applying the estimate $\|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n} \geq \lambda^{-1}\varepsilon$, and changing the order of integration, we obtain that

$$\begin{aligned}
& \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \frac{1}{\varepsilon \|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} \int_{s_1}^{s_1+\varepsilon} \int_{s_1}^{s_1+\varepsilon} \|\boldsymbol{\tau}(s_3) - \boldsymbol{\tau}(s_4)\|_{\mathbb{R}^n}^2 ds_3 ds_4 ds_1 \\
&\leq \frac{\lambda^2}{\varepsilon^3} \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \int_{s_4-\varepsilon}^{s_4+\varepsilon} \int_{s_3-\varepsilon}^{s_3} \|\boldsymbol{\tau}(s_3) - \boldsymbol{\tau}(s_4)\|_{\mathbb{R}^n}^2 ds_1 ds_3 ds_4 \\
&\leq \lambda^2 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \int_{s_4-\varepsilon}^{s_4+\varepsilon} \frac{\|\boldsymbol{\tau}(s_3) - \boldsymbol{\tau}(s_4)\|_{\mathbb{R}^n}^2}{\mathcal{D}(\mathbf{f}(s_3), \mathbf{f}(s_4))^2} ds_3 ds_4.
\end{aligned}$$

Because

$$[\mathbf{f}']_{H^{1/2}}^2 = \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \frac{\|\boldsymbol{\tau}(s_3) - \boldsymbol{\tau}(s_4)\|_{\mathbb{R}^n}^2}{\mathcal{D}(\mathbf{f}(s_3), \mathbf{f}(s_4))^2} ds_3 ds_4$$

is finite, the absolute continuity of the integration yields that

$$\lim_{\varepsilon \rightarrow +0} \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \frac{1}{\varepsilon \|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} \int_{s_1}^{s_1+\varepsilon} \int_{s_1}^{s_1+\varepsilon} \|\boldsymbol{\tau}(s_3) - \boldsymbol{\tau}(s_4)\|_{\mathbb{R}^n}^2 ds_3 ds_4 ds_1 = 0.$$

Similarly, we have that

$$\begin{aligned}
& \left| \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \frac{2}{\|\mathbf{f}(s_1 + \varepsilon) - \mathbf{f}(s_1)\|_{\mathbb{R}^n}^2} \int_{s_1}^{s_1+\varepsilon} (1 - \boldsymbol{\tau}(s_3) \cdot \boldsymbol{\tau}(s_1)) ds_3 ds_1 \right| \\
&\leq \lambda^2 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \int_{s_1}^{s_1+\varepsilon} \frac{\|\boldsymbol{\tau}(s_3) - \boldsymbol{\tau}(s_1)\|_{\mathbb{R}^n}^2}{\mathcal{D}(\mathbf{f}(s_3), \mathbf{f}(s_1))^2} ds_3 ds_1 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow +0.
\end{aligned}$$

Hence, we obtain that

$$\int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left[\frac{\partial}{\partial s_1} \log \frac{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right]_{s_2=s_1-\varepsilon}^{s_2=s_1+\varepsilon} ds_1 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow +0,$$

which leads to the expression

$$\mathcal{E}(\mathbf{f}) = \text{p.v.} \iint_{(\mathbb{R}/\mathbb{Z})^2} \left(\frac{1}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} - \frac{\partial^2}{\partial s_1 \partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n} \right) ds_1 ds_2 + 4.$$

By manipulating the log term in the above, we obtain that

$$\begin{aligned} \mathcal{E}(\mathbf{f}) - 4 = \text{p.v.} \iint_{(\mathbb{R}/\mathbb{Z})^2} & \left\{ \frac{1 + \boldsymbol{\tau}(s_1) \cdot \boldsymbol{\tau}(s_2)}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} \right. \\ & \left. - \frac{2\{(\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_1)\} \{(\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_2)\}}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^4} \right\} ds_1 ds_2. \end{aligned} \quad (3.4.1)$$

The density of this integral can be expressed as

$$\begin{aligned} & \frac{1 + \boldsymbol{\tau}(s_1) \cdot \boldsymbol{\tau}(s_2)}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} - \frac{2\{(\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_1)\} \{(\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_2)\}}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^4} \\ &= \frac{1 - \boldsymbol{\tau}(s_1) \cdot \boldsymbol{\tau}(s_2)}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} + \frac{2\boldsymbol{\tau}(s_1) \cdot \boldsymbol{\tau}(s_2) \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^4} \\ & \quad - \frac{\{(\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_1)\} \{(\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_2)\}}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^4} \\ &= \frac{\|\boldsymbol{\tau}(s_1) - \boldsymbol{\tau}(s_2)\|_{\mathbb{R}^n}^2}{2\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} + \frac{2}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^4} \\ & \quad \times \det \begin{pmatrix} \boldsymbol{\tau}(s_1) \cdot \boldsymbol{\tau}(s_2) & (\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_1) \\ (\mathbf{f}(s_1) - \mathbf{f}(s_2)) \cdot \boldsymbol{\tau}(s_2) & \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 \end{pmatrix} \\ &= \mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f}). \end{aligned}$$

It remains to remove p.v. from in front of the double integral in (3.4.1). Because $\mathcal{M}_1(\mathbf{f})$ is non-negative, the energy \mathcal{E}_1 is defined as the absolute integral for curves in $H^{\frac{3}{2}}(\mathbb{R}/\mathbb{Z})$ with bi-Lipschitz continuity. Therefore, the absolute integrability of $\mathcal{M}_2(\mathbf{f})$ follows from the fact that

$$\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f}) \geq 0,$$

which we will show below.

We use the notation Δ to denote the difference between values at $s = s_1$ and s_2 :

$$\Delta s = s_1 - s_2, \quad \Delta \mathbf{f} = \mathbf{f}(s_1) - \mathbf{f}(s_2), \quad \text{etc.}$$

For a function \mathbf{v} on \mathbb{R}/\mathbb{Z} , we define $R\mathbf{v} = R\mathbf{v}(s_1, s_2)$ by

$$R\mathbf{v} = \frac{|\Delta s|(\Delta \mathbf{v})}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n} \Delta s}$$

and observe that $R\mathbf{f}$ is clearly a unit vector field. By applying the Lagrange formula, we have that

$$\mathcal{M}_2(\mathbf{f}) = \frac{2}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} \langle (\boldsymbol{\tau}(s_1) \wedge R\mathbf{f}), (\boldsymbol{\tau}(s_2) \wedge R\mathbf{f}) \rangle,$$

where \wedge is the wedge product of vectors, and $\langle \cdot, \cdot \rangle$ is the inner product on $\wedge^2 \mathbb{R}^n$. It is easy to see that for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and any unit vector $\mathbf{e} \in \mathbb{R}^n$,

$$\langle (\mathbf{x} \wedge \mathbf{e}), (\mathbf{y} \wedge \mathbf{e}) \rangle = (P_{\mathbf{e}}^\perp \mathbf{x}) \cdot (P_{\mathbf{e}}^\perp \mathbf{y}),$$

where

$$P_{\mathbf{e}} \mathbf{x} = (\mathbf{x} \cdot \mathbf{e}) \mathbf{e}, \quad P_{\mathbf{e}}^\perp \mathbf{x} = \mathbf{x} - P_{\mathbf{e}} \mathbf{x}.$$

Therefore,

$$\mathcal{M}_2(\mathbf{f}) = \frac{2}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} (P_{Rf}^\perp \boldsymbol{\tau}(s_1)) \cdot (P_{Rf}^\perp \boldsymbol{\tau}(s_2)).$$

On the other hand, the first density is given by

$$\mathcal{M}_1(\mathbf{f}) = \frac{1}{2} \frac{\|\Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2}.$$

Consequently, we obtain that

$$\begin{aligned} & 2\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2 (\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f})) \\ &= \|\Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2 + 4(P_{Rf}^\perp \boldsymbol{\tau}(s_1)) \cdot (P_{Rf}^\perp \boldsymbol{\tau}(s_2)) \\ &= \|P_{Rf} \Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2 + \|P_{Rf}^\perp \Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2 + 4(P_{Rf}^\perp \boldsymbol{\tau}(s_1)) \cdot (P_{Rf}^\perp \boldsymbol{\tau}(s_2)) \\ &= \|P_{Rf} \Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2 + \|P_{Rf}^\perp (\boldsymbol{\tau}(s_1) + \boldsymbol{\tau}(s_2))\|_{\mathbb{R}^n}^2, \end{aligned}$$

which is non-negative. □

As a consequence, we can provide an alternative proof for the following fact, which has already been established. See, for example, [5].

Corollary 3.4.1. It holds that $\mathcal{E}(\mathbf{f}) \geq 4$. Equality holds if and only if \mathbf{f} is a right circle.

Proof. Theorem 3.4.1 and the non-negativity of $\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f})$ implies $\mathcal{E}(\mathbf{f}) \geq 4$, and equality holds if and only if $\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f}) \equiv 0$. This is equivalent to

$$P_{Rf} \Delta \boldsymbol{\tau} \equiv \mathbf{0}, \quad P_{Rf}^\perp (\boldsymbol{\tau}(s_1) + \boldsymbol{\tau}(s_2)) \equiv \mathbf{0}.$$

In particular, from the second relation we find a function μ such that

$$\mathbf{f}(s_1) - \mathbf{f}(s_2) = \mu(s_1, s_2)(\boldsymbol{\tau}(s_1) + \boldsymbol{\tau}(s_2)). \quad (3.4.2)$$

Because a minimizer is smooth, so is μ . We differentiate the above relation three times with respect to s_1 , to obtain that

$$\boldsymbol{\tau}(s_1) = \frac{\partial \mu}{\partial s_1}(s_1, s_2)(\boldsymbol{\tau}(s_1) + \boldsymbol{\tau}(s_2)) + \mu(s_1, s_2)\boldsymbol{\kappa}(s_1), \quad (3.4.3)$$

$$\boldsymbol{\kappa}(s_1) = \frac{\partial^2 \mu}{\partial s_1^2}(s_1, s_2)(\boldsymbol{\tau}(s_1) + \boldsymbol{\tau}(s_2)) + 2 \frac{\partial \mu}{\partial s_1}(s_1, s_2)\boldsymbol{\kappa}(s_1) + \mu(s_1, s_2)\boldsymbol{\kappa}'(s_1), \quad (3.4.4)$$

$$\begin{aligned} \boldsymbol{\kappa}'(s_1) &= \frac{\partial^3 \mu}{\partial s_1^3}(s_1, s_2)(\boldsymbol{\tau}(s_1) + \boldsymbol{\tau}(s_2)) + 3 \frac{\partial^2 \mu}{\partial s_1^2}(s_1, s_2)\boldsymbol{\kappa}(s_1) \\ &\quad + 3 \frac{\partial \mu}{\partial s_1}(s_1, s_2)\boldsymbol{\kappa}'(s_1) + \mu(s_1, s_2)\boldsymbol{\kappa}''(s_1). \end{aligned} \quad (3.4.5)$$

Setting $s_1 = s_2 = s$ in (3.4.2)–(3.4.5), we have that

$$\mu(s, s) = 0, \quad \frac{\partial \mu}{\partial s_1}(s, s) = \frac{1}{2}, \quad \frac{\partial^2 \mu}{\partial s_1^2}(s, s) = 0,$$

and

$$\boldsymbol{\kappa}'(s) = -4 \frac{\partial^3 \mu}{\partial s_1^3}(s, s)\boldsymbol{\tau}(s). \quad (3.4.6)$$

Taking the inner product between each side of (3.4.6) and $\boldsymbol{\kappa}(s)$, we know that $\|\boldsymbol{\kappa}(s)\|_{\mathbb{R}^n}$ is independent of s , and we write this as κ . If $\kappa = 0$, then $\boldsymbol{\tau}'(s) = \boldsymbol{\kappa}(s) = \mathbf{0}$, and therefore $\boldsymbol{\tau}(s)$ is a constant vector. This is impossible, because \mathbf{f} is a closed curve. Consequently, $\kappa > 0$. By taking the inner product between each side of (3.4.6) and $\boldsymbol{\tau}(s)$, we know that

$$-4 \frac{\partial^3 \mu}{\partial s_1^3}(s, s) = \boldsymbol{\kappa}'(s) \cdot \boldsymbol{\tau}(s) = -\kappa(s) \cdot \kappa(s) = -\kappa^2.$$

By inserting this into (3.4.6), we obtain that

$$\boldsymbol{\kappa}'(s) + \kappa^2 \boldsymbol{\tau}(s) = \mathbf{0}$$

for every $s \in \mathbb{R}/\mathcal{L}\mathbb{Z}$. Because $\boldsymbol{\tau}(s) = \mathbf{f}'(s)$, there exists a constant vector \mathbf{c} such that

$$\boldsymbol{\kappa}(s) + \kappa^2 (\mathbf{f}(s) - \mathbf{c}) = \mathbf{0}. \quad (3.4.7)$$

By integrating with respect to s on $\mathbb{R}/\mathcal{L}\mathbb{Z}$, and dividing by \mathcal{L} , we find that

$$\mathbf{c} = \frac{1}{\mathcal{L}} \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \mathbf{f}(s) ds.$$

We can rewrite (3.4.7) as the following second order differential equation:

$$(\mathbf{f}(s) - \mathbf{c})'' + \kappa^2 (\mathbf{f}(s) - \mathbf{c}) = \mathbf{0}.$$

The solution is given by

$$\mathbf{f}(s) - \mathbf{c} = (\mathbf{f}(0) - \mathbf{c}) \cos \kappa s + \frac{\sin \kappa s}{\kappa} \boldsymbol{\tau}(0).$$

That is, \mathbf{f} is a right circle with center \mathbf{c} and radius κ^{-1} . Because the total length is \mathcal{L} , the radius is $\kappa^{-1} = \frac{\mathcal{L}}{2\pi}$. \square

3.4.2 Variational formulae

Let $\mathcal{F}(\mathbf{f})$ be a geometric quantity determined by the closed curve \mathbf{f} , and let $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$ be functions from $\mathbb{R}/\mathcal{L}\mathbb{Z}$ to \mathbb{R}^n . We use δ and δ^2 to denote

$$\delta \mathcal{F}(\mathbf{f})[\boldsymbol{\varphi}] = \left. \frac{d}{d\varepsilon} \mathcal{F}(\mathbf{f} + \varepsilon \boldsymbol{\varphi}) \right|_{\varepsilon=0},$$

$$\delta^2 \mathcal{F}(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}] = \left. \frac{\partial^2}{\partial \varepsilon_1 \partial \varepsilon_2} \mathcal{F}(\mathbf{f} + \varepsilon_1 \boldsymbol{\varphi} + \varepsilon_2 \boldsymbol{\psi}) \right|_{\varepsilon_1=\varepsilon_2=0}.$$

Calculating the first and second variational formulae of the Möbius energy formally, we have that

$$\delta \mathcal{E}(\mathbf{f})[\boldsymbol{\varphi}] = \iint \left(\frac{***}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^4} + \cdots \right) ds_1 ds_2,$$

$$\delta^2 \mathcal{E}(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}] = \iint \left(\frac{***}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^8} + \cdots \right) ds_1 ds_2.$$

These expressions contain many terms that are not integrable. He [7] obtained these formulae as the integral of Cauchy's principal value. Ishizeki [8] investigates the absolute integrability in the regularity class $C^{3+\alpha}$.

Theorem 3.4.2 ([8]). If \mathbf{f} , $\boldsymbol{\varphi}$ and $\boldsymbol{\psi} \in C^{3+\alpha}(\mathbb{R}/\mathcal{L}\mathbb{Z})$, then the integrands of $\delta \mathcal{E}$ and $\delta^2 \mathcal{E}$ are absolutely integrable.

This result can be improved for the proper domain by using the decomposition. Indeed, using this, we can derive the explicit expressions of variational formulae and reasonable estimates in several function spaces, including absolute integrability ([11, 13]).

The first variation \mathcal{G}_i and the second variation \mathcal{H}_i of the decomposed energies \mathcal{E}_i are given by

$$\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}] ds_1 ds_2 = \delta(\mathcal{M}_i(\mathbf{f}) ds_1 ds_2)[\boldsymbol{\varphi}],$$

$$\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}] ds_1 ds_2 = \delta^2(\mathcal{M}_i(\mathbf{f}) ds_1 ds_2)[\boldsymbol{\varphi}, \boldsymbol{\psi}].$$

Next, we obtain the explicit expressions of variations. To present the corresponding statement, we first introduce several notations. For a function \mathbf{v} on $\mathbb{R}/\mathcal{L}\mathbb{Z}$, the operations Q , \tilde{Q}_i , S , and S_i are defined as

$$Q\mathbf{v} = \Delta \mathbf{v}', \quad \tilde{Q}_i \mathbf{v} = (-1)^{i-1} 2 \{ \mathbf{v}'_i - R\mathbf{f} \cdot \boldsymbol{\tau}_i \} R\mathbf{v},$$

$$R\mathbf{v} = \frac{|\Delta \mathbf{f}| \Delta \mathbf{v}}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n} \Delta s}, \quad \hat{R}\mathbf{v} = \frac{1}{2}(\mathbf{v}'_1 + \mathbf{v}'_2),$$

$$S(\mathbf{v}, \mathbf{w}) = \hat{R}\mathbf{v} \cdot Q\mathbf{w} + Q\mathbf{v} \cdot \hat{R}\mathbf{w}, \quad \tilde{S}_i(\mathbf{v}, \mathbf{w}) = R\mathbf{v} \cdot \tilde{Q}_i \mathbf{w} + \tilde{Q}_i \mathbf{v} \cdot R\mathbf{w},$$

where the operation R is the same one as in the proof of Theorem 3.4.1.

Proposition 3.4.1. We have that

$$\begin{aligned}\mathcal{G}_1(f)[\varphi] &= \frac{Qf \cdot Q\varphi}{\|\Delta f\|_{\mathbb{R}^n}^2} - \frac{2\mathcal{M}_1(f)\Delta f \cdot \Delta\varphi}{\|\Delta f\|_{\mathbb{R}^n}^2}, \\ \mathcal{G}_2(f)[\varphi] &= -\frac{\tilde{Q}_1 f \cdot \tilde{Q}_2 \varphi + \tilde{Q}_2 f \cdot \tilde{Q}_1 \varphi}{2\|\Delta f\|_{\mathbb{R}^n}^2} - \frac{2\mathcal{M}_2(f)\Delta f \cdot \Delta\varphi}{\|\Delta f\|_{\mathbb{R}^n}^2}\end{aligned}$$

and

$$\begin{aligned}\mathcal{H}_1(f)[\varphi, \psi] &= \frac{Q\varphi \cdot Q\psi}{\|\Delta f\|_{\mathbb{R}^n}^2} - \frac{S(f, \varphi)S(f, \psi)}{\|\Delta f\|_{\mathbb{R}^n}^2} \\ &\quad - \frac{2\mathcal{G}_1(f)[\varphi]\Delta f \cdot \Delta\psi}{\|\Delta f\|_{\mathbb{R}^n}^2} - \frac{2\mathcal{G}_1(f)[\psi]\Delta f \cdot \Delta\varphi}{\|\Delta f\|_{\mathbb{R}^n}^2} - \frac{2\mathcal{M}_1(f)\Delta\varphi \cdot \Delta\psi}{\|\Delta f\|_{\mathbb{R}^n}^2}, \\ \mathcal{H}_2(f)[\varphi, \psi] &= -\frac{\tilde{Q}_1 \varphi \cdot \tilde{Q}_2 \psi + \tilde{Q}_2 \varphi \cdot \tilde{Q}_1 \psi}{2\|\Delta f\|_{\mathbb{R}^n}^2} + \frac{\tilde{S}_1(f, \varphi)\tilde{S}_2(f, \psi) + \tilde{S}_2(f, \varphi)\tilde{S}_1(f, \psi)}{2\|\Delta f\|_{\mathbb{R}^n}^2} \\ &\quad - \frac{2\mathcal{G}_2(f)[\varphi]\Delta f \cdot \Delta\psi}{\|\Delta f\|_{\mathbb{R}^n}^2} - \frac{2\mathcal{G}_2(f)[\psi]\Delta f \cdot \Delta\varphi}{\|\Delta f\|_{\mathbb{R}^n}^2} - \frac{2\mathcal{M}_2(f)\Delta\varphi \cdot \Delta\psi}{\|\Delta f\|_{\mathbb{R}^n}^2}.\end{aligned}$$

This is proven as follows. Set

$$\mathcal{M}_i(f) = \frac{\mathcal{N}_i(f)}{\|\Delta f\|_{\mathbb{R}^n}^2}.$$

Then, it holds that

$$\mathcal{N}_1(f) = \frac{1}{2}Qf \cdot Qf, \quad \mathcal{N}_2(f) = -\frac{1}{2}\tilde{Q}_1 f \cdot \tilde{Q}_2 f.$$

The above relation for \mathcal{N}_1 is easy to prove. We have already demonstrated the one for \mathcal{N}_2 , in the proof of Corollary 3.4.1. Therefore, to obtain the expressions of \mathcal{G}_i and \mathcal{H}_i , we require those of variations of \mathcal{N}_i , $\|\Delta f\|_{\mathbb{R}^n}^2$, and $ds_1 ds_2$. We begin with the variations of basic quantities.

Lemma 3.4.1. The following first variational formulae hold:

1. $\delta\tau[\varphi] = \varphi' - (\tau \cdot \varphi')\tau.$
2. $\delta\|\Delta\tau\|_{\mathbb{R}^n}^2[\varphi] = 2\Delta\tau \cdot \Delta\varphi' - \|\Delta\tau\|_{\mathbb{R}^n}^2(\tau_1 \cdot \varphi'_1 + \tau_2 \cdot \varphi'_2).$
3. $\delta\left(\frac{1}{\|\Delta f\|_{\mathbb{R}^n}^2}\right)[\varphi] = -\frac{2\Delta f \cdot \Delta\varphi}{\|\Delta f\|_{\mathbb{R}^n}^4}.$
4. $\delta(ds_j)[\varphi] = \tau_j \cdot \varphi'_j ds_j.$

Because the proof is not difficult, we omit it. As a consequence of this lemma, we obtain that

$$\begin{aligned}\delta\mathcal{N}_1(f)[\varphi] &= Qf \cdot Q\varphi - (\tau_1 \cdot \varphi'_1 + \tau_2 \cdot \varphi'_2)\mathcal{N}_1(f), \\ \delta\mathcal{N}_2(f)[\varphi] &= -\frac{1}{2}(\tilde{Q}_1 f \cdot \tilde{Q}_2 \varphi + \tilde{Q}_2 f \cdot \tilde{Q}_1 \varphi) - (\tau_1 \cdot \varphi'_1 + \tau_2 \cdot \varphi'_2)\mathcal{N}_2(f).\end{aligned}$$

From Lemma 3.4.1, we have that

$$\begin{aligned}\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}] ds_1 ds_2 &= \delta \mathcal{M}_i(\mathbf{f})[\boldsymbol{\varphi}] ds_1 ds_2 + \mathcal{M}_i(\mathbf{f}) \delta(ds_1 ds_2)[\boldsymbol{\varphi}] \\ &= \left\{ \delta \mathcal{M}_i(\mathbf{f})[\boldsymbol{\varphi}] + \mathcal{M}_i(\mathbf{f})(\boldsymbol{\tau}_1 \cdot \boldsymbol{\varphi}'_1 + \boldsymbol{\tau}_2 \cdot \boldsymbol{\varphi}'_2) \right\} ds_1 ds_2,\end{aligned}$$

i.e.,

$$\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}] = \delta \mathcal{M}_i(\mathbf{f})[\boldsymbol{\varphi}] + \mathcal{M}_i(\mathbf{f})(\boldsymbol{\tau}_1 \cdot \boldsymbol{\varphi}'_1 + \boldsymbol{\tau}_2 \cdot \boldsymbol{\varphi}'_2).$$

Now, we are in the position to calculate \mathcal{G}_i . Using Lemma 3.4.1 again, we have that

$$\begin{aligned}\delta \mathcal{M}_i(\mathbf{f})[\boldsymbol{\varphi}] &= \frac{\delta \mathcal{N}_i(\mathbf{f})[\boldsymbol{\varphi}]}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} + \mathcal{N}_i(\mathbf{f}) \delta \left(\frac{1}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} \right) [\boldsymbol{\varphi}] \\ &= \frac{\delta \mathcal{N}_i(\mathbf{f})[\boldsymbol{\varphi}]}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} - \frac{2 \mathcal{M}_i(\mathbf{f}) \Delta \mathbf{f} \cdot \Delta \boldsymbol{\varphi}}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2}.\end{aligned}$$

By inserting the expressions for $\delta \mathcal{N}_i$, we obtain Proposition 3.4.1.

Similarly, we can obtain the expressions for \mathcal{H}_i . See [11] for details.

We can derive estimates of variational formulae in several function spaces.

Theorem 3.4.3. We set $X = H^{\frac{3}{2}}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathcal{L}\mathbb{Z})$, and $Y = H^{\frac{1}{2}}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap L^\infty(\mathbb{R}/\mathcal{L}\mathbb{Z})$. Assume that there exists a positive constant λ such that $\|\Delta \mathbf{f}\|_{\mathbb{R}^n} \geq \lambda^{-1} |\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))|$.

1. If \mathbf{f} , $\boldsymbol{\varphi}$, and $\boldsymbol{\psi} \in X$, then $\mathcal{M}_i(\mathbf{f})$, $\mathcal{G}_i(\mathbf{f})[\boldsymbol{\psi}]$, and $\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]$ belong to $L^1((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)$. Furthermore, there exists a positive constant C , depending on $\|\mathbf{f}'\|_Y$ and λ , such that

$$\|\mathcal{M}_i(\mathbf{f})\|_{L^1((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C,$$

$$\|\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}]\|_{L^1((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C \|\boldsymbol{\varphi}'\|_Y,$$

$$\|\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]\|_{L^1((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C \|\boldsymbol{\varphi}'\|_Y \|\boldsymbol{\psi}'\|_Y.$$

2. If \mathbf{f} , $\boldsymbol{\varphi}$, and $\boldsymbol{\psi} \in C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$, then $\mathcal{M}_i(\mathbf{f})$, $\mathcal{G}_i(\mathbf{f})[\boldsymbol{\psi}]$, and $\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]$ belong to $L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)$. Furthermore, there exists a positive constant C , depending on $\|\mathbf{f}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{R})}$, λ and \mathcal{L} , such that

$$\|\mathcal{M}_i(\mathbf{f})\|_{L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C,$$

$$\|\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}]\|_{L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C \|\boldsymbol{\varphi}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{R})},$$

$$\|\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]\|_{L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C \|\boldsymbol{\varphi}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{R})} \|\boldsymbol{\psi}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{R})}.$$

3. If \mathbf{f} , $\boldsymbol{\varphi}$, and $\boldsymbol{\psi} \in C^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$, then $\mathcal{M}_i(\mathbf{f})$, $\mathcal{G}_i(\mathbf{f})[\boldsymbol{\psi}]$, $\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]$ can be extended on the diagonal set $\{(s_1, s_2) \mid s_1 \equiv s_2 \pmod{\mathcal{L}\mathbb{Z}}\}$ such that these functions are continuous everywhere. The limits of sum vanish on the diagonal set:

$$\lim_{(s_1, s_2) \rightarrow (s, s)} (\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f})) = 0,$$

$$\begin{aligned} \lim_{(s_1, s_2) \rightarrow (s, s)} (\mathcal{G}_1(\mathbf{f})[\boldsymbol{\varphi}] + \mathcal{G}_2(\mathbf{f})[\boldsymbol{\varphi}]) &= 0, \\ \lim_{(s_1, s_2) \rightarrow (s, s)} (\mathcal{H}_1(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}] + \mathcal{H}_2(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]) &= 0. \end{aligned}$$

Furthermore, there exists a positive constant C , depending on $\|\mathbf{f}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}$, λ and \mathcal{L} , such that

$$\begin{aligned} \|\mathcal{M}_i(\mathbf{f})\|_{C^0((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} &\leq C, \\ \|\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}]\|_{C^0((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} &\leq C\|\boldsymbol{\varphi}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}, \\ \|\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]\|_{C^0((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} &\leq C\|\boldsymbol{\varphi}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}\|\boldsymbol{\psi}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}. \end{aligned}$$

The estimates in Theorem 3.4.3 follow from next lemma.

Lemma 3.4.2. 1. For $\mathbf{v} \in X$, the following estimate holds:

$$\left\| \frac{Q\mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{L^2((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq \|\mathbf{v}'\|_Y.$$

2. For $\mathbf{v} \in C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$, the following estimate holds:

$$\left\| \frac{Q\mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq \|\mathbf{v}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})}.$$

3. Assume that $\mathbf{v} \in C^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$. If we set $Q\mathbf{v}|_{s=s_1=s_2} = \mathbf{v}''$, then $Q\mathbf{v}$ is continuous everywhere, and

$$\left\| \frac{Q\mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{C^0((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq \|\mathbf{v}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}.$$

4. Assume that $\mathbf{f} \in X$ and that $\|\Delta\mathbf{f}\|_{\mathbb{R}^n} \geq \lambda^{-1}|\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))|$. Then, there exists a positive constant C , depending on $\|\mathbf{f}'\|_Y$ and λ , such that

$$\left\| \frac{\tilde{Q}_i\mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{L^2((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C\|\mathbf{v}'\|_Y$$

holds for all $\mathbf{v} \in X$.

5. Assume that $\mathbf{f} \in C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$ and that $\|\Delta\mathbf{f}\|_{\mathbb{R}^n} \geq \lambda^{-1}|\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))|$. Then, there exists a positive constant C , depending on $\|\mathbf{f}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})}$, λ , and \mathcal{L} , such that

$$\left\| \frac{\tilde{Q}_i\mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C\|\mathbf{v}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})}$$

holds for all $\mathbf{v} \in C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$.

6. Assume that $\mathbf{f} \in C^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$ and that \mathbf{f} has no self-intersections. For $\mathbf{v} \in C^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$, $\tilde{Q}_i \mathbf{v}$ is continuous everywhere by setting $\tilde{Q}_i \mathbf{v}|_{s=s_1=s_2} = \mathbf{v}''$. If we further assume that $\|\Delta \mathbf{f}\|_{\mathbb{R}^n} \geq \lambda^{-1} |\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))|$, then, there exists a positive constant C , depending on $\|\mathbf{f}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}$, λ , and \mathcal{L} , such that

$$\left\| \frac{\tilde{Q}_i \mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{C^0((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} \leq C \|\mathbf{v}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}$$

holds for all $\mathbf{v} \in C^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$.

Proof. Without loss of generality, we may assume that $|s_1 - s_2| \leq \frac{\mathcal{L}}{2}$, and for simplicity we use $|\Delta s|$ instead of $\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))$. The assertion for $Q\mathbf{v}$ almost matches the definition of norms. Indeed, we immediately have that

$$\left\| \frac{Q\mathbf{v}}{\Delta s} \right\|_{L^2((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} = [\mathbf{v}']_{H^{\frac{1}{2}}(\mathbb{R}/\mathcal{L}\mathbb{Z})} \leq \|\mathbf{v}'\|_Y$$

and

$$\left\| \frac{Q\mathbf{v}}{\Delta s} \right\|_{L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} = \|\mathbf{v}'\|_{\text{Lip}} \leq \|\mathbf{v}'\|_{C^{0,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})}.$$

If $\mathbf{v} \in C^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$, then it is easy to see that

$$\lim_{(s_1, s_2) \rightarrow (s, s)} \frac{Q\mathbf{v}}{\Delta s} = \mathbf{v}''(s),$$

and

$$\left\| \frac{Q\mathbf{v}}{\Delta s} \right\|_{C^0((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)} = \max_{|\Delta s| \leq \frac{\mathcal{L}}{2}} \left\| \frac{1}{s_1 - s_2} \int_{s_2}^{s_1} \mathbf{v}''(s) ds \right\|_{\mathbb{R}^n} \leq \|\mathbf{v}'\|_{C^1(\mathbb{R}/\mathcal{L}\mathbb{Z})}.$$

To show that the assertion for $\tilde{Q}_i \mathbf{v}$ holds, we decompose $\frac{(-1)^{i-1}}{2} \tilde{Q}_i \mathbf{v} = \mathbf{v}'_i - (R\mathbf{f} \cdot \boldsymbol{\tau}_i)R\mathbf{v}$ into

$$\mathbf{v}'_i - (R\mathbf{f} \cdot \boldsymbol{\tau}_i)R\mathbf{v} = \left(\mathbf{v}'_i - \frac{\Delta \mathbf{v}}{\Delta s} \right) + \left(\frac{\Delta \mathbf{v}}{\Delta s} - R\mathbf{v} \right) + (1 - R\mathbf{f} \cdot \boldsymbol{\tau}_i)R\mathbf{v} = V_1 + V_2 + V_3.$$

We must derive L^2 , L^∞ , and C^0 estimates for each $V_i/\Delta s$. Because these are rather complicated, we refer the reader to [11]. \square

We now present the proof of Theorem 3.4.3. Let \bar{Q} be Q or \tilde{Q}_i . Then, we have that

$$|\mathcal{M}_i(\mathbf{f})| \leq \frac{\lambda^2}{2} \left\| \frac{\bar{Q}\mathbf{f}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \left\| \frac{\bar{Q}\mathbf{f}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n},$$

for $i = 1, 2$. Similarly, from Proposition 3.4.1 it can be derived that

$$\begin{aligned} |\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}]| &\leq \lambda^2 \left\| \frac{\bar{Q}\mathbf{f}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \left\| \frac{\bar{Q}\boldsymbol{\varphi}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} + 2\lambda |\mathcal{M}_i(\mathbf{f})| \left\| \frac{\Delta \boldsymbol{\varphi}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \\ &\leq \lambda^2 \left\| \frac{\bar{Q}\mathbf{f}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \left\| \frac{\bar{Q}\boldsymbol{\varphi}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} + 2\lambda |\mathcal{M}_i(\mathbf{f})| \|\boldsymbol{\varphi}\|_{\text{Lip}}. \end{aligned}$$

Let \bar{R} be \hat{R} or R , and let \bar{S} be S or \tilde{S}_i . Then, the definitions of these operations yield that

$$\begin{aligned} & \left| \frac{\bar{S}(\mathbf{v}, \mathbf{w})}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right| \\ & \leq \|\bar{R}\mathbf{v}\|_{\mathbb{R}^n} \left\| \frac{\bar{Q}\mathbf{w}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} + \left\| \frac{\bar{Q}\mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \|\bar{R}\mathbf{w}\|_{\mathbb{R}^n} \\ & \leq \lambda \left(\|\mathbf{v}\|_{\text{Lip}} \left\| \frac{\bar{Q}\mathbf{w}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} + \left\| \frac{\bar{Q}\mathbf{v}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \|\mathbf{w}\|_{\text{Lip}} \right). \end{aligned}$$

Therefore, Proposition 3.4.1 implies that

$$\begin{aligned} |\mathcal{H}_i(\mathbf{f})[\boldsymbol{\varphi}, \boldsymbol{\psi}]| & \leq \lambda^2 \left\| \frac{\bar{Q}\boldsymbol{\varphi}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \left\| \frac{\bar{Q}\boldsymbol{\psi}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \\ & + \lambda^4 \left(\|\mathbf{f}\|_{\text{Lip}} \left\| \frac{\bar{Q}\boldsymbol{\varphi}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} + \left\| \frac{\bar{Q}\mathbf{f}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \|\boldsymbol{\varphi}\|_{\text{Lip}} \right) \\ & \quad \times \left(\|\mathbf{f}\|_{\text{Lip}} \left\| \frac{\bar{Q}\boldsymbol{\psi}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} + \left\| \frac{\bar{Q}\mathbf{f}}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))} \right\|_{\mathbb{R}^n} \|\boldsymbol{\psi}\|_{\text{Lip}} \right) \\ & + 2\lambda^2 |\mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}]| \|\boldsymbol{\psi}\|_{\text{Lip}} + 2\lambda^2 |\mathcal{G}_i(\mathbf{f})[\boldsymbol{\psi}]| \|\boldsymbol{\varphi}\|_{\text{Lip}} \\ & + 2\lambda^2 |\mathcal{M}_i(\mathbf{f})| \|\boldsymbol{\varphi}\|_{\text{Lip}} \|\boldsymbol{\psi}\|_{\text{Lip}}. \end{aligned}$$

Consequently, the estimates in Theorem 3.4.3 can easily be derived using Lemma 3.4.2. If $\mathbf{f} \in C^2(\mathbb{R}/\mathbb{L}\mathbb{Z})$, then Lemma 3.4.2 yields that

$$\lim_{(s_1, s_2) \rightarrow (s, s)} (\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f})) = \frac{1}{2} \|\mathbf{f}''(s)\|_{\mathbb{R}^n}^2 - \frac{1}{2} \|\mathbf{f}''(s)\|_{\mathbb{R}^n}^2 = 0.$$

Similarly, we can show that both of the limits of $\mathcal{G}_1(\mathbf{f}) + \mathcal{G}_2(\mathbf{f})$ and $\mathcal{H}_1(\mathbf{f}) + \mathcal{H}_2(\mathbf{f})$ vanish.

Theorem 3.4.3 shows that the first variation $\delta\mathcal{E}_i(\mathbf{f})[\cdot]$ is a linear form on the space $X = H^{\frac{3}{2}}(\mathbb{R}/\mathbb{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathbb{L}\mathbb{Z})$. If $\mathbf{f} \in H^3(\mathbb{R}/\mathbb{L}\mathbb{Z})$, then it seems that the first variation can be extended into $L^2(\mathbb{R}/\mathbb{L}\mathbb{Z})$ as a linear form by a formal integration by parts.

Indeed, the principal term of $\delta\mathcal{E}_1(\mathbf{f})$ is

$$\iint_{(\mathbb{R}/\mathbb{L}\mathbb{Z})^2} \frac{(\mathbf{f}'(s_1) - \mathbf{f}'(s_2)) \cdot (\boldsymbol{\varphi}'(s_1) - \boldsymbol{\varphi}'(s_2))}{\|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2} ds_1 ds_2.$$

By considering bi-Lipschitz continuity, we replace the denominator with $\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2$, so that

$$\begin{aligned} & \iint_{(\mathbb{R}/\mathbb{L}\mathbb{Z})^2} \frac{(\mathbf{f}'(s_1) - \mathbf{f}'(s_2)) \cdot (\boldsymbol{\varphi}'(s_1) - \boldsymbol{\varphi}'(s_2))}{\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2} ds_1 ds_2 \\ & = 2\pi \int_{(\mathbb{R}/\mathbb{L}\mathbb{Z})} (-\Delta_s)^{\frac{1}{4}} \mathbf{f}' \cdot (-\Delta_s)^{\frac{1}{4}} \boldsymbol{\varphi}' ds. \end{aligned}$$

Here, $\Delta_s = \partial_s^2$ is the Laplace operator with respect to s , not $\Delta s = s_1 - s_2$. Integrating by parts formally, we obtain that

$$2\pi \int_{(\mathbb{R}/\mathcal{L}\mathbb{Z})} (-\Delta_s)^{\frac{3}{2}} \mathbf{f} \cdot \boldsymbol{\varphi} \, ds.$$

This seems to be meaningful for $\mathbf{f} \in H^3(\mathbb{R}/\mathcal{L}\mathbb{Z})$ and $\boldsymbol{\varphi} \in L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$. Indeed, we can justify this for not only the principal term, but also all terms, including $\delta\mathcal{E}_2$.

Here, we define a new operation T_i^k in order to describe the L^2 -gradient of $\delta\mathcal{E}_i$ as

$$T_i^k \mathbf{f} := \left(\frac{|\Delta s|}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}} \right)^k \frac{\Delta \mathbf{f}}{\Delta s} - \boldsymbol{\tau}_i.$$

Theorem 3.4.4. Let $\mathbf{f} \in H^3$ be bi-Lipschitz. Then, for $\boldsymbol{\varphi} \in L^2$ it holds that

$$\delta\mathcal{E}_i(\mathbf{f})[\boldsymbol{\varphi}] = \langle L_i \mathbf{f} + \mathbf{N}_i(\mathbf{f}), \boldsymbol{\varphi} \rangle_{L^2},$$

where

$$\begin{aligned} L_1 \mathbf{f} &= 2\pi(-\Delta_s)^{\frac{3}{2}} \mathbf{f} - 4 \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^3 \text{si}(|k\pi|) \langle \mathbf{f}, \boldsymbol{\varphi}_k \rangle_{L^2} \boldsymbol{\varphi}_k + \frac{8}{\mathcal{L}} \Delta_s (\mathbf{f} - \check{\mathbf{f}}), \\ L_2 \mathbf{f} &= -\frac{4}{3} \pi (-\Delta_s)^{\frac{3}{2}} \mathbf{f} + \frac{8}{3} \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^3 \text{si}(|k\pi|) \langle \mathbf{f}, \boldsymbol{\varphi}_k \rangle_{L^2} \boldsymbol{\varphi}_k + \frac{16}{3\mathcal{L}} \Delta_s \check{\mathbf{f}} + \frac{128}{3\mathcal{L}^3} (\mathbf{f} - \check{\mathbf{f}}), \\ \text{si}(t) &= - \int_t^\infty \frac{\sin \lambda}{\lambda} \, d\lambda, \quad \boldsymbol{\varphi}_k(s) = \frac{1}{\mathcal{L}} \exp\left(\frac{2\pi i k s}{\mathcal{L}}\right), \quad \check{\mathbf{f}}(s) = \mathbf{f}(s + \frac{\mathcal{L}}{2}), \end{aligned}$$

$$\begin{aligned}
\mathbf{N}_1(\mathbf{f})(s_1) &= -2 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left\{ \frac{2}{(\Delta s)^2} (T_1^4 \mathbf{f} \cdot \boldsymbol{\tau}_1) \Delta \boldsymbol{\tau} - \mathcal{M}(\mathbf{f}) \boldsymbol{\kappa}_1 \right\} ds_2 \\
&\quad - 4 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left[\frac{\mathcal{M}_1(\mathbf{f})}{\Delta s} T_1^2 \mathbf{f} + \frac{1}{\Delta s} \left\{ \mathcal{M}_1(\mathbf{f}) - \frac{1}{2} \|\boldsymbol{\kappa}_1\|_{\mathbb{R}^n}^2 \right\} \boldsymbol{\tau}_1 \right] ds_2, \\
\mathbf{N}_2(\mathbf{f})(s_1) &= -4 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \frac{1}{(\Delta s) \|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} \left\{ (T_1^2 \mathbf{f} \cdot \boldsymbol{\tau}_1) T_2^0 \mathbf{f} + (T_2^2 \mathbf{f} \cdot \boldsymbol{\tau}_2) T_1^0 \mathbf{f} \right\} ds_2 \\
&\quad - 4 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \frac{1}{(\Delta s)^3} \left[(T_1^4 \mathbf{f} \cdot \boldsymbol{\tau}_1) T_2^0 \mathbf{f} + (T_1^0 \mathbf{f} \cdot \boldsymbol{\tau}_2) T_1^4 \mathbf{f} \right. \\
&\quad \quad \left. + 2 \left\{ (T_2^0 \mathbf{f} \cdot \boldsymbol{\tau}_2) + 1 \right\} (T_1^2 \mathbf{f} \cdot \boldsymbol{\tau}_1) T_1^4 \mathbf{f} \right] ds_2 \\
&\quad - 4 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \frac{1}{(\Delta s)^3} \left[T_1^4 \mathbf{f} \cdot \boldsymbol{\tau}_1 - T_2^4 \mathbf{f} \cdot \boldsymbol{\tau}_2 \right. \\
&\quad \quad \left. + 2 \left\{ (T_2^0 \mathbf{f} \cdot \boldsymbol{\tau}_2) + 1 \right\} (T_1^2 \mathbf{f} \cdot \boldsymbol{\tau}_1) + T_1^0 \mathbf{f} \cdot \boldsymbol{\tau}_2 \right. \\
&\quad \quad \left. - \frac{(\Delta s)^2}{6} \|\boldsymbol{\kappa}_1\|_{\mathbb{R}^n}^2 \right] \boldsymbol{\tau}_1 ds_2 \\
&\quad - 4 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left[\frac{\mathcal{M}_2(\mathbf{f})}{\Delta s} T_1^2 \mathbf{f} + \frac{1}{\Delta s} \left\{ \mathcal{M}_2(\mathbf{f}) + \frac{1}{2} \|\boldsymbol{\kappa}_1\|_{\mathbb{R}^n}^2 \right\} \boldsymbol{\tau}_1 \right] ds_2.
\end{aligned}$$

Furthermore, for $\alpha \in (0, \frac{1}{2})$ it holds that

$$\|\mathbf{N}_i(\mathbf{f})\|_{L^2} \leq C_\alpha (\|\mathbf{f}\|_{H^{3-\alpha}}).$$

Our strategy for proving the above theorem is as follows. Because $C^\infty(\mathbb{R}/\mathcal{L}\mathbb{Z})$ is dense in both $H^3(\mathbb{R}/\mathcal{L}\mathbb{Z})$ and $L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$, we can assume that \mathbf{f} and $\boldsymbol{\varphi}$ are sufficiently smooth. According to Theorem 3.4.3, the first variation $\delta \mathcal{E}_i(\cdot)[\cdot]$ can be expressed as

$$\begin{aligned}
\delta \mathcal{E}_i(\mathbf{f})[\boldsymbol{\varphi}] &= \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \mathcal{G}_i(\mathbf{f})[\boldsymbol{\varphi}] ds_1 ds_2 \\
&= \sum_{j=1}^2 \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} G_{ij}(\mathbf{f}, \boldsymbol{\varphi})(s_1, s_2) ds_1 ds_2,
\end{aligned}$$

where

$$\begin{aligned}
G_{i1}(\mathbf{f}, \boldsymbol{\varphi}) &= \frac{Q_{i1} \mathbf{f} \cdot Q_{i2} \boldsymbol{\varphi} + Q_{i2} \mathbf{f} \cdot Q_{i1} \boldsymbol{\varphi}}{2 \|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} \\
G_{i2}(\mathbf{f}, \boldsymbol{\varphi}) &= -\frac{2 \mathcal{M}_i(\mathbf{f}) \Delta \mathbf{f} \cdot \Delta \boldsymbol{\varphi}}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2},
\end{aligned}$$

$$Q_{11} = Q_{12} = Q, \quad Q_{2j} \mathbf{v} = 2 \{ \mathbf{v}'_j - (R \mathbf{f} \cdot \boldsymbol{\tau}_j) R \mathbf{v} \}.$$

We decompose these operations Q_{ij} as $Q_{ij}\mathbf{v} = \tilde{Q}_{ij}\mathbf{v} + \bar{Q}_{ij}\mathbf{v}$, where

$$\begin{aligned}\tilde{Q}_{1j} &= Q_{1j} = Q, & \bar{Q}_{1j} &= 0, & \tilde{Q}_{2j}\mathbf{v} &= 2\left(\mathbf{v}'_j - \frac{\Delta\mathbf{v}}{\Delta s}\right), \\ \bar{Q}_{2j}\mathbf{v} &= 2\left\{\frac{\Delta\mathbf{v}}{\Delta s} - (R\mathbf{f} \cdot \boldsymbol{\tau}_j)R\mathbf{v}\right\} = 2\left\{1 - (R\mathbf{f} \cdot \boldsymbol{\tau}_j)\frac{|\Delta s|}{\|\Delta\mathbf{f}\|_{\mathbb{R}^n}}\right\}\frac{\Delta\mathbf{v}}{\Delta s}.\end{aligned}$$

Then, we have that

$$\begin{aligned}G_{i1}(\mathbf{f}, \boldsymbol{\varphi}) &= \sum_{k=1}^3 G_{i1k}(\mathbf{f}, \boldsymbol{\varphi}), \\ G_{i11}(\mathbf{f}, \boldsymbol{\varphi}) &= \frac{\tilde{Q}_{i1}\mathbf{f} \cdot \tilde{Q}_{i2}\boldsymbol{\varphi} + \tilde{Q}_{i2}\mathbf{f} \cdot \tilde{Q}_{i1}\boldsymbol{\varphi}}{2(\Delta s)^2}, \\ G_{i12}(\mathbf{f}, \boldsymbol{\varphi}) &= \frac{1}{2}\mathcal{M}(\mathbf{f})(\tilde{Q}_{i1}\mathbf{f} \cdot \tilde{Q}_{i2}\boldsymbol{\varphi} + \tilde{Q}_{i2}\mathbf{f} \cdot \tilde{Q}_{i1}\boldsymbol{\varphi}), \\ G_{i13}(\mathbf{f}, \boldsymbol{\varphi}) &= \frac{\tilde{Q}_{i1}\mathbf{f} \cdot \bar{Q}_{i2}\boldsymbol{\varphi} + \bar{Q}_{i1}\mathbf{f} \cdot \tilde{Q}_{i2}\boldsymbol{\varphi} + \bar{Q}_{i1}\mathbf{f} \cdot \bar{Q}_{i2}\boldsymbol{\varphi}}{2\|\Delta\mathbf{f}\|_{\mathbb{R}^n}^2} \\ &\quad + \frac{\tilde{Q}_{i2}\mathbf{f} \cdot \bar{Q}_{i1}\boldsymbol{\varphi} + \bar{Q}_{i2}\mathbf{f} \cdot \tilde{Q}_{i1}\boldsymbol{\varphi} + \bar{Q}_{i2}\mathbf{f} \cdot \bar{Q}_{i1}\boldsymbol{\varphi}}{2\|\Delta\mathbf{f}\|_{\mathbb{R}^n}^2}.\end{aligned}$$

We have that G_{i11} is linear with respect to \mathbf{f} , but, G_{i12} , G_{i13} , and G_{i2} are not. Next, we would like to write

$$\begin{aligned}\iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} G_{i11}(\mathbf{f}, \boldsymbol{\varphi}) ds_1 ds_2 &= \langle L_i \mathbf{f}, \boldsymbol{\varphi} \rangle_{L^2}, \\ \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} (G_{i12}(\mathbf{f}, \boldsymbol{\varphi}) + G_{i13}(\mathbf{f}, \boldsymbol{\varphi}) + G_{i2}(\mathbf{f}, \boldsymbol{\varphi})) ds_1 ds_2 &= \langle \mathbf{N}_i(\mathbf{f}), \boldsymbol{\varphi} \rangle_{L^2},\end{aligned}$$

where L_i and \mathbf{N}_i are linear and nonlinear operations from H^3 to L^2 , which we will then estimate.

For the linear parts L_i we use the Fourier expansion

$$\mathbf{f} = \sum_{k \in \mathbb{Z}} \varphi_k \mathbf{a}_k, \quad \boldsymbol{\varphi} = \sum_{k \in \mathbb{Z}} \varphi_k \mathbf{b}_k,$$

where $\{\varphi_k\}$ is a complete orthogonal basis of $L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})$. Using the fact that

$$\varphi'_k(s) = \frac{2\pi i k}{\mathcal{L}} \varphi_k(s) \quad \text{and} \quad \varphi_k(s+h) = \exp\left(\frac{2\pi i k h}{\mathcal{L}}\right) \varphi_k(s),$$

we have that

$$\begin{aligned}\mathbf{f}'(s_1) - \mathbf{f}'(s_1+h) &= \sum_{k \in \mathbb{Z}} \frac{2\pi i k}{\mathcal{L}} \left\{1 - \exp\left(\frac{2\pi i k h}{\mathcal{L}}\right)\right\} \varphi_k(s_1) \mathbf{a}_k, \\ \boldsymbol{\varphi}'(s_1) - \boldsymbol{\varphi}'(s_1+h) &= \sum_{k \in \mathbb{Z}} \frac{2\pi i k}{\mathcal{L}} \left\{1 - \exp\left(\frac{2\pi i k h}{\mathcal{L}}\right)\right\} \varphi_k(s_1) \mathbf{b}_k.\end{aligned}$$

Consequently, orthogonality implies that

$$\begin{aligned}
 & \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} G_{111}(\mathbf{f}, \boldsymbol{\varphi}) ds_1 ds_2 \\
 &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z} - \frac{\mathcal{L}}{2}}^{\frac{\mathcal{L}}{2}} \int_{-\frac{\mathcal{L}}{2}}^{\frac{\mathcal{L}}{2}} \frac{(\mathbf{f}'(s_1) - \mathbf{f}'(s_1 + h)) \cdot (\boldsymbol{\varphi}'(s_1) - \boldsymbol{\varphi}'(s_1 + h))}{h^2} dh ds_1 \\
 &= 2 \sum_{k \in \mathbb{Z}} \left(\frac{2\pi k}{\mathcal{L}} \right)^2 \left\{ \int_{-\frac{\mathcal{L}}{2}}^{\frac{\mathcal{L}}{2}} \frac{1 - \cos\left(\frac{2\pi kh}{\mathcal{L}}\right)}{h^2} dh \right\} ds_1 \langle \mathbf{a}_k, \mathbf{b}_k \rangle_{\mathbb{C}} \\
 &= 2 \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^3 \left(\int_{-\pi|k|}^{\pi|k|} \frac{1 - \cos \lambda}{\lambda^2} d\lambda \right) \langle \mathbf{a}_k, \mathbf{b}_k \rangle_{\mathbb{C}}.
 \end{aligned}$$

Similarly, we have that

$$\begin{aligned}
 & \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} G_{211}(\mathbf{f}, \boldsymbol{\varphi}) ds_1 ds_2 \\
 &= \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^3 \left[\int_{-\pi|k|}^{\pi|k|} \frac{4\{\lambda^2 \cos \lambda - 2\lambda \sin \lambda + 2(1 - \cos \lambda)\}}{\lambda^4} d\lambda \right] \langle \mathbf{a}_k, \mathbf{b}_k \rangle_{\mathbb{C}}.
 \end{aligned}$$

Thus, we arrive at

$$L_i \mathbf{f} = \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^3 \left(\int_{-\pi|k|}^{\pi|k|} z_i(\lambda) d\lambda \right) \langle \mathbf{f}, \boldsymbol{\varphi}_k \rangle_{L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})} \boldsymbol{\varphi}_k,$$

where

$$z_1(\lambda) = \frac{2(1 - \cos \lambda)}{\lambda^2}, \quad z_2(\lambda) = \frac{4\{\lambda^2 \cos \lambda - 2\lambda \sin \lambda + 2(1 - \cos \lambda)\}}{\lambda^4}.$$

It is not difficult to see that

$$\int_{-\pi|k|}^{\pi|k|} z_i(\lambda) d\lambda = a_i(\pi + 2\text{si}(|k\pi|)) + 2b_i|k\pi|z_i(|k\pi|),$$

where

$$a_1 = 2, \quad b_1 = -1, \quad a_2 = -\frac{4}{3}, \quad b_2 = -\frac{1}{3}.$$

From the definition of z_i , it follows that

$$\begin{aligned} |k\pi|z_1(|k\pi|) &= \frac{2\{1 - \cos(|k\pi|)\}}{|k\pi|} = \frac{2\{1 - (-1)^k\}}{|k\pi|}, \\ |k\pi|z_2(|k\pi|) &= \frac{4[-|k\pi|^2 \cos(|k\pi|) + 2|k\pi| \sin(|k\pi|) - 2\{1 - \cos(|k\pi|)\}]}{|k\pi|^3} \\ &= \frac{4[-(-1)^k |k\pi|^2 - 2\{1 - (-1)^k\}]}{|k\pi|^3}. \end{aligned}$$

By combining these with the expression

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \langle \mathbf{f}, \varphi_k \rangle_{L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})} \varphi_k &= \mathbf{f}, \\ \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^2 \langle \mathbf{f}, \varphi_k \rangle_{L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})} \varphi_k &= -\Delta_s \mathbf{f}, \\ \sum_{k \in \mathbb{Z}} (-1)^k \langle \mathbf{f}, \varphi_k \rangle_{L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})} \varphi_k &= \check{\mathbf{f}}, \\ \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^2 (-1)^k \langle \mathbf{f}, \varphi_k \rangle_{L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})} \varphi_k &= -\Delta_s \check{\mathbf{f}} \end{aligned}$$

we obtain the same expressions for the linear parts L_i as in Theorem 3.4.4.

Because the derivation of the nonlinear parts \mathbf{N}_i is considerably more complicated, we only provide a sketch. For details, see [13].

We employ $G(\mathbf{f}, \boldsymbol{\varphi})$ as one of $G_{i1k}(\mathbf{f}, \boldsymbol{\varphi})$ ($k = 2, 3$) or $G_{i2}(\mathbf{f}, \boldsymbol{\varphi})$, which then take the form

$$G(\mathbf{f}, \boldsymbol{\varphi}) = \mathbf{G}_A(\mathbf{f}) \cdot \Delta \boldsymbol{\varphi}' + \mathbf{G}_B(\mathbf{f}) \cdot \Delta \boldsymbol{\varphi} + \mathbf{G}_C(\mathbf{f}) \cdot \boldsymbol{\varphi}'(s_1) + \mathbf{G}_D(\mathbf{f}) \cdot \boldsymbol{\varphi}'(s_2).$$

The following can be obtained through a simple calculation.

Lemma 3.4.3. The following relations hold:

1.

$$\begin{aligned} \iint_{|s_1 - s_2| \geq \varepsilon} \boldsymbol{\zeta}(s_1, s_2) \cdot \boldsymbol{\varphi}'(s_1) ds_1 ds_2 &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} (\boldsymbol{\zeta}(s, s + \varepsilon) - \boldsymbol{\zeta}(s, s - \varepsilon)) \cdot \boldsymbol{\varphi}(s) ds \\ &\quad - \iint_{|s_1 - s_2| \geq \varepsilon} \frac{\partial}{\partial s_1} \boldsymbol{\zeta}(s_1, s_2) \cdot \boldsymbol{\varphi}(s_1) ds_1 ds_2 \end{aligned}$$

2.

$$\begin{aligned} \iint_{|s_1 - s_2| \geq \varepsilon} \boldsymbol{\zeta}(s_1, s_2) \cdot \boldsymbol{\varphi}'(s_2) ds_1 ds_2 &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} (\boldsymbol{\zeta}(s + \varepsilon, s) - \boldsymbol{\zeta}(s - \varepsilon, s)) \cdot \boldsymbol{\varphi}(s) ds \\ &\quad - \iint_{|s_1 - s_2| \geq \varepsilon} \frac{\partial}{\partial s_2} \boldsymbol{\zeta}(s_1, s_2) \cdot \boldsymbol{\varphi}(s_2) ds_1 ds_2 \end{aligned}$$

3.

$$\iint_{|s_1-s_2|\geq\epsilon} \boldsymbol{\zeta}(s_1, s_2) \cdot \Delta \boldsymbol{\varphi} \, ds_1 ds_2 = \iint_{|s_1-s_2|\geq\epsilon} (\boldsymbol{\zeta}(s_1, s_2) - \boldsymbol{\zeta}(s_2, s_1)) \cdot \boldsymbol{\varphi}(s_1) \, ds_1 ds_2$$

4.

$$\begin{aligned} & \iint_{|s_1-s_2|\geq\epsilon} \boldsymbol{\zeta}(s_1, s_2) \cdot \Delta \boldsymbol{\varphi}' \, ds_1 ds_2 \\ &= \int_{\mathbb{R}/\mathbb{L}\mathbb{Z}} (\boldsymbol{\zeta}(s, s+\epsilon) - \boldsymbol{\zeta}(s+\epsilon, s) - \boldsymbol{\zeta}(s, s-\epsilon) + \boldsymbol{\zeta}(s-\epsilon, s)) \cdot \boldsymbol{\varphi}(s) \, ds \\ & \quad - \iint_{|s_1-s_2|\geq\epsilon} \frac{\partial}{\partial s_1} (\boldsymbol{\zeta}(s_1, s_2) - \boldsymbol{\zeta}(s_2, s_1)) \cdot \boldsymbol{\varphi}(s_1) \, ds_1 ds_2. \end{aligned}$$

From Lemma 3.4.3, we obtain that

$$\begin{aligned}
& \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} G(\mathbf{f}, \boldsymbol{\varphi}) ds_1 ds_2 = \lim_{\varepsilon \rightarrow +0} \iint_{|s_1 - s_2| \geq \varepsilon} G(\mathbf{f}, \boldsymbol{\varphi}) ds_1 ds_2 \\
&= \lim_{\varepsilon \rightarrow +0} \left(\iint_{|s_1 - s_2| \geq \varepsilon} \mathbf{G}_A(\mathbf{f}) \cdot \Delta \boldsymbol{\varphi}' ds_1 ds_2 + \iint_{|s_1 - s_2| \geq \varepsilon} \mathbf{G}_B(\mathbf{f}) \cdot \Delta \boldsymbol{\varphi} ds_1 ds_2 \right. \\
&\quad \left. + \iint_{|s_1 - s_2| \geq \varepsilon} \mathbf{G}_C(\mathbf{f}) \cdot \boldsymbol{\varphi}'(s_1) ds_1 ds_2 + \iint_{|s_1 - s_2| \geq \varepsilon} \mathbf{G}_D(\mathbf{f}) \cdot \boldsymbol{\varphi}'(s_2) ds_1 ds_2 \right) \\
&= \lim_{\varepsilon \rightarrow +0} \left\{ \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} (\mathbf{G}_A(\mathbf{f})(s, s + \varepsilon) - \mathbf{G}_A(\mathbf{f})(s + \varepsilon, s)) \right. \\
&\quad - \mathbf{G}_A(\mathbf{f})(s, s - \varepsilon) - \mathbf{G}_A(\mathbf{f})(s - \varepsilon, s)) \cdot \boldsymbol{\varphi}(s) ds \\
&\quad - \iint_{|s_1 - s_2| \geq \varepsilon} \frac{\partial}{\partial s_1} (\mathbf{G}_A(\mathbf{f})(s_1, s_2) - \mathbf{G}_A(\mathbf{f})(s_2, s_1)) \cdot \boldsymbol{\varphi}(s_1) ds_1 ds_2 \\
&\quad + \iint_{|s_1 - s_2| \geq \varepsilon} (\mathbf{G}_B(\mathbf{f})(s_1, s_2) - \mathbf{G}_B(\mathbf{f})(s_2, s_1)) \cdot \boldsymbol{\varphi}(s_1) ds_1 ds_2 \\
&\quad + \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} (\mathbf{G}_C(\mathbf{f})(s, s + \varepsilon) - \mathbf{G}_C(\mathbf{f})(s, s - \varepsilon)) \cdot \boldsymbol{\varphi}(s) ds \\
&\quad - \iint_{|s_1 - s_2| \geq \varepsilon} \frac{\partial}{\partial s_1} \mathbf{G}_C(\mathbf{f})(s_1, s_2) \cdot \boldsymbol{\varphi}(s_1) ds_1 ds_2 \\
&\quad + \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} (\mathbf{G}_D(\mathbf{f})(s + \varepsilon, s) - \mathbf{G}_D(\mathbf{f})(s - \varepsilon, s)) \cdot \boldsymbol{\varphi}(s) ds \\
&\quad \left. - \iint_{|s_1 - s_2| \geq \varepsilon} \frac{\partial}{\partial s_2} \mathbf{G}_D(\mathbf{f})(s_1, s_2) \cdot \boldsymbol{\varphi}(s_2) ds_1 ds_2 \right\} \\
&= (\dagger).
\end{aligned}$$

Here, we will prove that if $\mathbf{f} \in H^3$, then it holds that

$$\begin{aligned}
 (\dagger) &= \iint_{(\mathbb{R}/\mathbb{L}\mathbb{Z})^2} \left\{ -\frac{\partial}{\partial s_1} (\mathbf{G}_A(\mathbf{f})(s_1, s_2) - \mathbf{G}_A(\mathbf{f})(s_2, s_1)) \right. \\
 &\quad + \mathbf{G}_B(\mathbf{f})(s_1, s_2) - \mathbf{G}_B(\mathbf{f})(s_2, s_1) \\
 &\quad \left. - \frac{\partial}{\partial s_1} \mathbf{G}_C(\mathbf{f})(s_1, s_2) - \frac{\partial}{\partial s_1} \mathbf{G}_D(\mathbf{f})(s_2, s_1) \right\} \cdot \boldsymbol{\varphi}(s_1) ds_1 ds_2 \\
 &= (\ddagger).
 \end{aligned}$$

Furthermore, we shall verify that

$$\begin{aligned}
 \mathbf{N}(\mathbf{f})(s) &= \int_{\mathbb{R}/\mathbb{L}\mathbb{Z}} \left\{ -\frac{\partial}{\partial s} (\mathbf{G}_A(\mathbf{f})(s, s_2) - \mathbf{G}_A(\mathbf{f})(s_2, s)) \right. \\
 &\quad + \mathbf{G}_B(\mathbf{f})(s, s_2) - \mathbf{G}_B(\mathbf{f})(s_2, s) \\
 &\quad \left. - \frac{\partial}{\partial s} \mathbf{G}_C(\mathbf{f})(s, s_2) - \frac{\partial}{\partial s} \mathbf{G}_D(\mathbf{f})(s_2, s) \right\} ds_2
 \end{aligned}$$

is well-defined at \mathcal{L}^1 -a.e. $s \in \mathbb{R}/\mathbb{L}\mathbb{Z}$, so that we can apply Fubini's theorem, which implies that

$$(\ddagger) = \int_{\mathbb{R}/\mathbb{L}\mathbb{Z}} \mathbf{N}(\mathbf{f})(s) \cdot \boldsymbol{\varphi}(s) ds = \langle \mathbf{N}(\mathbf{f}), \boldsymbol{\varphi} \rangle_{L^2(\mathbb{R}/\mathbb{L}\mathbb{Z})}.$$

We will also show that $\mathbf{N}(\mathbf{f})$ is a lower order term, whose order is less than three, and

$$\|\mathbf{N}(\mathbf{f})\|_{L^2} \leq C(\|\mathbf{f}\|_{H^{3-a}}, \lambda).$$

We employ \mathbf{N}_{i1k} or \mathbf{N}_{i2} as counterparts of \mathbf{N} when $G = G_{i1k}$ or G_{i2} , respectively.

Here, we will present details for \mathbf{N}_{112} only. For this, we require the following facts, which we state without their proof.

Lemma 3.4.4. Let $\boldsymbol{\kappa} \in L^\infty$. Then, it holds that

$$T_i^0 \mathbf{f} = \mathcal{O}(\Delta s), \quad T_i^0 \mathbf{f} \cdot \boldsymbol{\tau}(s_j) = \mathcal{O}(\Delta s)^2, \quad T_i^0 \mathbf{f} \cdot \frac{\Delta \mathbf{f}}{\Delta s} = \mathcal{O}(\Delta s)^2.$$

Suppose that \mathbf{f} is bi-Lipschitz and $\boldsymbol{\kappa} \in L^\infty$. For $k \geq 1$, it holds that

$$T_i^k \mathbf{f} = \mathcal{O}(\Delta s), \quad T_i^k \mathbf{f} \cdot \boldsymbol{\tau}(s_j) = \mathcal{O}(\Delta s)^2, \quad T_i^k \mathbf{f} \cdot \frac{\Delta \mathbf{f}}{\Delta s} = \mathcal{O}(\Delta s)^2.$$

Lemma 3.4.5. It holds that

$$\frac{\partial}{\partial s_j} \mathcal{M}(\mathbf{f}) = \frac{2(-1)^j}{(\Delta s)^3} T_j^4 \mathbf{f} \cdot \boldsymbol{\tau}(s_j).$$

It is easy to see that

$$G_{112}(\mathbf{f}, \boldsymbol{\varphi}) = \mathbf{g}_{112}(\mathbf{f}) \cdot \Delta \boldsymbol{\varphi}',$$

where

$$\mathbf{g}_{112}(\mathbf{f})(s_1, s_2) = \mathcal{M}(\mathbf{f}) \Delta \mathbf{f}'.$$

In the following, we denote $\mathbf{g}_{112}(\mathbf{f})(s_1, s_2)$ by $\mathbf{g}_{112}(s_1, s_2)$, for convenience.

Lemma 3.4.6. Let $\alpha \in (0, \frac{1}{2})$. If $\mathbf{f} \in H^{3-\alpha}(\mathbb{R}/\mathcal{L}\mathbb{Z})$, it follows that

$$\iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} G_{112}(\mathbf{f}, \boldsymbol{\varphi})(s_1, s_2) ds_1 ds_2 = \langle \mathbf{N}_{112}(\mathbf{f}), \boldsymbol{\varphi} \rangle_{L^2},$$

where

$$\begin{aligned} \mathbf{N}_{112}(\mathbf{f})(s_1) &= 2 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left\{ \frac{2}{(\Delta s)^3} (T_1^4 \mathbf{f} \cdot \boldsymbol{\tau}(s_1)) \Delta \boldsymbol{\tau} - \mathcal{M}(\mathbf{f}) \boldsymbol{\kappa}(s_1) \right\} ds_2, \\ \|\mathbf{N}_{112}(\mathbf{f})\|_{L^2} &\leq C(\|\mathbf{f}\|_{H^{3-\alpha}}). \end{aligned}$$

Proof. Using Lemma 3.4.3, we have that

$$\begin{aligned} & \iint_{|s_1-s_2| \geq \varepsilon} G_{112}(\mathbf{f}, \boldsymbol{\varphi}) ds_1 ds_2 \\ &= \iint_{|s_1-s_2| \geq \varepsilon} \mathbf{g}_{112}(\mathbf{f}) \cdot \Delta \boldsymbol{\varphi}' ds_1 ds_2 \\ &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} (\mathbf{g}_{112}(s, s+\varepsilon) - \mathbf{g}_{112}(s+\varepsilon, s) - \mathbf{g}_{112}(s, s-\varepsilon) + \mathbf{g}_{112}(s-\varepsilon, s)) \cdot \boldsymbol{\varphi}(s) ds \\ &\quad - \iint_{|s_1-s_2| \geq \varepsilon} \frac{\partial}{\partial s_1} (\mathbf{g}_{112}(s_1, s_2) - \mathbf{g}_{112}(s_2, s_1)) \cdot \boldsymbol{\varphi}(s_1) ds_1 ds_2 \\ &= (*). \end{aligned}$$

We remark that

$$\mathbf{g}_{112}(s_1, s_2) = \mathcal{M}(\mathbf{f}) \Delta \mathbf{f}' = \mathcal{M}(\mathbf{f}) \Delta \boldsymbol{\tau} = \mathcal{O}(\Delta s),$$

which implies that

$$\mathbf{g}_{112}(s, s+\varepsilon) - \mathbf{g}_{112}(s+\varepsilon, s) - \mathbf{g}_{112}(s, s-\varepsilon) + \mathbf{g}_{112}(s-\varepsilon, s) = \mathcal{O}(\varepsilon) \quad (\varepsilon \rightarrow 0).$$

Note that $\mathcal{O}(\varepsilon)$ is uniform with regard to $s \in \mathbb{R}/\mathcal{L}\mathbb{Z}$, and in the following, we employ the above notation with this meaning. Because it holds that

$$\mathbf{g}_{112}(s_1, s_2) - \mathbf{g}_{112}(s_2, s_1) = 2 \mathcal{M}(\mathbf{f}) \Delta \boldsymbol{\tau},$$

we can apply Lemmas 3.4.4–3.4.5 to infer that

$$\begin{aligned}
 & \frac{\partial}{\partial s_1} \{ \mathbf{g}_{112}(s_1, s_2) - \mathbf{g}_{112}(s_2, s_1) \} \\
 &= 2 \left(\frac{\partial \mathcal{M}(\mathbf{f})}{\partial s_1} \Delta \boldsymbol{\tau} + \mathcal{M}(\mathbf{f}) \frac{\partial \Delta \boldsymbol{\tau}}{\partial s_1} \right) \\
 &= 2 \left\{ -\frac{2}{(\Delta s)^3} (T_1^4 \mathbf{f} \cdot \boldsymbol{\tau}(s_1)) \Delta \boldsymbol{\tau} + \mathcal{M}(\mathbf{f}) \boldsymbol{\kappa}(s_1) \right\} \\
 &= \mathcal{O}(\Delta s)^{-3+2+1} + \mathcal{O}(1) \\
 &= \mathcal{O}(1) \quad \text{as } \Delta s \rightarrow 0
 \end{aligned}$$

which implies that $\frac{\partial}{\partial s_1} \{ \mathbf{g}_{112}(s_1, s_2) - \mathbf{g}_{112}(s_2, s_1) \} \in L^\infty((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)$. Therefore, the expression is absolutely integrable, and we can use Fubini's theorem in order to obtain that

$$\begin{aligned}
 (*) & \rightarrow - \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} 2 \left\{ -\frac{2}{(\Delta s)^3} (T_1^4 \mathbf{f} \cdot \boldsymbol{\tau}(s_1)) \Delta \boldsymbol{\tau} + \mathcal{M}(\mathbf{f}) \boldsymbol{\kappa}(s_1) \right\} \cdot \boldsymbol{\varphi}(s_1) ds_1 ds_2 \\
 &= \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left[2 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \left\{ \frac{2}{(\Delta s)^3} (T_1^4 \mathbf{f} \cdot \boldsymbol{\tau}(s_1)) \Delta \boldsymbol{\tau} - \mathcal{M}(\mathbf{f}) \boldsymbol{\kappa}(s_1) \right\} ds_2 \right] \cdot \boldsymbol{\varphi}(s_1) ds_1 \\
 &= \langle \mathbf{N}_{112}(\mathbf{f}), \boldsymbol{\varphi} \rangle_{L^2},
 \end{aligned}$$

by letting $\varepsilon \rightarrow +0$ in (*). Because it holds that

$$\|\boldsymbol{\kappa}\|_{L^\infty} = \|\mathbf{f}''\|_{L^\infty} \leq C_\alpha \|\mathbf{f}\|_{H^{3-\alpha}}$$

for $\alpha \in (0, \frac{1}{2})$, the bound on the integrand of \mathbf{N}_{112} follows from Lemma 3.4.4. Thus, we have that

$$\|\mathbf{N}_{112}(\mathbf{f})\|_{L^2} \leq C \|\mathbf{f}\|_{H^{3-\alpha}}.$$

□

Although the proof is more complicated, a similar result holds for G_{212} . Because $G_{113} = 0$, no further work is necessary in this case. We must consider G_{213} very carefully. This is decomposed as

$$G_{213}(\mathbf{f}, \boldsymbol{\varphi}) = \mathbf{G}_{213B}(\mathbf{f}) \cdot \Delta \boldsymbol{\varphi} + \mathbf{G}_{213C}(\mathbf{f}) \cdot \boldsymbol{\varphi}'(s_1) + \mathbf{G}_{213D}(\mathbf{f}) \cdot \boldsymbol{\varphi}'(s_2),$$

where

$$\begin{aligned} \mathbf{G}_{213B}(\mathbf{f}) &= \frac{2}{(\Delta s)\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} \left\{ (T_1^2 \mathbf{f} \cdot \boldsymbol{\tau}(s_1)) \left(T_1^0 \mathbf{f} + \frac{\Delta \mathbf{f}}{\Delta s} \right) + (T_2^2 \mathbf{f} \cdot \boldsymbol{\tau}(s_2)) \left(T_2^0 \mathbf{f} + \frac{\Delta \mathbf{f}}{\Delta s} \right) \right. \\ &\quad \left. + 2(T_1^2 \mathbf{f} \cdot \boldsymbol{\tau}(s_1))(T_2^2 \mathbf{f} \cdot \boldsymbol{\tau}(s_2)) \frac{\Delta \mathbf{f}}{\Delta s} \right\}, \\ \mathbf{G}_{213C}(\mathbf{f}) &= -\frac{2}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} (T_2^2 \mathbf{f} \cdot \boldsymbol{\tau}(s_2)) \frac{\Delta \mathbf{f}}{\Delta s}, \\ \mathbf{G}_{213D}(\mathbf{f}) &= -\frac{2}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} (T_1^2 \mathbf{f} \cdot \boldsymbol{\tau}(s_1)) \frac{\Delta \mathbf{f}}{\Delta s}. \end{aligned}$$

Because none of \mathbf{G}_{213B} , \mathbf{G}_{213C} , and \mathbf{G}_{213D} are absolutely integrable, we must combine them appropriately. We require a similar treatment for G_{i2} , but the situation is less complex than that of G_{213} . For details, see [13].

3.4.3 The Möbius invariance

The Möbius invariance of each decomposed energy \mathcal{E}_i is discussed in [10, 12].

Theorem 3.4.5 ([10]). Each \mathcal{E}_i is Möbius invariant in the following sense:

1. It is invariant under rigid motion and dilation.
2. Let

$$\mathbf{f} \mapsto \mathbf{p} = \mathbf{c} + \frac{r^2(\mathbf{f} - \mathbf{c})}{\|\mathbf{f} - \mathbf{c}\|_{\mathbb{R}^n}^2}$$

be the inversion with respect to sphere with center \mathbf{c} and radius r .

- (1) If $\mathbf{f} \in W^{1,1}(\mathbb{R}/\mathbb{Z})$, then $\mathcal{E}_1(\mathbf{f}) + \mathcal{E}_2(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}) + \mathcal{E}_2(\mathbf{p})$.
- (2) If $\mathbf{c} \notin \text{Im } \mathbf{f}$ and if $\mathcal{E}(\mathbf{f}) < \infty$, then $\mathcal{E}_1(\mathbf{f}) = \mathcal{E}_1(\mathbf{p})$ and $\mathcal{E}_2(\mathbf{f}) = \mathcal{E}_2(\mathbf{p})$.

Remark 3.4.1. In the assertion (1) we assume that neither $\mathbf{c} \notin \text{Im } \mathbf{f}$ nor $\mathcal{E}(\mathbf{f}) < \infty$.

Because the invariance under rigid motion and dilation can be easily proved, we focus on the invariance under inversion with respect to the sphere. Although the invariance of $(\mathcal{M}_1 + \mathcal{M}_2) ds_1 ds_2$ was shown in [5], we provide the proof here. First, we note that

$$\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f}) = \frac{1 + \boldsymbol{\tau}(s_1) \cdot \boldsymbol{\tau}(s_2)}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} - \frac{2(\Delta \mathbf{f} \cdot \boldsymbol{\tau}(s_1))(\Delta \mathbf{f} \cdot \boldsymbol{\tau}(s_2))}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^4}.$$

Even if s is an arc-length parameter for \mathbf{f} , it is not necessarily one for \mathbf{p} . Therefore, we use a general parameter θ instead of s , and the energy density with respect to $d\theta_1 d\theta_2$ is

$$\begin{aligned} (\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f})) \|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n} &= \frac{\|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n} + \dot{\mathbf{f}}(\theta_1) \cdot \dot{\mathbf{f}}(\theta_2)}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \\ &\quad + \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right). \end{aligned}$$

Here $\dot{\mathbf{f}}$ means the differentiation of \mathbf{f} with respect to the general parameter, and similarly for other functions.

Theorem 3.4.6. Let

$$\mathbf{f} \mapsto \mathbf{p} = \mathbf{c} + \frac{r^2(\mathbf{f} - \mathbf{c})}{\|\mathbf{f} - \mathbf{c}\|_{\mathbb{R}^n}^2}$$

be an inversion with respect to the sphere with center \mathbf{c} and radius r . Then, it holds that

$$(\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f})) \|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n} - (\mathcal{M}_1(\mathbf{p}) + \mathcal{M}_2(\mathbf{p})) \|\dot{\mathbf{p}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{p}}(\theta_2)\|_{\mathbb{R}^n} = 0 \quad (3.4.8)$$

for θ_1 and θ_2 such that

$$\mathbf{f}(\theta_1) \neq \mathbf{f}(\theta_2), \quad \mathbf{f}(\theta_i) \neq \mathbf{c} \quad (i = 1, 2).$$

Proof. We decompose the difference between the densities for \mathbf{f} and for \mathbf{p} as follows:

$$\begin{aligned} & (\mathcal{M}_1(\mathbf{f}) + \mathcal{M}_2(\mathbf{f})) \|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n} - (\mathcal{M}_1(\mathbf{p}) + \mathcal{M}_2(\mathbf{p})) \|\dot{\mathbf{p}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{p}}(\theta_2)\|_{\mathbb{R}^n} \\ &= J_1 + J_2 + J_3, \end{aligned}$$

where

$$\begin{aligned} J_1 &= \frac{\|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n}}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} - \frac{\|\dot{\mathbf{p}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{p}}(\theta_2)\|_{\mathbb{R}^n}}{\|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2}, \\ J_2 &= \frac{\dot{\mathbf{f}}(\theta_1) \cdot \dot{\mathbf{f}}(\theta_2)}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} - \frac{\dot{\mathbf{p}}(\theta_1) \cdot \dot{\mathbf{p}}(\theta_2)}{\|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2}, \\ J_3 &= \frac{1}{2} \left\{ \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \right. \\ & \quad \left. - \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \right\}. \end{aligned}$$

It holds that

$$\begin{aligned} & \|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2 \\ &= \left\| \frac{r^2(\mathbf{f}(\theta_1) - \mathbf{c})}{\|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2} - \frac{r^2(\mathbf{f}(\theta_2) - \mathbf{c})}{\|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2} \right\|_{\mathbb{R}^n}^2 \\ &= r^4 \left\{ \frac{1}{\|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2} - \frac{2(\mathbf{f}(\theta_1) - \mathbf{c}) \cdot (\mathbf{f}(\theta_2) - \mathbf{c})}{\|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2} + \frac{1}{\|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2} \right\} \\ &= \frac{r^4 \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2}{\|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2}. \end{aligned}$$

If we define the projection $P_c(\theta)$ and $P_c^\perp(\theta)$ for a vector \mathbf{v} by

$$P_c(\theta)\mathbf{v} = \left(\mathbf{v} \cdot \frac{\mathbf{f}(\theta) - \mathbf{c}}{\|\mathbf{f}(\theta) - \mathbf{c}\|_{\mathbb{R}^n}} \right) \frac{\mathbf{f}(\theta) - \mathbf{c}}{\|\mathbf{f}(\theta) - \mathbf{c}\|_{\mathbb{R}^n}}, \quad P_c^\perp(\theta) = I - P_c(\theta),$$

then the derivative of $\mathbf{p}(\theta)$ may be expressed as

$$\begin{aligned} \dot{\mathbf{p}}(\theta) &= r^2 \left[\frac{\dot{\mathbf{f}}(\theta)}{\|\mathbf{f}(\theta) - \mathbf{c}\|_{\mathbb{R}^n}^2} - \frac{2\{\dot{\mathbf{f}}(\theta) \cdot (\mathbf{f}(\theta) - \mathbf{c})\}(\mathbf{f}(\theta) - \mathbf{c})}{\|\mathbf{f}(\theta) - \mathbf{c}\|_{\mathbb{R}^n}^4} \right] \\ &= \frac{r^2}{\|\mathbf{f}(\theta) - \mathbf{c}\|_{\mathbb{R}^n}^2} (P_c^\perp(\theta) - P_c(\theta)) \dot{\mathbf{f}}(\theta). \end{aligned}$$

Therefore, we have that

$$\begin{aligned} \|\dot{\mathbf{p}}(\theta)\|_{\mathbb{R}^n}^2 &= \frac{r^4}{\|\mathbf{f}(\theta) - \mathbf{c}\|_{\mathbb{R}^n}^4} \left(\|P_c^\perp(\theta) \dot{\mathbf{f}}(\theta)\|_{\mathbb{R}^n}^2 + \|P_c(\theta) \dot{\mathbf{f}}(\theta)\|_{\mathbb{R}^n}^2 \right) \\ &= \frac{r^4 \|\dot{\mathbf{f}}(\theta)\|_{\mathbb{R}^n}^2}{\|\mathbf{f}(\theta) - \mathbf{c}\|_{\mathbb{R}^n}^4}. \end{aligned}$$

As a straightforward consequence, we have that

$$\begin{aligned} &\frac{\|\dot{\mathbf{p}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{p}}(\theta_2)\|_{\mathbb{R}^n}}{\|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2} \\ &= \frac{r^2 \|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n}}{\|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2} \frac{r^2 \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n}}{\|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2} \frac{\|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2}{r^4 \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \\ &= \frac{\|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n}}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2}, \end{aligned}$$

which demonstrates that $J_1 = 0$. By a similar calculation, it holds that

$$\frac{\dot{\mathbf{p}}(\theta_1) \cdot \dot{\mathbf{p}}(\theta_2)}{\|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2} = \frac{\left\{ (I - 2P_c(\theta_1)) \dot{\mathbf{f}}(\theta_1) \right\} \cdot \left\{ (I - 2P_c(\theta_2)) \dot{\mathbf{f}}(\theta_2) \right\}}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2}.$$

By observing that

$$\begin{aligned} (I - 2P_c(\theta_i)) \dot{\mathbf{f}}(\theta_i) &= \dot{\mathbf{f}}(\theta_i) - 2 \left(\frac{\mathbf{f}(\theta_i) - \mathbf{c}}{\|\mathbf{f}(\theta_i) - \mathbf{c}\|_{\mathbb{R}^n}} \cdot \dot{\mathbf{f}}(\theta_i) \right) \frac{\mathbf{f}(\theta_i) - \mathbf{c}}{\|\mathbf{f}(\theta_i) - \mathbf{c}\|_{\mathbb{R}^n}} \\ &= \dot{\mathbf{f}}(\theta_i) - \left(\frac{\partial}{\partial \theta_i} \log \|\mathbf{f}(\theta_i) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) (\mathbf{f}(\theta_i) - \mathbf{c}), \end{aligned}$$

we may write

$$\begin{aligned}
 J_2 &= \frac{1}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \\
 &\times \left[\dot{\mathbf{f}}(\theta_1) \cdot \dot{\mathbf{f}}(\theta_2) - \left\{ \dot{\mathbf{f}}(\theta_1) - \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) (\mathbf{f}(\theta_1) - \mathbf{c}) \right\} \right. \\
 &\quad \left. \cdot \left\{ \dot{\mathbf{f}}(\theta_2) - \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) (\mathbf{f}(\theta_2) - \mathbf{c}) \right\} \right] \\
 &= \frac{1}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \left[\left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left\{ (\mathbf{f}(\theta_1) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_2) \right\} \right. \\
 &\quad + \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left\{ (\mathbf{f}(\theta_2) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_1) \right\} \\
 &\quad - \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \\
 &\quad \left. \times \{ (\mathbf{f}(\theta_1) - \mathbf{c}) \cdot (\mathbf{f}(\theta_2) - \mathbf{c}) \} \right].
 \end{aligned}$$

Using the facts that

$$\begin{aligned}
 (\mathbf{f}(\theta_1) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_2) &= (\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2) + \mathbf{f}(\theta_2) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_2) \\
 &= \frac{1}{2} \frac{\partial}{\partial \theta_2} \left(-\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 + \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right)
 \end{aligned}$$

and

$$\begin{aligned}
 (\mathbf{f}(\theta_2) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_1) &= (\mathbf{f}(\theta_2) - \mathbf{f}(\theta_1) + \mathbf{f}(\theta_1) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_1) \\
 &= \frac{1}{2} \frac{\partial}{\partial \theta_1} \left(-\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 + \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right),
 \end{aligned}$$

we arrive at

$$\begin{aligned}
 &\frac{(\mathbf{f}(\theta_1) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_2)}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \\
 &= -\frac{1}{2} \frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 + \frac{1}{2} \frac{\|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2, \\
 &\frac{(\mathbf{f}(\theta_2) - \mathbf{c}) \cdot \dot{\mathbf{f}}(\theta_1)}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \\
 &= -\frac{1}{2} \frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 + \frac{1}{2} \frac{\|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2.
 \end{aligned}$$

Thus, our previous expression for J_2 becomes

$$\begin{aligned}
 J_2 &= -\frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \\
 &\quad - \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \\
 &\quad + \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \\
 &\quad \times \frac{\|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 + \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 - 2(\mathbf{f}(\theta_1) - \mathbf{c}) \cdot (\mathbf{f}(\theta_2) - \mathbf{c})}{\|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2} \\
 &= -\frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \\
 &\quad - \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \\
 &\quad + \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right).
 \end{aligned}$$

Finally, we have that

$$\frac{\partial}{\partial \theta_i} \log \|\mathbf{p}(\theta_1) - \mathbf{p}(\theta_2)\|_{\mathbb{R}^n}^2 = \frac{\partial}{\partial \theta_i} \left(\log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 - \log \|\mathbf{f}(\theta_i) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right),$$

and therefore

$$\begin{aligned}
 J_3 &= \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \\
 &\quad + \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \\
 &\quad - \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \\
 &= -J_2.
 \end{aligned}$$

□

Corollary 3.4.2. Let $\mathbf{f} \in H^{1,1}(\mathbb{R}/2\pi\mathbb{Z})$. Then it holds that

$$\mathcal{E}_1(\mathbf{f}) + \mathcal{E}_2(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}) + \mathcal{E}_2(\mathbf{p}).$$

Remark 3.4.2. This corollary does not exclude that the case that both sides are infinite.

Proof. First, let $\mathbf{f} \in H^{\frac{3}{2}}(\mathbb{R}/2\pi\mathbb{Z})$. If the 2-dimensional Lebesgue measure of

$$\{(\theta_1, \theta_2) \in (\mathbb{R}/2\pi\mathbb{Z})^2 \mid \mathbf{f}(\theta_1) = \mathbf{f}(\theta_2)\}$$

is positive, then so is that of

$$\{(\theta_1, \theta_2) \in (\mathbb{R}/2\pi\mathbb{Z})^2 \mid \mathbf{p}(\theta_1) = \mathbf{p}(\theta_2)\}.$$

Therefore, $\mathcal{E}(\mathbf{f}) = \mathcal{E}(\mathbf{p}) = \infty$, and we have that

$$\mathcal{E}_1(\mathbf{f}) + \mathcal{E}_2(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}) + \mathcal{E}_2(\mathbf{p}) = \infty.$$

Thus, we assume that the 2-dimensional Lebesgue measure of

$$\{(\theta_1, \theta_2) \in (\mathbb{R}/2\pi\mathbb{Z})^2 \mid \mathbf{f}(\theta_1) = \mathbf{f}(\theta_2)\}$$

is 0, and we claim that the measure of

$$\{(\theta_1, \theta_2) \in (\mathbb{R}/2\pi\mathbb{Z})^2 \mid \mathbf{f}(\theta_1) = \mathbf{f}(\theta_2) \text{ or } \mathbf{f}(\theta_1) = \mathbf{c} \text{ or } \mathbf{f}(\theta_2) = \mathbf{c}\}$$

is also 0. In order to see this, considering \mathbf{f} as a function of s on $\mathbb{R}/\mathcal{L}\mathbb{Z}$, we need to prove that

$$S = \{s \in \mathbb{R}/\mathcal{L}\mathbb{Z} \mid \mathbf{f}(s) = \mathbf{c}\}$$

is a finite set. To argue by contradiction, we suppose that S is not a finite set. Then, by the compactness of $\mathbb{R}/\mathcal{L}\mathbb{Z}$, there exists a sequence such that $\mathbf{f}(s_j) = \mathbf{c}$ and $\lim_{j \rightarrow \infty} s_j = s^*$.

From the fact that

$$\begin{aligned} 0 &= \|\mathbf{f}(s_{j+1}) - \mathbf{f}(s_j)\|_{\mathbb{R}^n}^2 = \int_{s_j}^{s_{j+1}} \int_{s_j}^{s_{j+1}} \boldsymbol{\tau}(s) \cdot \boldsymbol{\tau}(s') ds ds' \\ &= (s_{j+1} - s_j)^2 - \frac{1}{2} \int_{s_j}^{s_{j+1}} \int_{s_j}^{s_{j+1}} \|\boldsymbol{\tau}(s) - \boldsymbol{\tau}(s')\|_{\mathbb{R}^n}^2 ds ds', \end{aligned}$$

it follows that

$$\int_{s_j}^{s_{j+1}} \int_{s_j}^{s_{j+1}} \frac{\|\boldsymbol{\tau}(s) - \boldsymbol{\tau}(s')\|_{\mathbb{R}^n}^2}{(s - s')^2} ds ds' \geq \int_{s_j}^{s_{j+1}} \int_{s_j}^{s_{j+1}} \frac{\|\boldsymbol{\tau}(s) - \boldsymbol{\tau}(s')\|_{\mathbb{R}^n}^2}{(s_{j+1} - s_j)^2} ds ds' = 2.$$

However, considering the fact that $\boldsymbol{\tau} \in H^{\frac{1}{2}}(\mathbb{R}/\mathcal{L}\mathbb{Z})$ and the absolute continuity of integral, this implies that

$$2 \leq \lim_{j \rightarrow \infty} \int_{s_j}^{s_{j+1}} \int_{s_j}^{s_{j+1}} \frac{\|\boldsymbol{\tau}(s) - \boldsymbol{\tau}(s')\|_{\mathbb{R}^n}^2}{(s - s')^2} ds ds' = 0,$$

which is obviously a contradiction. By these arguments, we find that (3.4.8) holds for \mathcal{L}^2 -a.e. (θ_1, θ_2) , and the desired conclusion follows by integrating this expression.

Finally, we consider the case that $\mathbf{f} \notin H^{\frac{3}{2}}(\mathbb{R}/2\pi\mathbb{Z})$, which implies that $\mathcal{E}(\mathbf{f}) = \infty$. Then, we will show that $\mathcal{E}(\mathbf{p}) = \infty$. Again, arguing by contradiction, we suppose that $\mathcal{E}(\mathbf{p}) < \infty$ from which we know that \mathbf{p} does not have self-intersections. Furthermore, it holds that $\mathbf{p} \in H_{\text{loc}}^{\frac{3}{2}}$, and we remark that $\mathbf{p} \in H^{\frac{3}{2}}$ if \mathbf{p} does not pass through the point

at infinity. If we turn \mathbf{p} back by inversion with respect to the sphere with center \mathbf{c} and radius r , then it returns to \mathbf{f} . Because \mathbf{f} does not pass through the point at infinity, \mathbf{p} does not pass through \mathbf{c} . Therefore the 2-dimensional Lebesgue measure of

$$\{(\theta_1, \theta_2) \mid \mathbf{p}(\theta_1) = \mathbf{p}(\theta_2), \text{ or } \mathbf{p}(\theta_1) = \mathbf{c} \text{ or } \mathbf{p}(\theta_2) = \mathbf{c}\}$$

is 0. This implies that (3.4.8) holds for \mathcal{L}^2 -a.e. (θ_1, θ_2) , and by integrating this, we obtain that

$$\mathcal{E}_1(\mathbf{f}) + \mathcal{E}_2(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}) + \mathcal{E}_2(\mathbf{p}) = \mathcal{E}(\mathbf{p}) - 4 < \infty.$$

However, from this we obtain that $\infty = \mathcal{E}(\mathbf{f}) = \mathcal{E}_1(\mathbf{f}) + \mathcal{E}_2(\mathbf{f}) + 4 < \infty$, which is obviously a contradiction. We can conclude that it holds that $\mathcal{E}(\mathbf{p}) = \infty$, and so $\mathcal{E}_1(\mathbf{f}) + \mathcal{E}_2(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}) + \mathcal{E}_2(\mathbf{p}) = \infty$, as desired. \square

We discuss the invariance of each energy \mathcal{E}_i under the inversion $\mathbf{f} \mapsto \mathbf{p}$.

Theorem 3.4.7. Assume that the center \mathbf{c} of the inversion is not in the image of \mathbf{f} . We also assume that the energy $\mathcal{E}(\mathbf{f})$ is finite. Then, we have that

$$\mathcal{E}_1(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}), \quad \mathcal{E}_2(\mathbf{f}) = \mathcal{E}_2(\mathbf{p}).$$

Proof. In view of Corollary 3.4.2, it is sufficient to prove that $\mathcal{E}_1(\mathbf{f}) = \mathcal{E}_1(\mathbf{p})$. Let J_1 and J_2 be defined as in the proof of Theorem 3.4.6. It follows that

$$\mathcal{M}_1(\mathbf{f}) \|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n} - \mathcal{M}_1(\mathbf{p}) \|\dot{\mathbf{p}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{p}}(\theta_2)\|_{\mathbb{R}^n} = J_1 - J_2 = -J_2.$$

We need to prove that the integration of this goes to 0. From this point on, we use the arc-length variable s_j .

Because we are assuming that $\mathcal{E}(\mathbf{f}) < \infty$, we know that it holds that $\mathcal{M}_1(\mathbf{f}) \in L^1((\mathbb{R}/\mathcal{L}\mathbb{Z})^2)$. Because $\mathcal{M}_1(\mathbf{p}) \geq 0$, we may write

$$\mathcal{E}_1(\mathbf{f}) - \mathcal{E}_1(\mathbf{p}) = - \lim_{\varepsilon \rightarrow +0} \iint_{|s_1 - s_2| \geq \varepsilon} \frac{J_2}{\|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n}} ds_1 ds_2,$$

where $|s_1 - s_2| \geq \varepsilon$ is in the sense of mod \mathcal{L} . We remark that

$$\frac{1}{\|\dot{\mathbf{f}}(\theta_j)\|_{\mathbb{R}^n}} \frac{\partial}{\partial \theta_j}(\cdots) = \frac{\partial}{\partial s_j}(\cdots),$$

so we obtain that

$$\begin{aligned}
& - \lim_{\varepsilon \rightarrow +0} \iint_{|s_1 - s_2| \geq \varepsilon} \frac{J_2}{\|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n}} ds_1 ds_2 \\
&= \lim_{\varepsilon \rightarrow +0} \frac{1}{2} \iint_{|s_1 - s_2| \geq \varepsilon} \left\{ \left(\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 \right) \right. \\
&\quad + \left(\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \\
&\quad \left. - \left(\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(s_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right) \right\} ds_1 ds_2 \\
&= \lim_{\varepsilon \rightarrow +0} \frac{1}{2} \int_0^{\mathcal{L}} \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \int_{s_1 + \varepsilon}^{s_1 + \mathcal{L} - \varepsilon} \frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 ds_2 ds_1 \\
&\quad + \lim_{\varepsilon \rightarrow +0} \frac{1}{2} \int_0^{\mathcal{L}} \frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \int_{s_2 + \varepsilon}^{s_2 + \mathcal{L} - \varepsilon} \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 ds_1 ds_2 \\
&\quad - \frac{1}{2} \int_0^{\mathcal{L}} \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 ds_1 \int_0^{\mathcal{L}} \frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(s_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 ds_2.
\end{aligned}$$

Furthermore, it holds that

$$\int_{s_1 + \varepsilon}^{s_1 + \mathcal{L} - \varepsilon} \frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 ds_2 = \log \frac{\|\mathbf{f}(s_1) - \mathbf{f}(s_1 - \varepsilon)\|_{\mathbb{R}^n}^2}{\|\mathbf{f}(s_1) - \mathbf{f}(s_1 + \varepsilon)\|_{\mathbb{R}^n}^2}.$$

Moreover, the fact that $\mathcal{E}(\mathbf{f}) < \infty$ implies that $\mathbf{f} \in H^{\frac{3}{2}}(\mathbb{R}/\mathcal{L}\mathbb{Z})$, and thus

$$\|\mathbf{f}(s_1) - \mathbf{f}(s_1 \pm \varepsilon)\|_{\mathbb{R}^n}^2 = \varepsilon^2 + o(\varepsilon^2)$$

uniformly with regard to s_1 as $\varepsilon \rightarrow +0$. Therefore, we have that

$$\int_{s_1 + \varepsilon}^{s_1 + \mathcal{L} - \varepsilon} \frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 ds_2 = \log(1 + o(1)) = o(1)$$

holds uniformly with regard to s_1 . From this, and the fact that \mathbf{f} does not pass through \mathbf{c} , it holds that

$$\begin{aligned}
& \left| \int_0^{\mathcal{L}} \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \int_{s_1 + \varepsilon}^{s_1 + \mathcal{L} - \varepsilon} \frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 ds_2 ds_1 \right| \\
&= o(1) \int_0^{\mathcal{L}} \left| \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \right| ds_1 \\
&= o(1).
\end{aligned}$$

Then, we arrive at

$$\lim_{\varepsilon \rightarrow +0} \int_0^{\mathcal{L}} \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 \int_{s_1+\varepsilon}^{s_1+\mathcal{L}-\varepsilon} \frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 ds_2 ds_1 = 0.$$

Similarly, we find that

$$\lim_{\varepsilon \rightarrow +0} \int_0^{\mathcal{L}} \frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 \int_{s_2+\varepsilon}^{s_2+\mathcal{L}-\varepsilon} \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 ds_1 ds_2 = 0.$$

Finally, because \mathbf{f} does not pass through \mathbf{c} , we have that

$$\int_0^{\mathcal{L}} \frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{c}\|_{\mathbb{R}^n}^2 ds_1 \int_0^{\mathcal{L}} \frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(s_2) - \mathbf{c}\|_{\mathbb{R}^n}^2 ds_2 = 0$$

from which the desired conclusion $\mathcal{E}_1(\mathbf{f}) = \mathcal{E}(\mathbf{p})$ follows. \square

Assume that $\text{Im } \mathbf{f} = S^1$ and $\mathbf{c} \in \text{Im } \mathbf{f}$. Then, \mathbf{p} is a straight line. Therefore $\mathcal{E}_1(\mathbf{f}) = 2\pi$ and $\mathcal{E}_1(\mathbf{p}) = 0$. This shows that $\mathcal{E}_1(\mathbf{f}) = \mathcal{E}_1(\mathbf{p})$ does not hold unless $\mathbf{c} \notin \text{Im } \mathbf{f}$. In fact, we have the following result.

Theorem 3.4.8 ([12]). Let $\mathbf{f} \in C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$. Then, for an inversion with center \mathbf{c} on $\text{Im } \mathbf{f}$, we have that

$$\mathcal{E}_1(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}) + 2\pi^2, \quad \mathcal{E}_2(\mathbf{f}) = \mathcal{E}_2(\mathbf{p}) - 2\pi^2.$$

Proof. We only provide a sketch of proof here. We refer the reader interested in the details to [12]. Let θ be a general parameter, as above. Under our assumption, \mathbf{f} has no self-intersections, and so \mathbf{f} passes through the center \mathbf{c} only once at most per period. Thus, we may assume that $\mathbf{c} = \mathbf{f}(0)$. Then, if $\theta \neq 0 \pmod{2\pi}$, it holds that $\mathbf{f}(\theta) \neq \mathbf{c}$. As shown in [10], we have that

$$\mathcal{M}_1(\mathbf{f}) \|\dot{\mathbf{f}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{f}}(\theta_2)\|_{\mathbb{R}^n} - \mathcal{M}_1(\mathbf{p}) \|\dot{\mathbf{p}}(\theta_1)\|_{\mathbb{R}^n} \|\dot{\mathbf{p}}(\theta_2)\|_{\mathbb{R}^n} = \frac{1}{2} J(\theta_1, \theta_2), \quad (3.4.9)$$

where

$$\begin{aligned} \frac{1}{2} J(\theta_1, \theta_2) = & \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \\ & + \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(\theta_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right) \\ & - \left(\frac{\partial}{\partial \theta_1} \log \|\mathbf{f}(\theta_1) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial \theta_2} \log \|\mathbf{f}(\theta_2) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right). \end{aligned}$$

Because

$$\frac{\partial}{\partial \theta_i}(\cdots) d\theta_i = \frac{\partial}{\partial s_i}(\cdots) ds_i,$$

it is easy to see that

$$\iint_{(\mathbb{R}/2\pi\mathbb{Z})^2} J(\theta_1, \theta_2) d\theta_1 d\theta_2 = \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \mathcal{J}(s_1, s_2) ds_1 ds_2, \quad (3.4.10)$$

where

$$\begin{aligned} \mathcal{J}(s_1, s_2) = & \left(\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 \right) \\ & + \left(\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(s_2)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_2) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right) \\ & - \left(\frac{\partial}{\partial s_1} \log \|\mathbf{f}(s_1) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right) \left(\frac{\partial}{\partial s_2} \log \|\mathbf{f}(s_2) - \mathbf{f}(0)\|_{\mathbb{R}^n}^2 \right) \end{aligned}$$

Here, we set $s = 0$ at $\theta = 0$. We introduce a function \mathcal{F} replacing the Euclidean distance $\|\cdot\|_{\mathbb{R}^n}$ in \mathcal{J} with the intrinsic distance \mathcal{D} :

$$\begin{aligned} \mathcal{F}(s_1, s_2) = & \left(\frac{\partial}{\partial s_1} \log \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(0))^2 \right) \left(\frac{\partial}{\partial s_2} \log \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2 \right) \\ & + \left(\frac{\partial}{\partial s_1} \log \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2 \right) \left(\frac{\partial}{\partial s_2} \log \mathcal{D}(\mathbf{f}(s_2), \mathbf{f}(0))^2 \right) \\ & - \left(\frac{\partial}{\partial s_1} \log \mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(0))^2 \right) \left(\frac{\partial}{\partial s_2} \log \mathcal{D}(\mathbf{f}(s_2), \mathbf{f}(0))^2 \right). \end{aligned}$$

We can show that

$$\iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \mathcal{F}(s_1, s_2) ds_1 ds_2 = 4\pi^2 \quad (3.4.11)$$

and

$$\iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} (\mathcal{J}(s_1, s_2) - \mathcal{F}(s_1, s_2)) ds_1 ds_2 = 0. \quad (3.4.12)$$

The assertion of Theorem easily follows from (3.4.9)–(3.4.12). There, we have used the existence of a bounded κ almost everywhere, so we must assume that $\mathbf{f} \in C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$. \square

Corollary 3.4.3. Right circles are the only global minimizers of \mathcal{E}_1 in the class $C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$.

Proof. Because $\mathcal{M}_1(\mathbf{p}) \geq 0$, it is clear that $\mathcal{E}_1(\mathbf{p}) \geq 0$. Hence, $\mathcal{E}_1(\mathbf{f}) = \mathcal{E}_1(\mathbf{p}) + 2\pi^2 \geq 2\pi^2$. Furthermore, $\mathcal{E}_1(\mathbf{f}) = 2\pi^2$ holds if and only if $\mathcal{E}_1(\mathbf{p}) = 0$, which is equivalent to $\mathcal{M}_1(\mathbf{p}) \equiv 0$. This implies that $\mathbf{p}'(s)$ is a constant vector. That is, \mathbf{p} is a straight line. Consequently, the pre-image \mathbf{f} is a right circle. \square

As stated above, the lower bound of \mathcal{E}_1 in $C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$ is $2\pi^2$. This estimate seems to be related to the first eigenvalue of $(-\Delta_s)^{3/2}$, which is the principal part of the L^2 -gradient, is $\left(\frac{2\pi}{\mathcal{L}}\right)^3$. Indeed, by the bi-Lipschitz continuity of \mathbf{f} , we have that

$$\begin{aligned}\mathcal{E}_1(\mathbf{f}) &= \frac{1}{2} \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \frac{\|\Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2}{\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} ds_1 ds_2 \simeq \frac{1}{2} \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \frac{\|\Delta \mathbf{f}'\|_{\mathbb{R}^n}^2}{|\Delta s|^2} ds_1 ds_2 \\ &\simeq \pi \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \|(-\Delta_s)^{1/4} \mathbf{f}'\|_{\mathbb{R}^n}^2 ds = \pi \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \|(-\Delta_s)^{3/4} \mathbf{f}\|_{\mathbb{R}^n}^2 ds \\ &= \pi \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \mathbf{f} \cdot (-\Delta_s)^{3/2} \mathbf{f} ds \geq \pi \left(\frac{2\pi}{\mathcal{L}}\right)^3 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \|\mathbf{f}\|_{\mathbb{R}^n}^2 ds.\end{aligned}$$

The equality in the above estimate implies that each component of \mathbf{f} is the first eigenfunction. This means that \mathbf{f} is a right circle with the center at the origin and perimeter \mathcal{L} . Therefore,

$$\|\mathbf{f}\|_{\mathbb{R}^n} \equiv \frac{\mathcal{L}}{2\pi}.$$

Consequently, we obtain that

$$\pi \left(\frac{2\pi}{\mathcal{L}}\right)^3 \int_{\mathbb{R}/\mathcal{L}\mathbb{Z}} \|\mathbf{f}\|_{\mathbb{R}^n}^2 ds = \pi \left(\frac{2\pi}{\mathcal{L}}\right)^3 \cdot \mathcal{L} \cdot \left(\frac{\mathcal{L}}{2\pi}\right)^2 = 2\pi^2$$

The above is not a rigorous argument, but seems to the author to be of interest.

It is still an open problem whether Theorem 3.4.8 holds for $\mathbf{f} \in H^{3/2}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathcal{L}\mathbb{Z})$. However we know a (non-sharp) lower estimate of \mathcal{E}_1 .

Proposition 3.4.2. We have that

$$\mathcal{E}_1(\mathbf{f}) \geq 2\pi \int_{-\pi}^{\pi} \frac{1 - \cos u}{u^2} du$$

for any $\mathbf{f} \in H^{3/2}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathcal{L}\mathbb{Z})$.

Remark 3.4.3. Note that the estimate for curves of $C^{1,1}(\mathbb{R}/\mathcal{L}\mathbb{Z})$ can be written as

$$\mathcal{E}_1(\mathbf{f}) \geq 2\pi^2 = 2\pi \int_{-\infty}^{\infty} \frac{1 - \cos u}{u^2} du.$$

Proof. Because the Euclidean distance cannot be longer than the intrinsic distance, we have that

$$\mathcal{E}_1(\mathbf{f}) = \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \frac{\|\Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2}{2\|\Delta \mathbf{f}\|_{\mathbb{R}^n}^2} ds_1 ds_2 \geq \iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \frac{\|\Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2}{2\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2} ds_1 ds_2.$$

We insert the Fourier series

$$\mathbf{f} = \sum_{k \in \mathbb{Z}} \varphi_k \mathbf{a}_k$$

into the above. In a similar manner to derive the L^2 -gradient, we can obtain

$$\iint_{(\mathbb{R}/\mathcal{L}\mathbb{Z})^2} \frac{\|\Delta \boldsymbol{\tau}\|_{\mathbb{R}^n}^2}{2\mathcal{D}(\mathbf{f}(s_1), \mathbf{f}(s_2))^2} ds_1 ds_2 = \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^3 \|\mathbf{a}_k\|_{\mathbb{C}^n}^2 \int_{-\pi|k|}^{\pi|k|} \frac{1 - \cos u}{u^2} du.$$

By applying Parseval's identity to

$$\boldsymbol{\tau} = \sum_{k \in \mathbb{Z}} \frac{2\pi i k}{\mathcal{L}} \varphi_k \mathbf{a}_k,$$

we obtain that

$$\|\boldsymbol{\tau}\|_{L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})}^2 = \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^2 \|\mathbf{a}_k\|_{\mathbb{C}^n}^2.$$

On the other hand, because $\boldsymbol{\tau}$ is a unit vector, it holds that

$$\|\boldsymbol{\tau}\|_{L^2(\mathbb{R}/\mathcal{L}\mathbb{Z})}^2 = \mathcal{L}.$$

Therefore it holds that

$$\sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^2 \|\mathbf{a}_k\|_{\mathbb{C}^n}^2 = \mathcal{L}.$$

Consequently, we arrive at

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^3 \|\mathbf{a}_k\|_{\mathbb{C}^n}^2 \int_{-\pi|k|}^{\pi|k|} \frac{1 - \cos u}{u^2} du &\geq \frac{2\pi}{\mathcal{L}} \int_{-\pi}^{\pi} \frac{1 - \cos u}{u^2} du \sum_{k \in \mathbb{Z}} \left| \frac{2\pi k}{\mathcal{L}} \right|^2 \|\mathbf{a}_k\|_{\mathbb{C}^n}^2 \\ &= 2\pi \int_{-\pi}^{\pi} \frac{1 - \cos u}{u^2} du. \end{aligned}$$

□

From the above proof, we can see that the given estimate is not sharp. The author expects that the lower bound of ε_1 in $H^{3/2}(\mathbb{R}/\mathcal{L}\mathbb{Z}) \cap W^{1,\infty}(\mathbb{R}/\mathcal{L}\mathbb{Z})$ is also $2\pi^2$, and that right circles are the only minimizers. Blatt [3] helpfully informed the author that the answer may be affirmatively.

Bibliography

- [1] A. Abrams, J. Cantarella, J. H. G. Fu, M. Ghomi & R. Howard, *Circles minimize most knot energies*, Topology **42** (2) (2003), 381–394.

- [2] S. Blatt, *Boundedness and regularizing effects of O'Hara's knot energies*, J. Knot Theory Ramifications **21** (2012), 1250010, 9 pp.
- [3] S. Blatt, private communication, 2015.
- [4] S. Blatt, Ph. Reiter & A. Schikorra, *Harmonic analysis meets critical knots. Critical points of the Möbius energy are smooth*, Trans. Amer. Math. Soc. **368** (9) (2016), 6391–6438.
- [5] M. H. Freedman, Z.-X. He & Z. Wang, *Möbius energy of knots and unknots*, Ann. of Math. (2) **139** (1) (1994), 1–50.
- [6] S. Fukuhara, *Energy of a knot* in “A fête of Topology”, Ed.: Y. Matsumoto, T. Mizutani & S. Morita, Academic Press, Boston, 1988, pp.443–451.
- [7] Z.-X. He, *The Euler-Lagrange equation and heat flow for the Möbius energy*, Comm. Pure Appl. Math. **53** (4) (2000), 399–431.
- [8] A. Ishizeki, “Some Properties of the Möbius Energy and its Variational Formulae”, Master Thesis, Saitama University, 2013.
- [9] A. Ishizeki, “Decomposition of the Möbius energy: the Möbius Invariance and Variational Formulae of Decomposed Energies”, Doctor Thesis, Saitama University, 2016.
- [10] A. Ishizeki & T. Nagasawa, *A decomposition theorem of the Möbius energy I: Decomposition and Möbius invariance*, Kodai Math. J. **37** (3) (2014), 737–754.
- [11] A. Ishizeki & T. Nagasawa, *A decomposition theorem of the Möbius energy II: Variational formulae and estimates*, Math. Ann. **363** (1–2) (2015), 617–635.
- [12] A. Ishizeki & T. Nagasawa, *The invariance of decomposed Möbius energies under inversions with center on curves*, J. Knot Theory Ramifications **26** (2016), 1650009, 22 pp.
- [13] A. Ishizeki & T. Nagasawa, *The L^2 -gradient of decomposed Möbius energies*, Calc. Var. Partial Differential Equations **55** (3) (2016), Art. No.56, 31 pp.
- [14] R. Kusner & J. M. Sullivan, *On distortion and thickness of knots*, in “Ideal Knots” (Ed.: A. Stasiak, V. Katrich, L. H. Kauffman), World Scientific, Singapore, 1998, pp. 315–352.
- [15] S. Mizohata, “The Theory of Partial Differential Equations”, Cambridge University Press, New York, 1973 (transl. from Japanese Ed.: Iwanami, Tokyo, 1965).
- [16] T. Nagasawa, *Analytic approaches to the Möbius energy: History and recent topics*, RIMS Kôkyûroku **1979** (2015), 177–193.
- [17] J. O'Hara, *Energy of a knot*, Topology **30** (2) (1991), 241–247.
- [18] J. O'Hara, *Family of energy functionals of knots*, Topology Appl. **48** (2) (1992), 147–161.
- [19] J. O'Hara, *Energy functionals of knots II*, Topology Appl. **56** (1) (1994), 45–61.
- [20] J. O'Hara & G. Solanes, *Möbius invariant energies and average linking with circles*, Tôhoku Math. J. (2) **67** (1) (2015), 51–82.
- [21] Ph. Reiter, *Regularity theory for the Möbius energy*, Comm. Pure Appl. Anal. **9** (5) (2010), 1463–1471.

Pseudogradient Flows of Geometric Energies

Abstract: For differentiable functions on Riemannian manifolds, the gradient vector field and its induced gradient flow are well-defined and well-understood concepts. Unfortunately, many nonquadratic, infinite-dimensional optimization problems cannot be formulated on Riemannian manifolds in a convenient way. We introduce a category of infinite-dimensional Banach manifolds that allows us to define a generalized gradient. Moreover, we show short-time existence of the induced flows and apply their discretizations to the numerical minimization of various geometric energies of immersed curves and surfaces.

Keywords: gradient flow, minimal surfaces, Euler elastica, Willmore energy

MSC: 49Q05, 49Q10

4.1 Introduction

Gradient flows are a popular tool for minimizing (or maximizing) continuously differentiable functions, both theoretically and numerically. Algorithms derived from gradient flows, often subsumed as methods of steepest descent, belong to virtually every library for (smooth) numerical optimization. The idea behind gradient flows is very intuitive and appealing:

Let (M, g) be a Riemannian manifold, $F: M \rightarrow \mathbb{R}$ be a function of class C^1 , and $a \in M$ some point. A tangent vector $u \in T_a M$ of length $r > 0$ is called a *direction of steepest descent of F at the point $a \in M$ of length r* if it minimizes the slope $S(u) := R(u)^{-1} \langle dF|_a, u \rangle$ among all tangent vectors of length r , where $R(u) := g(u, u)^{\frac{1}{2}}$. That means, u is a solution of the following constraint optimization problem:

$$\text{Minimize } S(u) \text{ over } u \in T_a M \text{ subject to the constraint } R(u) = r. \quad (4.1.1)$$

By introducing a scalar Lagrange multiplier $\lambda \in \mathbb{R}$, we have to solve the system

$$\begin{cases} \langle dS|_u, v \rangle + \lambda \langle dR|_u, v \rangle = 0 & \text{for all } v \in T_a M, \\ R(u) = r. \end{cases}$$

This system is equivalent to

$$\begin{cases} r^2 \langle dF|_a, v \rangle - \langle dF|_a, u \rangle g(u, v) + \lambda r^2 g(u, v) = 0 & \text{for all } v \in T_a M, \\ g(u, u) = r^2. \end{cases}$$

Testing with $v = u$ leads to $\lambda = 0$. Afterwards, testing with all $v \in \ker(dF|_a)$ leads to $u \in \ker(dF|_a)^\perp$. Thus, the system is solved by $u = \sigma \operatorname{grad}(F)|_a$ with $\sigma \in$

$\{-r|dF|_a|_g^{-1}, r|dF|_a|_g^{-1}\}$. Here, the *gradient* $\text{grad}(F)|_a \in T_aM$ is the vector defined by the equation

$$g(\text{grad}(F)|_a, v) := \langle dF|_a, v \rangle \quad \text{for all } v \in T_aM. \quad (4.1.2)$$

Moreover, $|dF|_a|_g$ is the Frobenius norm of the linear form $dF|_a: T_aM \rightarrow \mathbb{R}$ with respect to the inner product g . Summarized, the direction of steepest descent of length r is given by

$$u = -r|dF|_a|_g^{-1} \text{grad}(F)|_a.$$

Usually, one chooses $r = |dF|_a|_g$ and defines the *direction of steepest descent* by

$$u = -\text{grad}(F)|_a.$$

We point out a simple but often forgotten fact: One needs a metric d on T_aM in order to be able to talk about the slope of a linear function $\xi: T_aM \rightarrow \mathbb{R}$ in direction u , the slope being given by

$$\frac{\langle \xi, u \rangle - \langle \xi, 0 \rangle}{d(u, 0)}.$$

In particular, the notion of steepest descent requires the specification of a metric, at least on tangent spaces; different metrics will lead to different directions of steepest descent. *There is no such thing as a direction of steepest descent per se.*

A flow $\Phi: U \times I \subset M \times \mathbb{R} \rightarrow M$ is called a *downward gradient flow* of F if it is generated by the vector field $-\text{grad}(F)$, i.e., if it satisfies

$$\frac{d}{dt}\Phi(a, t) = -\text{grad}(F)|_{\Phi(a, t)} \quad \text{and} \quad \Phi(a, 0) = a \quad \text{for all } (a, t) \in U \times I. \quad (4.1.3)$$

Analogously, an *upward gradient flow* of F is generated by $\text{grad}(F)$. Under mild regularity assumptions on F (e.g., dF is locally Lipschitz continuous), both the downward and upward flow exist at least locally and for short times. Moreover, each of the flows is unique.

This theory carries over to infinite-dimensional manifolds as far as only Riemannian/Hilbert manifolds are considered. Unfortunately, many interesting variational problems can only be formulated as *differentiable* variational problems on Hilbert manifolds if one shrinks the feasible set to an extent that minimizers of the actual problem of interest might be excluded. Many geometric variational problems involving nonquadratic energies and nonlinear constraints belong to this class; we discuss several examples in Section 4.5.1 and Section 4.5.2. The essential reason for this phenomenon is that the Sobolev spaces $W^{s,2}(\mathbb{R}^n; \mathbb{R})$ do not embed continuously into $W^{1,\infty}(\mathbb{R}^n; \mathbb{R})$ for moderate $n \geq 2$ and small $s \in [0, 1 + \frac{n}{2}]$. Thus, a generic continuous functional on $W^{1,\infty}(\mathbb{R}^n; \mathbb{R})$ can only be made continuous on $W^{s,2}(\mathbb{R}^n; \mathbb{R})$ by choosing s overly large.

This is why the generalization of gradients and gradient flows to more general infinite-dimensional manifolds is a quite delicate business. As far as we know, there have been the following two main approaches so far:

1. One may generalize gradients as directions of steepest descent in the sense of (4.1.1), using $R(u) = \|u\|_{T_a M}$ and $r = \|dF|_a\|_{(T_a M)^*}$. If the Banach space $T_a M$ is reflexive, this leads again to a well-defined gradient $\text{grad}(F)|_a \in T_a M$. However, this gradient depends nonlinearly on $dF|_a$. In practice, this means that a costly minimization or root finding algorithm has to be used in order to determine $\text{grad}(F)|_a$. Moreover, even more costly algorithms have to be applied if the norm $\|\cdot\|_{T_a M}$ is not differentiable on $T_a M \setminus \{0\}$. For more details on this kind of gradient flows and for generalizations to more general metric spaces, we refer the reader to [1].
2. In some cases, one may exploit a *Gelfand triple* $T_a M \hookrightarrow H_a \hookrightarrow (T_a M)'$, i.e., a sequence of continuous, linear, and dense embeddings, where (H_a, g_a) is a Hilbert space. Then one may use (4.1.2) to define a “gradient” $\text{grad}^H(F)|_a \in H_a$. Since $\text{grad}^H(F)|_a \notin T_a M$ is not a tangent vector, the mapping $a \mapsto \text{grad}^H(F)|_a$ cannot be interpreted as a vector field so that (4.1.3) is no longer an ordinary differential equation. Under certain conditions on dF , (4.1.3) may still be formulated as a parabolic partial differential equation; existence and uniqueness of the “gradient flow” can be ensured for short *positive* times $t \in]0, \varepsilon[$. Prominent examples are the *mean curvature flow* (the L^2 -gradient flow of the surface area functional, see [5]) and the *Willmore flow* (the L^2 -gradient flow of the Willmore functional). An extensive overview and further literature on these flows can be found in [7].

In Section 4.4, we propose another way to define (pseudo)gradients for certain classes of Banach manifolds. In contrast to the first of the two approaches mentioned above, the pseudogradient $\text{grad}(F)$ depends linearly on dF . In contrast to the second approach, these pseudogradients form indeed a vector field $\text{grad}(F)$ and this vector field has the same regularity as dF . If F is of class $C_{\text{loc}}^{1,1}$, then the pseudogradient flow defined by (4.1.3) exists locally and for small times, both in forward and in backward direction (see 4.4.2). Of course, one cannot expect to obtain such a strong result for arbitrary Banach manifolds and arbitrary functions. It turns out that the suitable category of manifolds to consider is the category of *Riesz manifolds* which we introduce in Section 4.3. We will also see there that the concept of Riesz manifolds is a bit richer than the concept of Riemannian manifolds.

We aim at treating pseudogradient flows for quite general classes of infinite-dimensional Banach manifolds—and not only for open sets in Banach spaces. Such situations arise, e.g., when treating optimization problems with nonlinear constraints. The class of examples we have primarily in mind is the minimization of knot energies among knots in parameterization by arclength. The treatment of infinite-dimensional Banach manifolds will inevitably involve certain infinite-dimensional vector bundles so that we require the notion of a Banach bundle. This is why we give a very brief introduction to fiber bundles in general and to Banach bundles in particular (see Section 4.2).

Section 4.5 is devoted to applications of the developed theory. We consider volume functionals (Section 4.5.1) and certain functionals depending on extrinsic cur-

vature (Section 4.5.2) and show that these functionals together with suitably chosen domains belong to the category of Riesz manifolds (4.5.1 and 4.5.5). As a side effect, we verify that the category of Riesz manifolds (i) is strictly larger than the category of Riemannian manifolds and (ii) contains many examples of practical relevance. Moreover, we give several numerical examples of discretized pseudogradient descent for the minimization of these energies subject to various constraints. In particular, these examples will demonstrate the efficiency and robustness of pseudogradient methods when applied to numerical minimization of geometric energies subject to nonlinear equality constraints.

4.2 Banach Bundles

Fiber bundles of various types and in particular vector bundles are typical working horses in algebraic topology, differential geometry, and global analysis. Since the reader might not be familiar with the notion of a fiber bundle, we give a brief introduction, being well aware that a thorough treatment of this matter may easily fill a full-semester master course. A beautiful and short introduction to Banach manifolds and vector bundles can be found in [13]. Moreover, we refer to [12], maybe *the* classical text on fiber bundles, and to [14], a rather modern exposition which also covers great parts of the theory of partial differential equations for fields with values in fiber bundles.

4.2.1 General Fiber Bundles

The central idea of a fiber bundle is that it looks locally like a product of a base space and a fiber. The following definitions make this more precise.

Definition 4.2.1. Let M, E and X be topological spaces, let $\pi: E \rightarrow M$ be a continuous, surjective map, and let $U \subset M$ be an open set. We write $E|U := \pi^{-1}(U) \subset E$. A *local trivialization of $\pi: E \rightarrow M$ on U with typical fiber X* is a homeomorphism $\varphi: E|U \rightarrow U \times X$ such that the following diagram commutes

$$\begin{array}{ccc} E|U & \xrightarrow{\varphi} & U \times X \\ & \searrow \pi|_{E|U} & \swarrow \text{pr} \\ & U & \end{array}$$

Here, $\text{pr}: U \times X \rightarrow U$, $(a, x) \mapsto a$ is the projection onto the first factor.

Definition 4.2.2. Let M, E and X be topological spaces and let $\pi: E \rightarrow M$ be a continuous, surjective map. A *fiber bundle atlas on $\pi: E \rightarrow M$ with typical fiber X* is a

covering family of local trivializations $(\varphi_i: E|U_i \rightarrow U_i \times X)_{i \in I}$ in the sense that $(U_i)_{i \in I}$ is an open covering of M .

Two fiber bundle atlantes with typical fiber X are called *compatible* if their union is a fiber bundle atlas. A fiber bundle atlas is called *maximal* if it contains each compatible fiber bundle atlas.

A *fiber bundle with typical fiber X* is a continuous and surjective mapping $\pi: E \rightarrow M$ together with a maximal fiber bundle atlas with typical fiber X . With a slight abuse of notation, one usually omits mentioning the precise atlas and simply refers to $\pi: E \rightarrow M$ or even to E as the fiber bundle. One refers to E as the *total space*, to M as the *base space*, and to π as the *bundle projection* of the fiber bundle $\pi: E \rightarrow M$. For $a \in M$, one calls $E_a := \pi^{-1}(\{a\})$ the *fiber over a* . Note that E_a is homeomorphic to X .

For each pair $i, j \in I$ with $U_{ij} := U_i \cap U_j \neq \emptyset$, the following commutative diagram defines a homeomorphism $\Phi_{ij}: U_{ij} \times X \rightarrow U_{ij} \times X$, the so-called *transition map*:

$$\begin{array}{ccccc}
 & & \Phi_{ij} & & \\
 & \swarrow & & \searrow & \\
 U_{ij} \times X & \xleftarrow{\varphi_j|_{E|U_{ij}}} & E|U_{ij} & \xrightarrow{\varphi_i|_{E|U_{ij}}} & U_{ij} \times X \\
 & \searrow \text{pr}_j & \downarrow \pi & \swarrow \text{pr}_i & \\
 & & U_{ij} & &
 \end{array} \tag{4.2.1}$$

Note that one has $\Phi_{ij} = (\Phi_{ji})^{-1}$ and $\Phi_{ij} \circ \Phi_{jk} = \Phi_{ik}$.

Let $R \in \{C_{\text{loc}}^{k,\alpha} \mid k \in \mathbb{N} \cup \{0, \infty\}, \alpha \in [0, 1] \text{ with } k + \alpha \geq 1\}$. If M and X are Banach manifolds of class R and if each transition Φ^{ij} is of class R , we say that $\pi: E \rightarrow M$ is a *fiber bundle of class R* . Note that in this case, the total space E is a manifold of class R and π is a submersion of class R .

Example 4.2.3. Let M and X be topological spaces. Then $E := M \times X$ together with the projection $\pi: E \rightarrow M$, $(a, f) \mapsto a$ is a fiber bundle with typical fiber X . Such fiber bundles are called *trivial*, since there exists a global trivialization. By definition, every fiber bundle restricted to a sufficiently small set is a trivial fiber bundle.

Example 4.2.4. Let $E \subset \mathbb{R}^3$ be the embedded Möbius strip depicted in Figure 4.1. Let $M \subset E$ be the centerline of the Möbius strip and $\pi: E \rightarrow M$ be the shortest distance projection with respect to the intrinsic distance of E . Then $\pi: E \rightarrow M$ is a *nontrivial* smooth fiber bundle with a closed interval as typical fiber.

Example 4.2.5. Let $\pi_1: E_1 \rightarrow M$ and $\pi_2: E_2 \rightarrow M$ be fiber bundles with typical fibers X_1 and X_2 respectively. Then the *product bundle* $\pi: E_1 \times_M E_2 \rightarrow M$ defined by

$$\begin{aligned}
 E_1 \times_M E_2 &:= \{(e_1, e_2) \in E_1 \times E_2 \mid \pi_1(e_1) = \pi_2(e_2)\} \quad \text{and} \\
 \pi: E_1 \times_M E_2 &\rightarrow M, \quad \pi(e_1, e_2) := \pi_1(e_1) = \pi_2(e_2).
 \end{aligned}$$

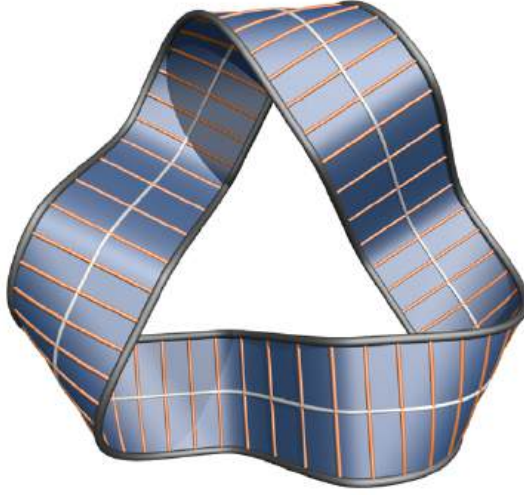


Fig. 4.1: An embedded Möbius strip as nontrivial fiber bundle over its centerline (white). Several fibers are depicted in orange.

is a fiber bundle with typical fiber $X_1 \times X_2$.

Definition 4.2.6. Let $\pi: E \rightarrow M$ be a fiber bundle with typical fiber X . A mapping $\sigma: M \rightarrow E$ is called a *section* of the bundle E if it satisfies $\pi \circ \sigma = \text{id}_M$, i.e., σ maps each point $a \in M$ into its fiber E_a . If both $\pi: E \rightarrow M$ and $\sigma: M \rightarrow E$ are of class R , we say that σ is a *section of class R* .

Definition 4.2.7. Let $\pi_1: E_1 \rightarrow M_1$ and $\pi_2: E_2 \rightarrow M_2$ be fiber bundles and let $F: E_1 \rightarrow E_2$ and $f: M_1 \rightarrow M_2$ be continuous maps. We say that F is a *fiber bundle morphism from E_1 to E_2 over f* if the following diagram is commutative:

$$\begin{array}{ccc} E_1 & \xrightarrow{F} & E_2 \\ \pi_1 \downarrow & & \downarrow \pi_2 \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

Thus one has well-defined and continuous *fiber mappings* $F_a := F|_{(E_1)_a}: (E_1)_a \rightarrow (E_2)_{f(a)}$ for each $a \in M_1$. We say that $F: E_1 \rightarrow E_2$ is an *isomorphism of fiber bundles over f* , if both f and F are homeomorphisms. In that case, the inverse $F^{-1}: E_2 \rightarrow E_1$ is a fiber bundle morphism over f^{-1} .

4.2.2 Banach Bundles and Hilbert Bundles

Note that the transition maps $\Phi_{ij} = U_{ij} \times X \rightarrow U_{ij} \times X$ of a fiber bundle $\pi: E \rightarrow M$ are of the special form

$$\Phi_{ij}(a, x) = (a, \varphi_{ij}(a)(x))$$

with some *transition function* $\varphi_{ij}: U_{ij} \rightarrow \text{Homeo}(X)$ into the group of homeomorphisms of X . Moreover, one has $(\varphi_{ij}(a))^{-1} = \varphi_{ji}(a)$ and $\varphi_{ij}(a)\varphi_{jk}(a) = \varphi_{ik}(a)$.

One may define various special types of fiber bundles in the following way: Let X be a space with an additional structure that is compatible with the topology of X . For example, X can be a topological vector space, a topological group, or a metric space. Denote by $G \subset \text{Homeo}(X)$ the group of continuous isomorphisms of X , i.e., those homeomorphisms of X that also respect the additional structure. By adding the further requirement to a bundle atlas that its transition functions take only values in this group G , one may define fiber bundles whose fibers exhibit the same structure as its typical fiber X . In the following, we outline this construction for the special case that the typical fiber X is a Banach space (or Hilbert space), leading to the notion of a Banach bundle (or Hilbert bundle, respectively).

Definition 4.2.8. Let X be a Banach space and let $\pi: E \rightarrow M$ be a fiber bundle with typical fiber X . A *Banach bundle atlas* on $\pi: E \rightarrow M$ with typical fiber X is a covering family of local trivializations $(\varphi_i: E|U_i \rightarrow U_i \times X)_{i \in I}$ such that each transition function is a continuous map $\varphi_{ij}: U_{ij} \rightarrow L(X; X)$, where $L(X; X)$ denotes the Banach algebra of continuous linear operators from X to X .

A fiber bundle $\pi: E \rightarrow M$ with typical fiber X is called a *Banach bundle with typical fiber X* if its maximal bundle atlas of $\pi: E \rightarrow M$ contains a (maximal) Banach bundle atlas.

Analogously, one defines the notion of a Hilbert bundle. The definitions are such that each Hilbert bundle is also a Banach bundle and that each Banach bundle is also a fiber bundle.

Proposition 4.2.9. Let $\pi: E \rightarrow M$ be a Banach bundle. Then each E_a , $a \in M$ carries the structure of a topological vector space in the following way: Choose a local trivialization $\varphi: E|U \rightarrow U \times X$ from a Banach bundle atlas of E such that $a \in U$. Then define for $v, v_1, v_2 \in E_a$ and $\lambda \in \mathbb{R}$:

$$v_1 (+)_a v_2 := \varphi^{-1}(a, \text{pr}_X(\varphi(v_1)) + \text{pr}_X(\varphi(v_2))) \quad \text{and} \quad \lambda (\cdot)_a v := \varphi^{-1}(a, \lambda \cdot \text{pr}_X(\varphi(v))).$$

By the special structure of Banach bundle atlantes (see 4.2.8), this is independent of the choice of φ and the induced mappings^{4.1}

$$\begin{aligned} + : E \times_M E &\rightarrow E, & v_1 + v_2 &:= v_1 (+)_{\pi(v_1, v_2)} v_2 \quad \text{and} \\ \cdot : \mathbb{R} \times E &\rightarrow E, & \lambda \cdot v &:= \lambda (\cdot)_{\pi(v)} v \end{aligned}$$

are continuous.

Example 4.2.10. Let M be a Banach manifold of class $C_{\text{loc}}^{k, \alpha}$, $k + \alpha \geq 2$. Then the tangent bundle $\pi_{TM} : TM \rightarrow M$ is a Banach bundle of class $C_{\text{loc}}^{k-1, \alpha}$. Sections of TM are precisely the vector fields on M .

Example 4.2.11. Let $\pi_X : X \rightarrow M$ and $\pi_Y : Y \rightarrow M$ be Banach bundles over M . Then the space $L(X; Y) := \coprod_{a \in M} \{a\} \times L(X_a; Y_a)$ together with $\pi_{L(X; Y)} : L(X; Y) \rightarrow M$, $(a, A) \mapsto a$ is a Banach bundle in a canonical way; a Banach bundle atlas for $L(X; Y)$ can be constructed from bundle atlantes for X and Y .

In particular, the *dual bundle* $X' := L(X; M \times \mathbb{R})$ is also a Banach bundle with fibers $(X')_a \cong (X_a)'$ for all $a \in M$.

Definition 4.2.12. Let $\pi_1 : E_1 \rightarrow M_1$ and $\pi_2 : E_2 \rightarrow M_2$ be Banach bundles, and let $F : E_1 \rightarrow E_2$ be a fiber bundle morphism over $f : M_1 \rightarrow M_2$ (see 4.2.7). We say that F is a *Banach bundle morphism from E_1 to E_2 over f* if for each $a \in M_1$ the fiber mapping $F_a := F|_{(E_1)_a} : (E_1)_a \rightarrow (E_2)_{f(a)}$ induced by F is linear. We say that $F : E_1 \rightarrow E_2$ is an *isomorphism of Banach bundles over f* , if it is also a fiber bundle isomorphism. Note that in this case, the inverse $F^{-1} : E_2 \rightarrow E_1$ is automatically a Banach bundle morphism over f^{-1} .

4.3 Riesz Structures

First, we introduce Riesz structures on Banach spaces (Section 4.3.1). To some extent, Banach spaces with Riesz structures look and feel quite much like Hilbert spaces. For example, 4.3.5 introduces the concept of a pseudoadjoint $A^* \in L(X_2; X_1)$ of a Riesz morphism $A \in L(X_1; X_2)$ (a concept which we utilize to introduce pseudogradients in Section 4.4). The generalization of Riesz structures to Banach bundles is straightforward. Hence, a great deal of Section 4.3.2 will occupy us with producing a practically relevant example. Finally, we introduce the category of Riesz manifolds (Section 4.3.3), a category that carries many desirable properties of the category of Riemannian manifolds.

^{4.1} see 4.2.5 for the definition of the product bundle $\pi : E \times_M E \rightarrow M$

4.3.1 Riesz Structures

Definition 4.3.1. Let X be a Banach space. A *Riesz structure* on X consists of

1. a continuous, linear and dense injection $i: X \hookrightarrow H$ into a Hilbert space H ;
2. a continuous, linear and dense injection $j: H \hookrightarrow Y$ into a Banach space Y ;
3. a fixed Hilbert norm $\|\cdot\|_H$ or equivalently, its Riesz isomorphism $I: H \rightarrow H'$;
4. and of an isomorphism of Banach spaces $J: X \rightarrow Y'$

such that the following diagram commutes:

$$\begin{array}{ccc}
 X & \xrightarrow{J} & Y' \\
 \downarrow i & \cong & \downarrow j' \\
 H & \xrightarrow{I} & H'
 \end{array} \quad (4.3.1)$$

We may call J a *pseudo-Riesz isomorphism* of X . Moreover, every Banach space X with Riesz structure (i, j, I, J) exhibits a pre-Hilbert metric, which can be expressed by the Riesz structure in the following ways

$$b(x_1, x_2) := \langle i x_1, i x_2 \rangle_H = \langle I i x_1, i x_2 \rangle = \langle J x_1, j i x_2 \rangle, \quad \text{for all } x_1, x_2 \in X.$$

Remark 1. At first glance, the notion of a Riesz structure seems to boil down to the notion of a *Gelfand triple*, i.e., a topological vector space X , a Hilbert space H together with linear, dense embeddings $X \hookrightarrow H \hookrightarrow X'$. However, this is not true, since Riesz structures involve a *third* Banach space Y which need not coincide with X' , in particular if X is nonreflexive.

The prototypical Riesz structure is given by the following example.

Example 4.3.2. Let $(\Omega, \mathcal{A}, \mu)$ be a finite measure space, let $p \in [2, \infty]$, and let $q \in [1, 2]$ be the Hölder conjugate of p . Put $X := L^p(\Omega, \mu)$, $H := L^2(\Omega, \mu)$, $Y := L^q(\Sigma; \mu)$, and denote by $i: L^p(\Omega, \mu) \hookrightarrow L^2(\Omega, \mu)$, $j: L^2(\Omega, \mu) \rightarrow L^q(\Omega, \mu)$ the canonical embeddings. The Riesz isomorphism $I: L^2(\Omega, \mu) \rightarrow (L^2(\Omega, \mu))'$ is given by $\langle I v_1, v_2 \rangle := \int_{\Omega} v_1 v_2 \, d\mu$ for $v_1, v_2 \in H$. Analogously, one may consider the operator $J: L^p(\Omega, \mu) \rightarrow (L^q(\Omega, \mu))'$ defined by $\langle J u, w \rangle := \int_{\Omega} u w \, d\mu$ for $u \in X$ and $w \in Y$. Observe that $I \circ i = j' \circ J$. By the Radon-Nikodym theorem, J is an isomorphism of Banach spaces, hence (i, j, I, J) is a Riesz structure on $X = L^p(\Omega, \mu)$.

It is essential for our applications that Riesz structures are able to encode elliptic operators.

Example 4.3.3. Let (Σ, g) be a compact Riemannian manifold all whose connected components have nontrivial boundary, let $p \in [2, \infty]$, and let $q \in [1, 2]$ be the

Hölder conjugate of p . Let $X := W^{2,p}(\Sigma; \mathbb{R}^m) \cap W_0^{1,p}(\Sigma; \mathbb{R}^m)$, $H := W_0^{1,2}(\Sigma; \mathbb{R}^m)$, $Y := L^q(\Sigma; \mathbb{R}^m)$ and let $i: X \hookrightarrow H$ and $j: H \hookrightarrow Y$ be the canonical embeddings. Let $0 < c \leq C < \infty$ be constants and let $b \in W^{1,r}(\Sigma; \text{Sym}^2(T\Sigma))$ with $r > \dim(\Sigma)$ be a tensor field of symmetric, bilinear forms with $c|g|_a \leq b|_a \leq C|g|_a$ for all $a \in \Sigma$. Thus, the Laplace-Beltrami operator $-\Delta_b: W^{2,p}(\Sigma; \mathbb{R}^m) \rightarrow L^p(\Sigma; \mathbb{R}^m)$ with respect to the metric b is well-defined and continuous (see 4.5.3). This allows us to define the operator $J_b: X \rightarrow Y'$ by

$$\langle J_b u, w \rangle := \int_{\Sigma} g_0(-\Delta_b u, w) \text{vol}_b, \quad u \in X, w \in Y',$$

where vol_b denotes the Riemannian volume density induced by b . Elliptic regularity for the Dirichlet problem (see Theorem 9.15 in [10]) implies that J_b is an isomorphism provided that $r \geq p$. Integration by parts leads to the weak formulation $I_b: W_0^{1,2}(\Sigma; \mathbb{R}^m) \rightarrow (W_0^{1,2}(\Sigma; \mathbb{R}^m))'$ of the Laplace-Beltrami operator:

$$\langle I_b v_1, v_2 \rangle := \int_{\Sigma} \langle dv_1, dv_2 \rangle_b \text{vol}_b \quad \text{for } v_1, v_2 \in W_0^{1,2}(\Sigma; \mathbb{R}^m).$$

The condition $c|g| \leq b \leq C|g|$ guarantees that I_b is both continuous and continuously invertible. The fact that I_b and J_b are connected via integration by parts is reflected in the equality $I_b \circ i = j' \circ J_b$, showing that (i, j, I_b, J_b) constitutes a Riesz structure on X .

Definition 4.3.4. Let $(X_1, i_1, j_1, I_1, J_1)$ and $(X_2, i_2, j_2, I_2, J_2)$ be Banach spaces with Riesz structure. We call a triple (A, B, C) with $A \in L(X_1; X_2)$, $B \in L(H_1; H_2)$, and $C \in L(Y_1; Y_2)$ a *morphism of Riesz structures* or shorter, a *Riesz morphism*, if the following diagram commutes:

$$\begin{array}{ccc} X_1 & \xrightarrow{A} & X_2 \\ \downarrow i_1 & & \downarrow i_2 \\ H_1 & \xrightarrow{B} & H_2 \\ \downarrow j_1 & & \downarrow j_2 \\ Y_1 & \xrightarrow{C} & Y_2. \end{array}$$

Proposition 4.3.5. Let $(X_1, i_1, j_1, I_1, J_1)$ and $(X_2, i_2, j_2, I_2, J_2)$ be Banach spaces with Riesz structure and let (A, B, C) be a Riesz morphism. We define the pseudoadjoint A^* of A by $A^* := J_1^{-1} C' J_2$, where $C' \in L(Y_2'; Y_1')$ denotes the dual map given by $C' \xi := \xi \circ C$. The pseudoadjoint has the following properties:

1. $A^* \in L(X_2; X_1)$.
2. A^* is an adjoint with respect to the induced pre-Hilbert metrics b_1 and b_2 , i.e., one has

$$b_2(A x_1, x_2) = b_1(x_1, A^* x_2) \quad \text{for all } x_1 \in X_1 \text{ and } x_2 \in X_2.$$

3. $A^* A$ is symmetric and positive semi-definite with respect to b_1 .
4. AA^* is symmetric and positive semi-definite with respect to b_2 .

Proof. *Claim 1.* Since C' , J_1^{-1} and J_2 are continuous, A^* is clearly continuous.

Claim 2. Observe

$$i_1 A^* = (i_1 J_1^{-1}) C' J_2 = I_1^{-1} (j_1' C') J_2 = I_1^{-1} B' (j_2' J_2) = I_1^{-1} B' I_2 i_2 = B^* i_2.$$

With $x_1 \in X_1$ and $x_2 \in X_2$, the second statement follows from

$$b_1(x_1, A^* x_2) = \langle i_1 x_1, B^* i_2 x_2 \rangle_{H_1} = \langle B i_1 x_1, i_2 x_2 \rangle_{H_1} = \langle i_2 A x_1, i_2 x_2 \rangle_{H_2} = b_2(A x_1, x_2).$$

Claim 3. This follows from

$$b_1(u, A^* A v) = b_2(A u, A v)$$

and

$$b_1(u, A^* A u) = b_2(A u, A u) = \|i_2 A u\|_{H_2}^2 \geq 0$$

for all $u, v \in X_1$.

Claim 4. One has $b_2(AA^* u, v) = b_1(A^* u, A^* v)$ and

$$b_2(AA^* u, u) = b_1(A^* u, A^* u) = \langle i_1 A^* u, i_1 A^* u \rangle_{H_1} = \|B^* i_2 u\|_{H_1}^2 \geq 0$$

for all $u, v \in X_2$. □

4.3.2 Riesz Bundle Structures

Analogously to 4.3.1 and 4.3.4, we define Riesz structures on Banach bundles and Riesz bundle morphisms as follows.

Definition 4.3.6. Let $\pi_X: X \rightarrow M$ be a Banach bundle. A *Riesz structure* on X consists of

1. a continuous, linear, and dense bundle injection $i: X \hookrightarrow H$ over id_M into a Hilbert bundle $\pi_H: H \rightarrow M$;
2. a continuous, linear, and dense bundle injection $j: H \hookrightarrow Y$ over id_M into a further Banach bundle $\pi_Y: Y \rightarrow M$;
3. a fixed Hilbert bundle norm $\|\cdot\|_H$ or equivalently, its Riesz isomorphism $I: H \rightarrow H'$;
4. and an isomorphism $J: X \rightarrow Y'$ of Banach bundles over id_M

such that the following diagram commutes.

$$\begin{array}{ccccc}
 X & \xrightarrow{J} & Y' & & \\
 \downarrow i & \searrow \pi_X & \swarrow \pi_{Y'} & & \downarrow j' \\
 & & M & & \\
 \swarrow \pi_H & & \nwarrow \pi_{H'} & & \\
 H & \xrightarrow{I} & H' & &
 \end{array}
 \quad (4.3.2)$$

Let $R \in \{C_{\text{loc}}^{k,\alpha} \mid k \in \mathbb{N} \cup \{0, \infty\}, \alpha \in [0, 1] \text{ with } k + \alpha \geq 1\}$ and suppose that $\pi: X \rightarrow M$ is a Banach bundle of class R . We say that a Riesz structure (i, j, I, J) is of class R , if

1. H and Y are Banach bundles of class R and
2. i, j, I , and J are Banach bundle morphisms of class R .

Example 4.3.7. Let Σ be a connected, compact, n -dimensional manifold with non-trivial boundary. Denote by g_0 the Euclidean metric on \mathbb{R}^m . For $f \in W^{1,\infty}(\Sigma; \mathbb{R}^m)$, denote by $f^\# g_0 := g_0(df \cdot, df \cdot)$ the *pullback* of the bilinear form g_0 and define the space of *Lipschitz immersions* as

$$\text{Imm}^{1,\infty}(\Sigma; \mathbb{R}^m) := \{f \in W^{1,\infty}(\Sigma; \mathbb{R}^m) \mid \exists C > 0: C^{-1} g \leq f^\# g_0 \leq C g\}, \quad (4.3.3)$$

where g may be any smooth Riemannian metric on Σ . Note that $\text{Imm}^{1,\infty}(\Sigma; \mathbb{R}^m) \subset W^{1,\infty}(\Sigma; \mathbb{R}^m)$ is an open set. For $p > n$, the Sobolev embedding theorem shows that $W^{2,p}(\Sigma; \mathbb{R}^m)$ embeds continuously into $W^{1,\infty}(\Sigma; \mathbb{R}^m)$ which enables us to define

$$\text{Imm}^{2,p}(\Sigma; \mathbb{R}^m) := \text{Imm}^{1,\infty}(\Sigma; \mathbb{R}^m) \cap W^{2,p}(\Sigma; \mathbb{R}^m).$$

The trace theorem for Sobolev spaces states that the so-called *trace operator* $\text{res}: C^\infty(\Sigma; \mathbb{R}^m) \rightarrow C^\infty(\partial\Sigma; \mathbb{R}^m)$, $\text{res}(f) = f|_{\partial\Sigma}$ can be continuously extended to $\text{res}: W^{2,p}(\Sigma; \mathbb{R}^m) \rightarrow W^{2-\frac{1}{p},p}(\partial\Sigma; \mathbb{R}^m)$. We fix an immersion $\gamma \in \text{Imm}^{2-\frac{1}{p},p}(\partial\Sigma; \mathbb{R}^m)$ and consider the configuration space of immersed surfaces

$$\mathcal{C} := \text{Imm}_{\gamma}^{2,p}(\Sigma; \mathbb{R}^m) := \{f \in \text{Imm}_{\gamma}^{2,p}(\Sigma; \mathbb{R}^m) \mid \text{res}(f) = \gamma\}.$$

Analogously to 4.3.3, we define the (trivial) Banach bundles

$$\mathcal{X} := \mathcal{C} \times (W^{2,p}(\Sigma; \mathbb{R}^m) \cap W_0^{1,p}(\Sigma; \mathbb{R}^m)), \quad \mathcal{H} := \mathcal{C} \times W_0^{1,2}(\Sigma; \mathbb{R}^m), \quad \mathcal{Y} := \mathcal{C} \times L^q(\Sigma; \mathbb{R}^m)$$

along with canonical injections $i: \mathcal{X} \hookrightarrow \mathcal{H}$ and $j: \mathcal{H} \hookrightarrow \mathcal{Y}$. These mappings are fiber-wise continuous and dense injections. We summarize the setting in the commutative diagram

$$\begin{array}{ccccc}
 \mathcal{X} & \xhookrightarrow{i} & \mathcal{H} & \xhookrightarrow{j} & \mathcal{Y} \\
 \downarrow \text{pr} & & \downarrow \text{pr} & & \downarrow \text{pr} \\
 \mathcal{C} & \xrightarrow{\text{id}_{\mathcal{C}}} & \mathcal{C} & \xrightarrow{\text{id}_{\mathcal{C}}} & \mathcal{C}.
 \end{array}$$

Each immersion $f \in \mathcal{C}$ induces a Riemannian metric $b := f^\# g_0$ of Sobolev class $W^{1,p}$.^{4.2} Thus, the Laplace-Betrami operator $\Delta_f := \Delta_b: W^{2,p}(\Sigma; \mathbb{R}^m) \rightarrow L^p(\Sigma; \mathbb{R}^m)$ is well-defined. With the definitions and results of 4.3.3, we observe that $J_f := J_b: \mathcal{X}_f \rightarrow (\mathcal{Y}_f)'$ and $I_f := I_b: \mathcal{H}_f \rightarrow (\mathcal{H}_f)'$ are isomorphisms satisfying $I_f \circ i_f = j_f' \circ J_f$. Hence the data (i, j, I, J) constitutes a Riesz structure on the Banach bundle \mathcal{X} . Without going into detail, we mention that (i, j, I, J) is even a Riesz structure of class C^∞ .

Definition 4.3.8. Let $(X_1, i_1, j_1, I_1, J_1,)$ and $(X_2, i_2, j_2, I_2, J_2,)$ be Banach bundles with Riesz bundle structures. We call a triple (A, B, C) of Banach bundle morphisms $A: X_1 \rightarrow X_2, B: H_1 \rightarrow H_2$, and $C: Y_1 \rightarrow Y_2$ a *Riesz bundle morphism*, if (A_a, B_a, C_a) is a Riesz morphism for each $a \in M$, i.e., if the following diagram commutes:

$$\begin{array}{ccc} X_1 & \xrightarrow{A} & X_2 \\ \downarrow i_1 & & \downarrow i_2 \\ H_1 & \xrightarrow{B} & H_2 \\ \downarrow j_1 & & \downarrow j_2 \\ Y_1 & \xrightarrow{C} & Y_2. \end{array}$$

4.3.3 Riesz Manifolds

Definition 4.3.9. Let $k \in \mathbb{N} \cup \{\infty\}$ and $\alpha \in [0, 1]$ with $k + \alpha \geq 2$. A *Riesz manifold of class $C_{\text{loc}}^{k,\alpha}$* is a Banach manifold M of class $C_{\text{loc}}^{k,\alpha}$ together with a Riesz structure (i, j, I, J) of class $C_{\text{loc}}^{k-1,\alpha}$ on TM .

Definition 4.3.10. Let $(M_1, i_1, j_1, I_1, J_1)$ and $(M_2, i_2, j_2, I_2, J_2)$ be Riesz manifolds of class $C_{\text{loc}}^{k,\alpha}$ and let $F: M_1 \rightarrow M_2$ be a mapping of class $C_{\text{loc}}^{k,\alpha}$. We say that F is a *Riesz morphism of class $C_{\text{loc}}^{k,\alpha}$* if the tangent map $TF: TM_1 \rightarrow TM_2$ induces a morphism of Riesz bundles of class $C_{\text{loc}}^{k-1,\alpha}$. In particular, for a Riesz morphism $F: M \rightarrow \mathbb{R}$ into the real numbers, we also say that F is a *Riesz function*.

Example 4.3.11. Let (M, g) be a Riemannian manifold, i.e., a (smooth) Banach manifold together with a (smooth) section g of $\text{Sym}^2(TM)$ such that for each $a \in M$, the symmetric bilinear form $g_a: T_a M \times T_a M \rightarrow \mathbb{R}$ is positive-definite and generates the topology on $T_a M$. In particular, $(T_a M, g_a)$ is a Hilbert space and $I_a: T_a M \rightarrow T_a M$, $\langle I_a u, v \rangle := g_a(u, v)$ is a Riesz isomorphism. With $TM = H = Y$, and $J = I, i = j = \text{id}_{TM}$, we observe that (M, i, j, I, J) is a Riesz manifold in a natural way. Note that a smooth map between Riemannian manifolds induces also a (unique) Riesz morphism between the induced Riesz manifolds.

^{4.2} Since $f^\# g_0$ depends quadratically on df , the condition $p > n$ is necessary here.

Example 4.3.12. Let (M, i, j, I, J) be a smooth Riesz manifold. Then one may define a section g of $\text{Sym}^2(TM)$ by

$$g_a(u, v) := \langle J_a u, j i v \rangle = \langle I_a i u, i v \rangle \quad \text{for } u, v \in T_a M.$$

Since I_a is a Riesz isomorphism, g_a is a symmetric and positive-definite bilinear form. Thus, $(T_a M, g_a)$ is a pre-Hilbert space with completion $(H_a, \langle I_a \cdot, \cdot \rangle)$. If $T_a M$ is not isomorphic to a Hilbert space, (TM, g) is *not* a Riemannian manifold. In general, there is no chance to complete (M, g) to a Hilbert manifold (\tilde{M}, \tilde{g}) . Moreover, in the rare cases where this is possible, many continuous (or even differentiable) functions on M cannot be extended continuously (differentiably) to \tilde{M} .

These examples show that Riemannian manifolds form a (proper) full subcategory of the category of Riesz manifolds in the sense that the functor described in 4.3.11 is injective (but not surjective) on objects and fully faithful. Still, the latter category shares many useful properties with the former (see Section 4.4 below for a few examples). Forgetting the underlying Riesz structure induces a forgetful functor from the category of Riesz manifolds to the category of Banach manifolds. In this sense, Riesz manifolds may be considered as lying somewhere between Riemannian manifolds and arbitrary Banach manifolds. More precisely, we have the following commutative diagram of functors

$$\begin{array}{ccc} \left\{ \begin{array}{c} \text{Riemannian} \\ \text{manifolds} \end{array} \right\} & \xrightarrow{(M, g) \mapsto (M, \text{id}, \text{id}, I, I)} & \left\{ \begin{array}{c} \text{Riesz} \\ \text{manifolds} \end{array} \right\} \\ & \searrow (M, g) \mapsto M & \swarrow (M, i, j, I, J) \mapsto M \\ & \left\{ \begin{array}{c} \text{Banach} \\ \text{manifolds} \end{array} \right\} & \end{array}$$

4.4 Pseudogradient Flow

Now we have all necessary ingredients for defining the pseudogradient of a Riesz function.

Definition 4.4.1. Let (M, i, j, I, J) be a Riesz manifold of class $C_{\text{loc}}^{k, \alpha}$ and let $F: M \rightarrow \mathbb{R}$ be a Riesz mapping of class $C_{\text{loc}}^{k, \alpha}$. We define the *pseudogradient field* $\text{grad}(F): M \rightarrow TM$ by setting $\text{grad}(F) := (TF)^*$.

Compared to the elaborate theory that is needed for the treatment of nonlinear flows of parabolic type, the existence theory of pseudogradient flows relies on quite elemen-

tary principles. We gather the central properties of pseudogradient fields and their flows in the following theorem.

Theorem 4.4.2. *Let $k \in \mathbb{N} \cup \{\infty\}$ and $\alpha \in [0, 1]$ with $k + \alpha \geq 2$.*

Let M be a Riesz manifold of class $C_{\text{loc}}^{k,\alpha}$ and let $F: M \rightarrow \mathbb{R}$ be a Riesz function of class $C_{\text{loc}}^{k,\alpha}$. Then one has

1. *For all $a \in M$: $\langle dF|_a, \text{grad}(F)|_a \rangle \geq 0$.*
2. *Let $a \in M$. Then $\text{grad}(F)|_a = 0$ if and only if $dF|_a = 0$.*
3. *The pseudogradient field $\text{grad}(F)$ is a vector field of class $C_{\text{loc}}^{k-1,\alpha}$.*
4. *For each interior point $a_0 \in M \setminus \partial M$ there is a neighborhood U of a_0 and an $\varepsilon > 0$ such that there is a unique flow $\Phi: U \times]-\varepsilon, \varepsilon[\rightarrow M$ with*

$$\frac{d}{dt}\Phi(a, t) = -\text{grad}(F)|_{\Phi(a, t)} \quad \text{and} \quad \Phi(a, 0) = a \quad \text{for all } (a, t) \in U \times]-\varepsilon, \varepsilon[.$$

5. *For each interior point $a \in M \setminus \partial M$, the function $t \mapsto F \circ \Phi(a, t)$ is monotonically decreasing.*

Proof. Denote the Riesz structure on M with

$$i: TM \hookrightarrow H, \quad j: H \hookrightarrow Y, \quad I: H \rightarrow H' \quad \text{and} \quad J: TM \rightarrow Y'$$

and let $B: H \rightarrow T\mathbb{R}$ and $C: Y \rightarrow T\mathbb{R}$ be the Banach bundle morphisms over F of class $C_{\text{loc}}^{k-1,\alpha}$ with $i_{\mathbb{R}} TF = B i$ and $j_{\mathbb{R}} i_{\mathbb{R}} TF = C j i$.

Claim 1. Denote by $b_{\mathbb{R}}$ the Euclidean inner product on $T\mathbb{R}$. By 4.3.5, we have for each $a \in M$ that

$$\langle dF|_a, \text{grad}(F)|_a \rangle = b_{\mathbb{R}}(1, T_a F \text{grad}(F)|_a \cdot 1) = \|B_a^* \cdot 1\|_{H_a}^2 \geq 0.$$

Claim 2. Note that $dF|_a = 0$ is equivalent to $T_a F = 0$. On the one hand, the definition of $\text{grad}(F)$ implies immediately that $\text{grad}(F)|_a$ vanishes if $T_a F$ vanishes (see 4.4.1). On the other hand, the inequality above shows that $\text{grad}(F)|_a = 0$ implies $B_a^* = 0$ and thus $B_a = 0$. We have $(i_{\mathbb{R}})_a T_a F = B_a i_a = 0$ and since $(i_{\mathbb{R}})_a$ is an isomorphism, we obtain $T_a F = 0$ and hence $dF|_a = 0$.

Claim 3. With J also J^{-1} is of class $C_{\text{loc}}^{k-1,\alpha}$. Moreover, continuous bilinear combinations of Banach-space valued mappings of class $C_{\text{loc}}^{k-1,\alpha}$ are again of class $C_{\text{loc}}^{k-1,\alpha}$. Hence $\text{grad}(F) = J^{-1} C$ is a section of class $C_{\text{loc}}^{k-1,\alpha}$ in the Banach bundle TM , hence a vector field of class $C_{\text{loc}}^{k-1,\alpha}$.

Claim 4. Note that $a \mapsto -\text{grad}(F)|_a$ is a vector field at least of class $C_{\text{loc}}^{0,1}$. Hence the statement follows from the Picard-Lindelöf theorem.

Claim 5. This follows from Claim 2 and from the chain rule:

$$\frac{d}{dt}F(\Phi(a, t)) = -\langle dF|_{\Phi(a, t)}, \text{grad} F|_{\Phi(a, t)} \rangle \leq 0.$$

If the initial condition is not a critical point, the function $t \mapsto F(\Phi(a, t))$ is even strictly monotonically decreasing, since stationary points cannot be reached within finite time. \square

4.5 Applications

Geometric functionals for curves and surfaces, such as those we discuss in this section, play a major role in the field of geometry processing and are applied to a variety of tasks. Since reducing computational time is a specific objective in this field, it is no wonder that the advantages of H^1 - and H^2 -gradient flows have already been observed and utilized within this community (see [9]). Our aim here is to give some more theoretical justification on these flows, even for the pre-discretized, infinite-dimensional setting, which is often not covered appropriately.

4.5.1 Minimal Surfaces

Let us return to the setting of 4.3.7. Observe that $\mathcal{X} = T\mathcal{C}$ is precisely the tangent bundle of $\mathcal{C} = \text{Imm}_{\gamma}^{2,p}(\Sigma; \mathbb{R}^m)$, the configuration space of immersions subject to Dirichlet boundary conditions. As we have seen in 4.3.7, the data (i, j, I, J) represents a smooth Riesz structure on the Banach bundle $T\mathcal{C}$, hence $(\mathcal{C}, i, j, I, J)$ is a smooth Riesz manifold. Now consider the *volume functional*

$$\mathcal{F}: \mathcal{C} \rightarrow \mathbb{R}, \quad \mathcal{F}(f) := \int_{\Sigma} \text{vol}_{f^{\#}g_0}.$$

Theorem 4.5.1. *The volume functional \mathcal{F} is a smooth Riesz function on the smooth Riesz manifold $\text{Imm}_{\gamma}^{2,p}(\Sigma; \mathbb{R}^m)$.*

Proof. One has

$$D(f \mapsto \text{vol}_{f^{\#}g_0})u = \langle df, du \rangle_{f^{\#}g_0} \text{vol}_{f^{\#}g_0} \quad (4.5.1)$$

for each Lipschitz immersion $f \in \text{Imm}^{1,\infty}(\Sigma; \mathbb{R}^m)$ and each $u \in T_f \text{Imm}(\Sigma; \mathbb{R}^m) = W^{1,\infty}(\Sigma; \mathbb{R}^m)$. Hence, the volume functional is differentiable and its differential $d\mathcal{F}: \mathcal{C} \rightarrow T'\mathcal{C}$ is given by

$$\langle d\mathcal{F}|_f, u \rangle := \int_{\Sigma} \langle df, du \rangle_{f^{\#}g_0} \text{vol}_{f^{\#}g_0} \quad \text{for } u \in T_f \mathcal{C} = W^{2,p}(\Sigma; \mathbb{R}^m) \cap W_0^{1,p}(\Sigma; \mathbb{R}^m).$$

Extending $d\mathcal{F}|_f$ continuously to \mathcal{H}_f leads to the smooth section $B: \mathcal{C} \rightarrow \mathcal{H}'$ given by

$$\langle B_f, v \rangle := \int_{\Sigma} \langle df, dv \rangle_{f^{\#}g_0} \text{vol}_{f^{\#}g_0} \quad \text{for } v \in \mathcal{H}_f = W_0^{1,2}(\Sigma; \mathbb{R}^m).$$

Via integration by parts, we obtain

$$\langle d\mathcal{F}|_f, u \rangle = \int_{\Sigma} g_0(-\Delta_f f, u) \text{vol}_{f^{\#}g_0}.$$

Since $\Delta_f f \in L^p(\Sigma; \mathbb{R}^m)$, this leads to the smooth section $C: \mathcal{C} \rightarrow \mathcal{Y}'$ given by

$$\langle C_f, w \rangle := \int_{\Sigma} g_0(-\Delta_f f, w) \operatorname{vol}_{f^\# g_0} \quad \text{for } w \in \mathcal{Y}_f = L^q(\Sigma; \mathbb{R}^m).$$

Moreover, the family $(C_f)_{f \in \mathcal{C}}$ induces a smooth Banach bundle morphism $C: \mathcal{Y} \rightarrow \mathcal{C} \times \mathbb{R}$ and one has $j_{\mathbb{R}} i_{\mathbb{R}} T\mathcal{F} = C j i$, hence \mathcal{F} is a Riesz function on \mathcal{C} . \square

Combining 4.4.2 and 4.5.1 allows us to discuss the pseudogradient flow of the volume function \mathcal{F} . For an immersion $f \in \mathcal{C}$, the defining equation for the pseudogradient $u = \operatorname{grad} \mathcal{F}|_f$ can be written as

$$\int_{\Sigma} \langle du, dv \rangle_{f^\# g_0} \operatorname{vol}_{f^\# g_0} = \int_{\Sigma} \langle df, dv \rangle_{f^\# g_0} \operatorname{vol}_{f^\# g_0}, \quad \text{for all } v \in \mathcal{H}_f = W_0^{1,2}(\Sigma; \mathbb{R}^m). \quad (4.5.2)$$

Note that the weak formulation of the Laplace-Beltrami operator occurs on both sides of the equation. For numerical optimization, we have to discretize the configuration space \mathcal{C} and the objective function. To this end, we may fix a triangulation \mathcal{T} of Σ and define the *discrete configuration space* $\mathcal{C}_{\mathcal{T}}$ as the set of all immersions that are piecewise-linear with respect to \mathcal{T} and that restrict to γ on all boundary vertices of \mathcal{T} . This way, the image $f(\Sigma)$ of an element $f \in \mathcal{C}_{\mathcal{T}}$ is an immersed simplicial mesh whose boundary complex is inscribed into $\gamma(\partial \Sigma)$. A discrete volume function $\mathcal{F}_{\mathcal{T}}$ can be defined straight-forwardly as the restriction of the volume function to $\mathcal{C}_{\mathcal{T}}$. Since $\mathcal{C}_{\mathcal{T}}$ is finite dimensional, it is a Riemannian manifold when equipped with the discretization of the weak formulation of the Laplace-Beltrami operator and the discretized pseudogradient $u = \operatorname{grad}(\mathcal{F}_{\mathcal{T}})|_f$ can be defined by

$$\int_{\Sigma} \langle du, dv \rangle_{f^\# g_0} \operatorname{vol}_{f^\# g_0} = \langle d\mathcal{F}_{\mathcal{T}}|_f, v \rangle = \int_{\Sigma} \langle df, dv \rangle_{f^\# g_0} \operatorname{vol}_{f^\# g_0}, \quad \text{for all } v \in T_f \mathcal{C}_{\mathcal{T}}. \quad (4.5.3)$$

The pseudogradient descent algorithm consists now in choosing an initial guess $f_0 \in \mathcal{C}_{\mathcal{T}}$ and by computing recursively

$$f_{n+1} = f_n - \tau_n \operatorname{grad}(\mathcal{F}_{\mathcal{T}})|_{f_n}, \quad \text{for } n \in \mathbb{N} \cup \{0\}, \quad (4.5.4)$$

where $\operatorname{grad} \mathcal{F}_{\mathcal{T}}|_{f_n}$ has to be computed from (4.5.3) and $\tau_n \geq 0$ is a step size parameter that has to be chosen appropriately. Again, the (discretized) Laplace-Beltrami operator occurs on both sides of the equation. This comes in handy, as it has to be reassembled in each iteration and this way, it can be used at least twice: once for computing $d\mathcal{F}_{\mathcal{T}}|_f$ and once for solving for the pseudogradient. Since assembling the discrete Laplace-Beltrami operator is a standard task, the implementation is pleasantly easy. We point out that this algorithm has already been introduced in [15], however, from a different



Fig. 4.2: From left to right: initial surface (512 faces); surface after a single explicit Euler step in negative discrete mean curvature direction; surface after a single explicit Euler step in negative discrete pseudogradient direction.



Fig. 4.3: The same as in Figure 4.2 with the same step sizes, but with refined mesh (32 768 faces).

perspective. The mentioned article is also a good reference for the assembly details of discrete Laplace-Beltrami operators on discrete surfaces.

When τ_n is chosen to be independent of n , (4.5.4) boils down to an explicit Euler scheme for the (discrete) pseudogradient flow. In principle, one may apply also more sophisticated step size rules for line search, such as the Armijo algorithm or the Wolfe-Powell algorithm.

In the case of the area functional, it turns out that one can go quite well with constant step size of magnitude $\tau = 1$ and this is essentially independent of the mesh resolution. In our experiments, we rarely needed more than two dozen pseudogradient steps in order to reach a configuration from which Newton's algorithm converged within several steps to a critical point—no matter if the discrete surface consisted of a few thousand or a million triangles. As Figure 4.4 suggest, the discrete pseudogradient flow performs well and is very robust even in cases where minimizers do not exist.

These findings are in strong contrast to the step size rules for discrete mean curvature flow, the L^2 -gradient flow of the volume functional: As discrete mean curvature flow is a parabolic partial differential equation (in time and space), the step size has to be decreasing along with the mesh size, even if more sophisticated integration schemes such as (semi-)implicit methods are employed. As a comparison, we include Figure 4.2 and Figure 4.3, each depicting a step of discrete L^2 -gradient descent and discrete pseudogradient descent at different mesh resolutions. We can also see there

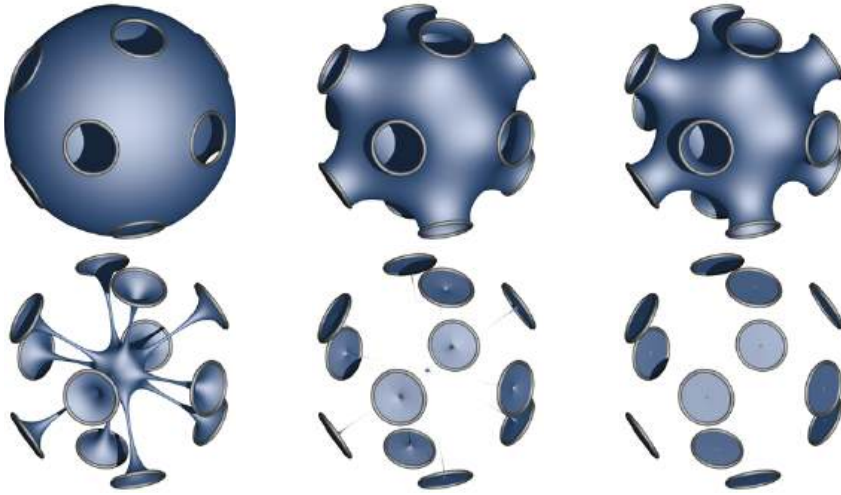


Fig. 4.4: Diverging pseudogradient flow of area functional in the absence of minimizers. It seems to converge to a minimizer in another topological class.

that the L^2 -gradient is increasingly localized with raising L^∞ -norm as the mesh size is decreased. Of course, the reason is that the initial surface is not twice differentiable so that its mean curvature is a vector-valued distribution with support concentrated on the creases. Hence the discrete mean curvature, trying to approximate this distribution, blows up under mesh refinement. In order to obtain a stable algorithm, the step size has to be decreased accordingly.

On the other hand, the information contained in the mean curvature distribution is spread out over the whole surface when solving for the pseudogradient.^{4.3} In particular, points in medium distance to an “incident” get informed immediately during the next pseudogradient step so that they have the opportunity to “react” just-in-time. See also Figure 4.5 for a typical example of an “incident” that occurs quite frequently with discrete mean curvature flows but considerably less often with discrete pseudogradient flows. We point out that this behavior is primarily an artifact of the time discretization: Mean curvature flows with infinitesimal step size have infinite propagation speed and this infinite speed is hard to capture by a discrete time stepping algorithm.

^{4.3} Note that the pseudo-Riesz isomorphism is still an elliptic operator such its inverse is smoothing.

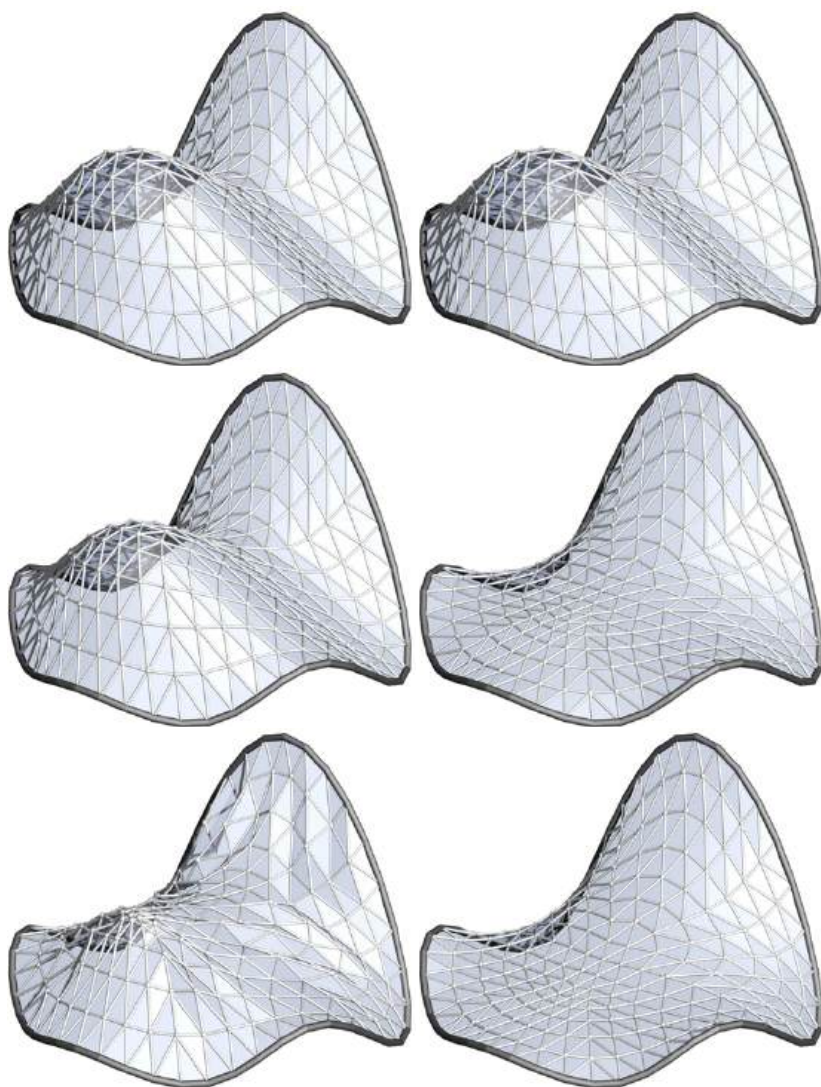


Fig. 4.5: L^2 -gradient descent (first column) and pseudogradient descent (second column), both for the area functional, with explicit Euler scheme and constant step sizes.

First row: initial surfaces; second row: after first step; third row: after nine steps.

4.5.2 Elasticae

Curvature dependent energies such as the elastica functional can also be well-formulated on the space $\text{Imm}^{2,p}(\Sigma; \mathbb{R}^m)$, when p is suitably chosen. While the discussion of the volume functional was dominated by the Laplacian operator, it turns out that a suitable operator for defining a Riesz structure for curvature energies is the bi-Laplacian. If $p > \dim(\Sigma)$, the discussion is even simplified as we do *not* need to resort to the space $\text{Imm}^{4,p}(\Sigma; \mathbb{R}^m)$, where the strong formulation of the bi-Laplacian resides.

Let Σ be an n -dimensional, compact, and connected smooth manifold. In order to keep the exposition as brief as possible, we focus our attention to manifolds with nontrivial smooth boundary and to the space

$$\mathcal{C} := \text{Imm}_\gamma^{2,p}(\Sigma; \mathbb{R}^m), \quad p \in [2, \infty[\cap]n, \infty[$$

for a given $\gamma \in \text{Imm}^{2-\frac{1}{p},p}(\Sigma; \mathbb{R}^m)$.^{4.4} From now on we use the Banach bundles

$$\begin{aligned} \mathcal{X} &:= T\mathcal{C} = \mathcal{C} \times (W^{2,p}(\Sigma; \mathbb{R}^m) \cap W_0^{1,p}(\Sigma; \mathbb{R}^m)), \\ \mathcal{H} &:= \mathcal{C} \times (W^{2,2}(\Sigma; \mathbb{R}^m) \cap W_0^{1,2}(\Sigma; \mathbb{R}^m)), \quad \text{and} \\ \mathcal{Y} &:= \mathcal{C} \times (W^{2,q}(\Sigma; \mathbb{R}^m) \cap W_0^{1,q}(\Sigma; \mathbb{R}^m)), \end{aligned}$$

where q is the Hölder conjugate of p .^{4.5} Note that p is chosen such that the canonical injections $i: \mathcal{X} \hookrightarrow \mathcal{H}$ and $j: \mathcal{H} \hookrightarrow \mathcal{Y}$ are dense in each fiber. From 4.3.3, we recall that the Laplace-Beltrami operator Δ_f maps $W^{2,r}(\Sigma; \mathbb{R}^m) \cap W_0^{1,r}(\Sigma; \mathbb{R}^m)$ isomorphically to $L^r(\Sigma; \mathbb{R}^m)$ for $r \in]1, p]$ such that we obtain a Riesz structure (i, j, I, J) with the isomorphisms

$$I: \mathcal{H} \rightarrow \mathcal{H}', \quad \langle I_f v_1, v_2 \rangle := \int_{\Sigma} g_0(\Delta_f v_1, \Delta_f v_2) \text{vol}_{f^\# g_0}, \quad (4.5.5)$$

$$J: \mathcal{X} \rightarrow \mathcal{Y}', \quad \langle J_f u, w \rangle := \int_{\Sigma} g_0(\Delta_f u, \Delta_f w) \text{vol}_{f^\# g_0}. \quad (4.5.6)$$

In order to define the elastica functional, we have to introduce the second fundamental form of an immersion. Therefore, we make a short excursion to the Hessian of a function on a Riemannian manifold.

Proposition 4.5.2. *Let $f \in \mathcal{C}$, $r \in [1, \infty]$, and $u \in W^{2,r}(\Sigma; \mathbb{R}^m)$. Define the Hessian $\text{Hess}_f(u)$ of u with respect to f by*

$$\text{Hess}_f(u)(X, Y) = (d(du \, df^\dagger) X) \cdot (df Y), \quad \text{for all } X, Y \in T_a \Sigma.$$

4.4 The case of manifolds without boundary can be treated, e.g., by imposing the nonlinear barycenter constraint $\int_{\Sigma} f \text{vol}_{f^\# g_0} = 0$.

4.5 We point out that the analogous Banach bundles for treating the case $\partial \Sigma = \emptyset$ would not be mere products anymore.

One has $\text{Hess}_f(u) \in L^{\min(p,r)}(\Sigma; \text{Sym}^2(T\Sigma; \mathbb{R}^m))$ and Hess_f is a continuous linear operator^{4.6}

$$\text{Hess}_f^f : W^{2,r}(\Sigma; \mathbb{R}^m) \rightarrow L^{\min(p,r)}(\Sigma; \text{Sym}^2(T\Sigma; \mathbb{R}^m)).$$

Proof. Let $u \in W^{2,r}(\Sigma; \mathbb{R}^m)$ so that du is an element of $W^{1,r}(\Sigma; \text{Hom}(T\Sigma; \mathbb{R}^m))$. The Moore-Penrose pseudoinverse restricted to linear maps of fixed rank is a smooth transformation. A concise treatment can be found, e.g., in [11] which allows us to deduce

$$D(f \mapsto df^\dagger) u = -df^\dagger du df^\dagger + df^\dagger (du df^\dagger)^* (\text{id}_{\mathbb{R}} - df df^\dagger). \quad (4.5.7)$$

Hence one has $df^\dagger \in W^{1,p}(\Sigma; \text{Hom}(\mathbb{R}^m; T\Sigma))$ and 4.5.3 below completes the proof. \square

Lemma 4.5.3. *Let E_1, E_2 , and E_3 be smooth Banach bundles over the compact, smooth manifold Σ , let $\mu : E_1 \times_M E_2 \rightarrow E_3$ be a locally Lipschitz continuous bilinear bundle map and let $p > \dim(\Sigma)$.*

Then $\mu \circ (\sigma_1, \sigma_2) \in W^{1, \min(p,r)}(\Sigma; E_3)$ holds for all sections $\sigma_1 \in W^{1,p}(\Sigma; E_1)$ and $\sigma_2 \in W^{1,r}(\Sigma; E_2)$. Moreover, the induced bilinear map

$$A : W^{1,p}(\Sigma; E_1) \times W^{1,r}(\Sigma; E_2) \rightarrow W^{1, \min(p,r)}(\Sigma; E_3)$$

is continuous.

Proof. It suffices to perform the regularity analysis locally. Thus, we may focus our attention to an open set $U \subset \Sigma$ and we may assume that $E_i|_U \cong U \times \mathbb{R}^{m_i}$, $i \in \{1, 2, 3\}$ are trivial vector bundles. Moreover, we may write $\sigma_1(x) = (x, f_1(x))$, $\sigma_2(x) = (x, f_2(x))$, and $\mu_x = B_x$ for all $x \in U$ with $f_1 \in W^{1,p}(\Sigma; \mathbb{R}^{m_1})$, $f_2 \in W^{1,r}(\Sigma; \mathbb{R}^{m_2})$, and $B \in W^{1,\infty}(U; \text{Bil}(\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}; \mathbb{R}^{m_3}))$.

The Sobolev embedding $W^{1,p}(\Sigma; \mathbb{R}^{m_1}) \hookrightarrow L^\infty(\Sigma; \mathbb{R}^{m_1})$ shows that $B(f_1, f_2) \in L^r(\Sigma; \mathbb{R}^{m_3})$. With $n := \dim(\Sigma)$, one has the Sobolev embedding $W^{1,r}(\Sigma; \mathbb{R}^{m_2}) \hookrightarrow L^{\tilde{r}}(\Sigma; \mathbb{R}^{m_2})$ where

$$\tilde{r} \in \begin{cases} [1, \frac{n}{n-r}], & r < n, \\ [1, \infty[, & r = n, \\ [1, \infty], & r > n. \end{cases}$$

For each smooth vector field X on U , we obtain

$$d(B(f_1, f_2))X = (dBX)(f_1, f_2) + B(df_1 X, f_2(x)) + B(f_1, df_2 X)$$

4.6 We point out for geometers that the section $\text{Hess}_f(u)$ coincides with the Hessian $\nabla^{f^\# g_0} du$, provided that f and u are sufficiently smooth. Here, $\nabla^{f^\# g_0}$ denotes the Levi-Civita connection of the Riemannian metric $f^\# g_0$

and this together with the Hölder inequality implies $d(B(f_1, f_2)) \in L^{\min(s, r)}(\Sigma; \text{Hom}(T'\Sigma; \mathbb{R}^{m_3}))$, hence $B(f_1, f_2) \in W^{1, \min(s, r)}(\Sigma; \mathbb{R}^{m_3})$, where $s = (\frac{1}{p} + \frac{1}{r})^{-1}$. We analyse the following three cases:

Case 1.: $r < n$. Because of $p > n > r$, we have

$$s = \frac{1}{\frac{1}{p} + \frac{1}{r} - \frac{1}{n}} > \frac{1}{\frac{1}{n} + \frac{1}{r} - \frac{1}{n}} = r,$$

so that $\min(s, r) = r = \min(p, r)$.

Case 2.: $r = n$. One may write $r = n = (1 + \varepsilon)^{-1}p < p$ with some $\varepsilon > 0$. Choosing $\bar{r} = \frac{p}{\varepsilon} < \infty$, we obtain

$$s = \frac{1}{\frac{1}{p} + \frac{\varepsilon}{p}} = \frac{p}{1 + \varepsilon} = r.$$

This shows $\min(s, r) = r = \min(p, r)$.

Case 3.: $r > n$. Then one has $\bar{r} = \infty$ and $s = p$, leading directly to $\min(s, r) = \min(p, r)$. The continuity of A follows from the already mentioned Hölder and Sobolev inequalities. \square

Definition 4.5.4. The *second fundamental form* $\Pi(f)$ of f can be written as $\Pi(f) := \text{Hess}_f(f)$ so that one has

$$\Pi(f) \in L^p(\Sigma; \text{Bil}(T\Sigma \times_\Sigma T\Sigma; \mathbb{R}^m)) \subset L^2(\Sigma; \text{Bil}(T\Sigma \times_\Sigma T\Sigma; \mathbb{R}^m)).$$

We define the *elastica functional* $\mathcal{F}: \mathcal{C} \rightarrow \mathbb{R}$ by

$$\mathcal{F}(f) = \frac{1}{2} \int_{\Sigma} |\Pi(f)|_{f^\# g_0}^2 \text{vol}_{f^\# g_0}.$$

Theorem 4.5.5. For $p \in [2, \infty[\cap]\dim(\Sigma), \infty]$, the *elastica functional* \mathcal{F} and the *Willmore energy* \mathcal{W} are smooth Riesz functions on the smooth Riesz manifold $(\mathcal{C}, i, j, I, J)$ from (4.5.5) and (4.5.6).

Proof. Fix $f \in \mathcal{C}$ and let $u \in T_f \mathcal{C} \subset W^{2, p}(\Sigma; \mathbb{R}^m)$. From (4.5.7), we may deduce

$$D \Pi(f) u = (\text{id}_{\mathbb{R}^m} - df df^\dagger) \text{Hess}_f(u) + (du df^\dagger)^* \Pi(f).$$

Equation (4.5.1) provides us with a formula for the derivative of $f \mapsto \text{vol}_{f^\# g_0}$. Moreover, we have $|S|_{f^\# g_0}^2 = |S(df^\dagger \cdot, df^\dagger \cdot)|_{g_0}^2$ for $S \in L^r(\Sigma; \text{Bil}(T\Sigma \times_\Sigma T\Sigma; \mathbb{R}^m))$. This would allow us to compute a precise expression for $\langle d\mathcal{F}|_f, u \rangle$, but it already suffices for our considerations to observe that $d\mathcal{F}$ is of the form

$$\langle d\mathcal{F}|_f, u \rangle = \frac{1}{2} \int_{\Sigma} \left(\langle \Pi(f), \text{Hess}_f(u) \rangle_{f^\# g_0} + \mu(\Pi(f), df, df^\dagger, du) \right) \text{vol}_{f^\# g_0},$$

where $\mu(\Pi(f), df, df^\dagger, du)$ is a polynomial expression in $\Pi(f), df, df^\dagger, du$ with constant coefficients in which $\Pi(f)$ occurs with order two and du occurs with order one.

Now let $w \in \mathcal{Y}_f = W^{2,q}(\Sigma; \mathbb{R}^m)$. Note that we have $\Pi(f) \in L^p$, $df \in L^\infty$, and $df^\dagger \in L^\infty$, hence $\mu(\Pi(f), df, df^\dagger, \cdot) \in L^{p/2}$. In the case $n = 1$, we have $dw \in W^{1,q} \hookrightarrow L^\infty$. For $n \geq 2$, we have $dw \in W^{1,q} \hookrightarrow L^r$ with $r \geq \frac{p}{p-1-\frac{p}{n}}$. Because of $p > n$, we obtain

$$r \geq \frac{p}{p-1-\frac{p}{n}} > \frac{p}{p-2} = (p/2)',$$

thus $dw \in L^{(p/2)'}$. In any case, we obtain $\mu(\Pi(f), df, df^\dagger, dw) \in L^1$. This shows that $d\mathcal{F}|_f$ can be continuously extended to $C_f: \mathcal{Y}_f \rightarrow \mathbb{R}$ so that $j_{\mathbb{R}} i_{\mathbb{R}} T\mathcal{F} = C j i$ holds. Hence, \mathcal{F} is a Riesz function. The statement for \mathcal{W} follows from the identity $H_f = \frac{1}{\dim(\Sigma)} \sum_{i=0}^m \Pi(f)(df^\dagger e_i, df^\dagger e_i)$ for any g_0 -orthonormal basis e_1, \dots, e_m of \mathbb{R}^m and from the above discussion. \square

4.5.3 Euler-Bernoulli Energy and Euler Elastica

The elastica functional is well-known under different names in dimensions one and two. For $\dim(\Sigma) = 1$, the functional F is identical to the *Euler-Bernoulli bending energy*

$$\mathcal{F}(f) = \frac{1}{2} \int_{\Sigma} |\kappa_f|_{f^\#g_0}^2 \operatorname{vol}_{f^\#g_0},$$

where κ_f is the curvature of the regular curve $f: \Sigma \rightarrow \mathbb{R}^m$. Let Σ be a compact interval equipped with the Euclidean metric g , let $\gamma: \partial\Sigma \rightarrow \mathbb{R}^m$, and $\nu: \partial\Sigma \rightarrow S^{m-1} \subset \mathbb{R}^m$ be given.

The critical points of the Euler-Bernoulli bending energy \mathcal{F} on the set

$$\{f \in \mathcal{C} \mid \operatorname{vol}_{f^\#g_0} = \operatorname{vol}_g, f|_{\partial\Sigma} = \gamma, n(f) = \nu\} \quad (4.5.8)$$

are called *Euler elasticae (with clamped ends)* where the outward unit normals are denoted with $n(f): \partial\Sigma \rightarrow S^{m-1}$. While $f|_{\partial\Sigma} = \gamma$ is clearly an inhomogeneous Dirichlet boundary condition, one may call $n(f) = \nu$ a *Neumann-type boundary condition*. Note that $n(f)$ depends nonlinearly on f . The constraint $\operatorname{vol}_{f^\#g_0} = \operatorname{vol}_g$ amounts to the requirement that all curves f in the feasible set shall be in arclength parameterization (with respect to the given metric g).

With Figure 4.6 and Figure 4.7, we provide two numerical examples of discretized pseudogradient flows for the Euler-Bernoulli bending energy subject to arclength parameterization constraints. We discretize the space \mathcal{C} by polygonal lines. Since polygonal lines are almost never elements of $W^{2,2}(\Sigma; \mathbb{R}^m)$ (unless they are straight lines), we have to discretize the energy \mathcal{F} as well. Our choice is

$$\mathcal{F}(f) = \frac{1}{2} \sum_{i=1}^n \left(\frac{2 \arctan(\varphi_i/2)}{\ell_i} \right)^2 \ell_i,$$

where φ_i denotes the turning angle of the polygonal line f at vertex i (i.e., π minus the angle enclosed by the neighboring edges) and where ℓ_i is the average of the lengths of contiguous edges. As discrete pseudo-Riesz isomorphism, we use (4.5.5) and (4.5.6), where Δ_f is replaced by the discrete Laplace-Beltrami operator. Strictly speaking, this is not justified by the smooth theory of pseudogradient flows as the discretization is nonconforming. However, pseudogradient search using this discretization performs surprisingly well as the figures attest. For comparison, we refer the reader to [3] and [4], where the L^2 -gradient flow of the Euler-Bernoulli energy is treated.

We point out that the presented flows preserve the parameterization by arclength. This was achieved by

- *projecting* the pseudogradient to the tangent space of the constraint manifold;
- performing an Armijo line search on the *osculating circle* of the constraint manifold in pseudogradient direction;
- and projecting the result onto the constraint manifold via a Newton-type algorithm involving a *pseudoinverse* of the linearized constraint map.

Of course, each of these steps involves operations that are usually only available for Riemannian manifolds. The methods go through numerically as (i) the discretized pseudo-Riesz isomorphism actually induces a *Riemannian* structure on the discrete configurations space \mathcal{C}_T and (ii) the Euler-Bernoulli energy can be formulated on the Riemannian manifold $\text{Imm}^{2,2}(\Sigma; \mathbb{R}^m)$. However, we believe that at least some of these operations can also be established for general Riesz manifolds.

4.5.4 Willmore Energy

As mentioned before, the case $\dim(\Sigma) = 1$ is very special in that the elastica functional can be well-formulated on the Riemannian manifold $\text{Imm}^{2,2}(\Sigma; \mathbb{R}^m)$. Therefore, we also consider the case $\dim(\Sigma) = 2$. One has the relation $|\Pi(f)|^2 = \kappa_1^2 + \kappa_2^2$, $K_f = \kappa_1 \kappa_2$, and $|H_f|^2 = \frac{1}{4}(\kappa_1 + \kappa_2)^2$ for the *principle curvatures* κ_1 and κ_2 , the *Gauss curvature* K_f , and the *mean curvature vector* H_f of the immersed surface f . This leads to the identity

$$\mathcal{F}(f) = \int_{\Sigma} (2 |H_f|^2 - K_f) \text{vol}_{f^*g_0}.$$

In the case that $\partial\Sigma = \emptyset$, the Gauss-Bonnet theorem and denseness considerations imply that

$$\int_{\Sigma} K_f \text{vol}_{f^*g_0} = 2\pi\chi(E) \quad (4.5.9)$$

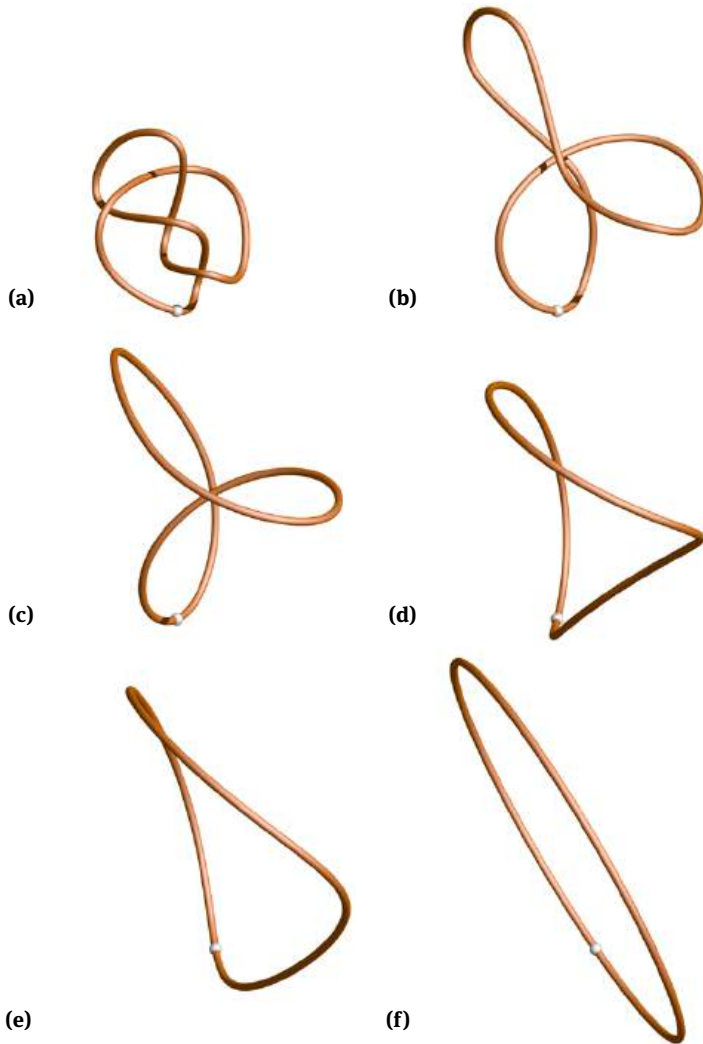


Fig. 4.6: Discrete elastic figure eight knot (2000 edges): (a) initial condition; (b)–(e) the first four iterations of pseudogradient flow with Armijo line search; (f) ultimate minimizer (the round circle), obtained after 6 pseudogradient steps in total and 3 Newton iterations.

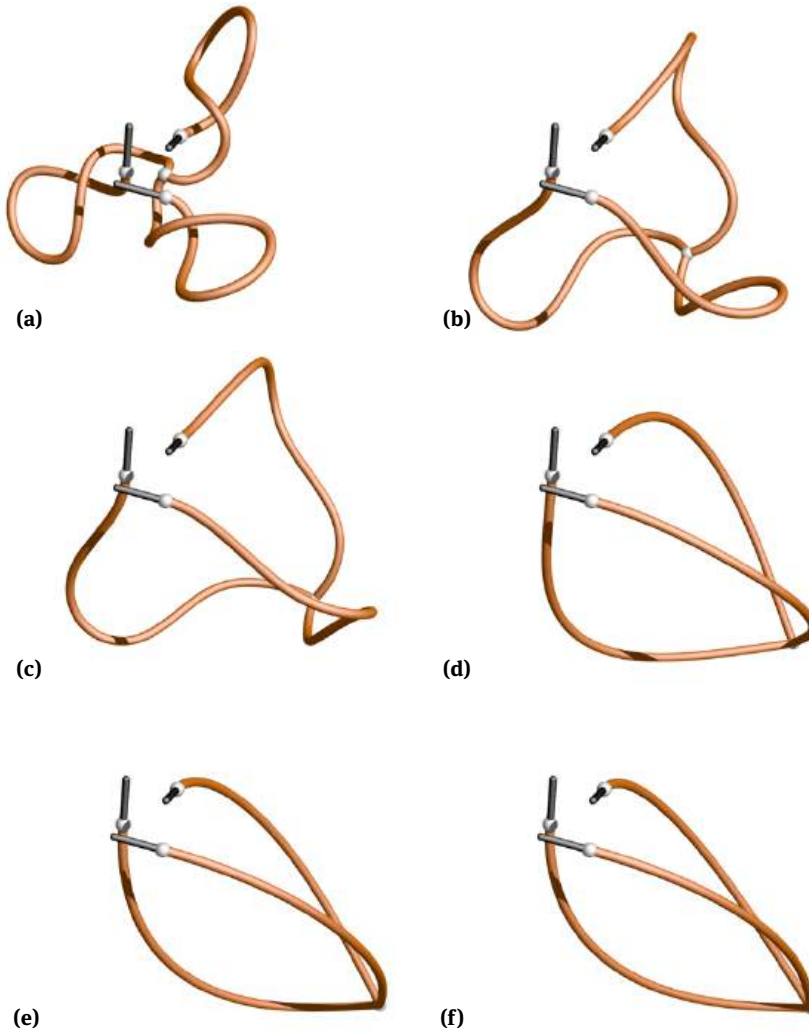


Fig. 4.7: Three discrete elastic threads (6000 edges in total) joint together with clamped boundary conditions imposed at the free ends: (a) initial condition; (b)–(e) the first four iterations of pseudogradient flow with Armijo line search; (f) ultimate minimizer, obtained after 4 pseudogradient steps in total and 4 Newton iterations.

with the Euler characteristic of Σ . Hence, \mathcal{F} is essentially identical to the *Willmore energy*

$$\mathcal{W}(f) := \frac{1}{2} \int_{\Sigma} |H_f|^2 \operatorname{vol}_{f^\# g_0}.$$

Different scaling conventions for H_f and equation (4.5.9) lead to a confusing variety of definitions for the Willmore energy in the literature, another favorite variant being $\int_{\Sigma} |H_f|^2 \operatorname{vol}_{f^\# g_0} = \int_{\Sigma} (\kappa_1 - \kappa_2)^2 \operatorname{vol}_{f^\# g_0}$. In general, all these energies differ by a total Gauss curvature expressions and by boundary integrals involving Dirichlet and Neumann boundary data and thus lead to essentially the same optimization problems when subject to both Dirichlet and Neumann-type boundary conditions.

With Figure 4.8 and Figure 4.9, we provide two numerical examples for discretized pseudogradient flows for the Willmore functional subject to constrained total area. We use triangle meshes as discrete surfaces, so that the Willmore energy has also to be discretized (see [8] for details on the discretized Willmore energy that we used).

The numerical experiments we have conducted so far indicate that pseudogradient search provides a quite efficient and robust method for minimizing curvature dependent energies—at least if compared to the Willmore flow, the L^2 -gradient flow of the Willmore energy. The latter is a forth-order parabolic flow and it suffers even more severely from the time discretization issues discussed at the end of Section 4.5.1. More details on the Willmore flow along with numerical examples for comparison can be found, e.g., in [2] and [8].

Concerning the minimization of the Willmore energy of surfaces, we point out that there is a further very efficient method which involves an L^2 -gradient descent in “curvature space” (see [6]). However, this method is heavily based on the very special relationship between mean curvature and conformal geometry and it is not clear if and how this method can be carried over to other geometric energies.

4.6 Final Remarks

Another interesting feature of pseudogradient flows is that all points of a flow trajectory have regularity not below the initial condition, provided that the used Riesz structure is defined by an elliptic operator. This does not directly imply that critical points are arbitrarily smooth (since they will usually be obtained only as limit of the trajectory for $t \rightarrow \infty$), but, maybe, this can be exploited for regularity theory in cases where long time existence and a priori bounds for the behavior along the flow (e.g., in the spirit of Grönwall’s inequalities) can be provided.

We plan to apply the presented techniques to knot energies such as the Möbius energy and integral Menger curvatures. Their L^2 -gradient flows (if existent) would be parabolic of fractional order somewhere between 2 and 4. The Möbius energy is best

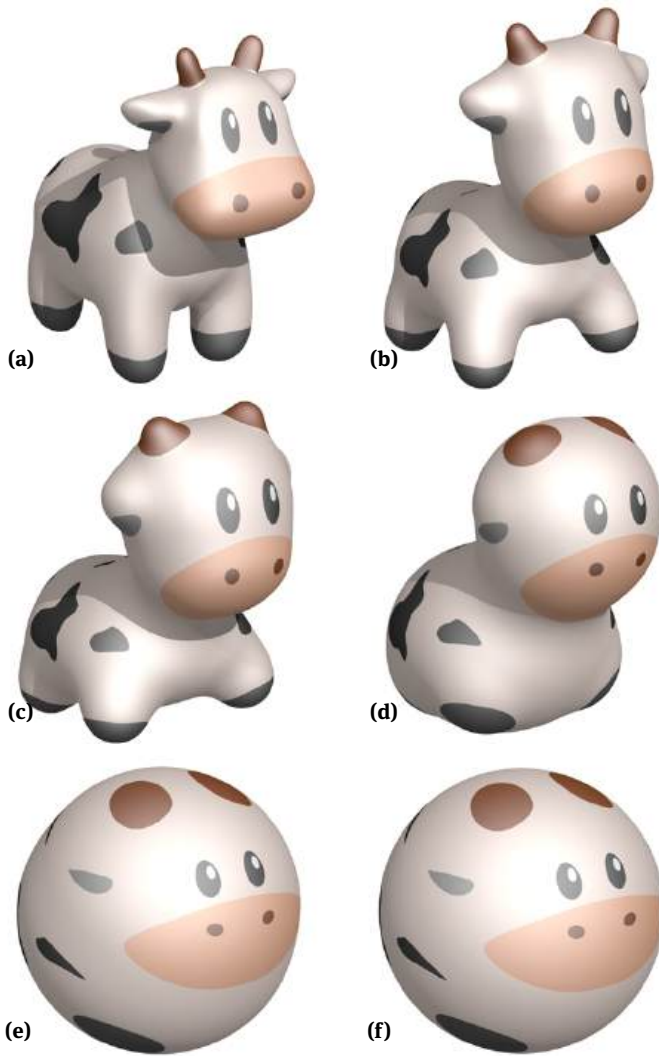


Fig. 4.8: Discrete surface ($\sim 93k$ faces) relaxing under the pseudogradient flow of the Willmore energy subject to an equality constraint on total area: (a) initial condition; (b)–(e) iterations 6, 12, 18, 24, and 30 of the flow. Initial model kindly provided to the public by Keenan Crane.

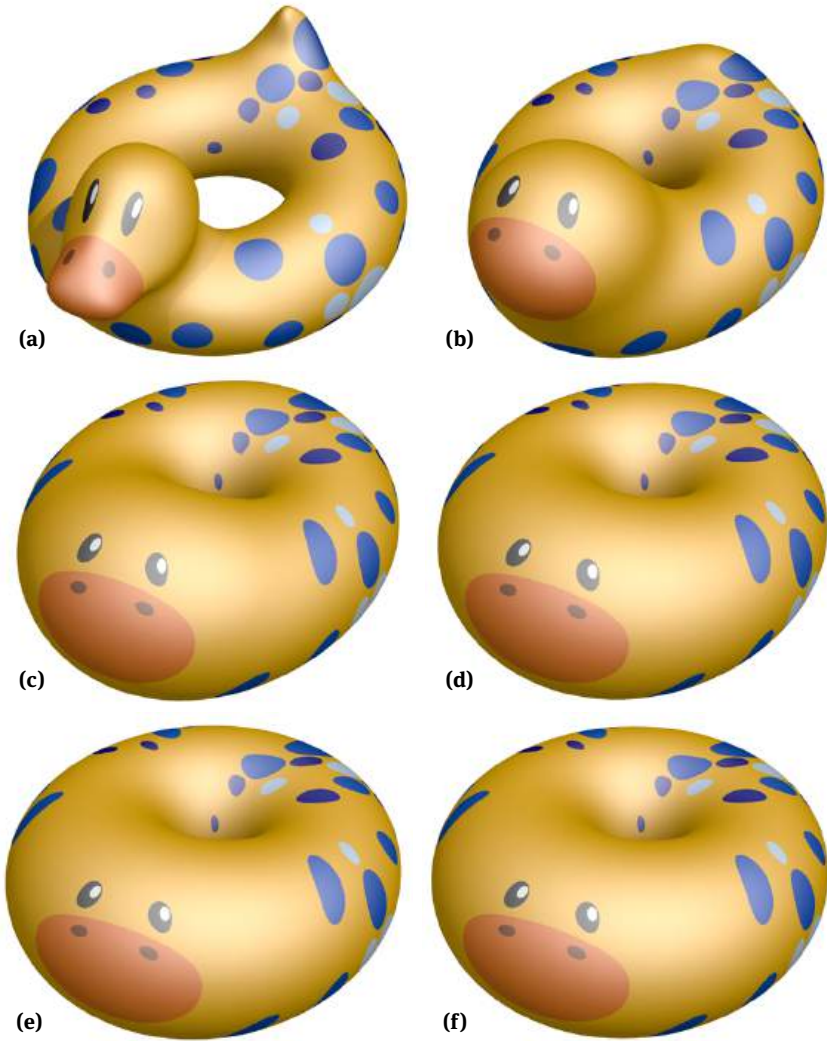


Fig. 4.9: Discrete surface ($\sim 42k$ faces) relaxing under the pseudogradient flow of the Willmore energy subject to an equality constraint on total area: (a) initial condition; (b)–(e) iterations 6, 12, 18, 24, and 30 of the flow. Initial model kindly provided to the public by Keenan Crane.

described on the constraint manifold of an arclength parametrization constraint.^{4,7} Moreover, integral Menger curvatures become arbitrarily small when scaling the knot with large factors (the same occurs for the Euler-Bernoulli energy), and constraints such as total length or parameterization by arclength have to be imposed in order to obtain a well-posed optimization problem. Thus, both energies require, in one way or the other, a nonlinear constraint for using smooth optimization techniques. We expect that their analysis might benefit significantly from pseudogradient descent with respect to suitable (yet to be found) Riesz structures.

Bibliography

- [1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [2] John W. Barrett, Harald Garcke, and Robert Nürnberg. Parametric approximation of Willmore flow and related geometric evolution equations. *SIAM J. Sci. Comput.*, 31(1):225–253, 2008.
- [3] John W. Barrett, Harald Garcke, and Robert Nürnberg. Parametric approximation of isotropic and anisotropic elastic flow for closed and open curves. *Numer. Math.*, 120(3):489–542, 2012.
- [4] Sören Bartels. A simple scheme for the approximation of the elastic flow of inextensible curves. *IMA J. Numer. Anal.*, 33(4):1115–1125, 2013.
- [5] Kenneth A. Brakke. *The motion of a surface by its mean curvature*, volume 20 of *Mathematical Notes*. Princeton University Press, Princeton, N.J., 1978.
- [6] Keenan Crane, Ulrich Pinkall, and Peter Schröder. Robust fairing via conformal curvature flow. *ACM Trans. Graph.*, 32(4), 2013.
- [7] Klaus Deckelnick, Gerhard Dziuk, and Charles M. Elliott. Computation of geometric partial differential equations and mean curvature flow. *Acta Numer.*, 14:139–232, 2005.
- [8] Gerhard Dziuk. Computational parametric Willmore flow. *Numer. Math.*, 111(1):55–80, 2008.
- [9] I. Eckstein, J.-P. Pons, Y. Tong, C.-C. J. Kuo, and M. Desbrun. Generalized surface flows for mesh processing. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, SGP '07, pages 183–192, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.
- [10] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

^{4,7} Indeed, this is one of the reasons why we also discussed arbitrary Riesz manifolds.

- [11] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10:413–432, 1973. Collection of articles dedicated to the memory of George E. Forsythe.
- [12] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry. Vol I*. Interscience Publishers, a division of John Wiley & Sons, New York-London, 1963.
- [13] Serge Lang. *Introduction to differentiable manifolds*. Interscience Publishers (a division of John Wiley & Sons, Inc.), New York-London, 1962.
- [14] Liviu I. Nicolaescu. *Lectures on the geometry of manifolds*. World Scientific Publishing Co. Inc., River Edge, NJ, 1996.
- [15] Ulrich Pinkall and Konrad Polthier. Computing discrete minimal surfaces and their conjugates. *Experiment. Math.*, 2(1):15–36, 1993.

Sebastian Scholtes

Discrete knot energies

Abstract: The present chapter gives an overview on results for discrete knot energies. These discrete energies are designed to make swift numerical computations and thus open the field to computational methods. Additionally, they provide an independent, geometrically pleasing and consistent discrete model that behaves similarly to the original model. We will focus on Möbius energy, integral Menger curvature and thickness.

Keywords: Möbius energy, thickness, ropelength, ideal knot, Menger curvature, polygonal knot, knot energy, Γ -convergence, discrete energy, knots

MSC: 49J45, 57M25, 49Q10, 53A04

5.1 Introduction


In classic knot theory, mathematicians are often interested in knot classes and how to distinguish between them, for example via *knot invariants*. Contrary to this approach, geometric knot theory deals with the specific shape of knots and how to find or compute particularly nice representatives of a given knot class. The exact meaning of nice depends on the context and can vary from applied (see the section on thickness) to theoretical considerations (see the section on integral Menger curvature). To capture the “quality” of a knot, Fukuhara introduced the concept of a *knot energy* (see [27]). An optimal representative is then said to be a minimizer of this energy among all curves of a given knot class. Later on, this approach was further developed by other authors (see [76, 17, 88]) and by now, a functional on the space of knotted curves that is bounded from below and gets infinite as curves approach a self intersection is called a knot energy (see [53]). This definition already includes some helpful ingredients to show that, indeed, minimizers of the energy exist. The *Möbius energy*

$$\mathcal{E}(\gamma) := \int_{\mathbb{S}_L} \int_{\mathbb{S}_L} \left(\frac{1}{|\gamma(t) - \gamma(s)|^2} - \frac{1}{d_{\mathbb{S}_L}(t, s)^2} \right) ds dt,$$

for arc length curves γ , is a particular example of such an energy (see [49]). In this chapter, we additionally consider *integral Menger curvature* and *thickness*. Each of these three energies is connected to curvature. For an overview of regularizing and knot theoretic properties of different curvature energies see [84, 86, 87, 11, 12].

Sebastian Scholtes, Lehrstuhl I für Mathematik, RWTH Aachen University, 52062 Aachen, Germany,
E-mail: sebastian.scholtes@rwth-aachen.de

<https://doi.org/10.1515/9783110571493-004>

Open Access.  © 2018 Sebastian Scholtes, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

One way to think about knot invariants vs. knot energies is that the first one gives rough information about the shape of the knot, for example, by restricting the knot to belong to a certain knot class, while the energy resolves the finer details. For instance, an energy bound can in turn give bounds on the curvature, bi-Lipschitz constant, average crossing number or stick number.

The explicit shape of energy minimizing knots is only known in case of the unknot. For most energies, the energy optimal unknot is proven or expected to be the round circle (see [1]). Therefore, it is important to know how to make the energies accessible to computers and approximate such minimizers. One approach is to investigate so-called *discrete knot energies*, that is energies which are defined on polygonal knots. Here, the goal is to minimize these discrete energies in the class of polygonal knots of a fixed knot class and fixed number of vertices and then prove that these minimizers converge to the minimizer of a “smooth” knot energy within the same knot class. When trying to find such a discrete energy, the most obvious approach would be to restrict the original energy to the space of polygonal knots. However, such an approach does not work, as polygons have infinite energy. This is related to the fact that all three energies are regularising, i.e. curves of finite energy are more regular than the curves for which the energy is well-defined. By borrowing appropriate concepts of curvature from discrete geometry and replacing smooth notions by discrete ones (for example integrals by sums), it is possible to define discrete energies in the same spirit as the original energies. Hence, these energies are not merely discretisations but discrete versions of the original energy. A suitable convergence of energies, which to some extent already includes the convergence of minimizers, is Γ -convergence.

For each of the three energies, there is a section in which we first consider the history of the energies and comment on recent developments. Then we introduce the appropriate discrete energies and explain the connections to the original energies. For the sake of clarity, we present most results for curves of length 1. In the appendix, we give a short introduction to Γ -convergence.

5.1.1 Notation

In the following sections, \mathcal{C} is the space of closed arc length curves with length one and \mathcal{P}_n the subspace of equilateral polygons with n segments. Furthermore, we abbreviate $\mathcal{C}_{k,p} := \mathcal{C} \cap (C^k \cup \bigcup_{n \in \mathbb{N}} \mathcal{P}_n)$, where C^k is the class of k -times continuously differentiable functions. We write $C^{k,\alpha}$ for the class of functions in C^k whose derivatives are Hölder continuous with exponent α and L^q for the Lebesgue spaces of q -integrable functions. By $W^{k,q}$ we denote the standard Sobolev spaces of k -times weakly differentiable closed curves with q -integrable weak derivative. Adding a knot class \mathcal{K} in brackets to a set

of curves restricts this set to the subset of curves that belong to the knot class \mathcal{K} . The circle of length $L > 0$ is denoted by \mathbb{S}_L .

5.2 Möbius Energy

The *Möbius energy*

$$\mathcal{E}(\gamma) := \int_{\mathbb{S}_L} \int_{\mathbb{S}_L} \left(\frac{1}{|\gamma(t) - \gamma(s)|^2} - \frac{1}{d_{\mathbb{S}_L}(t, s)^2} \right) ds dt,$$

is defined on closed rectifiable arc length curves γ of length L . Here, $d_{\mathbb{S}_L}$ is the intrinsic metric on \mathbb{S}_L . This energy was introduced by O'Hara (see [49]) and has the interesting property that it is invariant under Möbius transformations, hence its name. O'Hara could show that finite energy prevents the curve from having self intersections. Later on, the existence of energy minimizers in prime knot classes was proven by Freedman, He and Wang (see [26]), while there is a folklore conjecture, usually attributed to Kusner and Sullivan, questioning the existence in composite knot classes based on computer experiments (see [41]). Additionally, it was shown that the unique absolute minimizer is the round circle (see [26]). Similar uniqueness results are known for broader classes of energies (see [1]). The regularity of minimizers and, more generally, of critical points was investigated and smoothness could be proven (see [26, 32, 63, 64, 14]). Furthermore, it was shown that the Möbius energy of a curve is finite if and only if the curve is simple and the arc length parametrisation belongs to the fractional Sobolev space $W^{3/2,2}$ (see [4]). The gradient flow of the Möbius energy was investigated (see [5, 8]) and results for the larger class of O'Hara's knot energies (see [51, 50, 52]) are available (see [9, 65, 10, 7]).

A discrete version of the Möbius energy, called *minimum distance energy*, was introduced by Simon (see [74]). If p is a polygon with n consecutive segments X_i this energy is defined by

$$\mathcal{E}_{\text{md},n}(p) := \mathcal{U}_{\text{md},n}(p) - \mathcal{U}_{\text{md},n}(g_n) \quad (5.2.1)$$

with

$$\mathcal{U}_{\text{md},n}(p) := \sum_{i=1}^n \sum_{\substack{j=1 \\ X_i, X_j \text{ not adjacent}}}^n \frac{|X_i| |X_j|}{\text{dist}(X_i, X_j)^2},$$

where g_n is the regular n -gon. Note, that this energy is scale invariant. Similar energy functionals for polygons were considered previously (see [27, 16]). There is also a variant of the Möbius energy for graphs (see [36]) and for curves with self intersections (see [24]). It is known that the minimum distance energy is Lipschitz continuous on sublevel sets:

Theorem 5.2.1 (Minimum distance energy is Lipschitz continuous, [75]). *The minimum distance energy $\mathcal{E}_{\text{md},n}$ is continuous on the space of nonsingular polygons with n segments. Moreover, it is Lipschitz continuous on the subspace of polygons of length at least ℓ , whose energy is bounded by a constant E .*

Furthermore, the existence of minimizers for every tame knot class was established:

Theorem 5.2.2 (Existence of discrete minimizers for $\mathcal{E}_{\text{md},n}$, [75]). *For every tame knot class \mathcal{K} there is a minimizer of $\mathcal{E}_{\text{md},n}$ in the set of all polygons with n vertices that belong to \mathcal{K} .*

For this energy, approximation results for suitably inscribed polygons could be shown:

Theorem 5.2.3 (Convergence for inscribed polygons, [61]). *Let $\gamma \in \mathcal{C} \cap C^2$ and let p_n be inscribed polygons that divide γ in n arcs of length $\frac{1}{n}$. Then*

$$|\mathcal{E}(\gamma) - \mathcal{E}_{\text{md},n}(p_n)| \leq \frac{C(\gamma)}{n^{\frac{1}{4}}}.$$

If n is large enough, the constant $C(\gamma)$ can be chosen as

$$C(\gamma) = \frac{290}{\Delta[\gamma]^{\frac{1}{4}}},$$

where $\Delta[\gamma]$ is the thickness of γ (see Section 5.4). Hence,

$$\mathcal{E}_{\text{md},n}(p_n) \rightarrow \mathcal{E}(\gamma).$$

Furthermore, an explicit error bound on the difference between the minimum distance energy of an equilateral polygonal knot and the Möbius energy of a smooth knot, appropriately inscribed in the polygonal knot, could be established in terms of thickness and the number of segments:

Theorem 5.2.4 (Energy bound for inscribed knots, [62]). *Let p be an equilateral polygon of length 1. Then there is an “inscribed” C^2 knot γ_p such that*

$$|\mathcal{E}_{\text{md},n}(p) - \mathcal{E}(\gamma_p)| \leq \frac{C_1(p)}{n^{\frac{1}{4}}} + \frac{C_2(p)}{n} + \frac{C_3(p)}{n^{\frac{5}{4}}} + \frac{C_4(p)}{n^{\frac{7}{4}}} + \frac{C_5(p)}{n^2}$$

and the constants $C_i(p)$ depend in an explicit way on negative powers of $\Delta_n[p]$.

However, from these results it is not possible to infer that the minimal minimum distance energy converges to the minimal Möbius energy in a fixed knot class. For the overall minimizers of the minimum distance energy the following result is known:

Theorem 5.2.5 (Minimizers of md energy, [89, 78, 79]). *The minimizers of the minimum distance energy $\mathcal{E}_{\text{md},n}$ are convex and for $n \in \{4, 5\}$ these minimizers are the regular n -gon.*

This evidence supports the conjecture that the regular n -gon minimizes the minimum distance energy in the class of n -gons. Numerical experiments regarding the minimum distance energy under the elastic flow were carried out (see [33]).

Another, more obvious, discrete version of the Möbius energy was used for numerical experiments (see [39]). This energy, defined on the class of arc length parametrisations of polygons of length L with n segments, is given by

$$\mathcal{E}_n(p) := \sum_{\substack{i,j=1 \\ i \neq j}}^n \left(\frac{1}{|p(a_j) - p(a_i)|^2} - \frac{1}{d(a_j, a_i)^2} \right) d(a_{i+1}, a_i) d(a_{j+1}, a_j), \quad (5.2.2)$$

where the a_i are consecutive points on \mathbb{S}_L and $p(a_i)$ the vertices of the polygon. Also this energy is scale invariant and it is easily seen to be continuous on the space of nonsingular polygons with n segments. The relationship between this discrete Möbius energy and the classic Möbius energy could be established in terms of a Γ -convergence result:

Theorem 5.2.6 (Möbius energy is Γ -limit of discrete Möbius energies, [68]). *For $q \in [1, \infty]$, $\|\cdot\| \in \{\|\cdot\|_{L^q(\mathbb{S}_1, \mathbb{R}^d)}, \|\cdot\|_{W^{1,q}(\mathbb{S}_1, \mathbb{R}^d)}\}$ and every tame knot class \mathcal{K} holds*

$$\mathcal{E}_n \xrightarrow{\Gamma} \mathcal{E} \quad \text{on } (\mathcal{C}_{1,p}(\mathcal{K}), \|\cdot\|).$$

Since we already know that minimizers of \mathcal{E} in prime knot classes exist and are smooth, Theorem A.1.2 implies the convergence of discrete almost minimizers:

Corollary 5.2.7 (Convergence of discrete almost minimizers, [68]). *Let \mathcal{K} be a tame prime knot class, $p_n \in \mathcal{P}_n(\mathcal{K})$ with*

$$\left| \inf_{\mathcal{P}_n(\mathcal{K})} \mathcal{E}_n - \mathcal{E}_n(p_n) \right| \rightarrow 0 \quad \text{and} \quad p_n \rightarrow \gamma \in \mathcal{C}(\mathcal{K}) \text{ in } L^1(\mathbb{S}_1, \mathbb{R}^d).$$

Then γ is a minimizer of \mathcal{E} in $\mathcal{C}(\mathcal{K})$ and $\lim_{n \rightarrow \infty} \mathcal{E}_n(p_n) = \mathcal{E}(\gamma)$.

The result remains true for subsequences, where the number of edges is allowed to increase by more than one for two consecutive polygons. Since all curves are parametrised by arc length, it is not hard to find a subsequence of the almost minimizers that converges in C^0 , but generally this does not guarantee that the limit curve belongs to the same knot class or is parametrised by arc length. For polygons inscribed in a $C^{1,1}$ curve, there is an estimate on the order of convergence:

Proposition 5.2.8 (Order of convergence for Möbius energy, [68]). *Let $\gamma \in C^{1,1}(\mathbb{S}_L, \mathbb{R}^d)$ be parametrised by arc length and $c, \bar{c} > 0$. Then for every $\varepsilon \in (0, 1)$ there is a constant $C_\varepsilon > 0$ such that*

$$|\mathcal{E}(\gamma) - \mathcal{E}_n(p_n)| \leq \frac{C_\varepsilon}{n^{1-\varepsilon}}$$

for every inscribed polygon p_n given by a subdivision $b_k, k = 1, \dots, n$ of \mathbb{S}_L such that

$$\frac{c}{n} \leq \min_{k=1, \dots, n} |\gamma(b_{k+1}) - \gamma(b_k)| \leq \max_{k=1, \dots, n} |\gamma(b_{k+1}) - \gamma(b_k)| \leq \frac{\bar{c}}{n}.$$

This is in accordance with the data from computer experiments, which suggests that the order of convergence should be roughly 1 (see [39]). If no regularity is assumed, the order of convergence might not be under control, but still, the energies converge:

Corollary 5.2.9 (Convergence of Möbius energies of inscribed polygons, [68]). *Let $\gamma \in \mathcal{C}$ with $\mathcal{E}(\gamma) < \infty$ and p_n as in Proposition 5.2.8. Then $\lim_{n \rightarrow \infty} \mathcal{E}_n(p_n) = \mathcal{E}(\gamma)$.*

In contrast to the situation for the minimum distance energy, the overall minimizers of the discrete Möbius energy are known:

Lemma 5.2.10 (Minimizers of discrete Möbius energy, [68]). *The unique minimizer of \mathcal{E}_n in \mathcal{P}_n is the regular n -gon.*

An immediate consequence is the convergence of overall discrete minimizers to the round circle:

Corollary 5.2.11 (Convergence of minimizers to the round circle, [68]). *Let $p_n \in \mathcal{P}_n$ bounded in L^∞ with $\mathcal{E}_n(p_n) = \inf_{\mathcal{P}_n} \mathcal{E}_n$. Then there is a subsequence with $p_{n_k} \rightarrow \gamma$ in $W^{1,\infty}(\mathbb{S}_1, \mathbb{R}^d)$, where γ is a round unit circle.*

One of the main differences between the discrete Möbius energy (5.2.2) and the minimum distance energy (5.2.1) is that bounded minimum distance energy avoids double point singularities, while for (5.2.2) this is only true in the limit. This avoidance of singularities permits to prove the existence of minimizers of the minimum distance energy (5.2.1) via the direct method. This might be harder or even impossible to achieve for the energy (5.2.2). Nevertheless, the relation between the discrete Möbius energy (5.2.2) and the smooth Möbius energy is more clearly visible than for the minimum distance energy (5.2.1), as reflected in Theorem 5.2.6 and Corollaries 5.2.7-5.2.11.

5.3 Integral Menger Curvature

The *integral Menger curvature* was first considered by Mel'nikov (see [47]) as a concept of curvature of a measure that is naturally connected with the Cauchy transform of this measure. For closed arc length curves γ of length L , the integral Menger curvature is

given by

$$\mathcal{M}_s(\gamma) := \int_{\mathbb{S}_L} \int_{\mathbb{S}_L} \int_{\mathbb{S}_L} \kappa^s(\gamma(t), \gamma(u), \gamma(v)) \, dt \, du \, dv,$$

where $s \in (0, \infty)$ and $\kappa(x, y, z)$ is the inverse of the circumradius $r(x, y, z)$ of the three points x, y and z . The integral Menger curvature for $s = 2$ played an important role in the solution of the Painlevé problem, i.e. to find geometric characterisations of removable sets for bounded analytic functions, see [54, 23, 90] for a detailed presentation and references. There is a remarkable theorem, which states that one-dimensional Borel sets in \mathbb{R}^d with finite integral Menger curvature \mathcal{M}_2 are 1-rectifiable (see [42]). These results for \mathcal{M}_2 were later extended to sets of fractional dimension and metric spaces (see [43, 31]). As a consequence, this theorem also ensures that an \mathcal{H}^1 measurable set $E \subset \mathbb{R}^d$ with $\mathcal{M}_2(E) < \infty$ has approximate 1-tangents at \mathcal{H}^1 a.e point.

Complementary to this research, where highly irregular sets are permitted, is the investigation of rectifiable curves with finite \mathcal{M}_s energy. These curves have a classic tangent \mathcal{H}^1 a.e. to begin with. It turns out that for $s > 3$ this guarantees that the curve is simple and that the arc length parametrisation is of class $C^{1,1-3/p}$, which can be interpreted as a geometric Morrey-Sobolev imbedding (see [83]). It could be shown that the space of curves with finite \mathcal{M}_s for $s > 3$ is that of Sobolev-Slobodeckij embeddings of class $W^{2-2/s,s}$ (see [6]) and that polygons have finite integral Menger curvature \mathcal{M}_s exactly for $s \in (0, 3)$ (see [66]). Furthermore, results regarding optimal Hölder regularity could be obtained (see [40]). Related energies have been investigated with regard to their regularizing properties (see [85, 82, 13]).

The movement of quadrilaterals according to their Menger curvature was investigated (see [35]) and computer experiments for the gradient flow of integral Menger curvature were carried out (see [34]). Besides the ad hoc method used there, the only theoretical results regarding discrete versions of the integral Menger curvature that we are aware of can be found in the thesis of the author (see [70]). These results are still subject to ongoing research and are soon to be extended in an article of the author (see [71]). There, the *discrete integral Menger curvature* of a polygon p with consecutive vertices $p(a_i) = x_i$ is defined by

$$\mathcal{M}_{s,n}(p) := \sum_{\substack{i,j,k=1 \\ \#\{i,j,k\}=3}}^n \kappa^s(x_i, x_j, x_k) \prod_{l \in \{i,j,k\}} \frac{|x_l - x_{l-1}| + |x_{l+1} - x_l|}{2}$$

if $x_i \neq x_j$ for $i \neq j$ and $\mathcal{M}_{s,n}(p) = \infty$ else. It is easily seen, that this energy is continuous on the space of nonsingular polygons with n segments. As for the Möbius energy, there is a Γ convergence result:

Theorem 5.3.1 (Menger curvature is Γ -limit of discrete energies, [70]). *For $q \in [1, \infty]$, $\|\cdot\| \in \{\|\cdot\|_{L^q(\mathbb{S}_1, \mathbb{R}^d)}, \|\cdot\|_{W^{1,q}(\mathbb{S}_1, \mathbb{R}^d)}\}$ and every tame knot class \mathcal{K} holds*

$$\mathcal{M}_{s,n} \xrightarrow{\Gamma} \mathcal{M}_s \quad \text{on } (\mathcal{C}_{2,p}(\mathcal{K}), \|\cdot\|).$$

Additionally, it could be shown that the energies of inscribed polygons converge to the energy of the curve if the number of vertices increases:

Corollary 5.3.2 (Convergence for inscribed polygons, [70]). *Let $s \in (0, \infty)$, $\gamma \in \mathcal{C} \cap C^2$ embedded and p_n be inscribed equilateral polygons with n segments. Then $\lim_{n \rightarrow \infty} \mathcal{M}_{s,n}(p_n) = \mathcal{M}_s(\gamma)$.*

5.4 Thickness

The *thickness* $\Delta[\gamma]$ of a curve γ was introduced by Gonzalez and Maddocks (see [29]) as

$$\Delta[\gamma] := \inf_{s \neq t \neq u \neq s} r(\gamma(s), \gamma(t), \gamma(u))$$

and is equivalent to Federer's reach (see [25]). As in the previous section, $r(x, y, z)$ is the circumradius of the three points x, y and z . Geometrically, the thickness of a curve gives the radius of the largest uniform tubular neighbourhood about the curve that does not intersect itself. The *ropelength*, which is length divided by thickness, is scale invariant and a knot is called *ideal* if it minimizes ropelength in a fixed knot class or, equivalently, minimizes this energy amongst all curves in this knot class with fixed length. Ideal knots are of great interest, not only to mathematicians but also to biologists, chemists and physicists, since they exhibit interesting physical features and resemble the time-averaged shapes of knotted DNA molecules in solution (see [80, 37, 38] and [81, 77] for an overview of physical knot theory with applications). The existence of ideal knots in every knot class was settled by different teams of authors (see [19, 30, 28]) and it was found that the unique absolute minimizer is the round circle. Furthermore, this energy is self-repulsive, meaning that finite energy prevents the curve from having self intersections. By now it is well-known that thick curves, or in general manifolds of positive reach, are of class $C^{1,1}$ and vice versa (see [45, 25, 72, 46, 67]). It was shown that ideal links must not be of class C^2 (see [19]) and computer experiments suggest that $C^{1,1}$ regularity is optimal for knots, too (see [88]). Further computer experiments were carried out with the software packages SONO, libbiarc and ridgerunner (see [55, 20, 2]). A previous conjecture (see [19, Conjecture 24]) that ropelength minimizers are piecewise analytic seems to be reversed by numerical results, which indicate that there might be more singularities than previously expected (see [3, 56]). Further interesting properties of critical points as well as the Euler-Lagrange equation were investigated (see [72, 73, 18]).

Another way to write the thickness of a thick arc length curve is

$$\Delta[\gamma] = \min \left\{ \min \text{Rad}(\gamma), 2^{-1} \text{dcsd}(\gamma) \right\} \quad (5.4.1)$$

(see [44, Theorem 1]). The minimal radius of curvature $\min \text{Rad}(\gamma)$ of γ is the inverse of the maximal curvature

$$\max \text{Curv}(\gamma) := \|\kappa\|_{L^\infty} \quad \text{and} \quad \text{dcsd}(\gamma) := \min_{(x,y) \in \text{dcrit}(\gamma)} |y - x|$$

is the doubly critical self distance. The set of doubly critical points $\text{dcrit}(\gamma)$ of a C^1 curve γ consists of all pairs (x, y) where $x = \gamma(t)$ and $y = \gamma(s)$ are distinct points on γ so that

$$\langle \gamma'(t), \gamma(t) - \gamma(s) \rangle = \langle \gamma'(s), \gamma(t) - \gamma(s) \rangle = 0,$$

i.e. s is critical for $u \mapsto |\gamma(t) - \gamma(u)|^2$ and t for $v \mapsto |\gamma(v) - \gamma(s)|^2$.

The *discrete thickness* Δ_n , derived from the representation in (5.4.1), was introduced by Rawdon (see [57]). The curvature of a polygon, localized at a vertex y , is defined by

$$\kappa_d(x, y, z) := \frac{2 \tan(\frac{\varphi}{2})}{\frac{|x-y|+|z-y|}{2}},$$

where x and z are the vertices adjacent to y and $\varphi = \angle(y-x, z-y)$ is the exterior angle at y . We then set

$$\min \text{Rad}(p) := \max \text{Curv}(p)^{-1} := \min_{i=1, \dots, n} \kappa_d^{-1}(x_{i-1}, x_i, x_{i+1})$$

if the polygon p has the consecutive vertices $x_i, x_0 := x_n, x_{n+1} := x_1$. The doubly critical self distance of a polygon p is given as for a smooth curve if we define $\text{dcrit}(p)$ to consist of pairs (x, y) where $x = p(t)$ and $y = p(s)$ and s locally extremizes $u \mapsto |p(t) - p(u)|^2$ and t locally extremizes $v \mapsto |p(v) - p(s)|^2$. Now, Δ_n is defined analogous to (5.4.1) by

$$\Delta_n[p] = \min \left\{ \min \text{Rad}(p), 2^{-1} \text{dcsd}(p) \right\}$$

if all vertices are distinct and $\Delta_n[p] = 0$ if two vertices of p coincide. In a series of works (see [57, 58, 59, 60, 48]) alternative representations and properties of the discrete thickness were established. For example, it was shown that the discrete thickness is continuous:

Theorem 5.4.1 (Discrete thickness is continuous, [57]). *The discrete thickness Δ_n is continuous on the space of simple closed polygons with n -segments with regard to the metric*

$$d(p, q) := \sum_{i=1}^n |x_i - y_i|,$$

where p and q are polygons with vertices x_i and y_i .

Another important result is the existence of discrete minimizers:

Theorem 5.4.2 (Existence of discrete minimizers for Δ_n , [60]). *For every tame knot class \mathcal{K} there is an ideal polygonal knot in $\mathcal{P}_n(\mathcal{K})$.*

More generally, the result remains true if instead of equilateral polygons one takes the space of polygons with a uniform bound on longest to shortest segment length. These results were then used to prove the convergence of ideal polygonal to smooth ideal knots, a result that could later be improved from C^0 convergence to $C^{0,1}$ convergence.

Corollary 5.4.3 (Convergence of ideal polygonal knots, [60, 69]). *Let \mathcal{K} be a tame knot class and $p_n \in \mathcal{P}_n(\mathcal{K})$ bounded in L^∞ with $|\inf_{\mathcal{P}_n(\mathcal{K})} \Delta_n^{-1} - \Delta_n[p_n]^{-1}| \rightarrow 0$. Then, there is a subsequence*

$$p_{n_k} \xrightarrow[k \rightarrow \infty]{W^{1,\infty}(\mathbb{S}_1, \mathbb{R}^3)} \gamma \in \mathcal{C}(\mathcal{K}) \quad \text{with} \quad \Delta^{-1}[\gamma] = \inf_{\mathcal{C}(\mathcal{K})} \Delta^{-1} = \lim_{k \rightarrow \infty} \Delta_{n_k}^{-1}[p_{n_k}].$$

The relationship on a functional level is again captured by a Γ -convergence result:

Theorem 5.4.4 (Inverse thickness is Γ -limit of discrete energies, [69]). *For every tame knot class \mathcal{K} holds*

$$\Delta_n^{-1} \xrightarrow{\Gamma} \Delta^{-1} \quad \text{on } (\mathcal{C}(\mathcal{K}), \|\cdot\|_{W^{1,\infty}(\mathbb{S}_1, \mathbb{R}^3)}).$$

Similar questions for more general energies were considered (see [21, 60]). If the knot class is not fixed, the unique absolute minimizers of Δ_n^{-1} is the regular n -gon:

Proposition 5.4.5 (Minimizers of inverse discrete thickness, [69]). *The unique minimizer of Δ_n^{-1} in \mathcal{P}_n is the regular n -gon.*

A.1 Appendix: Postlude in Γ -convergence

In this section, we repeat some relevant facts on Γ -convergence. For more details, we refer to the books by Dal Maso and Braides (see [22, 15]). This notion of convergence for functionals was introduced by DeGiorgi and is devised in a way, as to allow the convergence of minimizers and even almost minimizers.

Definition A.1.1 (Γ -convergence). Let X be a topological space, $\mathcal{F}, \mathcal{F}_n : X \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. Then \mathcal{F}_n Γ -converges to \mathcal{F} , if

- for every $x_n \rightarrow x$ holds $\mathcal{F}(x) \leq \liminf_{n \rightarrow \infty} \mathcal{F}_n(x_n)$,
- for every $x \in X$ there are $x_n \rightarrow x$ with $\limsup_{n \rightarrow \infty} \mathcal{F}_n(x_n) \leq \mathcal{F}(x)$.

The first inequality is usually called \liminf inequality and the second one \limsup inequality. If the functionals are only defined on subspaces Y and Y_n of X and we extend the functionals by plus infinity on the rest of X , it is enough to show that the \liminf inequality holds for every $x_n \in Y_n$, $x \in X$ and the \limsup inequality for $x \in Y$ and $x_n \in Y_n$ to establish Γ -convergence. We want to use Γ -convergence to ensure that minimizers of the discrete functional \mathcal{F}_n converge to minimizers of the “smooth” functional \mathcal{F} .

Theorem A.1.2 (Convergence of minimizers, [22, Corollary 7.17, p.78]). *Let $\mathcal{F}_n, \mathcal{F} : X \rightarrow \overline{\mathbb{R}}$ with $\mathcal{F}_n \xrightarrow{\Gamma} \mathcal{F}$. Let $\varepsilon_n > 0$, $\varepsilon_n \rightarrow 0$ and $x_n \in X$ with $|\inf \mathcal{F}_n - \mathcal{F}_n(x_n)| \leq \varepsilon_n$. If $x_{n_k} \rightarrow x$, then*

$$\mathcal{F}(x) = \inf \mathcal{F} = \lim_{k \rightarrow \infty} \mathcal{F}_{n_k}(x_{n_k}).$$

Note, that the previous theorem does not imply convergence of a subsequence, but instead assumes that we already have such a sequence to begin with. This fact is often taken care of by an accompanying compactness result.

We prove that the integral Menger curvature converges to the inverse thickness, to acquaint the reader with a particularly simple case of such a convergence theorem:

Lemma A.1.3 (Convergence of integral Menger curvature to inverse thickness). *The integral Menger curvature converges to inverse thickness*

$$\mathcal{M}_s^{\frac{1}{s}} \xrightarrow{\Gamma} \Delta^{-1} \quad \text{on } (\mathcal{C}(\mathcal{K}), \|\cdot\|_{L^\infty}).$$

Let \mathcal{K} be a tame knot class and let $s \in (3, \infty)$. Then, there are minimizers γ_s of \mathcal{M}_s in $\mathcal{C}(\mathcal{K})$. Moreover, we have that

$$\inf_{\mathcal{C}(\mathcal{K})} \mathcal{M}_s^{1/s} \xrightarrow{s \rightarrow \infty} \inf_{\mathcal{C}(\mathcal{K})} \Delta^{-1}$$

and, after translating the minimizers if necessary, there is subsequence such that

$$\gamma_{s_k} \xrightarrow[k \rightarrow \infty]{C^1} \gamma \in \mathcal{C}(\mathcal{K}),$$

where γ is an ideal knot, i.e.

$$\Delta[\gamma]^{-1} = \inf_{\mathcal{C}(\mathcal{K})} \Delta^{-1}.$$

Proof. Using the Hölder inequality, it is easy to see the monotonically increasing convergence

$$\mathcal{M}_s^{\frac{1}{s}}(\gamma) = \|\kappa(\gamma, \gamma, \gamma)\|_{L^s([0,1]^3)} \xrightarrow{s \rightarrow \infty} \|\kappa(\gamma, \gamma, \gamma)\|_{L^\infty([0,1]^3)} = \Delta[\gamma]^{-1}.$$

An application of Fatou's Theorem shows that the functionals are lower semi-continuous with regard to uniform convergence. Hence, we immediately have Γ convergence (see [15, Remark 1.40 (ii), p.35]). For $s > 3$, curves of finite energy are of class C^1 (see [83]), so that we can restrict to this topology. Minimizers of the energies are known to exist and the monotonicity from above gives a uniform bound on, say, $\mathcal{M}_4(\gamma_s)$. These facts together with the L^∞ bound imply C^1 subconvergence to a simple arc length curve of the same length which belongs to the same knot class (see [83]). Now, the result is a consequence of Theorem A.1.2. \square

Bibliography

- [1] Abrams, A., Cantarella, J., Fu, J. H. G., Ghomi, M., and Howard, R. (2003). Circles minimize most knot energies. *Topology*, 42(2):381–394.
- [2] Ashton, T., Cantarella, J., Piatek, M., and Rawdon, E. J. (2011). Knot tightening by constrained gradient descent. *Exp. Math.*, 20(1):57–90.
- [3] Baranska, J., Przytyl, S., and Pieranski, P. (2008). Curvature and torsion of the tight closed trefoil knot. *Eur. Phys. J. B*, 66(4):547–556.
- [4] Blatt, S. (2012a). Boundedness and regularizing effects of O'Hara's knot energies. *J. Knot Theory Ramifications*, 21(1):1250010, 9.
- [5] Blatt, S. (2012b). The gradient flow of the Möbius energy near local minimizers. *Calc. Var. Partial Differential Equations*, 43(3-4):403–439.
- [6] Blatt, S. (2013). A note on integral Menger curvature for curves. *Math. Nachr.*, 286(2-3):149–159.
- [7] Blatt, S. (2016a). The gradient flow of O'Hara's knot energies. arxiv:1601.02840.
- [8] Blatt, S. (2016b). The gradient flow of the Möbius energy: ε -regularity and consequences. arxiv:1601.07023.
- [9] Blatt, S. and Reiter, P. (2008). Does finite knot energy lead to differentiability? *J. Knot Theory Ramifications*, 17(10):1281–1310.
- [10] Blatt, S. and Reiter, P. (2013). Stationary points of O'Hara's knot energies. *Manuscripta Math.*, 140(1-2):29–50.
- [11] Blatt, S. and Reiter, P. (2014a). How nice are critical knots? Regularity theory for knot energies. *Journal of Physics: Conference Series*, 544(1):012020.
- [12] Blatt, S. and Reiter, P. (2014b). Modeling repulsive forces on fibres via knot energies. *Molecular Based Mathematical Biology*, 2:56–72.
- [13] Blatt, S. and Reiter, P. (2015). Towards a regularity theory for integral Menger curvature. *Ann. Acad. Sci. Fenn. Math.*, 40(1):149–181.
- [14] Blatt, S., Reiter, P., and Schikorra, A. (2015). Harmonic analysis meets critical knots. Critical points of the Möbius energy are smooth. *Trans. Amer. Math. Soc.*
- [15] Braides, A. (2002). *Γ -convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford.

- [16] Buck, G. and Simon, J. (1993). Knots as dynamical systems. *Topology Appl.*, 51(3):229–246.
- [17] Buck, G. and Simon, J. K. (1997). Energy and length of knots. In *Lectures at KNOTS '96 (Tokyo)*, volume 15 of *Ser. Knots Everything*, pages 219–234. World Sci. Publ., River Edge, NJ.
- [18] Cantarella, J., Fu, J. H. G., Kusner, R. B., and Sullivan, J. M. (2014). Ropelength criticality. *Geom. Topol.*, 18(4):1973–2043.
- [19] Cantarella, J., Kusner, R. B., and Sullivan, J. M. (2002). On the minimum ropelength of knots and links. *Invent. Math.*, 150(2):257–286.
- [20] Carlen, M. (2010). *Computation and Visualization of Ideal Knot Shapes*. PhD thesis, EPF Lausanne.
- [21] Dai, X. and Diao, Y. (2000). The minimum of knot energy functions. *J. Knot Theory Ramifications*, 9(6):713–724.
- [22] Dal Maso, G. (1993). *An introduction to Γ -convergence*. Progress in Nonlinear Differential Equations and their Applications, 8. Birkhäuser Boston Inc., Boston, MA.
- [23] Dudziak, J. J. (2010). *Vitushkin's conjecture for removable sets*. Universitext. Springer, New York.
- [24] Dunning, R. P. (2011). Optimally immersed planar curves under Möbius energy. *J. Knot Theory Ramifications*, 20(10):1381–1390.
- [25] Federer, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491.
- [26] Freedman, M. H., He, Z.-X., and Wang, Z. (1994). Möbius energy of knots and unknots. *Ann. of Math. (2)*, 139(1):1–50.
- [27] Fukuhara, S. (1988). Energy of a knot. In *A fête of topology*, pages 443–451. Academic Press, Boston, MA.
- [28] Gonzalez, O. and de la Llave, R. (2003). Existence of ideal knots. *J. Knot Theory Ramifications*, 12(1):123–133.
- [29] Gonzalez, O. and Maddocks, J. H. (1999). Global curvature, thickness, and the ideal shapes of knots. *Proc. Natl. Acad. Sci. USA*, 96(9):4769–4773 (electronic).
- [30] Gonzalez, O., Maddocks, J. H., Schuricht, F., and von der Mosel, H. (2002). Global curvature and self-contact of nonlinearly elastic curves and rods. *Calc. Var. Partial Differential Equations*, 14(1):29–68.
- [31] Hahlomaa, I. (2008). Menger curvature and rectifiability in metric spaces. *Adv. Math.*, 219(6):1894–1915.
- [32] He, Z.-X. (2000). The Euler-Lagrange equation and heat flow for the Möbius energy. *Comm. Pure Appl. Math.*, 53(4):399–431.
- [33] Hermes, T. (2008). Repulsive Potentiale in geometrischen Flüssen. Diplomarbeit, RWTH Aachen University.
- [34] Hermes, T. (2012). *Analysis of the first variation and a numerical gradient flow for integral Menger curvature*. PhD thesis, RWTH Aachen University.
- [35] Jecko, T. and Léger, J.-C. (2002). Polygon shortening makes (most) quadrilaterals circular. *Bull. Korean Math. Soc.*, 39(1):97–111.

- [36] Karpenkov, O. N. (2006). The Möbius energy of graphs. *Mat. Zametki*, 79(1):146–149.
- [37] Katritch, V., Bednar, J., Michoud, D., Scharein, R. G., Dubochet, J., and Stasiak, A. (1996). Geometry and physics of knots. *Nature*, 384(6605):142–145.
- [38] Katritch, V., Olson, W. K., Pieranski, P., Dubochet, J., and Stasiak, A. (1997). Properties of ideal composite knots. *Nature*, 388(6638):148–151.
- [39] Kim, D. and Kusner, R. (1993). Torus knots extremizing the Möbius energy. *Experiment. Math.*, 2(1):1–9.
- [40] Kolasinski, S. and Szumańska, M. (2013). Minimal Hölder regularity implying finiteness of integral Menger curvature. *Manuscripta Math.*, 141(1-2):125–147.
- [41] Kusner, R. B. and Sullivan, J. M. (1997). Möbius energies for knots and links, surfaces and submanifolds. In *Geometric topology (Athens, GA, 1993)*, volume 2 of *AMS/IP Stud. Adv. Math.*, pages 570–604. Amer. Math. Soc., Providence, RI.
- [42] Léger, J.-C. (1999). Menger curvature and rectifiability. *Ann. of Math. (2)*, 149:831–869.
- [43] Lin, Y. and Mattila, P. (2001). Menger curvature and C^1 regularity of fractals. *Proc. Amer. Math. Soc.*, 129(6):1755–1762 (electronic).
- [44] Litherland, R. A., Simon, J. K., Durumeric, O. C., and Rawdon, E. J. (1999). Thickness of knots. *Topology Appl.*, 91(3):233–244.
- [45] Lucas, K. R. (1957). Submanifolds of dimension $n-1$ in \mathcal{E}^n with normals satisfying a lipschitz condition. Technical Report 18, University of Kansas. Contract Nonr 58304.
- [46] Lytchak, A. (2005). Almost convex subsets. *Geom. Dedicata*, 115:201–218.
- [47] Mel’nikov, M. S. (1995). Analytic capacity: discrete approach and curvature of measure. *Sbornik: Mathematics*, 186(6):827–846.
- [48] Millett, K. C., Piatek, M., and Rawdon, E. J. (2008). Polygonal knot space near ropelength-minimized knots. *J. Knot Theory Ramifications*, 17(5):601–631.
- [49] O’Hara, J. (1991). Energy of a knot. *Topology*, 30(2):241–247.
- [50] O’Hara, J. (1992a). Energy functionals of knots. In *Topology Hawaii (Honolulu, HI, 1990)*, pages 201–214. World Sci. Publ., River Edge, NJ.
- [51] O’Hara, J. (1992b). Family of energy functionals of knots. *Topology Appl.*, 48(2):147–161.
- [52] O’Hara, J. (1994). Energy functionals of knots. II. *Topology Appl.*, 56(1):45–61.
- [53] O’Hara, J. (2003). *Energy of knots and conformal geometry*, volume 33 of *Series on Knots and Everything*. World Scientific Publishing Co. Inc., River Edge, NJ.
- [54] Pajot, H. (2002). *Analytic capacity, rectifiability, Menger curvature and the Cauchy integral*, volume 1799 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin.
- [55] Pierański, P. (1998). In search of ideal knots. In *Ideal knots*, volume 19 of *Ser. Knots Everything*, pages 20–41. World Sci. Publ., River Edge, NJ.
- [56] Przybył, S. and Pieranski, P. (2014). High resolution portrait of the ideal trefoil knot. *J. Phys. A: Math. Theor.*, 47(28):285201.
- [57] Rawdon, E. J. (1997). *Thickness of polygonal knots*. PhD thesis, University of Iowa.

- [58] Rawdon, E. J. (1998). Approximating the thickness of a knot. In *Ideal knots*, volume 19 of *Ser. Knots Everything*, pages 143–150. World Sci. Publ., River Edge, NJ.
- [59] Rawdon, E. J. (2000). Approximating smooth thickness. *J. Knot Theory Ramifications*, 9(1):113–145.
- [60] Rawdon, E. J. (2003). Can computers discover ideal knots? *Experiment. Math.*, 12(3):287–302.
- [61] Rawdon, E. J. and Simon, J. K. (2006). Polygonal approximation and energy of smooth knots. *J. Knot Theory Ramifications*, 15(4):429–451.
- [62] Rawdon, E. J. and Worthington, J. (2010). Error analysis of the minimum distance energy of a polygonal knot and the Möbius energy of an approximating curve. *J. Knot Theory Ramifications*, 19(8):975–1000.
- [63] Reiter, P. (2009). *Repulsive Knot Energies and Pseudodifferential Calculus. Rigorous Analysis and Regularity Theory for O’Hara’s Knot Energy Family $E^{(\alpha)}$, $\alpha \in [2, 3]$* . PhD thesis, RWTH Aachen University.
- [64] Reiter, P. (2010). Regularity theory for the Möbius energy. *Commun. Pure Appl. Anal.*, 9(5):1463–1471.
- [65] Reiter, P. (2012). Repulsive knot energies and pseudodifferential calculus for O’Hara’s knot energy family $E^{(\alpha)}$, $\alpha \in [2, 3]$. *Math. Nachr.*, 285(7):889–913.
- [66] Scholtes, S. (2011). For which positive p is the integral Menger curvature \mathcal{M}_p finite for all simple polygons? arxiv:1202.0504.
- [67] Scholtes, S. (2013). On hypersurfaces of positive reach, alternating Steiner formulæ and Hadwiger’s Problem. arxiv:1304.4179.
- [68] Scholtes, S. (2014a). Discrete Möbius energy. *J. Knot Theory Ramifications*, 23(9):1450045, 16.
- [69] Scholtes, S. (2014b). Discrete thickness. *Molecular Based Mathematical Biology*, 2(1):73–85.
- [70] Scholtes, S. (2014c). *Geometric Curvature Energies*. PhD thesis, RWTH Aachen University.
- [71] Scholtes, S. (2018). Discrete integral Menger curvature. In preparation.
- [72] Schuricht, F. and von der Mosel, H. (2003). Global curvature for rectifiable loops. *Math. Z.*, 243(1):37–77.
- [73] Schuricht, F. and von der Mosel, H. (2004). Characterization of ideal knots. *Calc. Var. Partial Differential Equations*, 19(3):281–305.
- [74] Simon, J. K. (1994a). Energy functions for polygonal knots. *J. Knot Theory Ramifications*, 3(3):299–320. Random knotting and linking (Vancouver, BC, 1993).
- [75] Simon, J. K. (1994b). Energy functions for polygonal knots. In *Random knotting and linking (Vancouver, BC, 1993)*, volume 7 of *Ser. Knots Everything*, pages 67–88. World Sci. Publ., River Edge, NJ.
- [76] Simon, J. K. (1996). Energy functions for knots: beginning to predict physical behavior. In *Mathematical approaches to biomolecular structure and dynamics (Min-*

- neapolis, MN, 1994), volume 82 of *IMA Vol. Math. Appl.*, pages 39–58. Springer, New York.
- [77] Simon, J. K. (2002). Physical knots. In *Physical knots: knotting, linking, and folding geometric objects in \mathbb{R}^3* (Las Vegas, NV, 2001), volume 304 of *Contemp. Math.*, pages 1–30. Amer. Math. Soc., Providence, RI.
 - [78] Speller, R. (2007). Convexity and Minimum Distance Energy. Reu project, Department of Mathematics, California State University, San Bernardino.
 - [79] Speller, R. (2008). Knots and Minimum Distance Energy. Technical report, Smith College, Mathematics and Statistics Department, Northampton, MA.
 - [80] Stasiak, A., Katritch, V., Bednar, J., Michoud, D., and Dubochet, J. (1996). Electrophoretic mobility of DNA knots. *Nature*, 384(6605):122.
 - [81] Stasiak, A., Katritch, V., and Kauffman, L. H. (1998). *Ideal knots*, volume 19 of *Series on Knots and Everything*. World Scientific Publishing Co. Inc., River Edge, NJ.
 - [82] Strzelecki, P., Szumańska, M., and von der Mosel, H. (2009). A geometric curvature double integral of Menger type for space curves. *Ann. Acad. Scient. Fenn. Math.*, 34:195–214.
 - [83] Strzelecki, P., Szumańska, M., and von der Mosel, H. (2010). Regularizing and self-avoidance effects of integral Menger curvature. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (5)*, Vol. IX:1–43.
 - [84] Strzelecki, P., Szumańska, M., and von der Mosel, H. (2013). On some knot energies involving Menger curvature. *Topology Appl.*, 160(13):1507–1529.
 - [85] Strzelecki, P. and von der Mosel, H. (2007). On rectifiable curves with L^p -bounds on global curvature: self-avoidance, regularity, and minimizing knots. *Math. Z.*, 257(1):107–130.
 - [86] Strzelecki, P. and von der Mosel, H. (2013). Menger curvature as a knot energy. *Phys. Rep.*, 530(3):257–290.
 - [87] Strzelecki, P. and von der Mosel, H. (2014). How averaged menger curvatures control regularity and topology of curves and surfaces. *J. Phys.: Conf. Ser.*, 544(1):012018.
 - [88] Sullivan, J. M. (2002). Approximating ropelength by energy functions. In *Physical knots: knotting, linking, and folding geometric objects in \mathbb{R}^3* (Las Vegas, NV, 2001), volume 304 of *Contemp. Math.*, pages 181–186. Amer. Math. Soc., Providence, RI.
 - [89] Tam, J. (2006). The minimum distance energy for polygonal unknots. REU Project, California State University San Bernardino (CSUSB).
 - [90] Tolsa, X. (2014). *Analytic capacity, the Cauchy transform, and non-/homogeneous Calderón-Zygmund theory*, volume 307 of *Progress in Mathematics*. Birkhäuser/Springer, Cham.

Radmila Sazdanovic

Khovanov homology and torsion

Abstract: This article offers an introduction to Khovanov homology with a brief overview of the developments in low dimensional topology it has inspired, and relations with other link homology theories. The emphasis is placed on the role of torsion in Khovanov homology and several open problems.

Keywords: Knot theory, Jones polynomial, Khovanov homology, torsion, categorification.

MSC: 57M25, 57M27

6.1 Introduction

Categorification is a powerful approach to solving problems in many different areas of mathematics. The core idea is that classical objects endowed with some integral structure (integers, polynomials with integer coefficients, vector spaces, etc.) can sometimes be viewed as shadows of new, algebraically richer objects [1, 23]. These newly constructed objects, which we refer to as categorifications, should contain more information than the ones we started with. Often categorifications are interesting on their own, leading to beautiful and structurally deep mathematics.

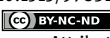
Since the discovery of the Jones polynomial in 1984 [18], the theory of quantum invariants for knots, links and 3-manifolds has been developed rapidly, first through many generalizations such as the HOMFLY-PT polynomial, then via representation-theoretical interpretations and Reshetikhin-Turaev invariants. At the end of the twentieth century, M. Khovanov [20] constructed a new kind of link invariant: a homology theory which lifts the properties of the Jones polynomial and carries an additional rich algebraic structure.

Khovanov homology is a stronger knot invariant than the Jones polynomial, better at distinguishing knots than the Jones polynomial it categorifies. It turns out that the Khovanov homology of an alternating knot is determined by the Jones polynomial and the signature of the knot. Khovanov homology contains more information than the Jones polynomial when it comes to non-alternating knots.

In 2010 P. Kronheimer and T. Mrowka proved that Khovanov homology distinguishes the unknot [27], that is, any knot K whose Khovanov homology is trivial is an unknot. The corresponding question for the Jones polynomial remains open. Although the Jones polynomial is known not to distinguish the 2-component unlink, Khovanov homology does [13].

Radmila Sazdanovic, Department of Mathematics, Box 8205 North Carolina State University, Raleigh NC 27695, E-mail: rsazdanovic@math.ncsu.edu

<https://doi.org/10.1515/9783110571493-005>

 Open Access.  © 2018 Radmila Sazdanovic, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

Among the first of many notable applications of Khovanov homology we emphasize J. Rasmussen's [37] concordance invariant, called the s -invariant, that enabled him to give a purely combinatorial proof of the Milnor conjecture, also known as the Kronheimer-Mrowka theorem.

Theorem 6.1.1 (J. Rasmussen). *The s -invariant provides a lower bound on the slice genus of a knot.*

Theorem 6.1.2 (Milnor conjecture). *The slice genus of the $T(p, q)$ torus knot is equal to $\frac{1}{2}(p-1)(q-1)$.*

Using the same s -invariant A. Shumakovitch improved the slice-Bennequin inequality for some classes of knots [40].

Khovanov's construction led to many significant advances and new discoveries. There are several extensions of Khovanov homology, including odd and reduced Khovanov homology, as well as symplectic and bordered Khovanov homology which have been studied since. Other knot polynomials such as HOMFLYPT, Kauffman 2-variable, and $\mathfrak{sl}(N)$ link polynomial have also been categorified. Motivated by the Floer homology, Ozsváth, Szabó, and Rasmussen constructed knot Floer homology that categorifies the Alexander polynomial. Soon after, this categorification was given a purely combinatorial description by Manolescu, Ozsvath and Sarkar [30]. This abundance of new structures has provided many answers to classical questions in 3- and 4-dimensional topology.

6.2 Definition and structure of Khovanov link homology

Khovanov link homology is a functor that assigns to every link L bigraded homology groups $Kh(L)$ whose Euler characteristic is the Jones polynomial of L , and to link cobordisms homomorphisms of homology groups. Here we define Khovanov homology in broad strokes, mostly relying on Bar Natan's cube construction [3]. For details check one of the following excellent resources [1, 3, 44, 45, 43], each of which emphasizes different aspects of Khovanov link homology.

This construction is based on the Kauffman state sum expression for the Jones polynomial, known as the Kauffman bracket polynomial [19]. This formula can be used instead of the skein relation to express the Jones polynomial as a sum of powers of binomials $(q + \frac{1}{q})$, which is the Jones polynomial of the unknot.

Let D be an n -crossing diagram of a knot K . A Kauffman state s of the knot diagram D is a function from the set of crossings $\{C_i\}_{i=1}^n$ of D to the set $\{0, 1\}$. Diagrammatically, we assign to each crossing of D a marker according to the convention in Figure



Fig. 6.1: The smoothings of a crossing and the corresponding arcs in the Kauffman state circles.

6.1, where each crossing is replaced by two parallel arcs. The Kauffman state can be viewed as a collection of circles D_s embedded in the plane obtained by smoothing each crossing in D . Let $|D_s|$ be the number of circles in D_s .

There are 2^n Kauffman states that can be conveniently organized as the vertices of an n -dimensional unit cube $I^n = [0, 1]^n$. Each vertex $V \in I^n$ of the cube corresponds to one Kauffman state s : the state that associates a 0 or 1 smoothing to a crossing C_i depending on whether the i -th coordinate of that vertex is 0 or 1.

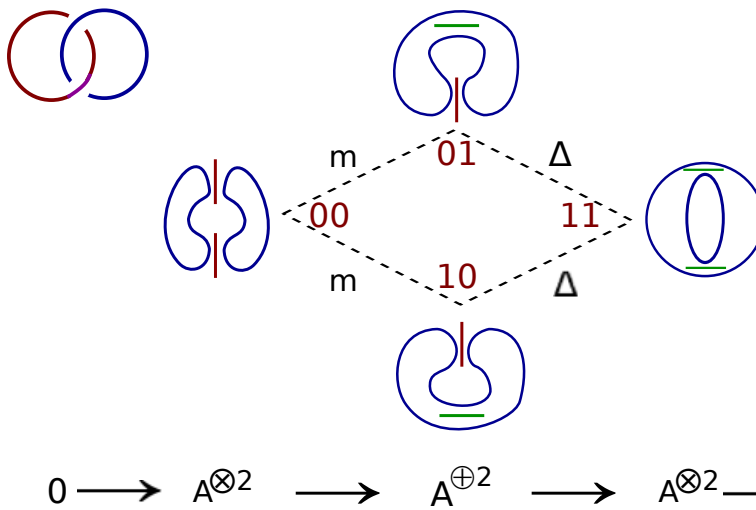


Fig. 6.2: Cube of resolutions in Khovanov homology of a Hopf link.

Let $\mathcal{A}_2 = \mathbb{Z}[x]/(x^2 = 0)$ be a graded module with 1 in bidegree $(0, 1)$ and x in bidegree $(0, -1)$. \mathcal{A}_2 has the Frobenius algebra structure including the multiplication

and comultiplication maps:

$$\begin{array}{ll}
 m : \mathcal{A}_2 \otimes \mathcal{A}_2 \rightarrow \mathcal{A}_2 & \begin{array}{l} 1 \otimes 1 \mapsto 1 \\ 1 \otimes x, x \otimes 1 \mapsto x \\ x \otimes x \mapsto 0 \end{array} \\
 \Delta : \mathcal{A}_2 \rightarrow \mathcal{A}_2 \otimes \mathcal{A}_2 & \begin{array}{l} 1 \mapsto 1 \otimes x + x \otimes 1 \\ x \mapsto x \otimes x \end{array}
 \end{array}$$

To each vertex assign a module $C(D_s) = C(D(V)) = \mathcal{A}_2^{\otimes |D_s|} [h(s)] \{q(s)\}$ with appropriate homological $[h(s)]$ and height (quantum) $\{q(s)\}$ shifts [3]. Next, let e be the edge of the cube I^n between vertices V and W whose coordinates are the same except in one place where W has entry 1 and V has 0. We define the *per edge map* $d_e : C(D(V)) \rightarrow C(D(W))$ for each such edge e in the following way:

- if the number of Kauffman circles in the Kauffman state of vertex W is smaller than the number of circles in the state associated to V then d_e is a multiplication on the corresponding copies of \mathcal{A}_2
- otherwise d_e is a comultiplication Δ on the respective tensor factors and all other tensor factors are mapped by the identity map.

The structure we have described is referred to as the *cube of resolutions* and the Khovanov chain complex is its total complex, Figure 6.2. The Khovanov chain groups $KhC^i(D)$ are obtained by taking direct sums of modules along the hyperplanes orthogonal to the main diagonal and defining the differential $d^i : KhC^i(D) \rightarrow KhC^{i+1}(D)$ to be a signed sum of edge maps d_e connecting vertices in I^n with i ones in their Kauffman states to the ones with $i + 1$ ones that differ only in one value of the Kauffman state. The homology of the Khovanov chain complex for a link L is called Khovanov link homology and denoted by $Kh(L)$.

Theorem 6.2.1. *Khovanov homology $Kh(L)$ of a link L is independent of the choice of diagram i.e. Khovanov homology is a link invariant.*

$Kh(L)$ recovers information about L known from the Jones polynomial, for example:

- The span of the Jones polynomial of L gives a lower bound on the crossing number of L . The span of $Kh(L)$, the difference between the highest and the lowest non-trivial quantum grading, gives a lower bound on the crossing number as well.
- Not only does the Euler characteristic of the Khovanov homology equal the Jones polynomial; in fact, there is a long exact sequence in Khovanov homology which lifts the skein relation which defines the Jones polynomial.

Khovanov homology is indeed much richer in structure than the Jones polynomial. In particular, it carries the following additional structures:

- torsion, applications of which are discussed in the remainder of this article;
- the Lee and other spectral sequences;
- Steenrod squares and other higher homological operations by Lipshitz-Sarkar, Kriz-Kriz-Po;
- due to the functorial nature of Khovanov homology, invariants of cobordisms exploited by M. Jacobson, D. Bar Natan, M. Khovanov, and Clark-Morrison-Walker.

Khovanov homology also appears in mathematical physics in the work of Gukov, Schwarz, Vafa, and Witten, where it admits a description in terms of so-called BPS states.

6.3 Torsion of Khovanov link homology

Recall that the Jones polynomial equals the Euler characteristic of Khovanov homology, so the Jones polynomial depends only on the free part of Khovanov homology. This section is devoted to results concerning torsion of Khovanov homology intertwined with the ideas and methods that could be useful in advancing our understanding of torsion and the additional information it may carry about knots and their cobordisms. Although computations hint at the abundance of torsion, describing and understanding torsion in Khovanov homology is still elusive. A great deal has been written about torsion of order 2; however, torsion of higher order is still a mystery [3, 5, 11]. Torsion in Khovanov homology is understood only for some classes of knots and links [2, 34, 36, 41, 42].

Conjecture 6.3.1 (A. Shumakovitch). *The Khovanov homology of every link except the unknot, the Hopf link, their disjoint unions and connect sums, has torsion of order 2.*

As another illustration of the importance of torsion in Khovanov homology, note that the affirmative resolution of Shumakovitch's conjecture would provide a new proof that the Khovanov homology is an unknot-detector. The only currently known proof of this fact due to P. Kronheimer and T. Mrowka [27] relies on the relations with the instanton Floer homology.

Goal 1. Characterize and compute torsion in Khovanov homology.

Approach 1. Spectral sequences for constructing odd torsion in link homology

A first step to understanding the torsion in Khovanov homology is to construct links whose Khovanov homology has torsion of a prescribed order. A. Shumakovitch recently used the Bockstein spectral sequence to compute all of the torsion in the Khovanov homology of alternating knots. In particular, he proves Conjecture 6.3.1 for alternating knots and shows that Khovanov homology of thin knots has only \mathbb{Z}_2 -torsion.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
43															1_3
41													1	1	
39													1_2	1_5	1
37											1_2	2	1		
35										2	1_2	1_2	1_5	1	
33								1		$1, 1_2$	1				
31						1		$1, 1_2$	2						
29						1	1	1_2	1						
27				1	1		1								
25				1_2	1										
23			1												
21	1														
19	1														

Fig. 6.3: Khovanov homology of torus knot $T(6, 5)$. Integer entry represents the rank of the group; integer with subscript stands for torsion of that order. For example, $Kh^{11,35}(T(6, 5)) = \mathbb{Z}_2 \oplus \mathbb{Z}_5$.

Example 1. Examples of knots with \mathbb{Z}_5 -torsion in their Khovanov homology include the 5-strand torus knots: $T(6, 5)$, $T(7, 5)$, $T(8, 5)$ and $T(9, 5)$ but in general appear to be quite rare, see Figure 6.3. Based on the results in [36], J. Przytycki and the author predicted in 2012 that the Khovanov link homology of the positive adequate 36-crossing knot K given by the closure of the braid $s_1^2 s_2^2 s_3^3 s_2^2 s_1 s_3 s_2^2 s_4 s_3 s_1^2 s_2^2 s_1^3 s_2^3 s_1^3 s_2^2 s_3^2 s_4^2 s_3^2$ will contain \mathbb{Z}_5 torsion. This conjecture was confirmed by A. Shumakovitch using *JavaKh* [5].

D. Bar-Natan showed that Khovanov homology of the $T(8, 7)$ torus knot contains \mathbb{Z}_7 , \mathbb{Z}_5 , \mathbb{Z}_4 , and \mathbb{Z}_2 -torsion. This 48 crossing knot reaches the limits of current computational resources [26, 6]. Extending theoretical results to non-alternating links in order to find \mathbb{Z}_p -torsion for odd primes p will require either a more sophisticated spectral sequence or perhaps arguments of a different nature altogether.

Approach 2. Relations between Khovanov link, Hochschild algebra and chromatic graph homology theories

Hochschild homology is a well-understood cyclic homology theory of associative algebras. Przytycki observed, somewhat surprisingly, that Khovanov homology and Hochschild homology share a common structure [35]. The limit of the Khovanov homology of $T(2, n)$ torus links when n is large can be interpreted as the Hochschild homology of the algebra $\mathcal{A}_2 = \mathbb{Z}[x]/(x^2 = 0)$ and some of the torsion of Hochschild homology can be seen in Khovanov homology of $T(2, n)$.

The cyclic nature of Hochschild homology can be interpreted pictorially. The n -th chain group can be pictured as an n -cycle. The differential corresponds to contracting the edges of n -cycle one by one to get an $(n - 1)$ -cycle and multiplying the elements in the copies of \mathcal{A} labelling the endpoints of the contracted edge.

For every planar diagram D of a link L there exists an associated planar graph $\Gamma(D)$ obtained via the checkerboard black and white coloring of regions of D . Vertices of $\Gamma(D)$ are black regions of D and edges correspond to crossings of D . We generalize this construction in the following way.

Let D be a knot diagram. The Kauffman state graph $\Gamma(D_A)/\Gamma(D_B)$ contains a vertex for each circle in the all- A /all- B Kauffman state D_A/D_B and an edge if two circles in D_A/D_B have arcs obtained by smoothing a single crossing in D . Checkerboard colorings of an alternating knot diagram correspond to $\Gamma(D_A)$ and $\Gamma(D_B)$.

The graph $\Gamma(D_A)$ corresponding to the $T(2, n)$ torus link and Kauffman state sending all crossings to 1 is just an n -cycle. The chromatic graph homology $HGR_{\mathcal{A}_2}(\Gamma)$ is one of several categorifications of polynomial graph invariants, and one of two categorifications of the chromatic polynomial [10, 15]. Given a graph Γ the chromatic graph polynomial $P(\Gamma)$ satisfies the recursive deletion-contraction relation

$$P(\Gamma) = P(\Gamma - e) - P(\Gamma/e), \quad (6.3.1)$$

where $\Gamma - e$ is the graph obtained by deleting an edge e from a graph Γ , and Γ/e is the graph obtained by contracting the same edge. After categorification, this relation lifts to long exact sequence of chromatic graph cohomology groups [15]:

$$\dots \rightarrow HGR_{\mathcal{A}_2}^i(\Gamma) \rightarrow HGR_{\mathcal{A}_2}^i(\Gamma - e) \rightarrow HGR_{\mathcal{A}_2}^i(\Gamma/e) \rightarrow HGR_{\mathcal{A}_2}^{i+1}(\Gamma) \rightarrow \dots \quad (6.3.2)$$

The construction of $HGR_{\mathcal{A}_2}(\Gamma)$ follows that of Khovanov: a bigraded homology theory $HGR_{\mathcal{A}_2}^{i,j}(\Gamma)$ is associated to a graph Γ and a commutative (graded) algebra \mathcal{A}_2 , in such a way that its (graded) Euler characteristic is the value of the chromatic polynomial at the graded dimension of the algebra \mathcal{A}_2 .

In terms of the chromatic graph homology, the relation between Khovanov and Hochschild homology can be expressed by saying that the chromatic graph homology of the cycle of length n is the Hochschild homology of algebra $\mathcal{A}_2 = \mathbb{Z}[x]/(x^2 = 0)$ through a range of dimensions increasing with n .

Theorem 6.3.2 (Correspondence between Khovanov and chromatic homology). *Let $\Gamma(D_A)$ be the graph associated to diagram D of a link L whose girth, the length of the shortest cycle, $\ell = \ell(G) > 1$. Then the first ℓ Khovanov homology $Kh(L)$ groups of L and chromatic graph homology $HGR_{\mathcal{A}_2}(\Gamma(D_A))$ have isomorphic torsion [34].*

This point of view enables us to describe some \mathbb{Z}_2 -torsion in the Khovanov homology of certain classes of links via torsion in chromatic homology [2, 14, 34, 36] expressed in terms of combinatorial data of one of the corresponding graphs. For example, the torsion depends on the *cyclomatic number* that is the rank of the first homology of that graph as a 1-dimensional CW-complex.

Proposition 6.3.1. Given a diagram D of a knot or link L such that the corresponding graph $\Gamma(D_A)$ is a connected graph with no loops. Let $\Gamma'(D_A)$ be the graph obtained from

$\Gamma(D_A)$ by removing all multiple edges and $p_1 = p_1(\Gamma'(D_A))$ be the cyclomatic number of $\Gamma'(D_A)$. Then

$$\begin{aligned} \text{tor}H_{n-2, n+2|D_{s+}|-4}(L) &= \begin{cases} \mathbb{Z}_2, & \text{if } \Gamma(D_A) \text{ has an odd cycle;} \\ 0, & \text{if } \Gamma(D_A) \text{ is a bipartite graph.} \end{cases} \\ \text{tor}H_{n-4, n+2|D_{s+}|-8}(L) &= \begin{cases} \mathbb{Z}_2^{p_1-1}, & \text{if } \Gamma'(D_A) \text{ has an odd cycle;} \\ \mathbb{Z}_2^{p_1}, & \text{if } \Gamma'(D_A) \text{ is a bipartite graph.} \end{cases} \end{aligned}$$

This proposition establishes Shumakovitch's conjecture for all adequate and semi-adequate knots whose corresponding graphs have odd cycles. Chromatic homology, like Khovanov homology of alternating links, is supported only along two diagonals [9], with torsion appearing only on one. An interesting computational problem would be to explicitly compute all torsion along the whole diagonal in chromatic homology; pulling this information back via Theorem 6.3.2 would greatly extend our current knowledge of the torsion in Khovanov homology.

Approach 3. Compute torsion via Khovanov tangle invariants

In what we have described so far, computing Khovanov homology requires global knowledge of the topology of Kauffman states. Instead, one can take a different approach and decompose a link diagram into local pieces called *tangles*. While it is straightforward to piece together tangles to obtain a link, it is quite challenging to piece together any of the existing Khovanov tangle invariants [21, 22, 4] to obtain the Khovanov homology of the link. Roberts [38, 39] constructs a Khovanov tangle invariant whose algebraic structure is inspired by the bordered Heegaard Floer homology. Most importantly, Roberts obtains gluing formulas which he uses to construct the Khovanov homology of a link from the Khovanov invariants of the tangles in its decomposition. This approach may be most suitable for studying the Khovanov homology of classes of links with especially nice tangle decompositions, such as almost-alternating links and closed 3-braids.

Goal 2. Describe topology of the link in terms of torsion of Khovanov homology

With a discovery of the new knot invariant one of the most intriguing questions is determining how much it can tell us about knots. In our framework, this translates into the question: what does torsion in Khovanov homology reveal about topological properties of the link. As mentioned above, it is conjectured that the existence of torsion in the Khovanov homology of a knot implies the knot is nontrivial. Moreover, computations indicate a potential relationship between the existence of \mathbb{Z}_p -torsion in the Khovanov homology of a link and the braid index of the link. However, the infinite family of links with braid index four containing \mathbb{Z}_{2^s} -torsion with $s \leq 23$ constructed in [31] provides a counterexample to the PS-conjecture [36] and hints that such a relation, if it exists, will be more subtle.

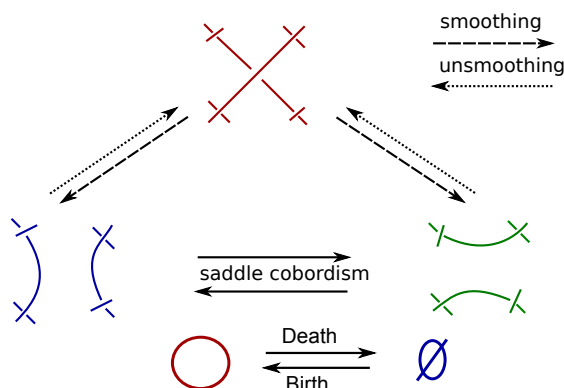


Fig. 6.4: Cobordisms of smoothings of a crossing.

6.4 Homological invariants of alternating and quasi-alternating cobordisms

Khovanov homology is functorial, up to overall minus sign, under link cobordisms. However, in general, it is extremely difficult to compute homomorphisms of Khovanov homology induced by link cobordisms.

Notice that the number of spanning trees $s(\Gamma)$ in a graph Γ satisfies a recursive relation similar to the deletion-contraction formula for the chromatic polynomial (6.3.2):

$$s(\Gamma) = s(\Gamma - e) + s(\Gamma/e). \quad (6.4.1)$$

Spanning trees of $\Gamma(D_A)$ play a fundamental role in link homology [8, 46]. The reduced Khovanov homology of L can be computed via a complex generated by spanning trees of $\Gamma(D_A)$ but the combinatorial form of the differential is currently unknown. The same property holds for the knot Floer homology of L . When D is a minimal alternating diagram, differentials in both complexes are zero, and the rank of reduced Khovanov and knot Floer homologies of L equals the number of spanning trees in $\Gamma(D_A)$, and also the determinant of L [32].

Goal 3. Compute homomorphisms of Khovanov homology induced by “alternating” cobordisms between alternating links

The spanning tree model gives a good inductive realization of Khovanov homology, and maps induced by alternating cobordisms could be effectively computable.

The *alternating cobordisms* are those generated by the cobordisms illustrated in 6.4: they arise from smoothing a crossing, saddles, and births and deaths of circles.

Khovanov-Rozansky $\mathfrak{sl}(n)$ link homology [24, 25] and its variation due to Mackaay, Stošić, Vaz [28], are based on homology theories of planar graphs, functorial under foam cobordisms (in $\mathbb{R}^2 \times [0, 1]$) between these graphs. Similarly one could search for

an interpretation of alternating cobordisms, which are realized by surfaces embedded in \mathbb{R}^4 , in terms of cobordisms between checkerboard graphs of alternating diagrams. Cobordisms between planar graphs induced by alternating cobordisms will resemble foams. The same idea could be employed towards computing maps between link Floer homologies of alternating links induced by alternating cobordisms, or the maps induced by lifts of alternating cobordisms to the double branch cover.

Goal 4. Find a combinatorial framework for quasi-alternating knots

Quasi-alternating links are generalizations of alternating links, defined in an inductive way.

The set \mathcal{Q} of quasi-alternating links is the smallest set of links including the unknot and satisfying the following property: If a link L has a diagram D with a crossing c such that both smoothings, L_0 and L_1 , of c are in \mathcal{Q} , and $\det(L) = \det(L_0) + \det(L_1)$, then L is in \mathcal{Q} [33, 7, 29].

The notion of quasi-alternating link is hard to capture combinatorially. The initial search for quasi-alternating link families was performed using the knot theory software *LinKnot* [17, 16] and the computational results have already been used to show that certain knots and links are not quasi-alternating [12]. It would be very exciting to find a combinatorial model for some large class of quasi-alternating links, incorporating inductive steps in their construction. Such a combinatorial model should also give a generalization of planar graphs (checkerboard graphs of alternating link diagrams) to more subtle higher-dimensional structures.

Furthermore, the Khovanov and knot Floer homology of quasi-alternating links are easy to compute, due to the vanishing of the differentials in short exact sequences that build up to homology groups of these links. Quasi-alternating cobordisms between quasi-alternating links can be defined as compositions of cobordisms built out of quasi-alternating crossings.

Acknowledgment: The author would like to thank S. Jablan, M. Khovanov, and J. Przytycki for their support and guidance that have been invaluable, and A. Lowrance for helpful conversations and sharing his ideas about Goal 1, Approach 6.3. Many thanks to the anonymous referee for corrections.

Funding: While writing this paper the author's research was partially supported by the Simons Collaboration Grant 318086.

Bibliography

- [1] M. M. Asaeda and M. Khovanov. Notes on link homology. In *Low-Dimensional Topology*, IAS/Park City Mathematics Series, pages 139–196. AMS, 2009.
- [2] M. M. Asaeda and J.H. Przytycki. Khovanov homology: torsion and thickness.

- In *Advances in Topological Quantum Field Theory*, pages 135–166. Kluwer Acad. Publ., 2004.
- [3] D. Bar-Natan. On Khovanov’s categorification of the Jones polynomial. *Alg. Geom. Topol.*, 2:337–370, 2002.
 - [4] D. Bar-Natan. Khovanov’s homology for tangles and cobordisms. *Geom. Topol.*, 9:1443–1499, 2005.
 - [5] D. Bar-Natan and J. Green. *JavaKh* - a fast program for computing Khovanov homology. <http://katlas.math.utoronto.ca/wiki/KhovanovHomology>, 2005. Part of the KnotTheory Mathematica Package.
 - [6] C. Cha and C. Livingston. Knotinfo: Table of knot invariants. <http://www.indiana.edu/~knotinfo>, 2016.
 - [7] A. Champanerker and I. Kofman. Twisting quasi-alternating links. *Proceedings of the American Mathematical Society*, 137:2451–2458, 2008.
 - [8] A. Champanerker and I. Kofman. Spanning trees and Khovanov homology. *Proceedings of the American Mathematical Society*, 137:2157–2167, 2009.
 - [9] M. Chmutov, S. Chmutov, and Y. Rong. Knight move in chromatic cohomology, *European Journal of Combinatorics*, 29(1):311–321, 2008.
 - [10] M. Eastwood and S. Huggett. Euler characteristics and chromatic polynomials. *European Journal of Combinatorics*, 28(6):1553–1560, 2007.
 - [11] W. D. Gillam. Knot homology of $(3, m)$ torus knots. *J. Knot Th. and Ramif.*, 21(8):1250072–1–21, 2012.
 - [12] J. Greene. Homologically thin, non-quasi-alternating links. *Math. Res. Lett.*, 17(1):3949, 2010.
 - [13] M. Hedden and Y. Ni. Khovanov module and the detection of unlinks. *Geometry and Topology*, 17(5):3027–3076, 2013.
 - [14] L. Helme-Guizon, J.H. Przytycki, and Y. Rong. Torsion in graph homology. *Fundamenta Mathematicae*, 190:139–177, 2006.
 - [15] L. Helme-Guizon and Y. Rong. A categorification for the chromatic polynomial. *Algebraic and Geometric Topology (AGT)*, 5:1365–1388, 2005.
 - [16] S. Jablan and R. Sazdanovic. Linknot. <http://www.mi.sanu.ac.rs/vismath/linknot/index.html>, 2007. *Mathematica* package and e-book *LinKnot- Knot Theory by Computer*.
 - [17] S. Jablan and R. Sazdanovic. *LinKnot: Knot Theory by Computer*, volume 21 of “*Knots and Everything*”. World Scientific, 2007.
 - [18] V. Jones. A polynomial invariant for knots via von Neumann algebras. *Bull. Amer. Math. Soc.*, 12:103–111, 1985.
 - [19] L. Kauffman. State models and the Jones polynomial. *Topology*, 26(3):395–407, 1987.
 - [20] M. Khovanov. A categorification of the Jones polynomial. *Duke Math. J.*, 101(3):359–426, 2000.
 - [21] M. Khovanov. A functor-valued invariant of tangles. *Alg. Geom. Topol.*, 2:665–741, 2002.

- [22] M. Khovanov. Triply-graded link homology and Hochschild homology of Soergel bimodules. *Int. Journal of Math.*, 18(8):869–885, 2007.
- [23] M. Khovanov. Linearization and categorification. 2016. <http://front.math.ucdavis.edu/1603.08223>.
- [24] M. Khovanov and L. Rozansky. Matrix factorizations and link homology I. *Fund. Math.*, 199:1–91, 2008.
- [25] M. Khovanov and L. Rozansky. Matrix factorizations and link homology II. *Geometry & Topology*, 12(1):97–208, 2011.
- [26] The Knot Atlas. http://katlas.org/wiki/Main_Page, 2016.
- [27] P. Kronheimer and T. Mrowka. Khovanov homology is an unknot-detector. *Publications mathématiques de l’IHES*, 113:1387–1425, 2008.
- [28] M. Mackaay, M. Stosić, and P. Vaz. $Sl(N)$ link homology using foams and the Kapustin-Li formula. *Geometry & Topology*, 13:1075–1128, 2009.
- [29] C. Manolescu and P. Ozsváth. On the Khovanov and knot Floer homologies of quasi-alternating links. In *Proceedings of the 14th Gökova Geometry-Topology Conference*, pages 60–81. International Press, Berlin, 2007.
- [30] C. Manolescu, P. Ozsváth, and S. Sarkar. A combinatorial description of knot Floer homology. *Annals of Mathematics*, 169(2):633–660, 2009.
- [31] Sujoy Mukherjee, Jozef Przytycki, Marithania Silvero, Xiao Wang, and Seung Yeop Yang. Search for torsion in Khovanov homology. *Experimental Mathematics*, 2017.
- [32] P. Ozsváth and Z. Szabó. Heegaard Floer homology and alternating knots. *Geom. Topol.*, 7:225–254, 2003.
- [33] P. Ozsváth and Z. Szabó. On the Heegaard Floer homology of branched double-covers. *Adv. Math.*, 194(1):1–33, 2005.
- [34] M. Pabiniak, J. Przytycki, and R. Sazdanovic. On the first group of the chromatic cohomology of graphs. *Geom. Dedicata*, 140(1):19–48, 2009.
- [35] J. Przytycki. When two theories meet: Khovanov homology as Hochschild homology of links. *Quantum Topology*, 1(2):93–109, 2010.
- [36] J. Przytycki and R. Sazdanovic. Torsion in Khovanov homology of semi-adequate links. *Fund. Math.*, 225:277–304, 2014.
- [37] J. Rasmussen. Khovanov homology and the slice genus. *Inventiones Mathematicae*, 182(2):419–447, 2010.
- [38] L. Roberts. A type A structure in Khovanov homology. 2013. <http://arxiv.org/abs/math.GT/1304.0465>.
- [39] L. Roberts. A type D structure in Khovanov homology. 2013. <http://arxiv.org/abs/math.GT/1304.0463>.
- [40] A. Shumakovitch. Rasmussen invariant, Slice-Bennequin inequality, and sliceness of knots. *J. Knot Theory*, 16(10):1403–1412, 2007.
- [41] A. Shumakovitch. Torsion of the Khovanov homology. *Fund. Math.*, 225:343–364, 2014.

- [42] M. Stosić. Khovanov homology of torus links. *Topology and its Applications*, 156(3):–, 2009.
- [43] P. Turner. Five lectures on Khovanov homology. 2006. <http://arxiv.org/abs/math/0606464>.
- [44] O. Viro. Remarks on definition of Khovanov homology. 1999. arxiv.org/abs/math.GT/0202199.
- [45] O. Viro. Khovanov homology, its definitions and ramifications. *Fund. Math.*, 184:317–342, 2004.
- [46] S.M. Wehrli. A spanning tree model for Khovanov homology. *J. Knot Th. and Ramif.*, 17(12):1561–1574, 2008.

Elizabeth Denne

Quadriseccants and essential secants of knots

with applications to the geometry of knots

Abstract: A quadriseccant line is one which intersects a curve in at least four points, while an essential secant captures something about the knottedness of a knot. This survey article gives a brief history of these ideas, and shows how they may be applied to questions about the geometry of a knot via the total curvature, ropelength and distortion of a knot.

Keywords: Knots, quadriseccants, triseccants, essential secants, total curvature, distortion, ropelength.

MSC: 57M25


7.1 Introduction

In this survey article, we consider the ways in which lines intersect knots. Some of these capture the knottedness of a knot. Triseccant and quadriseccant lines are straight lines which intersect a knot in at least three, respectively four distinct places. It is clear that any closed curve has a 2-parameter family of secants. A simple dimension count shows there is a 1-parameter family of triseccants and that quadriseccants are discrete (0-parameter family). A planar circle unknot does not have any triseccants and quadriseccants, however nontrivial tame knots must have them. In general, we do not expect knots to have quintiseccants (or higher order secants), they exist only for a codimension 1 (or higher) set of knots. Note that these statements are only valid for generic knots (for example, polygonal knots with extra conditions like no four points are coplanar).

The existence of quadriseccants and essential secants gives insight into the geometry of knots. Section 7.2 gives definitions of quadriseccants and essential secants and a brief history of known results. It ends with a discussion of the open question of finding bounds on the number of quadriseccants for a given knot type. Section 7.3 gives a brief outline of the tools used in many of the proofs of results about quadriseccants. Section 7.4 shows how the ideas may be applied to results about the geometry of knots via the total curvature, second hull, ropelength and distortion of a knot.

Elizabeth Denne, Department of Mathematics, Washington & Lee University, Lexington VA 24450, USA, E-mail: dennee@wlu.edu

<https://doi.org/10.1515/9783110571493-006>

Open Access.  © 2018 Elizabeth Denne, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

7.2 Quadrisecants

Recall that a *knot* is a homeomorphic image of S^1 in \mathbb{R}^3 , modulo reparametrizations, and a tame knot is one that is ambient isotopic to a polygonal knot.

Definition 7.2.1. Given a knot K , an n -secant line is an oriented line which intersects K in at least n components. An n -secant is an ordered n -tuple of points of K (no two of which lie in a common straight subarc of K) which lie in order along an n -secant line.

As previously described, by a secant we mean a 2-secant, by a trisecant a 3-secant, and by a quadrisecant a 4-secant. For a closed curve K , any two distinct points determine a straight line. These points are a secant if and only if they do not lie on a common straight subarc of K . Thus the set of secants $S = K^2 \setminus \tilde{\Delta}$ is topologically an annulus. (Here $\tilde{\Delta}$ denotes the set of n -tuples in which some pair of points lie in a common straight subarc of K .)

We now consider the set of trisecants of a knot, denoted $\mathcal{T} \subset K^3 \setminus \tilde{\Delta}$. For any trisecant abc , there are $|S_3/C_3| = 2$ cyclic orderings of the oriented knot and trisecant. We could label these by their lexicographically least elements (abc or acb), but we call them *direct* and *reversed* respectively. Figure 7.1 illustrates the two types of trisecant. Flipping the orientation of the knot or the trisecant would change its type.

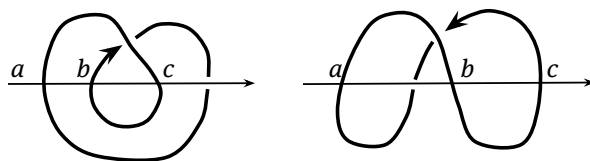


Fig. 7.1: These trisecants are *reversed* (left) and *direct* (right) because the cyclic order of the points along K is acb and abc respectively.

Just as with trisecants, we may compare the points of a quadrisecant line with their ordering along the knot. For quadrisecant $abcd$, the order along K is a cyclic order, and ignoring the orientation of K is just a dihedral order. Thus there are $|S_4/D_4| = 3$ dihedral orderings of a quadrisecant and non-oriented knot. We can represent these equivalence classes by $abcd$, $abdc$ and $acbd$, where we have again chosen the lexicographically least order as the name for each. Figure 7.2 illustrates these orderings.

Definition 7.2.2. Quadrisecants of type $acbd$ are called *alternating quadrisecants*. Quadrisecants of types $abcd$ and $abdc$ are called *simple* and *flipped* respectively.

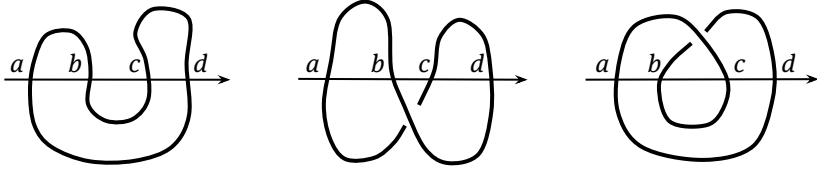


Fig. 7.2: From left to right, quadrisecants $abcd$ are simple, flipped and alternating.

When discussing quadrisecant $abcd$, we will usually choose to orient K so that $b \in \gamma_{ad}$. This means that the cyclic order of points along K will be $abcd$, $abdc$, or $acbd$, depending on the type of quadrisecant.

7.2.1 Essential secants

Before discussing known results about quadrisecants, we pause to introduce the notion of an essential secant. (The definitions in this subsection can all be found in Denne et al. [14].) G. Kuperberg [38] first introduced this idea (which he called “topologically nontrivial”) in his paper about quadrisecants. Being essential captures part of the knottedness of the knot.

Generically, the knot K together with the secant segment between two points on K forms a knotted Θ -graph in space (that is, a graph with three edges connecting the same two vertices).

Definition 7.2.3. Suppose α, β and γ are three disjoint simple arcs from points a to b , forming a knotted Θ -graph, as illustrated in Figure 7.3. Let $X := \mathbb{R}^3 \setminus (\alpha \cup \gamma)$, and let δ be a parallel curve to $\alpha \cup \beta$ in X . (By *parallel* we mean that $\alpha \cup \beta$ and δ cobound an annulus embedded in X .) We choose δ to be homologically trivial in X (that is, so that δ has zero linking number with $\alpha \cup \gamma$). Let $h = h(\alpha, \beta, \gamma) \in \pi_1(X)$ denote the (free) homotopy class of δ . Then (α, β, γ) is *inessential* if h is trivial. We say that (α, β, γ) is *essential* if it is not inessential.

In other words, the ordered triple (α, β, γ) is *inessential* if there is a disk D bounded by $\alpha \cup \beta$ having no interior intersections with the knot $\alpha \cup \gamma$. (We allow self-intersections of D , and interior intersections with β , as will be necessary if $\alpha \cup \beta$ is knotted.)

This notion is clearly a topological invariant of the (ambient isotopy) class of the knotted Θ -graph. We apply this definition to arcs of a knot K below. But first, some useful notation. Let $a, b \in K$. The arc from a to b following the orientation of the knot is denoted γ_{ab} and has length ℓ_{ab} . The arc from b to a , γ_{ba} , with length ℓ_{ba} is similarly defined. The secant segment from a to b is denoted \overline{ab} .

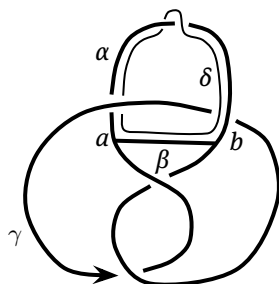


Fig. 7.3: In the knotted θ -graph $\alpha \cup \beta \cup \gamma$, the ordered triple (α, β, γ) is essential. To see this, we find the parallel δ to $\alpha \cup \beta$ which has linking number zero with $\alpha \cup \gamma$ and note that it is homotopically nontrivial in the knot complement $\mathbb{R}^3 \setminus (\alpha \cup \gamma)$. In this illustration, β is the straight segment \overline{ab} , so we may equally say that the arc $\alpha = \gamma_{ab}$ of the knot $\alpha \cup \gamma$ is essential.

Definition 7.2.4. Assume K is a nontrivial tame knot, $a, b \in K$, and $\ell = \overline{ab}$. We say γ_{ab} is *essential* if for every $\varepsilon > 0$ there exists some ε -perturbation of ℓ (with endpoints fixed) to a tame curve ℓ' such that $K \cup \ell'$ forms an embedded θ in which $(\gamma_{ab}, \ell', \gamma_{ba})$ is essential.

Note that this definition is quite flexible, as it allows for the situation where K intersects ℓ . In [14] Proposition 6.2, we show that it also ensures that the set of essential secants is closed in S .

Definition 7.2.5. A secant ab of K is *essential* if both subarcs γ_{ab} and γ_{ba} are essential. Otherwise it is *inessential*. Let ES be the set of essential secants in S .

It is straightforward to see that if K is an unknot, then any arc γ_{ab} is inessential. (Because the homology and homotopy groups of $X := \mathbb{R}^3 \setminus K$ are equal for an unknot, then any curve having zero linking number with K is homotopically trivial in X .) We can use Dehn's Lemma to prove a converse statement: if $a, b \in K$ and both γ_{ab} and γ_{ba} are inessential, then K is unknotted. (See for instance [10, 14].)

Later, in applications of these ideas, we need to find the least length of an essential arc, and use this to get better bounds on ropelength and distortion. This leads us to consider what happens when arcs change from inessential to essential.

Theorem 7.2.6 ([14] Theorem 7.1). *Suppose γ_{ac} is in the boundary of the set of essential arcs for a knot K . (That is, γ_{ac} is essential, but there are inessential arcs of K with endpoints arbitrarily close to a and c .) Then K must intersect the interior of segment \overline{ac} , and in fact there is some essential trisecant abc .*

Finally, to call a quadrisequant $abcd$ essential, we could follow Kuperberg and require that the secants ab , bc and cd all be essential. But instead, we require this only of those secants whose endpoints are consecutive along the knot

Definition 7.2.7. An n -secant $a_1a_2 \dots a_n$ is essential if we have $a_i a_{i+1}$ essential for each i such that one of the arcs $\gamma_{a_i a_{i+1}}$ and $\gamma_{a_{i+1} a_i}$ includes no other a_j .

That is, for simple quadrisequants, all three secants must be essential; for flipped quadrisequants the end secants ab and cd must be essential; for alternating quadrisequants, the middle secant bc must be essential.

7.2.2 Results about quadrisequants

The simple dimension count outlined in Section 7.1 means that we expect nontrivial tame knots to have quadrisequants. Indeed, in 1933, E. Pannwitz [44] first showed that every nontrivial generic polygonal knot in \mathbb{R}^3 has at least $2u^2$ quadrisequants, where u is the unknotting number of a knot^{7.1}. (It is entirely possible that Heinz Hopf suggested this problem to her. In the endnote to his paper, I. Fáry [22] mentions that Hopf used quadrisequants to prove that knots have total curvature greater than or equal to 4π .)

In the early 1980's, H.R. Morton and D.M.Q. Mond [40] rediscovered Pannwitz' result. They independently proved that every nontrivial generic knot has a quadrisequant, and they conjectured that a generic knot with crossing number n has at least $\binom{n}{2}$ quadrisequants. It was not until 1994 that Kuperberg [38] managed to extend the result and showed that *all* (nontrivial tame) knots in \mathbb{R}^3 have a quadrisequant. To do this, he introduced the notion of an essential secant. In 1998, C. Schmitz [48] nearly proved that alternating quadrisequants exist for nontrivial tame knots in Hadamard manifolds, but in his proof some quadrisequants may degenerate to trisequants.

In 2004, the following result was proved in my PhD thesis.

Theorem 7.2.8 ([16]). *All nontrivial tame knots have at least one alternating quadrisequant. All nontrivial knots of finite total curvature have at least one essential alternating quadrisequant.*

Also in 2004, R. Budney, J. Conant, K.P. Scannell and D. Sinha [6] gave a geometric interpretation of the second order Vassiliev invariant. To do this, they used the techniques of compactified configuration spaces and Goodwillie calculus to show that this invariant can be computed by counting alternating quadrisequants with appropriate

^{7.1} She actually stated her result in terms of the knottedness of K , which is the minimal number of singular points on the boundary of locally flat singular spanning disks of K , and is twice the unknotting number.

multiplicity (both for long knots and closed knots). This result implies the existence of alternating quadrisecants for many knots, while Theorem 7.2.8 shows existence for all nontrivial tame knots.

Later in 2013, Budney's Master's student G. Flowers [25] gave a natural extension of the quadrisecant count formula found in [6] to closed knots. Here, Flowers counts five- and six-point cocircularities instead of counting quadrisecant lines. The relative ordering of the knot and intersecting circle are again central to the arguments, and the idea of an alternating quadrisecant becomes that of either a 'satanic' or 'thelemic' circle.

In 2007, M. Sommer developed a wonderful jReality application *Visualization in Geometric Knot Theory* [50] as part of his Diploma Thesis [51]. This program allows the user to view the set of trisecants for polygonal knots, and even see how it changes as a vertex is moved. The program also shows the different projections of this set onto different planes, from which the user can read off the different quadrisecants.

In 2008, T. Fiedler and V. Kurlin [24] viewed quadrisecants in a different way. They fixed a straight line in \mathbb{R}^3 and a fibration around it by half-planes, and studied knots in general position with respect to this fibration using secants and quadrisecants lying in fibers. They give the minimum number of all fiber quadrisecants and fiber extreme secants which occur during an isotopy from one knot to another in terms of a sum of unordered and coordinated writhes of a particular kind of projection.

In 2009, J. Viro [54] estimated from below the number of lines meeting each of four disjoint smooth curves in both $\mathbb{R}P^3$ and \mathbb{R}^3 . In \mathbb{R}^3 her arguments may be translated to quadrisecant lines, however her count involves linking numbers and appears to be different from the one in [6].

7.2.3 Counting quadrisecants and quadrisecant approximations.

For a generic nontrivial knot, we expect that there are a finite number of quadrisecants. (Indeed, this was one of the results of Budney et al. [6].) The open and very challenging problem is to give bounds on the number of quadrisecants for a given knot type. Many of the papers cited in the previous section relate the number of quadrisecants to some topological invariant of knots (unknotting number, finite type invariants, writhe), or give conjectures about what this might be. However none of the known results seem to be close to the number of observed quadrisecants for knots with large crossing number. Indeed, there are not even conjectures about asymptotic bounds for the number of quadrisecants for particular families of knots.

There have been several papers which have grappled with this question for knots with small crossings. In 2005, G.T. Jin [35], gave trigonometric parametrizations of the 3_1 , 4_1 , 5_1 and 5_2 knots and found all quadrisecants for these parametrizations. He also defined the quadrisecant approximation of a knot K .

Definition 7.2.9. Let K be a knot with finitely many quadrisecants intersecting K in finitely many points. These points divide K into a finite number of subarcs. Replacing the subarcs by straight lines gives a polygonal closed curve called the *quadrisecant approximation* of K .

Jin then conjectured the following.

Conjecture 7.2.10 ([35]). *If a knot K has finitely many quadrisecants, then the quadrisecant approximation \hat{K} has the knot type of K . Furthermore, K and \hat{K} have the same set of quadrisecants.*

In 2011, Jin and S. Park [36] proved that every hexagonal (6-sided) trefoil knot has exactly three alternating quadrisecants. They then proved that the quadrisecant approximation of a hexagonal trefoil knot is also a trefoil knot, and moreover has the same three alternating quadrisecants as the original knot.

Most recently, S. Bai, C. Wang and J. Wang [3] proved that the Conjecture 7.2.10 is false. They gave two examples of an unknot: the first has a quadrisecant approximation which is not even an embedded curve, and the second has a quadrisecant approximation which is a left-handed trefoil knot. They then used a particular type of connected sum operation to show that there is a polygonal knot in every knot type whose quadrisecant approximation is either not embedded, or contains a trefoil summand.

There are still many interesting questions to explore here. For example, it is entirely possible that Conjecture 7.2.10 is true for polygonal knots which have minimum stick number, or for ideal knots (those which minimize some kind of energy like ropelength).

Returning to the question of counting quadrisecants, A. Cruz-Cota and T. Ramirez-Rosas [12] recently proved the first result giving an upper (rather than a lower) bound on the number of quadrisecants.

Theorem 7.2.11 ([12]). *Let K be a polygonal knot in general position, with exactly n edges. Then K has at most $\frac{n}{12}(n-3)(n-4)(n-5)$ generic quadrisecants.*

We finish this section by noting that the results of [6] are the only ones relating a count of *alternating* quadrisecants to a knot invariant. We have work-in-progress [13] showing that there is at least $u(K)$ number of alternating quadrisecants of a knot K , where $u(K)$ is the unknotting number of K . Examples from [35] have lead us to conjecture the following (say for knots of finite total curvature).

Conjecture 7.2.12. *The figure-8 knot has at least one essential flipped quadrisecant. The 5_2 knot has at least one essential simple quadrisecant.*

7.3 Key ideas in showing quadrisecants exist

Observe that quadrisecants are formed when several trisecants share common points. Quadrisecant $abcd$ includes four trisecants $(1)abc$, $(2)abd$, $(3)acd$, $(4)bcd$. Pannwitz [44] showed quadrisecants exist by looking for pairs of trisecants like $(1)abc$ and $(3)acd$, where the first and third points of trisecant abc are the same as the first and second points of trisecant acd . Kuperberg [38] showed that quadrisecants exist by looking for pairs of trisecants like $(2)abd$ and $(3)acd$, where the first and third points of the trisecants are the same. We proved [16] that alternating quadrisecants exist by looking at families $(1)abc$ and $(2)abd$, where the first and second points of the trisecants are the same. The quadrisecant count in Budney et al. [6] boiled down to looking at linking numbers of these sets of trisecants in the 3rd associahedron (Stasheff polytope) — a compactified configuration space.

All of these arguments presuppose that nontrivial knots have trisecants. Fortunately this is easy to prove, as was originally shown by Pannwitz.

Lemma 7.3.1 ([44]). *Each point of a nontrivial tame knot K is the first point of at least one trisecant.*

Proof. Suppose there is a point $a \in K$ which is not the start point of any trisecant. The union of all chords \overline{ab} for $b \in K$ is a disk with boundary K . If two chords ab and ac intersect at a place other than a , then they overlap and one is a subinterval of another. They form a trisecant $(abc$ or $acb)$, contrary to the assumption. Thus the disk is embedded and K unknotted, a contradiction. \square

The structure of the set of trisecants (and quadrisecants) can be understood more easily when we restrict the class of knots considered, such that the restricted class is dense in the set of tame knots. All of the previous work on quadrisecants we have discussed does this, exactly how depends on the context chosen. We choose to work with polygonal knots with some extra assumptions.

Definition 7.3.2. We say that the polygonal knot K in \mathbb{R}^3 is *generic* if the following conditions are satisfied

- No four vertices of K are coplanar, no three vertices of K are collinear.
- Given three pairwise skew edges of K , no other edge of K is contained in the quadric surface generated by those edges.
- There are no n -secants for $n \geq 5$.

Once the existence of *essential* quadrisecants is established for generic knots, a limit argument (see [38]) is used to show that any nontrivial tame knot has a quadrisecant. The key point here is that the limit of an essential secant remains a secant — it does not degenerate in the limit.

7.3.1 Trisecants and quadrisecants.

We now discuss how trisecants and quadrisecants can arise in generic polygonal knots. The first condition in Definition 7.3.2 means that three (or more) adjacent edges cannot be coplanar. Suppose e_i and e_{i+1} are adjacent edges. A 1-parameter family of trisecants arises when a third edge e_j of the knot intersects certain parts of the plane spanned by e_i and e_{i+1} . This family is either homeomorphic to $[0, 1]$ or $[0, 1)$ depending on which region e_j intersects (see Figure 7.4). A quadrisecant is formed when a fourth edge intersects one of the trisecant lines. Genericity implies that two non-adjacent edges cannot be coplanar, thus there can be at most one quadrisecant in this case.

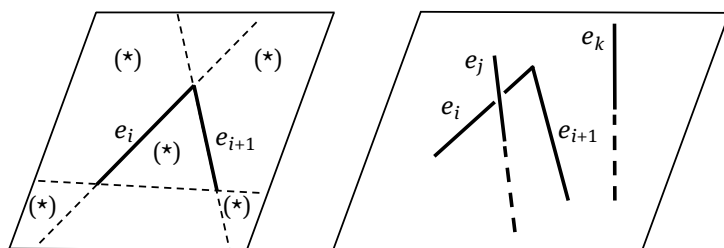


Fig. 7.4: The planes are spanned by edges e_i, e_{i+1} . On the left, there are trisecant lines intersecting e_i, e_{i+1} if the third edge intersects any one of the five regions marked (*), otherwise there are no trisecants. On the right, the family of trisecants intersecting e_i, e_j, e_{i+1} is homeomorphic to $[0, 1]$, while the family intersecting e_i, e_{i+1}, e_k is homeomorphic to $[0, 1)$.

Before moving on to the next case, we pause to remind the reader about some well known facts about *doubly ruled surfaces* (see for instance [31, 47, 43]). A triple of pairwise skew lines l_1, l_2, l_3 determines a unique quadric, a doubly ruled surface H (as in the blue lines in Figure 7.5). This is either a hyperbolic paraboloid, if the three lines are parallel to one plane, or a hyperboloid of one sheet, otherwise. Each point of H lies on a unique line from each ruling. The lines l_1, l_2, l_3 belong to one of the rulings of the surface, and every line intersecting all three lines belongs to the other. Thus a fourth line l_4 intersecting H yields two (or one) quadrisecant line(s) intersecting l_1, l_2, l_3 , and l_4 . There are an infinite number of quadrisecants only when l_4 is contained in H .

Returning our attention to generic polygonal knots, we see that a pairwise skew triple of edges e_i, e_j, e_k either has 0 or 1 trisecant, or has a 1-parameter family of trisecants. A quadrisecant is formed when a fourth edge intersects the quadric in a trisecant line. There are 1 or 2 quadrisecants in this case, since the second condition in Definition 7.3.2 means that there cannot be an infinite number of quadrisecants.

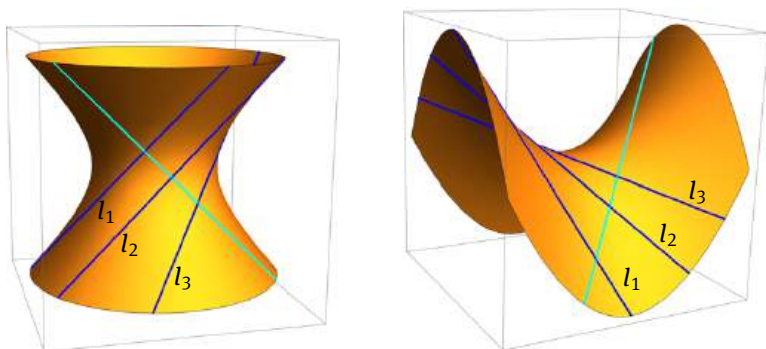


Fig. 7.5: Doubly ruled surfaces produced in *Mathematica*. On the left a hyperboloid of one sheet, on the right a hyperbolic paraboloid. The three blue lines l_1, l_2, l_3 lie in one ruling, the cyan trisecant line in the other ruling.

7.3.2 Structure of the set of trisecants.

Using the ideas from above, we can piece together the structure of the set of trisecants $\mathcal{T} \subset K^3$. Indeed this approach has been used to understand (and compute) \mathcal{T} and the set of quadrisecants in many papers (such as [3, 12, 16, 36, 51]). We expect that generically, \mathcal{T} is a 1-manifold. Of course, we can add more (generic) conditions to Definition 7.3.2 to give even more control over \mathcal{T} (for example those in [16, 51]). We omit these details here, and also the many technical but straightforward details needed to prove the following two results.

Proposition 7.3.3 (c.f. [16, 51]). *Let K be a nontrivial generic polygonal tame knot. In K^3 , $\bar{\mathcal{T}}$ is a compact 1-manifold with boundary, embedded in K^3 in a piecewise smooth way with $\mathcal{T} \subset K^3 \setminus \tilde{\Delta}$ and $\partial\mathcal{T} \subset \Delta$. Moreover, each component of \mathcal{T} is either a simple closed curve or a simple open arc.*

Proposition 7.3.4 (c.f. [16]). *Let K be a nontrivial generic polygonal tame knot. The projection $\pi_{ij} : K^3 \rightarrow K^2$ ($i < j$ and $i, j = 1, 2, 3$) is a piecewise smooth immersion of \mathcal{T} into the set of secants, and $T = \pi_{ij}(\mathcal{T})$ intersects itself (transversally) at double points.*

The astute reader will have realized that we really need to work with *essential* trisecants if we are to find *essential* (alternating) quadrisecants. Recall that the set of essential secants ES is a subset of S . In order to find essential alternating quadrisecants, we restrict our attention to the trisecants abc which are essential in the second segment bc . Thus $\mathcal{ET} = \pi_{23}^{-1}(ES) \cap \mathcal{T}$, and we define \mathcal{ET}^d and \mathcal{ET}^r to be the sets of essential trisecants of direct and reversed orderings in K^3 . Since we are interested in finding essential trisecants which share the first two points, we project \mathcal{ET} to S in a particular way.

Definition 7.3.5 ([16]). Let $ET = \pi_{12}(\mathcal{ET})$ be the projection of the set of essential trisecants to the set of secants S and similarly define $ET^d := \pi_{12}(\mathcal{ET}^d)$ and $ET^r := \pi_{12}(\mathcal{ET}^r)$.

We can easily prove that a version of Proposition 7.3.4 holds for ET : For a nontrivial generic polygonal tame knot, ET is a piecewise immersed 1-manifold which intersects itself transversally at double points ([16]). Finally, we observe that (essential) alternating quadrisecants occur precisely when direct and reversed (essential) trisecants intersect in S .

Lemma 7.3.6 ([16]). *Let $ab \in ET^d \cap ET^r$ in S . This means that there exists c and d such that $abc \in \mathcal{ET}^r$ and $abd \in \mathcal{ET}^d$. Then either $abcd$ or $abdc$ is an essential alternating quadrisecant.* \square

The proof of the existence of alternating quadrisecants, Theorem 7.2.8, follows the same pattern as the proof of Lemma 7.3.1. We proceed by contradiction, and assume that $ET^d \cap ET^r = \emptyset$ in S . We then show that K has to be unknotted, a contradiction. The details of proving that K is unknotted are quite involved and so we omit them here. [However, very briefly, since K is a polygonal knot, we can show that ET^d , respectively ET^r , stay at least the minimum edge length away from the bottom, respectively top, boundary of S . Since ET^d and ET^r do not intersect, we use a Meyer-Vietoris argument to construct a loop winding once around S which avoids the set of essential trisecants. We then use this loop to find a spanning disk whose boundary is the knot. After further technicalities, the loop theorem is invoked to show that this disk is embedded.]

7.4 Applications of essential secants and quadrisecants

7.4.1 Total curvature

For smooth closed curves, the total curvature can be thought of as the total angle through which the unit tangent vector turns (or the length of the tangent indicatrix). J.W. Milnor [39] defined the *total curvature* $\kappa(\gamma)$ of an arbitrary closed curve γ to be the supremal total curvature of inscribed polygons (where the total curvature is the sum of the exterior angles), and showed the two definitions are equivalent. In 1929, W. Fenchel [23] proved that the total curvature of a closed curve in \mathbb{R}^3 is greater than or equal to 2π , equality holding only for plane convex curves. In 1947, K. Borsuk [4] extended this result to \mathbb{R}^n and conjectured the following.

Theorem 7.4.1. *A nontrivial tame knot in \mathbb{R}^3 has total curvature greater than 4π .*

This result was first proved around 1949 by both Milnor [39] and Fáry [22]. It has since become known as the Fáry-Milnor theorem. In Cantarella et al. [10], the theorem was proved using the nonemptiness of the second hull for a knotted curve (defined below). In 1998, the Fáry-Milnor theorem was independently extended to knotted curves in Hadamard^{7.2} manifolds by Schmitz [48] and S.B. Alexander and R.L. Bishop [2]. Interestingly, in their proof, Alexander and Bishop show that a simple curve of finite total curvature is isotopic to an inscribed (geodesic) polygonal curve, fixing a mistake in [39]. They show that for any nontrivial knot K , some doubly covered bigon is inscribed in a polygon inscribed in K . This doubly covered bigon is an idea somewhat close to the midsegment of an alternating quadrisecant.

A new proof of the Fáry-Milnor theorem is given by the existence of alternating quadrisecants for any nontrivial tame knot K . An alternating quadrisecant can be thought of as an inscribed polygon in K , and has total curvature 4π . Thus by definition, $\kappa(K) \geq 4\pi$. To get a strict inequality, simply observe that a knot is not coplanar.

More recently, H. Gerlach, P. Reiter, and H. von der Mosel [26] have written about *elastic knots*, which are limit configurations of energy minimizers of an energy consisting of the classic bending energy and a small multiple of ropelength (defined below). One of the many results in this paper is an extension of the classic Fáry-Milnor theorem on total curvature to the C^1 -closure of the knot class. The proof of this result also relies on the existence of alternating quadrisecants.

7.4.2 Second Hull

The convex hull of a connected set K in \mathbb{R}^3 is characterized by the fact that every plane through a point in the hull must intersect K . If K is a closed curve, then a generic plane must intersect K an even number of times. Thus every generic plane through each point of the convex hull is cut by K at least twice. In proving the total curvature result above, Milnor observed that for a nontrivial tame knot, there are planes in every direction which cut the knot four times. More generally, there are points through which *every* plane cuts the knots four times. This idea was formalized by Cantarella, Kuperberg, R. Kusner and J.M. Sullivan in [10], where they defined these points as the *second hull* of a knot.

Definition 7.4.2 ([10]). Let K be a closed curve in \mathbb{R}^3 . Its n th hull $h_n(K)$ is the set of points $p \in \mathbb{R}^3$ such that K cuts every generic plane P through p at least $2n$ -times.

Cantarella et al. [10] proved that the second hull of a nontrivial tame knot in \mathbb{R}^3 is nonempty. This paper also conjectured the existence of alternating quadrisecants for

7.2 A Hadamard manifold is a complete simply-connected Riemannian manifold with non-positive sectional curvature.

nontrivial tame knots in \mathbb{R}^3 . The existence of alternating quadrisecants gives another way of proving that the second hull is nonempty. This is because the mid-segment bc of alternating quadrisecant $abcd$ is in the second hull.

7.4.3 Ropelength

The ropelength problem asks to minimize the length of a knotted curve subject to maintaining an embedded tube of fixed diameter around the tube; this is a mathematical model of tying the knot tight in a rope of fixed thickness.

More technically, the *thickness* $\tau(K)$ of a space curve K is defined (O. Gonzalez and J.H. Maddocks [27]) to be twice the infimal radius $r(a, b, c)$ of circles through any three distinct points of K . It is known ([11]) that $\tau(K) = 0$ unless K is $C^{1,1}$ (meaning that its tangent direction is a Lipschitz function of arclength). When K is C^1 , we can define normal tubes around K , and then indeed $\tau(K)$ is the supremal diameter of such a tube that remains embedded. We note that in the existing literature thickness is sometimes defined to be the radius rather than diameter of this thick tube. Following others lead, we define the ropelength as follows.

Definition 7.4.3. The *ropelength* of a knot K is the (scale-invariant) quotient of length over thickness, $\text{Rop}(K) = \text{len}(K)/\tau(K)$.

Cantarella, Kusner and Sullivan [11] proved that any (tame) knot or link type has a ropelength minimizer and gave certain lower bounds for the ropelength of links; these are sharp in certain simple cases where each component of the link is planar. In this section we outline our joint work [14], showing how essential alternating quadrisecants are used to prove that nontrivial knots have ropelength at least 15.66. This is an improvement on the bound of 12 from Y. Diao [18] and is greater than the conjectured bound of 15.25 from [11].

7.4.3.1 Ropelength basics

Because the ropelength problem is scale invariant, we find it most convenient to rescale any knot K to have thickness (at least) 1. This implies that K is a $C^{1,1}$ curve with curvature bounded above by 2. For any point $a \in \mathbb{R}^3$, let $B(a)$ denote the open unit ball centered at a . We now give several well known results about the local structure of thick knots (see for instance [19, 11] and [14]). Most proofs are elementary and we omit them here.

Lemma 7.4.4. *Let K be a knot of unit thickness. If $a \in K$, then $B(a)$ contains a single unknotted arc of K ; this arc has length at most π and is transverse to the nested spheres*

centered at a . If ab is a secant of K with $|a-b| < 1$, then the ball of diameter ab intersects K in a single unknotted arc (either γ_{ab} or γ_{ba}) whose length is at most $\arcsin |a-b|$. \square

As an immediate corollary, we see that if K has unit thickness, $a, b \in K$ and $p \in \gamma_{ab}$ with $a, b \notin B(p)$, then the complementary arc γ_{ba} lies outside $B(p)$.

Lemma 7.4.5. *Let K be a knot of unit thickness. If $a \in K$, then the radial projection of $K \setminus \{a\}$ to the unit sphere $\partial B(a)$ does not increase length.* \square

Lemma 7.4.6. *Suppose K has unit thickness, and $p, a, b \in K$ with $p \notin \gamma_{ab}$. Let $\angle apb$ be the angle between the vectors $a-p$ and $b-p$. Then $\ell_{ab} \geq \angle apb$. In particular, if apb is a reversed trisecant in K , then $\ell_{ab} \geq \pi$.* \square

Given a thick knot K with quadrisecant $abcd$, we can bound its ropelength in terms of the distances along the quadrisecant line. Whenever we discuss such a quadrisecant, we will abbreviate these three distances as $r := |a-b|$, $s := |b-c|$ and $t := |c-d|$. We start with some lower bounds for r , s and t for alternating quadrisecants.

Lemma 7.4.7 ([14] Lemma 4.2). *If $abcd$ is an alternating quadrisecant for a knot of unit thickness, then $r \geq 1$ and $t \geq 1$. With the usual orientation, the entire arc γ_{da} thus lies outside $B(b) \cup B(c)$. If $s \geq 1$ as well, then γ_{ac} lies outside $B(b)$ and γ_{bd} lies outside $B(c)$.* \square

As suggested by the discussion above, we will often find ourselves in the situation where we have an arc of a knot known to stay outside a unit ball. We can compute exactly the minimum length of such an arc in terms of the following functions.

Definition 7.4.8. For $r \geq 1$, let $f(r) := \sqrt{r^2 - 1} + \arcsin(1/r)$. For $r, s \geq 1$ and $\theta \in [0, \pi]$, the minimum length function is defined by

$$m(r, s, \theta) = \begin{cases} \sqrt{r^2 + s^2 - 2rs \cos \theta}, & \text{if } \theta \leq \arccos(1/r) + \arccos(1/s). \\ f(r) + f(s) + (\theta - \pi), & \text{if } \theta \geq \arccos(1/r) + \arccos(1/s). \end{cases}$$

The function $f(r)$ will arise again in other situations. The function m was defined exactly to make the following bound sharp:

Lemma 7.4.9. *Any arc γ from a to b staying outside $B(p)$ has length at least*

$$\ell_{ab} \geq m(|a-p|, |b-p|, \angle apb).$$

\square

An important special case is when $\theta = \pi$. If a and b lie at distances r and s along opposite rays from p (so that $\angle apb = \pi$) then the length of any arc from a to b avoiding $B(p)$ is at least

$$\ell_{ab} \geq f(r) + f(s) = \sqrt{r^2 - 1} + \arcsin(1/r) + \sqrt{s^2 - 1} + \arcsin(1/s).$$

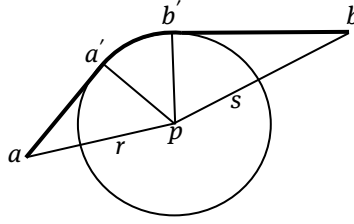


Fig. 7.6: The shortest arc from a to b avoiding the ball $B(p)$ consists of straight segments and an arc of the ball.

7.4.3.2 Length of essential secants and quadrisecants

We will improve our previous ropelength bounds by getting bounds on the length of an essential arc. Intuitively, we expect an essential arc γ_{ab} of a knot to “wrap at least halfway around” some point on the complementary arc γ_{ba} . Although when $|a - b| = 2$ we can have $\ell_{ab} = \pi$, when $|a - b| < 2$ we expect a better lower bound for ℓ_{ab} . Even though in fact an essential γ_{ab} might instead “wrap around” some point on itself, we can still derive the desired bound.

Lemma 7.4.10 ([14]). *If γ_{ab} is an essential arc in a knot K of unit thickness, then $|a - b| \geq 1$ and $\ell_{ab} \geq g(|a - b|)$, where*

$$g(|a - b|) = \begin{cases} 2\pi - 2 \arcsin(|a - b|/2), & \text{if } 0 \leq |a - b| \leq 2. \\ \pi, & \text{if } |a - b| \geq 2. \end{cases}$$

Proof. (Sketch) If $|a - b| < 1$, then by Lemma 7.4.4 the ball of diameter \overline{ab} contains a single unknotted arc of K , and thus γ_{ab} is inessential.

Knowing that sufficiently short arcs starting at any point a are inessential, consider the shortest arc γ_{aq} which is essential. From Theorem 7.2.6 there is a trisecant apq with both secants ap and pq essential, thus a and q are outside $B(p)$. Since ap is essential, apq is reversed and by Lemma 7.4.6 we get $\ell_{ab} \geq \ell_{aq} \geq \pi$.

Note that $|a - b| \in [1, 2]$, so $2\pi - \arcsin(|a - b|/2) \leq 5\pi/3$. Considering again γ_{aq} with reversed trisecant apq , we have $b \notin \gamma_{aq}$ and $\ell_{aq} \geq \pi$, so we may assume $\ell_{qb} \leq 2\pi/3$ or the bound is trivially satisfied.

The remainder of the proof involves finding a bound on ℓ_{qb} using the fact that γ_{qp} is essential. There are two cases. The first is where $b \notin B(p)$ and the whole arc γ_{aqb} stays outside $B(p)$. Thus ℓ_{ab} is greater than the length of the radial projection of the arc onto $\partial B(p)$, giving the result. The second case is where $b \in B(p)$. Here, we let γ_{qy} be the shortest essential arc starting at q and use a short argument (omitted) to show that $\ell_{ab} \geq 5\pi/3$. \square

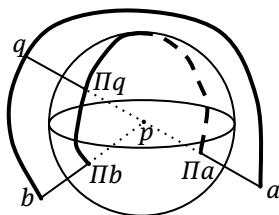


Fig. 7.7: The projection of γ_{ab} to the ball $B(p)$ does not increase the length ℓ_{ab} .

It is now straightforward to compute a lower bound of the ropelength of a unit thickness knot K with an essential quadrisecant $abcd$. There are three cases, depending on the type of quadrisecant.

1. When K has an essential simple quadrisecant, it has length $\ell_{ab} + \ell_{bc} + \ell_{cd} + \ell_{da} \geq (g(r) + f(r)) + (g(s) + s) + (g(t) + f(t))$. By minimizing each term separately, we find that K has ropelength at least $10\pi/3 + 2\sqrt{3} + 2 > 15.936$.
2. When K has an essential flipped quadrisecant, it has length $\ell_{ab} + \ell_{bd} + \ell_{dc} + \ell_{ca} \geq (g(r) + f(r)) + 2f(s) + (g(t) + f(t))$. By minimizing each term separately, we find that K has ropelength at least $10\pi/3 + 2\sqrt{3} > 13.936$.
3. When K has an essential alternating quadrisecant, the length of K is $\ell_{ac} + \ell_{cb} + \ell_{bd} + \ell_{da} \geq 2f(r) + (2f(s) + g(s) + s) + 2f(t)$. Again, by minimizing each term separately, we find K has ropelength at least 15.66.

Together with Theorem 7.2.8, we conclude the following.

Theorem 7.4.11 ([14]). *Any nontrivial knot has ropelength at least 15.66.*

Several independent numerical simulations (see for instance [46, 52]) have found a trefoil knot with ropelength less than 16.374, which is presumably close to the minimizer. This is about 5% greater than our bound, so there is not much room for improvement, although a careful analysis based on tangent directions at b and c could yield a slightly better bound.

As a final remark, we note that the ropelength problem is still open for all knot and most link types — it is a rich source of open questions. For example, there are many results relating ropelength to other knot invariants, in particular to crossing number, for instance [5, 7, 20, 17, 21, 32]. More recently, there have been several papers giving a set of necessary and sufficient conditions for ropelength criticality, for example [49, 8, 53, 9].

7.4.4 Distortion

Gromov introduced the notion of distortion for curves (see [28, 29, 30]).

Definition 7.4.12. If γ is a rectifiable curve in \mathbb{R}^3 , then its distortion is defined to be the quantity:

$$\delta(\gamma) = \sup_{p, q \in \gamma} \frac{d_\gamma(p, q)}{d_{\mathbb{R}^3}(p, q)} \geq 1.$$

where d_γ denotes the arclength along γ and $d_{\mathbb{R}^3}$ denotes the Euclidean distance in \mathbb{R}^3 .

Gromov showed that for any simple closed curve γ , we have $\delta(\gamma) \geq \frac{1}{2}\pi$, with equality if and only if γ is a circle, thus determining $\delta(\text{unknot})$. He then asked whether every knot type can be built with say, $\delta \leq 100$. In [15], we proved that for any nontrivial tame knot we have $\delta(K) \geq \frac{5}{3}\pi$. To do this, we first showed that *any* nontrivial tame knot has a shortest essential secant. Then, borderline-essential arcs and their lengths (namely Theorem 7.2.6 and Lemma 7.4.9) were key tools used in our distortion computations. Our bound is of course not sharp, but numerical simulations [41] have found a trefoil knot with distortion less than 7.16, so we are not too far off. We expect the true minimum distortion for a trefoil is closer to that upper bound than to our lower bound.

In 2011, J. Pardon [45] proved that the distortion of $T_{p,q}$, a (p, q) torus knot, is $\delta(T_{p,q}) \geq \frac{1}{160} \min(p, q)$. This shows the answer to Gromov's question is no. Pardon's main theorem considers isotopy classes of simple loops in a piecewise-linear embedded surface of genus $g \geq 1$, and his inequality involves certain minimum geometric intersection numbers of these loops.

In general, distortion is quite tricky to get a handle on. In [15] we give an example of a wild knot, the connected sum of infinitely many trefoils, with distortion less than 10.7. Unlike the ropelength case, having finite distortion does not put any regularity conditions on the curve. Because of this, it is an open problem to establish the existence of minimizers of δ in any nontrivial knot class. In 2007, C. Mullikin [42] started to develop a calculus of variations theory for distortion. However, it is not clear whether the techniques developed for ropelength criticality can be easily applied to this situation.

7.4.5 Final Remarks

Quadriseccants have made an appearance in other parts of knot theory as well. For example, they give a starting place to finding information about two “super-invariants” of knots.

Recall that the bridge index of a knot is defined to be the minimum number of bridges in all possible diagrams of a knot in its knot type. The *superbridge index* was

first defined by N.H. Kuiper [37], and is

$$sb[K] = \min_{K' \in [K]} \max_{\vec{v} \in S^2} b_{\vec{v}}(K'),$$

where $b_{\vec{v}}(K)$ is the number of bridges (or local maxima) of an orthogonal projection $K \rightarrow \mathbb{R}\vec{v}$. Kuiper then computed the superbridge index for all torus knots. Later, [33, 34] used quadrisecants to show that there are only finitely many knot types with superbridge index 3.

The supercrossing index of a knot is related to crossing index in the same way that superbridge index is related to bridge index. Namely, the *supercrossing index* of K is

$$sc[K] = \min_{K \in [K]} \max_{\vec{v} \in S^2} (\# \text{ of crossings}).$$

More simply, we maximize the number of crossings we see generated by a given conformation of a knot, and then minimize over all conformations. This invariant was studied by C. Adams et al. [1] and was related to other knot invariants like stick index. The existence of quadrisecants means that every nontrivial tame knot has supercrossing index at least 6. To see this, simply perturb the projection along the quadrisecant line to find six crossings.

In summary, quadrisecants and essential secants appear in many different parts of knot theory. They are a useful tool in understanding many phenomena of knots and links, as they form a bridge between topological and geometric properties. For all this, quadrisecants are still not completely understood. One of the more important open problems is to give bounds on the number of quadrisecants for knot and link types.

Acknowledgment: I first worked on quadrisecants as part of my PhD thesis, and I am forever grateful to my advisor John M. Sullivan for introducing me to them. I am also deeply appreciative of the many helpful conversations I have had over the years with Stephanie Alexander, Ryan Budney, Jason Cantarella, and Cliff Taubes about this material. Thanks also go to my coauthor Yuanan Diao for first realizing that quadrisecants could be applied to the ropelength problem. I am very thankful to the referee for helping me find and correct many small mistakes — any remaining are my own.

Bibliography

- [1] Colin Adams, Jonathan Othmer, Andrea Stier, Carmen Lefever, Sang Pahk, and James Tripp. An introduction to the supercrossing index of knots and the crossing map. *J. Knot Theory Ramifications*, 11(3):445–459, 2002. Knots 2000 Korea, Vol. 1 (Yongpyong).
- [2] Stephanie B. Alexander and Richard L. Bishop. The Fáry-Milnor theorem in Hadamard manifolds. *Proc. Amer. Math. Soc.*, 126(11):3427–3436, 1998.

- [3] Sheng Bai, Chao Wang, and Jiajun Wang, Counterexamples to the quadrisecant approximation conjecture. *J. Knot Theory Ramifications*, 27(2):1850022, 16, 2018.
- [4] Karol Borsuk. Sur la courbure totale des courbes fermées. *Ann. Soc. Polon. Math.*, 20:251–265 (1948), 1947.
- [5] Gregory Buck and Jonathan Simon. Thickness and crossing number of knots. *Topology Appl.*, 91(3):245–257, 1999.
- [6] Ryan Budney, James Conant, Kevin P. Scannell, and Dev Sinha. New perspectives on self-linking. *Adv. Math.*, 191(1):78–113, 2005.
- [7] Jason Cantarella, X. W. C. Faber, and Chad A. Mullikin. Upper bounds for ropelength as a function of crossing number. *Topology Appl.*, 135(1-3):253–264, 2004.
- [8] Jason Cantarella, Joseph H. G. Fu, Rob Kusner, John M. Sullivan, and Nancy C. Wrinkle. Criticality for the Gehring link problem. *Geom. Topol.*, 10:2055–2116 (electronic), 2006.
- [9] Jason Cantarella, Joseph H. G. Fu, Robert B. Kusner, and John M. Sullivan. Ropelength criticality. *Geom. Topol.*, 18(4):1973–2043, 2014.
- [10] Jason Cantarella, Greg Kuperberg, Robert B. Kusner, and John M. Sullivan. The second hull of a knotted curve. *Amer. J. Math.*, 125(6):1335–1348, 2003.
- [11] Jason Cantarella, Robert B. Kusner, and John M. Sullivan. On the minimum ropelength of knots and links. *Invent. Math.*, 150(2):257–286, 2002.
- [12] Aldo-Hilario Cruz-Cota and Teresita Ramirez-Rosas. An Elementary Upper Bound For the Number Of Generic Quadrisecants Of Polygonal Knots. *App. Math. E-Notes*, 17:268-276, 2017.
- [13] Elizabeth Denne. Quadrisecants and the unknotting number of knots. In preparation, 2015.
- [14] Elizabeth Denne, Yuanan Diao, and John M. Sullivan. Quadrisecants give new lower bounds for the ropelength of a knot. *Geom. Topol.*, 10:1–26, 2006.
- [15] Elizabeth Denne and John M. Sullivan. The distortion of a knotted curve. *Proc. Amer. Math. Soc.*, 137(3):1139–1148, 2009.
- [16] Elizabeth Denne. *Alternating quadrisecants of knots*. ProQuest LLC, Ann Arbor, MI, 2004. Thesis (Ph.D.)—University of Illinois at Urbana-Champaign.
- [17] Y. Diao and C. Ernst. Total curvature, ropelength and crossing number of thick knots. *Math. Proc. Cambridge Philos. Soc.*, 143(1):41–55, 2007.
- [18] Yuanan Diao. Minimal knotted polygons on the cubic lattice. *J. Knot Theory Ramifications*, 2(4):413–425, 1993.
- [19] Yuanan Diao. The lower bounds of the lengths of thick knots. *J. Knot Theory Ramifications*, 12(1):1–16, 2003.
- [20] Yuanan Diao and Claus Ernst. Hamiltonian cycles and rope lengths of Conway algebraic knots. *J. Knot Theory Ramifications*, 15(1):121–142, 2006.
- [21] Yuanan Diao, Claus Ernst, and Uta Ziegler. The linearity of the ropelengths of Conway algebraic knots in terms of their crossing numbers. *Kobe J. Math.*, 28(1-2):1–19, 2011.

- [22] István Fáry. Sur la courbure totale d'une courbe gauche faisant un nœud. *Bull. Soc. Math. France*, 77:128–138, 1949.
- [23] W. Fenchel. Über Krümmung und Windung geschlossener Raumkurven. *Math. Ann.*, 101:238–252, 1929.
- [24] T. Fiedler and V. Kurlin. Fiber quadrisecants in knot isotopies. *J. Knot Theory Ramifications*, 17(11):1415–1428, 2008.
- [25] G. Flowers. Satanic and Thelemic circles on knots. *J. Knot Theory Ramifications*, 22(5):1350017, 16, 2013.
- [26] Henryk Gerlach, Philipp Reiter, and Heiko von der Mosel. The elastic trefoil is the twice covered circle. *H. Arch Rational Mech Anal*, 225(1):89–139, 2017.
- [27] Oscar Gonzalez and John H. Maddocks. Global curvature, thickness, and the ideal shapes of knots. *Proc. Natl. Acad. Sci. USA*, 96(9):4769–4773 (electronic), 1999.
- [28] Mikhael Gromov. Homotopical effects of dilatation. *J. Differential Geom.*, 13(3):303–310, 1978.
- [29] Mikhael Gromov. *Structures métriques pour les variétés riemanniennes*, volume 1 of *Textes Mathématiques [Mathematical Texts]*. CEDIC, Paris, 1981. Edited by J. Lafontaine and P. Pansu.
- [30] Mikhael Gromov. Filling Riemannian manifolds. *J. Differential Geom.*, 18(1):1–147, 1983.
- [31] D. Hilbert and S. Cohn-Vossen. *Geometry and the imagination*. Chelsea Publishing Company, New York, N. Y., 1952. Translated by P. Néményi.
- [32] Youngsik Huh, Kyungpyo Hong, Hyoungjun Kim, Sungjong No, and Seungsang Oh. Minimum lattice length and ropelength of 2-bridge knots and links. *J. Math. Phys.*, 55(11):113503, 11, 2014.
- [33] Choon Bae Jeon and Gyo Taek Jin. A computation of superbridge index of knots. *J. Knot Theory Ramifications*, 11(3):461–473, 2002. Knots 2000 Korea, Vol. 1 (Yongpyong).
- [34] Gyo Taek Jin. Superbridge index of knots. *Kobe J. Math.*, 18(2):181–197, 2001.
- [35] Gyo Taek Jin. Quadrisecants of knots with small crossing number. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 507–523. World Sci. Publ., Singapore, 2005.
- [36] Gyo Taek Jin and Seojung Park. Quadrisecant approximation of hexagonal trefoil knot. *J. Knot Theory Ramifications*, 20(12):1685–1693, 2011.
- [37] Nicolaas H. Kuiper. A new knot invariant. *Math. Ann.*, 278(1-4):193–209, 1987.
- [38] Greg Kuperberg. Quadrisecants of knots and links. *J. Knot Theory Ramifications*, 3(1):41–50, 1994.
- [39] J. W. Milnor. On the total curvature of knots. *Ann. of Math. (2)*, 52:248–257, 1950.
- [40] H. R. Morton and D. M. Q. Mond. Closed curves with no quadrisecants. *Topology*, 21(3):235–243, 1982.
- [41] Chad A. S. Mullikin. *On length minimizing curves with distortion thickness bounded below and distortion bounded above*. PhD thesis, University of Georgia, 2006.

- [42] Chad A. S. Mullikin. A class of curves in every knot type where chords of high distortion are common. *Topology Appl.*, 154(14):2697–2708, 2007.
- [43] J.-P. Otal. Une propriété de géométrie élémentaire des noeds. *Le journal de maths des élèves*, 1(2):34–39, 1994.
- [44] Erika Pannwitz. Eine elementargeometrische Eigenschaft von Verschlingungen und Knoten. *Math. Ann.*, 108(1):629–672, 1933.
- [45] John Pardon. On the distortion of knots on embedded surfaces. *Ann. of Math. (2)*, 174(1):637–646, 2011.
- [46] Piotr Pierański. In search of ideal knots. In *Ideal knots*, volume 19 of *Ser. Knots Everything*, pages 20–41. World Sci. Publ., River Edge, NJ, 1998.
- [47] Helmut Pottmann and Johannes Wallner. *Computational Line Geometry*. Springer Verlag, 2 edition, 2010.
- [48] Carsten Schmitz. The theorem of Fáry and Milnor for Hadamard manifolds. *Geom. Dedicata*, 71(1):83–90, 1998.
- [49] Friedemann Schuricht and Heiko von der Mosel. Characterization of ideal knots. *Calc. Var. Partial Differential Equations*, 19(3):281–305, 2004.
- [50] M. Sommer. Visualization in geometric knot theory. <http://www3.math.tu-berlin.de/geometrie/lab/curvesnsurfaces.shtml#knots>, 2007.
- [51] M. Sommer. Visualization in geometric knot theory. Master’s thesis, TU Berlin, 2007.
- [52] John M. Sullivan. Approximating ropelength by energy functions. In *Physical knots: knotting, linking, and folding geometric objects in \mathbb{R}^3* (Las Vegas, NV, 2001), volume 304 of *Contemp. Math.*, pages 181–186. Amer. Math. Soc., Providence, RI, 2002.
- [53] John M. Sullivan and Nancy C. Wrinkle. Some ropelength-critical clasps. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 565–580. World Sci. Publ., Singapore, 2005.
- [54] Julia Viro. Lines joining components of a link. *J. Knot Theory Ramifications*, 18(6):865–888, 2009.

Gyo Taek Jin

Polygonal approximation of unknots by quadrisecants

Abstract: If a knot K in \mathbb{R}^3 has only finitely many quadrisecants which meet K at finitely many points, we use these points to form a polygonal approximation called the quadrisecant approximation which is denoted by \hat{K} . The quadrisecant approximation conjecture states that \hat{K} has the knot type of K if K is nontrivial. We give examples of unknots, some polygonal and some smooth, having finitely many but at least two quadrisecants, for which the conjecture holds.

Keywords: knots, quadrisecants, polygonal approximation

MSC: 57M25

8.1 Introduction

A *knot* K is the image of a locally flat embedding of S^1 in \mathbb{R}^3 , i.e., we do not consider wild knots. A *quadrisecant* of a knot K is a straight line which meets K at four distinct places. The existence of quadrisecants for nontrivial knots was first announced by Pannwitz in 1933 for polygonal knots and later by Morton and Mond for smooth knots and by Kuperberg for tame knots.



Fig. 8.1: A quadrisecant


Theorem 8.1.1 ([10, 9, 8]).

Every nontrivial knot has a quadrisecant.

It is not hard to see that every knot can be perturbed to have infinitely many quadrisecants. On the other hand, every knot can be perturbed to have only finitely many quadrisecants which meet the knot at finitely many points.

Gyo Taek Jin, Korea Advanced Institute of Science and Technology, 291 Daehak-ro Yuseong-gu Daejeon Korea 34141, E-mail: trefoil@kaist.ac.kr

<https://doi.org/10.1515/9783110571493-007>

Open Access.  © 2018 Gyo Taek Jin, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

Theorem 8.1.2 ([1, 6]).

A polygonal knot satisfying the following conditions has only finitely many quadrisecants^{8.1}.

1. No four vertices are coplanar.
2. No three edges are parallel to a single plane.
3. No four edges lie on a single hyperboloid of one sheet.

Definition 8.1.3. A polygonal knot is said to be in *general position* if it satisfies the three conditions mentioned in Theorem 8.1.2.

8.2 Quadrisecant approximation of knots

Definition 8.2.1. Let $\alpha : [a, b] \rightarrow \mathbb{R}^3$ be a parametrization of a knot K and let $a \leq t_1 < \dots < t_n = t_0 < b$, for some $n \geq 3$. Then the polygonal closed curve

$$K' = \overline{\alpha(t_0)\alpha(t_1)} \cup \overline{\alpha(t_1)\alpha(t_2)} \cup \dots \cup \overline{\alpha(t_{n-1})\alpha(t_n)}$$

is called a *polygonal approximation*, or simply, an *approximation* of the knot K . It may happen that K' has one or more self intersection points. We say that the polygonal approximation K' of K is a *good approximation* if K' is a knot which has the knot type of K .

Definition 8.2.2. Suppose that K is a knot having only finitely many quadrisecants which intersect K in finitely many points. The polygonal approximation of K using the points on the quadrisecants is called the *quadrisecant approximation* of K and is denoted by \hat{K} .

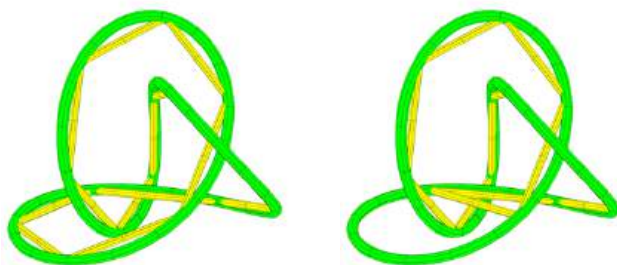


Fig. 8.2: A good approximation and a bad approximation

8.1 Possibly no quadrisecants in the case of an unknot

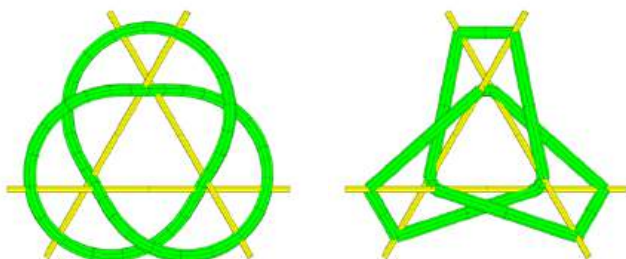


Fig. 8.3: The quadrisecant approximation of a smooth trefoil knot

Figure 8.3 shows that the quadrisecant approximation of a smooth trefoil is a good approximation and Figure 8.4 shows that the quadrisecant approximations of a polygonal trefoil and a polygonal figure eight knot are good approximations.

Conjecture 8.2.3 (Quadrisecant approximation conjecture^{8.2}).

Let K be a nontrivial knot having only finitely many quadrisecants which intersect K in finitely many points. Then

1. \hat{K} is a good approximation of K .
2. The quadrisecants of \hat{K} are those of K .

Theorem 8.2.4 ([7, 11]).

The quadrisecant approximation conjecture holds for

1. hexagonal trefoil knots and
2. heptagonal figure eight knots in general position.

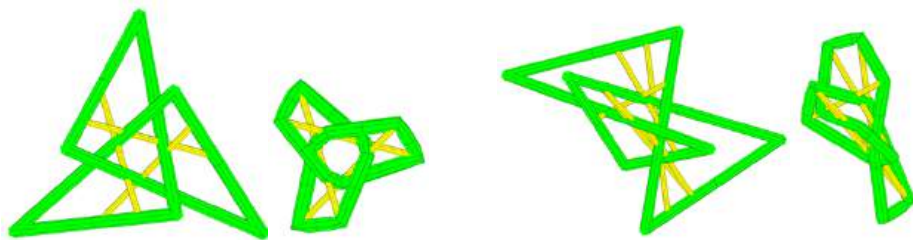


Fig. 8.4: The quadrisecant approximations of a polygonal 3_1 and a polygonal 4_1

8.2 In the original statement of the conjecture in [6], the word ‘nontrivial’ is missing.

8.3 Quadrisecants of Polygonal Unknots

In this part, we examined the part 1 of the quadrisecant approximation conjecture for 133 polygonal unknots of which the vertex data were supplied by Claus Ernst [4]. It is explained in Section 8.5 how we obtained quadrisecants, and in Section 8.6 how we confirmed quadrisecant approximations are good.

Example 8.3.1. We examined thirty 10-gonal unknots. They have 9.1 quadrisecants in average. Except the four cases of less than 2 quadrisecants, the quadrisecant approximations are unknots. The quadrisecant approximation is not defined or not embedded if there exist less than two quadrisecants. Figure 8.6 shows two unknots which have no quadrisecant and only one quadrisecant. In each of Figures 8.7 and 8.8, the quadrisecants are shown on the left and the quadrisecant approximation is shown on the right.

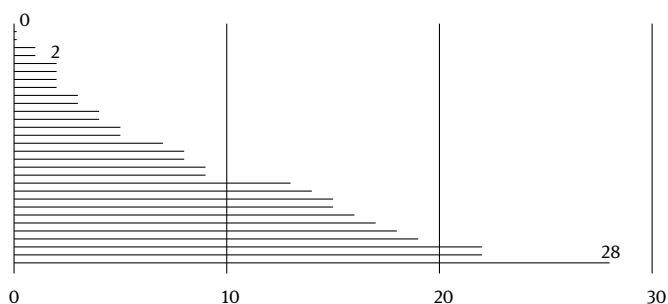


Fig. 8.5: Distribution of quadrisecants of thirty 10-gonal unknots

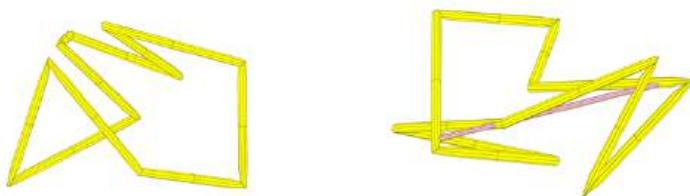


Fig. 8.6: 10-gonal unknots with 0 and 1 quadrisecant

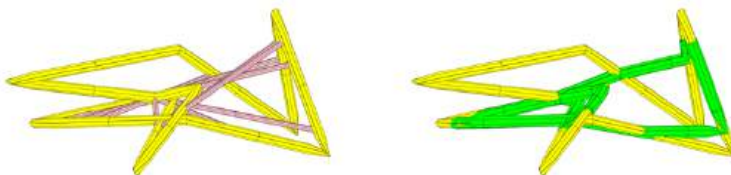


Fig. 8.7: A 10-gonal unknot with 9 quadrisecants and its quadrisecant approximation

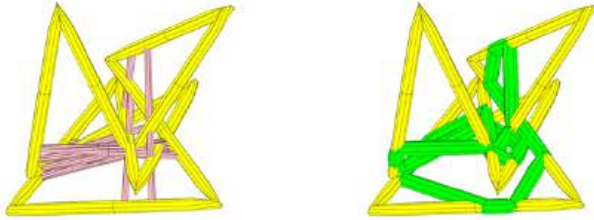


Fig. 8.8: A 10-gonal unknot with 28 quadrisecants and its quadrisecant approximation

Example 8.3.2. We examined thirty 20-gonal unknots. They have 305.5 quadrisecants in average. All quadrisecant approximations are unknots. Figures 8.10 and 8.11 show two of them.

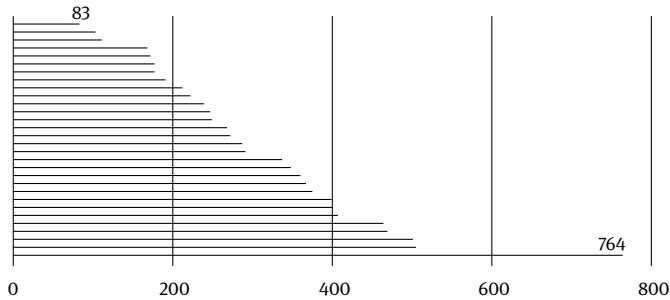


Fig. 8.9: Distribution of quadrisecants of thirty 20-gonal unknots

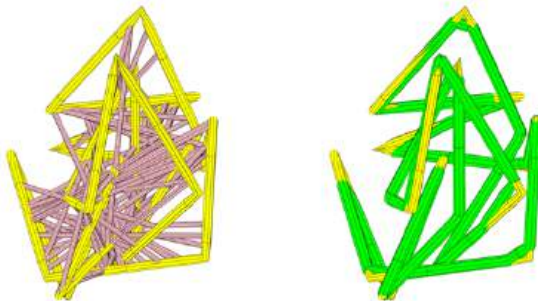


Fig. 8.10: A 20-gonal unknot with 83 quadrisecants and its quadrisecant approximation

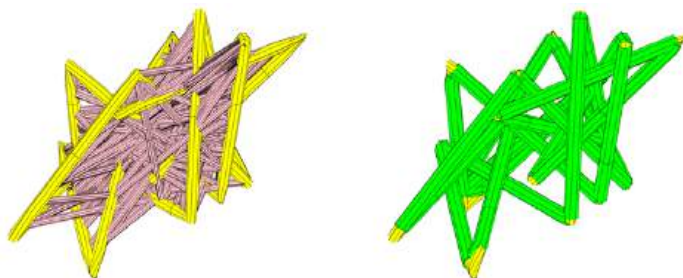


Fig. 8.11: A 20-gonal unknot with 764 quadriseccants and its quadriseccant approximation

Example 8.3.3. We examined thirty 30-gonal unknots. They have 1577.4 quadriseccants in average. All quadriseccant approximations are unknots. Figure 8.13 shows one of them; the knot, its quadriseccants, and its quadriseccant approximation with those quadriseccants incident to the new vertices.

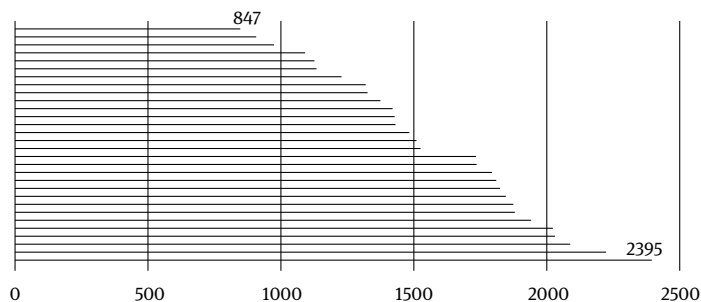


Fig. 8.12: Distribution of quadriseccants of thirty 30-gonal unknots



Fig. 8.13: A 30-gonal unknot with 1,431 quadriseccants and its quadriseccant approximation

Example 8.3.4. We examined thirty 40-gonal unknots. They have 5317.1 quadrisecants in average. All quadrisecant approximations are unknots. Figure 8.15 shows one of them.

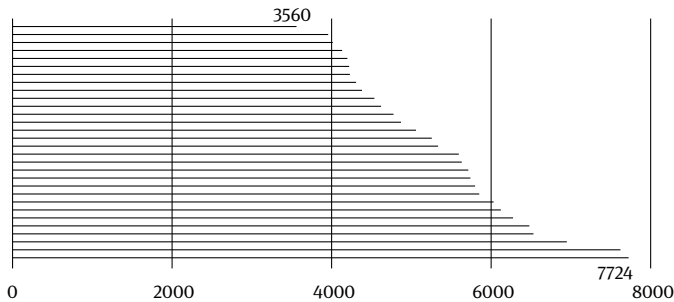


Fig. 8.14: Distribution of quadrisecants of thirty 40-gonal unknots

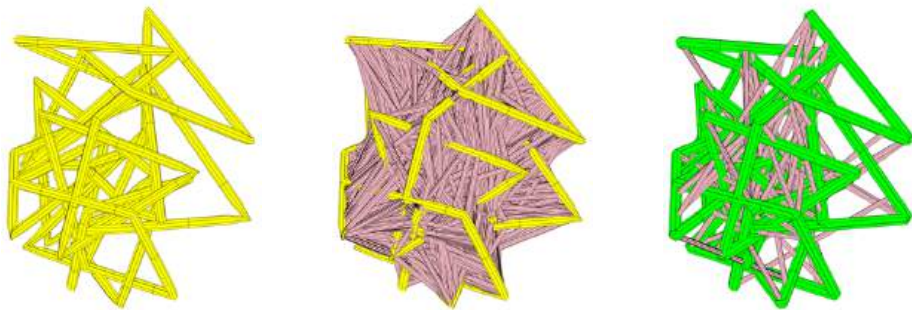


Fig. 8.15: A 40-gonal unknot with 6,275 quadrisecants and its quadrisecant approximation

Example 8.3.5. We examined thirteen 50-gonal unknots. They have 13161.4 quadrisecants in average. All quadrisecant approximations are unknots. Figure 8.17 shows one of them.

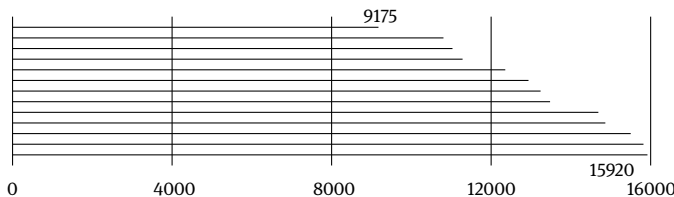


Fig. 8.16: Distribution of quadrisecants of thirteen 50-gonal unknots

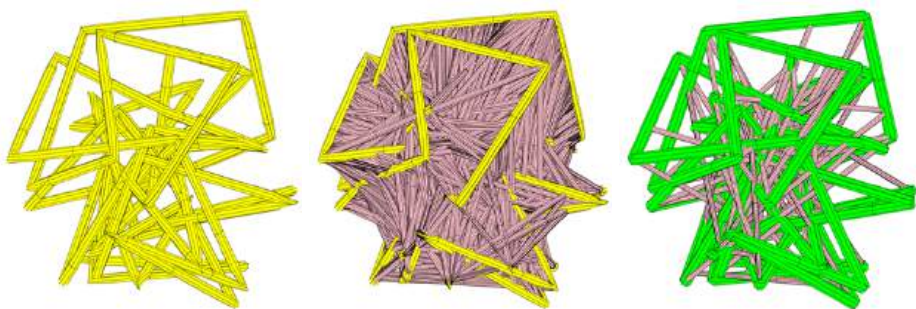


Fig. 8.17: A 50-gonal unknot with 15,920 quadriseccants and its quadriseccant approximation

8.4 Quadriseccants of Smooth Unknots

We constructed seven smooth unknots to examine the part 1 of the quadriseccant approximation conjecture. It is explained in Section 8.5 how we obtained quadriseccants, and in Section 8.6 how we confirmed quadriseccant approximations are good.

Example 8.4.1. Figure 8.18 shows a smooth unknot, its quadriseccants and the quadriseccant approximation. With $a = 1,000$ random choices of four points on this knot, we performed Newton's method and obtained $b = 184$ quadriseccants. Reducing duplicates, we obtained $c = 99$ quadriseccants as shown in the middle. Among the $a - b = 816$ failures, $d = 27$ cases diverged and $e = 789$ cases converged with less than four secant points. The approximation shown on the right is a good approximation.

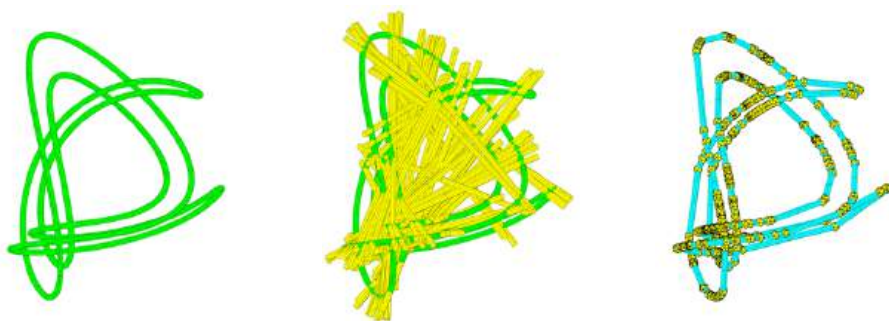


Fig. 8.18: A smooth unknot with 99 quadriseccants and its quadriseccant approximation

Example 8.4.2. Figure 8.19 shows a smooth unknot. We performed Newton's method with $a = 1,000$, $b = 143$, $c = 107$, $d = 28$, and $e = 829$ as in Example 8.4.1, and obtained a good approximation.

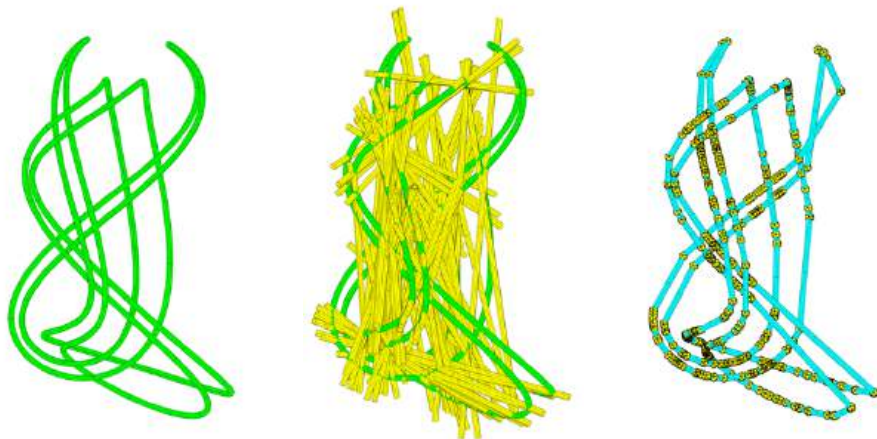


Fig. 8.19: A smooth unknot with 107 quadrisecants and its quadrisecant approximation

Example 8.4.3. Figure 8.20 shows a smooth unknot. We performed Newton's method with $a = 1,000$, $b = 54$, $c = 31$, $d = 19$, and $e = 927$ as in Example 8.4.1, and obtained a good approximation.

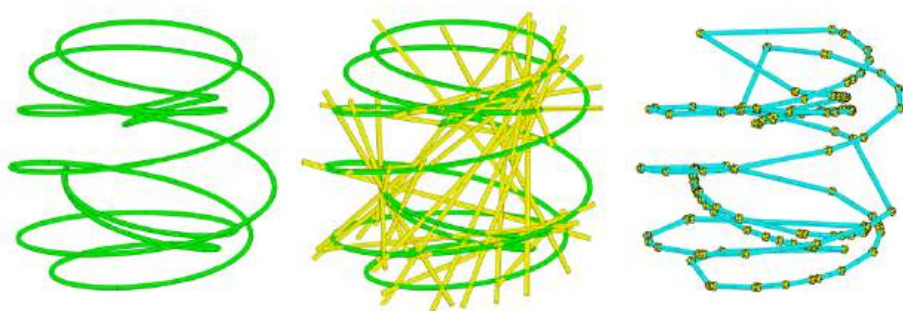


Fig. 8.20: A smooth unknot with 31 quadrisecants and its quadrisecant approximation

Example 8.4.4. Figure 8.21 shows the smooth unknot parametrized by

$$\begin{cases} x = \left(\frac{1}{2} \cos 3t + 4 + \cos 2t\right) \cos 6t \\ y = \left(\frac{1}{2} \cos 3t + 4 + \cos 2t\right) \sin 6t \\ z = 4 \sin 2t - \frac{1}{2} \sin 3t \end{cases} \quad 0 \leq t \leq 2\pi$$

We performed Newton's method with $a = 1,000$, $b = 129$, $c = 51$, $d = 6$, and $e = 865$ as in Example 8.4.1, and obtained a good approximation.

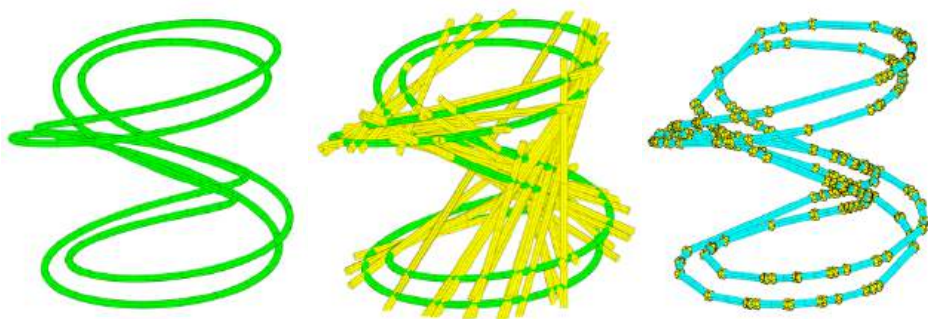


Fig. 8.21: A smooth unknot with 51 quadriseccants and its quadriseccant approximation

Example 8.4.5. Figure 8.22 shows the smooth unknot parametrized by

$$\begin{cases} x = \left(\frac{1}{2} \cos 5t + 4 + \cos 3t\right) \cos 9t \\ y = \left(\frac{1}{2} \cos 5t + 4 + \cos 3t\right) \sin 9t \\ z = 4 \sin 3t - \frac{1}{2} \sin 5t \end{cases} \quad 0 \leq t \leq 2\pi$$

We performed Newton's method with $a = 1,000$, $b = 152$, $c = 121$, $d = 44$, and $e = 804$ as in Example 8.4.1, and obtained a good approximation.

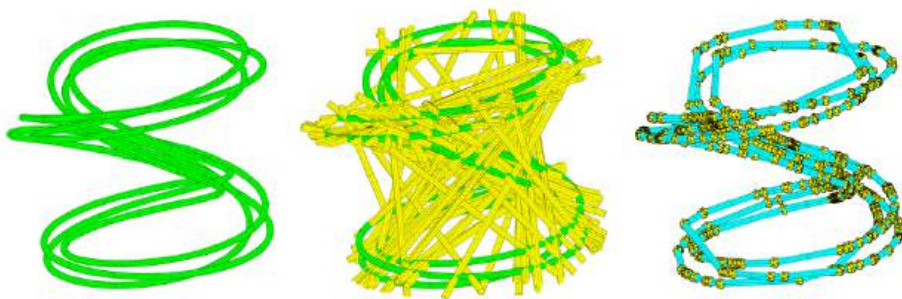


Fig. 8.22: A smooth unknot with 121 quadriseccants and its quadriseccant approximation

Example 8.4.6. Figure 8.23 shows the smooth unknot parametrized by

$$\begin{cases} x = \left(\frac{1}{2} \cos 8t + 4 + \cos 3t\right) \cos 12t \\ y = \left(\frac{1}{2} \cos 8t + 4 + \cos 3t\right) \sin 12t \\ z = 4 \sin 3t - \frac{1}{2} \sin 8t \end{cases} \quad 0 \leq t \leq 2\pi$$

We performed Newton's method with $a = 1,000$, $b = 217$, $c = 197$, $d = 37$, and $e = 746$ as in Example 8.4.1, and obtained a good approximation.

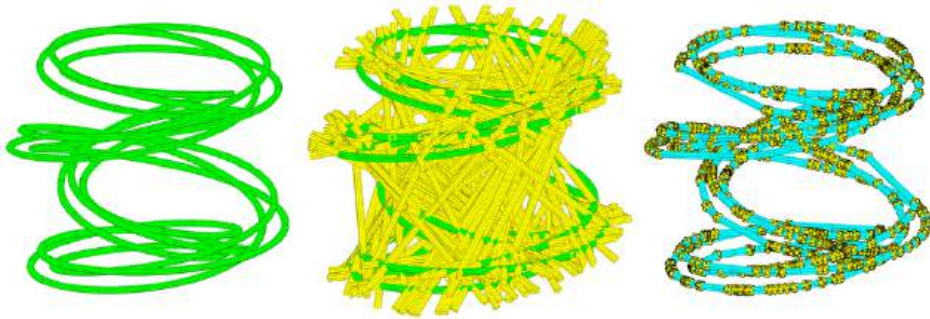


Fig. 8.23: A smooth unknot with 197 quadriseccants and its quadriseccant approximation

Example 8.4.7. Figure 8.24 shows the smooth unknot parametrized by

$$\begin{cases} x = \left(\frac{1}{2} \cos 7t + 4 + \cos 2t\right) \cos 10t \\ y = \left(\frac{1}{2} \cos 7t + 4 + \cos 2t\right) \sin 10t \\ z = 4 \sin 2t - \frac{1}{2} \sin 7t \end{cases} \quad 0 \leq t \leq 2\pi$$

We performed Newton's method with $a = 1,000$, $b = 145$, $c = 117$, $d = 46$, and $e = 809$ as in Example 8.4.1, and obtained a good approximation.

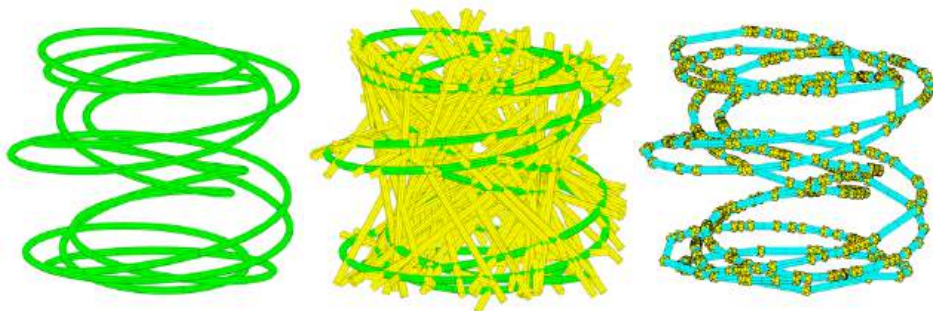


Fig. 8.24: A smooth unknot with 117 quadriseccants and its quadriseccant approximation

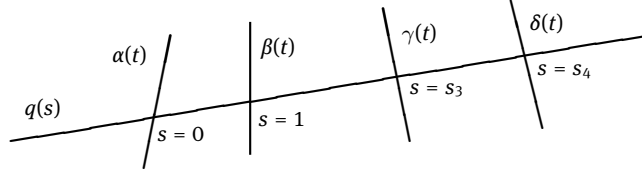
8.5 Finding Quadrisecants

For a polygonal knot K , we parametrize four distinct edges by

$$\alpha, \beta, \gamma, \delta : [0, 1] \rightarrow \mathbb{R}^3$$

which have constant tangent vectors coherent with an orientation of K . A line passing through $\alpha(t_1)$ and $\beta(t_2)$ can be parametrized by

$$q(s) = (1 - s)\alpha(t_1) + s\beta(t_2), \quad s \in \mathbb{R}. \quad (8.5.1)$$



It is a quadrisecant of K if it also intersects γ and δ . If the system

$$q(s_3) = (1 - s_3)\alpha(t_1) + s_3\beta(t_2) = \gamma(t_3)$$

$$q(s_4) = (1 - s_4)\alpha(t_1) + s_4\beta(t_2) = \delta(t_4)$$

can be solved for the six variables $s_3, s_4 \in \mathbb{R}, t_1, t_2, t_3, t_4 \in [0, 1]$ then each solution determines a quadrisecant parametrized by (8.5.1).

For a smooth knot K , we consider a parametrization by a differentiable map $f : [a, b] \rightarrow \mathbb{R}^3$ which is one-to-one for $a \leq t < b$ and $f(a) = f(b)$. If there exist $t_1, t_2, t_3, t_4 \in [a, b]$ such that the four points $f(t_1), f(t_2), f(t_3), f(t_4)$ are distinct and collinear, then there exist nonzero real numbers λ and μ satisfying

$$f(t_3) - f(t_1) = \lambda(f(t_2) - f(t_1)), \quad f(t_4) - f(t_1) = \mu(f(t_2) - f(t_1))$$

Therefore, finding a quadrisecant of K is equivalent to finding a zero of the vector valued map $F = (F_1, F_2, F_3, F_4, F_5, F_6) : \mathbb{R}^6 \rightarrow \mathbb{R}^6$ given by

$$F_1(t_1, t_2, t_3, t_4, \lambda, \mu) = x(t_3) - \lambda x(t_2) + (\lambda - 1)x(t_1)$$

$$F_2(t_1, t_2, t_3, t_4, \lambda, \mu) = x(t_4) - \mu x(t_2) + (\mu - 1)x(t_1)$$

$$F_3(t_1, t_2, t_3, t_4, \lambda, \mu) = y(t_3) - \lambda y(t_2) + (\lambda - 1)y(t_1)$$

$$F_4(t_1, t_2, t_3, t_4, \lambda, \mu) = y(t_4) - \mu y(t_2) + (\mu - 1)y(t_1)$$

$$F_5(t_1, t_2, t_3, t_4, \lambda, \mu) = z(t_3) - \lambda z(t_2) + (\lambda - 1)z(t_1)$$

$$F_6(t_1, t_2, t_3, t_4, \lambda, \mu) = z(t_4) - \mu z(t_2) + (\mu - 1)z(t_1)$$

where $f(t) = (x(t), y(t), z(t))$ for $t \in [a, b]$. With an initial guess $\mathbf{v}_0 \in \mathbb{R}^6$ for a solution of the equation $F(\mathbf{v}) = \mathbf{0}$, the sequence

$$\mathbf{v}_{n+1} = \mathbf{v}_n - (D_F(\mathbf{v}_n))^{-1}F(\mathbf{v}_n)$$

may converge to a solution. This is Newton's method we used to find quadrisecants in Examples 6–12.

8.6 Test for Good Approximations

Let K be a polygonal knot. If there is a triangular disk Δ which intersects K along two of its edges and nowhere else, then the portion of K along the two edges of Δ can be replaced by the third edge of Δ without changing the knot type. Such a move is called a *triangular move* if Δ contains three vertices of K , and a *corner move* if Δ contains a single vertex of K . If Δ contains two vertices of K , then it can be described as a combination of a corner move followed by a triangular move.



Fig. 8.25: A triangular move



Fig. 8.26: A corner move

After finding each quadrisecant approximation, we obtained a sequence of triangular moves and corner moves to deform the polygonal curve into a triangle; which shows that the quadrisecant approximation is an unknot. We first scanned through the vertices at which the corner moves can be performed. If we ran out of such vertices, we broke the edges to perform corner moves. Such processes were successfully performed to result in triangles.

We used Maple 18.02 to draw knots and unknots, to find quadrisecants, to construct quadrisecant approximations, and to find sequences of triangular moves and corner moves to show quadrisecant approximations are unknots.

Acknowledgment: The author would like to thank Claus Ernst for supplying these random unknot data [4], and the reviewer for valuable comments and suggestions.

Funding: This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (No. 2011-0027989).

Appendix

The vertices of the 10-gonal unknots of Figure 8.6

(Left Figure)

[0, 0, 0],

[0.26079016715637726, -0.6651136228131733, 0.6997230576898867],

[0.3851146726381711, 0.32376039919446953, 0.6180432809080222],

[0.29607875594678595, -0.26743750167257857, -0.18355365493275347],

[-0.5165348788026222, -0.4202409485147839, -0.7459682593417223],

[-0.2605699687081926, 0.5310040460410358, -0.5738783841809526],

[-0.21143605902512672, 0.4897806885652303, 0.4240627431996551],

[-0.9792224753363977, 0.030222606418956677, -0.02237917694442637],

[-0.3139988528585777, 0.30446438577681256, 0.6720767097410592],

[0.6397169293922946, 0.5647064005590116, 0.5214105210067966]

(Right Figure)

[0, 0, 0],

[0.8284426241853787, -0.13711967475775255, -0.5430294773096314],

[0.45649991860657285, 0.5694289821222773, 0.05900675019478009],

[0.5844719512915223, -0.22266670353456428, -0.5378245355253136],

[-0.20529732546788465, 0.28931020656749296, -0.19997230005946198],

[-0.5958278971220673, 0.28647492442354117, 0.7206132927990302],

[0.3237832719482289, 0.3557541454361506, 0.3339407087475498],

[-0.5730602105134658, 0.5135412952370683, -0.07930859049406143],

[-0.5605933955343954, -0.3916939036304804, 0.3454193380148927],

[0.33521831640868194, -0.5627213045367457, 0.7556278275477895]

The vertices of the 10-gonal unknot of Figure 8.7

[0, 0, 0],
 [0.2599759088879847, 0.7063479811963731, -0.658395820352525],
 [0.3232460020389661, -0.0708267405320456, -0.03229970375458519],
 [-0.3890093612580253, -0.577239084649801, 0.45374436336261653],
 [0.5985360102269767, -0.43015793235373406, 0.39787600113666116],
 [-0.335615146452009, -0.4896887492626049, 0.7497533921601192],
 [0.16214824879081988, 0.2326112149861075, 0.269634277770117],
 [-0.17941933018462342, 0.5192489431721007, -0.6254469485849145],
 [0.10970548156622802, -0.07398668530301361, 0.1258707820309488],
 [0.2805221340395753, 0.8722168474076153, 0.40068079990461647]

The vertices of the 20-gonal unknot of Figure 8.10

[0, 0, 0],
 [0.8482169711790621, 0.40732568763520016, -0.338546531508343],
 [-0.08769695736784196, 0.0951362437632568, -0.501650392550402],
 [-0.41095356729969773, -0.3391937227912085, 0.33910084550290054],
 [0.16559924648720026, -0.2840335049210709, -0.4760949612079347],
 [0.5670984753772971, 0.3329930052222135, 0.2007186431035161],
 [-0.11690921121550236, -0.21262685423432814, 0.684900895616545],
 [-0.05162718976454717, 0.670660755319105, 0.22063668583437912],
 [0.22992403518773627, 0.24603905626447556, -0.6398430178308662],
 [-0.5451058080348717, 0.13296592839047053, -0.01811702523675838],
 [-0.38299286673662986, -0.2901069529831363, 0.8733585668631414],
 [0.07820219100325038, -0.5735108573538344, 0.032536739301705614],
 [-0.8321749557919547, -0.1981170757345217, -0.1415119875381466],
 [-0.15270625050450992, -0.7460520028999029, -0.6294562406319523],
 [-0.658560049051659, 0.0931762686547083, -0.4299372121688003],
 [-0.3760613614914784, 0.4254692083888686, 0.46993827573472363],
 [-0.18227772625757413, 0.6421489541499391, -0.48687821162972966],
 [0.6200378545169324, 0.059516981894417724, -0.6166075211634959],
 [0.15289516739144102, -0.8197458643172172, -0.5234702185560548],
 [0.4288551720589901, 0.06413567871531958, -0.9010936999635384]

The vertices of the 30-gonal unknot of Figure 8.13

$[0, 0, 0]$,
 $[0.9559164467618607, -0.25443934965439446, 0.1465754555088525]$,
 $[0.02804655254704769, -0.3985822747142186, -0.19734354158928644]$,
 $[0.2230273288845479, 0.34936016038878287, 0.43713625370053855]$,
 $[-0.1423832631217396, 0.7875966440994483, -0.3840968944337329]$,
 $[0.30538251681348216, -0.10105737443120202, -0.4830910441009188]$,
 $[0.3012109144597282, 0.8341007496167119, -0.12888505775824197]$,
 $[0.3461536690536176, 0.2106052374462082, -0.9094191666303365]$,
 $[0.13644196659122435, 0.867317947629769, -0.18502326787490653]$,
 $[0.6159415857389964, 0.06346331321990045, -0.5370145205715016]$,
 $[-0.07459829470376207, -0.5849247254798481, -0.21647181703702034]$,
 $[-0.6456744097860264, -0.11645100799238642, 0.4576234631764124]$,
 $[0.013065275565096495, 0.6158901155248228, 0.6300720228951164]$,
 $[0.3257452740151201, -0.3163186077957065, 0.44781364933054646]$,
 $[-0.3692232349170414, 0.3650512940457395, 0.21813163336676566]$,
 $[0.5962648581644622, 0.522606432033769, 0.4255180374809578]$,
 $[0.4216649068013401, -0.4357824399672629, 0.19967378360544916]$,
 $[0.6891481900165033, 0.5134219687382546, 0.03395207804324586]$,
 $[0.48005323271764483, -0.34986327834578396, 0.4933189042282012]$,
 $[-0.07222465637064997, -0.5869713688224925, -0.30591127954532604]$,
 $[-0.1301082479344901, 0.2085509519106162, 0.29724187434104754]$,
 $[0.485784516741387, 0.19721675867335445, -0.49050658708783085]$,
 $[-0.41900114810318617, 0.2874375363705996, -0.07430577982832054]$,
 $[-0.5142445788646942, -0.6836888594256834, -0.2930344068630529]$,
 $[-0.13620686745409896, -0.6200714061477229, 0.6305674058191582]$,
 $[0.49847904544111116, -0.324346799305082, -0.0833798953734013]$,
 $[0.1741444901061093, 0.6016937970515704, -0.2763977963957214]$,
 $[-0.7272768925103164, 0.5408356549693376, 0.1522463188143096]$,
 $[-0.24741374056874957, -0.3060880701790464, 0.38126937265275906]$,
 $[0.646115226201013, -0.7282664692363255, 0.2283923471950623]$

Bibliography

- [1] Budney, R., Conant, J., Scannell, K. P., and Sinha, D. (2005). *New perspectives on self-linking*. *Advances in Mathematics* **191**, 78–113.
- [2] Burde, G. and Zieschang, H. (1985). *Knots*. de Gruyter Studies in Mathematics vol. 5, Walter de Gruyter, Berlin, New York.

- [3] Denne, E. (2004). *Alternating quadrisecants of knots*. Ph.D. Thesis, Univ. Ill. Urbana-Champaign, (arXiv:math/0510561).
- [4] Ernst, C. (2013). *Random polygons that are unknots*. An email message to the author.
- [5] Jin, G. T. (1997). *Polygon indices and superbridge indices of torus knots and links*. J. Knot Theory Ramifications **6**, 281– 289.
- [6] Jin, G. T. (2005). *Quadrisecants of knots with small crossing number*. Physical and Numerical Models in Knot Theory, Series on Knots and Everything, vol. 36, World Scientific, 507– 523.
- [7] Jin, G. T. and Park, S. (2011). *Quadrisecant approximation of hexagonal trefoil knot*. J. Knot Theory Ramifications **20**, 1685– 1693.
- [8] Kuperberg, G. (1994). *Quadrisecants of knots and links*. J. Knot Theory Ramifications **3**, 41– 50.
- [9] Morton, H. R. and Mond, D. M. Q. (1982). *Closed curves with no quadrisecants*. Topology **21**, 235– 243.
- [10] Pannwitz, E. (1933). *Eine elementargeometrische Eigenschaft von Verschlingungen und Knoten*. Math. Ann. **108**, 629– 672.
- [11] Park, S. (2012). *Quadrisecant approximation of minimal polygonal knots*, Ph.D. Thesis, KAIST.
- [12] Rolfsen, D. (1976). *Knots and Links*. Mathematics Lecture Series 7, Publish or Perish.

Erin Brine-Doyle, Madeline Shogren, Emily Vecchia, and Eric J. Rawdon

Open knotting

Abstract: Traditional topological knot theory mainly studies knotting within closed curves. However, much of what people consider knotting in everyday life, e.g. in shoelaces, involve tangled open rope-like materials with two free ends. Furthermore, knotting has been discovered in open DNA and protein chains. Several notions of open knotting have been defined and applied in different situations. In this paper, we survey the different techniques used for measuring knotting in open chains, including some discussion of different properties of the techniques. We present the knotting fingerprint, a means for visualizing the knotting of subchains of open and closed chains via image matrices. Finally, we show image matrices for some random closed knots and relate features of the matrices with features of the configurations.



9.1 Introduction

From tying shoelaces to detangling headphone cables, knotting is ever-present in modern day life. Knotting can be decorative (as in Celtic art and architecture), provide utility (as in boating applications and the manufacturing of clothing), and be used for safety (as in recreational climbing and surgical sutures). In addition to the human scale, knotting is present at microscopic scales (in DNA [35, 37] and proteins [3, 18, 26, 38, 44]), and may even be present at the subatomic scale in tightly knotted glueballs [5, 6, 7]. In many of these applications, the knotting is in an open arc (which we will call an open chain), i.e. the knotting occurs in a non self-intersecting arc with two free ends. Traditional knot theory, however, works predominantly with closed curves, i.e. curves that form loops. So while there is a long and rich history of studying knots mathematically, little of that work involves open chain knotting which might, indeed, be most relevant scientifically.

Knotting has been studied mathematically since the 1800s with Gauss's integral to measure the amount of interlinking between two loops [15]. Perhaps most famously, Lord Kelvin proposed that atoms were vortices in the æther (a hypothetical material that was thought to permeate the universe) forming different types of knots [45]. This hypothesis (which, ultimately, was discredited) inspired Peter Guthrie Tait to develop the first table of mathematical knots [40, 41, 42]. A mathematical knot is a simple closed curve, a one-dimensional non self-intersecting loop with no free ends. Given a mathematical knot, one can deform space (including the knot) continuously to obtain

Erin Brine-Doyle, Madeline Shogren, Emily Vecchia, Eric J. Rawdon, Department of Mathematics, University of St. Thomas, Saint Paul, MN 55105, USA

<https://doi.org/10.1515/9783110571493-008>

 Open Access.  © 2018 Erin Brine-Doyle et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

equivalent knots. These topological deformations are akin to manipulating a piece of knotted rope, although the “rope” is infinitely stretchy, infinitely thin, and cannot be fully pulled tight. This equivalence then divides mathematical knots into classes of curves called *knot types*, which are the objects studied in knot tables.

Applying the standard knot theory equivalence to open chains (with two free ends) turns out to be quite unrewarding since any open chain can be deformed topologically to a straight line segment, and thus all open curves are equivalent. So while we may perceive knotting in an open curve, the traditional mathematical techniques are not capable of classifying that entanglement. Still, as humans, we know knots when we see them, even if the curve happens to be open. This human intuitive notion of knotting was originally referred to as the “reasonable-person test” by Mansfield [25]. The question is how to take this human intuitive notion of knotting and make it consistent with the mathematical theory. In other words, how can we define knotting in open chains so that it coincides with the opinion of a reasonable person?

While it is interesting, intellectually, to consider how one might measure knotting in open curves, the desire to formally analyze these structures has been enhanced by the discovery of entanglements within proteins. In particular, anyone with an internet connection now has access to the geometric structure of a dizzying number of proteins via the Protein Data Bank (PDB) [2]. Currently, the KnotProt database [18] lists over 1200 structures that have been classified as containing some sort of knotting (although we wait for the next section to discuss more specifically about how knotting is detected in proteins). Proteins are chains of (typically hundreds of) amino acids. Each amino acid contains an alpha carbon. When researchers talk about knotting in proteins, they are really talking about knotting within the alpha carbon backbone (i.e. the open chain polygon with alpha carbons as vertices and edges joining alpha carbons whose amino acids are joined by peptide bonds).

To provide a brief history of knotting in proteins, the first knotted protein was discovered by Richardson in 1977 [34]. In 1994, Mansfield surveyed known protein structures (≈ 400 from the Brookhaven Protein Database) [25] in an attempt to find knots, but concluded “In summary, none of the 400 protein structures analyzed were found to have knots. Only one, human carbonic anhydrase B, comes close.” He also questioned how one should define knotting in proteins, since proteins are open chains. In 1997, Mansfield used a different technique [26] to identify two knotted proteins (one being Richardson’s knot [34], the other having been identified by Takusagawa and Kamitori in 1996 [43]). However, both of these knots were “shallow”, i.e. the removal of a few amino acids on either end would remove the knot. Later, Taylor [44] discovered a “deep” knot (meaning that the knotting occurs away from the endpoints, deep into the protein), which cemented the idea that knotting could exist in proteins. The most complicated knot seen to date is the Stevedore knot, 6_1 , from [3]. See [12] for a recent survey on knotting in proteins.

In Mansfield’s second paper on protein knotting [26], he made use of the fact that the ends of the proteins are on or near the “exterior” (i.e. the endpoints lie on the

convex hull), in which case one can unobtrusively close the protein chain via a simple arc that avoids the entangled portion (as in Figure 9.1A). If one were interested only in classifying knotting in such open chains, then the classification is straightforward and relatively unambiguous.

However, the knotting of an entire structure may not reveal all of the entanglement within a curve. The primary example of this behavior is with slipknots, where the entire chain is unknotted but there are subarcs of the chain (which we call subchains) which are knotted. Yeates’s group did the first PDB search for slipknots, detecting 37 slipknotted protein chains [20]. Currently, the KnotProt [18] database has cataloged over 400 slipknotted structures in the PDB. When analyzing subchains, we necessarily enter into situations where the endpoints are no longer on the “outside”, but rather are embedded directly in the complexity that defines the entanglement. In such cases, the classification of the knotting in the open chain is more ambiguous, oftentimes very ambiguous.

The purpose of this note is to survey the different notions that have been proposed for measuring knotting in open chains. In Section 2, we introduce the several different definitions and discuss some of the advantages and disadvantages of each approach. In Section 3, we present image matrices which communicate the knotting seen within subchains of open and closed chains. In Section 4, we show some examples of image matrices for random closed chains and discuss some of the relationships between features of the image matrices and features of the corresponding configurations.

9.2 Defining open knotting

We should begin by saying that there is no universally agreed upon definition for knotting in open chains. Knotting traditionally has been studied on closed curves and analyzed topologically, in which case the equivalence allows a sort of movement of the configurations. For open chains, we enter into more of a static state where small perturbations of a configuration can immensely change our perception of the knotting present in the curve. This paradigm shift is not unprecedented; the study of knotting in physical systems is replete with examples of searching for and studying certain special configurations, e.g. the minimizing configurations of the Möbius Energy (see e.g. [14, 22, 31]) or ropelength (see e.g. [4, 16, 23]). One of the reasons that there is no agreed upon definition is because different research groups from different disciplines have had different goals at different times. Currently, a main (and classic) tension is the balance between a definition having attractive properties versus being quick to compute. Having a “bad” definition that can be computed easily is not optimal, nor is having a “good” definition which is incomputable. So a compromise must be reached between the speed of computability and acceptance of limitations in the approach,

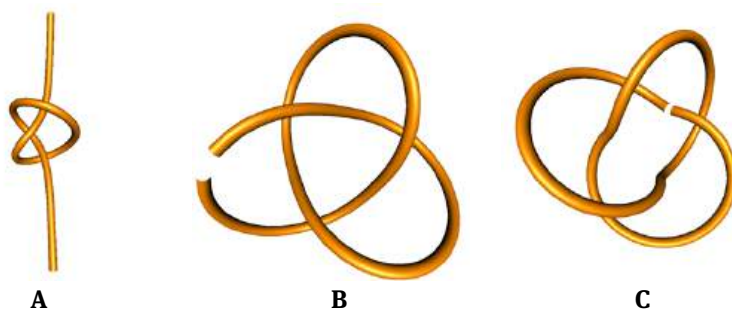


Fig. 9.1: Three examples of open chains that should be classified as being knotted by a reasonable classification scheme. The example in A is what most people would consider to be a knot. This is known as an overhand knot and when closed in a simple fashion becomes a trefoil knot, 3_1 . In B and C, we have closed knots from which a small part of the curve has been removed. Here again, we would expect a reasonable classification scheme for open chains to count these as being knotted, although the more precarious positioning of the opening in C may make us think that it is not as robustly knotted as B or A.

and that delicate balance leads to different opinions depending on the goals of the research.

In an attempt to define a mathematical notion of knotting for open chains, we begin by imagining properties that one might like the definition to satisfy. We discuss four such properties. First, the definition should be consistent with our intuition. In particular, the classification of the knotting in open entangled chains should roughly match the opinion of a reasonable person. For example, an overhand knot pulled tightly should correspond to a trefoil knot (as in Figure 9.1A). Second, open knotting should converge in some fashion to the classical knot theory of closed curves. In other words, by removing an arbitrarily short portion of a closed knot (see Figure 9.1, B and C), the definition should classify the open chain as the same knot type as the knot type of the original closed knot. Third, the definition should be well-defined, computable on a computer (at least for polygons), and not rely on any human interaction. Fourth, the classification should be stable in some sense. For example, an arbitrarily small perturbation of the endpoints of a given chain should not change the classification. Furthermore, a good classification scheme should minimize the effect of small perturbations on the classification. Millett mentioned two of these properties in [29] (he called the second property “continuity” and the fourth property “robustness”).

Specifically, our goal should be to define a notion of open knotting that is applicable to all open chains, not just a subclass of open chains (like full protein-like chains).

Researchers have defined several notions of knotting for open curves. In the remainder of this section, we introduce several different techniques used to measure knotting in open chains and discuss some of the properties of these techniques. In particular, we focus on how these algorithms behave on polygons, since this is the class of curves on which the algorithms are typically applied. All of the techniques de-

fined to date have relied (to some extent, at least) on closing the open chain in one or more ways and using the knot type classifications from traditional mathematical knot theory. We discuss two different classes of techniques, single closure and stochastic, as well as some related techniques and the traditional topological approach to studying knotted arcs in the subsections below. Figures 9.6 and 9.7 show examples of the single closure and stochastic techniques applied to a common open chain.

9.2.1 Single closure techniques

For the single closure techniques, a single closed knot is generated for a given open chain, and the knot type of the open chain is classified as being the knot type of this closed knot. The techniques in this class have the advantage that they are quick to compute. For chains with the endpoints near the “outside”, e.g. most proteins, the single closure techniques (with the exception of direct closure) are consistent with the more involved techniques of the following section. When the endpoints are well within the convex hull of an open chain, the classifications via single closure techniques could very well anger our reasonable person, i.e. the knot type classifications of open chains do not match our intuition. Since there is no agreed upon definition of open knotting, these classifications are not “wrong” per se, but we might be able to do better in some situations. However, for the evaluation of knotting within large samples of open chains, the statistical differences between single closure techniques and more complicated schemes may justify (depending on one’s goals) the speedup gained by employing single closure techniques. For example, in [19], the authors estimated the error in using a single closure technique for 500-edge freely fluctuating polygons by measuring the probability of finding a “knotted domain” within subarcs of random closed unknots. They report an estimated $\approx 3\%$ error, although we note that some of these knotted domains within the unknots may not actually be “errors” but rather detected slipknotting. Using a single “topological neutral closure” and the same error estimation on lattice polygons of length 500, an error of $\approx 0.2\%$ was reported in [27]. Because of the self-avoiding nature of lattice polygons (i.e. the chains have an inherent thickness and, thus, tend to be more spread out), one might expect smaller errors in analyzing lattice polygons. As with the case of freely fluctuating chains, the algorithms may be detecting slipknotting. Also, lattice polygons with 500 edges are considerably more constrained than 500-edge freely fluctuating polygons, so these errors should not be compared directly. It is not clear exactly how to extrapolate the errors to other types of knots or to polygons with more edges, but the reported errors suggest that single closure techniques can be effective if one is willing to tolerate some error.

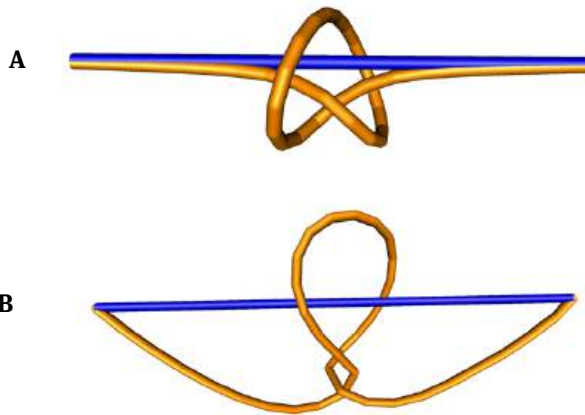


Fig. 9.2: Direct closure oftentimes does not match our intuition for knotting, even when the endpoints lie on the convex hull. The closing segment (in blue) can seemingly remove (A) or add (B) knotting.

9.2.1.1 Direct closure

The easiest way to close an open chain is to simply connect the two endpoints with a straight line segment. The main disadvantage is that the closing line segment can interact with the knotting (and can be much longer than the segment length of the polygons), and thus can introduce knotting or remove knotting that, intuitively, one expects the technique to capture. Arguably the simplest example of real-life knotting is shown in Figure 9.1A where an overhand knot has been tied and the ends have been pulled away from the entangled portion of the curve. Most reasonable people would say that this configuration is knotted. However, by joining the endpoints with a straight line segment, one obtains an unknot (see Figure 9.2A). Similarly, one can construct configurations that most people would classify as unknotted but whose direct closure is arbitrarily complex (see Figure 9.2B). In particular, this technique often does not match our intuition for open chains whose endpoints lie on the “outside” of the chain. Note that no one has used this technique as is and we mention it here only for its simplicity. However, in a future paper we show that direct closure actually coincides with other notions more often than one might expect.

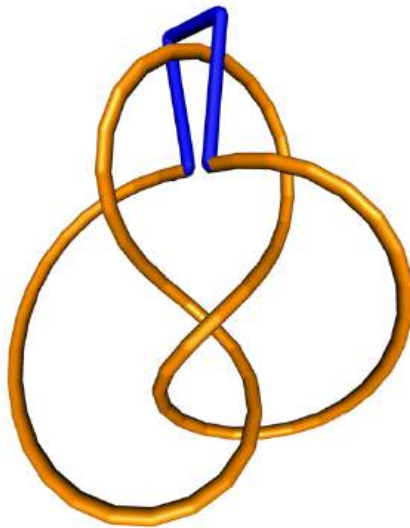


Fig. 9.3: For the center of mass technique, a subchain obtained by removing an arbitrarily short arc from a closed chain can yield a different knot type than the closed chain. In this example, the subchain of the KnotPlot 4_1 knot yields an unknot when the center of mass technique is employed.

9.2.1.2 Center of mass

For this technique [24], the center of mass of the open chain (the mass of which is assumed to be equally distributed at the vertices of the polygon) is computed. Two rays are then created, starting at the two endpoints of the chain. The direction in which the ray extends from each endpoint is given by the vector from the center of mass to that endpoint. The closed knot is defined by connecting the rays at infinity. On a practical note, we really do not need to connect the rays at infinity: once the rays have pierced the convex hull of the chain, we can cut the rays and connect via a simple polygonal arc that lies outside of the convex hull. This technique, at least in part, is an attempt to remedy the problems of direct closure with the examples of Figure 9.2, and indeed this technique is successful in doing so.

One disadvantage of this technique is that when an endpoint of the chain is near the center of mass, a small perturbation of the endpoint will wildly affect the direction of its associate ray, and thus the perceived knot type. In the non-generic situation where the endpoint is at the center of mass, the technique is undefined. Also, there are a number of situations where this technique might not match our intuition or converge to standard knot theory. For example, in Figure 9.3, a closed 4_1 knot has a small portion of the curve removed and is classified as an unknot.

9.2.1.3 Minimal interference

The minimal interference technique [46] might be thought of as an attempt to wed the best of the previous two techniques. A main problematic situation for direct closure is when the two endpoints are far away from each other. On the other hand, it seems somewhat silly to extend rays to infinity (as in the center of mass technique) when the two endpoints are close together. In the minimal interference technique, the minimum distance from each endpoint to the convex hull is computed. If the sum of those two distances is less than the distance between the endpoints, then edges are added from the endpoints to the closest points on the convex hull and the chain is closed by adding a simple arc on the outside of the convex hull between these points on the convex hull. If the distance between the endpoints is smaller than the sum of the two distances to the convex hull, then direct closure is used.

From a practical standpoint, this technique seems quite reasonable. The main problem comes from instability and not being well-defined. First, the distance from a point to the convex hull need not be realized uniquely, and the knots obtained by connecting to the different faces, in these cases, need not all have the same knot type. So defining the knot type via the closure on the convex hull is not well-defined. Note, however, this property is non-generic, and so this is not likely to be a problem computationally. Second, when the distance between the endpoints is equal to the sum of the distances to the convex hull, one has to decide which of the two choices to make, although, again this is a non-generic property. Third, when the sum of the distances to the convex hull is very close to the distance between the endpoints, a small perturbation of the chain can create a transition between the two protocols. For example, in Figure 9.4, we have removed a small portion of a KnotPlot 9_2 knot [36] where the straight line distance and sum of the distances to the convex hull are nearly equal. The direct closure creates a 9_2 knot, but the arc proceeding to the convex hull creates an unknot. Similarly, we could create configurations where a small perturbation creates an arbitrarily large difference in the complexity detected by the two protocols.

9.2.1.4 Simplification schemes

For the analysis of protein knotting, one of the earliest (and most popular) techniques involves simplifying the chain and is referred to as the KMT algorithm [21, 44] (although the ideas trace back at least to [33]). Protein chains are complicated. Trying to find a knot visually in the alpha carbon backbone will quickly leave the viewer with a headache. The goal of the simplification schemes is to eliminate unnecessary detail, reducing the number of vertices and edges to create an equivalent, yet much shorter, chain which is much easier on the human visual cortex. This simplification is accomplished by searching for triangles formed by consecutive edges that are not pierced by any edge of the chain. When such a triangle is not pierced, one can remove the mid-

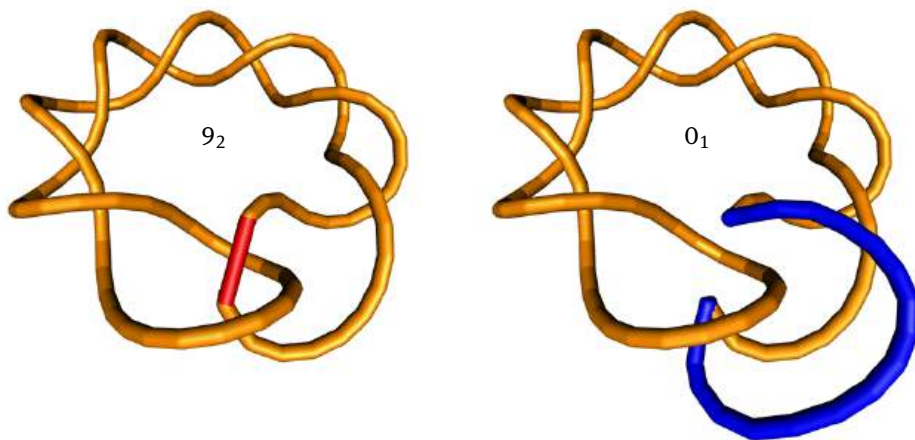


Fig. 9.4: For the minimal interference technique, a small perturbation of the vertices can lead to large differences in the knot detected. In this example of a KnotPlot 9_2 knot [36], a 9_2 knot is detected when closed directly (left, red arc), but an unknot, 0_1 , is detected when the arc is closed via proceeding to the convex hull (right, blue arc).

dle vertex of the consecutive edges and replace the pair of consecutive edges with one edge between the two remaining endpoints (see Figure 9.5). For polygons whose endpoints are on the “outside” (as is typically the case in proteins), this scheme seems to work quite well. One (usually) obtains a simple chain and connecting the endpoints via a large external loop provides a closed knot for classification. However, when the endpoints are not on the “outside”, such as in the search for slipknots or any sort of analysis of knotting in subchains, simplification is not effective. Furthermore, the order in which the delta moves are performed can affect the knot type classification of the chain, even when the endpoints are on the outside [28]. As such, classification by simplification is not well-defined mathematically and, thus, is not a good choice for classifying knotting in all open chains (although, to be fair, it was not designed to perform that task).

9.2.2 Stochastic techniques

For the stochastic techniques, the overall philosophy is that the open chain is an incomplete form, having the potential to be mathematically knotted but not, at this stage, having achieved its potential. The open chain is seen as a subchain of a class of closed knot configurations, where the class of configurations depends on the technique. As such, the “knot type” of the chain is a probability distribution of knot types

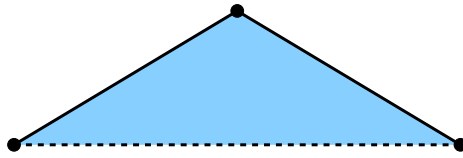


Fig. 9.5: A delta move can simplify a configuration to make it easier to see the underlying knotting in a chain. If the shaded triangle is not pierced by any edge from the polygon, then the two edges can be replaced with the dashed edge. Note, however, that Millett et al. [28] showed that the order of the delta moves can affect the knot type classification of open chains.

from the given class, and this probability distribution is approximated by sampling from the distribution. The stochastic techniques match our four criteria quite well. The main disadvantages are that 1) we must approximate the probability distribution, and 2) the algorithms can be orders of magnitude slower than the single closure techniques. Still, the authors hope to convince the reader that the stochastic techniques are the right way to think about open knotting.

9.2.2.1 Uniform closure

The idea here is to extend parallel rays from the endpoints of an open chain to infinity (like in the center of mass technique) and connect the rays at infinity. In particular, for a given open chain and a given direction (seen as a point on the unit sphere S^2), parallel rays are extended from both of the endpoints in the given direction and the configuration is closed at infinity. The “knot type” of the open chain is then the probability distribution of knot types obtained by completing this closure over all points on S^2 . Millett, Dobay, and Stasiak [28] defined the original version of this technique with a slight alteration: in their computations, they chose points on a large sphere containing the open chain. The differences between connecting at infinity and via a large sphere are likely negligible. Still, connecting to a large sphere requires the choice of a center for the sphere and a radius and, therefore, requires some human choices that we would like to avoid [30].

Note that one could choose to send the rays to infinity in independent directions ($S^2 \times S^2$ worth of choices) instead of one common direction (S^2 worth of choices). In fact, $S^2 \times S^2$ was Mansfield’s choice in his first knotted protein paper [25]. In [17, 18, 32, 38], we use one common direction. This decision is due mostly to computation speed. Typically, we have used 100 closure directions to obtain our approximation of the probability distribution. Sampling from a probability distribution, instead of using a single closure, already puts the stochastic techniques at a significant disadvantage speed-wise. To get a similar density of sampling on $S^2 \times S^2$, we would need 100^2 sample directions. There might be some advantages of sampling from $S^2 \times S^2$ in terms

of matching our intuition for knotting, although, to date, no one has systematically analyzed the differences.

Also, there are some historical reasons for using common directions. In particular, the average crossing number is computed by averaging the number of crossings seen as a configuration is projected over all directions in S^2 (writhe is another example). Therefore, using common directions is akin to computing the average knotting of the given chain.

Because this is the technique we have used in [17, 18, 32, 38], we discuss briefly its implementation. We use a fixed set of 100 directions (i.e. 100 points on S^2) that are roughly uniformly spread over the unit sphere. The early papers (e.g. [24, 25, 28]) used random directions. Unfortunately, random is seldom uniform and the goal in sampling the directions is to get a good approximation of the probability distribution of knot types. So we have chosen uniform fixed directions (which also guarantees consistent results). We have computed the spherical Voronoi cells associated with these fixed directions and we weight the knot types obtained from the closure in a given direction by the relative surface area of the Voronoi cell for the corresponding point on S^2 . The choice of 100 directions is somewhat arbitrary, but we did do some analysis of the effect. More specifically, in the summer of 2012 undergraduate University of St. Thomas students Nicole Lopez and Elizabeth Annoni explored the effect of using different numbers of closures and different algorithms for generating roughly uniform sets of points on the sphere in measuring the knotting of open chains. Without getting into the details, we found that the roughly uniform set of points using Martin's Polyhedra [1] has good uniformity and using more than 100 closures has a negligible effect on the approximation of the probability distribution. The students also computed the relative surface area of the Voronoi cells that we use. In addition, as mentioned above, we do not really close the knot at infinity. We compute the maximum distance between any two vertices on the knot, and then extend the rays twice that distance in the given directions and connect with an arc of line segments that avoids piercing through any of the entanglement.

Once one obtains an approximation of the probability distribution, there is still the question of what “knot type” to call a given chain. Certainly, it is not efficient to report a probability distribution as the knot type classification for an open chain. For the sake of definiteness, it is desirable in many situations to choose a classification of the knot type of an open chain. In such cases, we typically classify the knot type of a given open chain as the knot type that appears most frequently [17, 18, 32, 38]. The advantage of assigning the knot type with the highest probability to the open chain is that every open chain gets a classification as a knot type. The main disadvantage is that sometimes there are two (or more knot types) with similar probability values. In such cases, our approximation of the probability distribution might not be precise enough to truly classify the predominant knot type within the closures. One could sample more directions for these open chains. This adaptive approach has some merit, although it is not something that we have implemented. However, most open chains

have one knot type that appears more than 50% of the time, and typically even at much higher percentages. For example, in a sample of 1000 random equilateral open chains with 300 edges, Millett [29] reported that 99.6% of the chains had one knot type which appeared with over 50% probability (note that Millett used the original strategy of closing to points on a large sphere with random directions for closure, although this approach is not likely to alter the statistics considerably). Using the 50% cutoff also seems like a reasonable approach to classifying the knot type of an open chain, although it leaves the classification of some open chains as being ambiguous (i.e. no classification). While we have used the predominant knot type for classification, there certainly are contexts where a 50% cutoff might be preferable.

9.2.2.2 Random arc closure

Recall that the strategy with the stochastic techniques is to approximate a probability distribution one obtains by closing the configuration in some fashion. Perhaps the most natural way to generate such an ensemble is by closing the configuration by connecting the endpoints with random arcs. Cantarella et al. [9] have created an algorithm (and software) for efficiently generating random equilateral polygons and random equilateral arcs with a given distance between the endpoints. The main problem with this approach is that one must choose the length of the random chain (i.e. the number of edges) with which to close a given open chain.

Note that in an upcoming sequence of yet untitled papers, the authors of this paper, in collaboration with Cantarella and Shonkwiler, are analyzing how well some of the different strategies presented here predict the knot type of a closed knot given some subchain portion of the knot. In such a case, we know precisely how many edges have been removed to extract the open chain from the closed knot, so we can replace the missing portion of the knot with random arcs with the correct number of edges. For a random open chain (or a protein, for example), it is not so clear how one might determine an appropriate number of edges with which to connect the endpoints with random arcs. However, this issue might not be insurmountable. For example, Zirbel and Millett [47] determined the average distance between the endpoints of differing length subchains for closed polygons. One might use that information to devise a scheme for choosing the number of edges in the replacement arcs.

9.2.3 Other closure techniques

For the sake of completeness, we mention other techniques (mainly variations of the ideas presented above) that have been employed in some papers. In [19], the authors attach an “almost complete circle” to the endpoints of an open chain in a random way to obtain a closed knot. In [27], the authors connect the endpoints at infinity us-

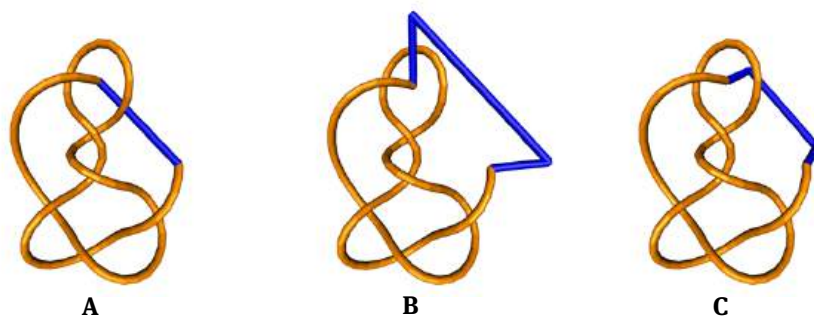


Fig. 9.6: The closed chains coming from single closure techniques associated with a subchain of the 98-edge KnotPlot $+7_5$ knot [36]. From left to right, we have (A) direct closure, (B) center of mass, and (C) minimal interference.

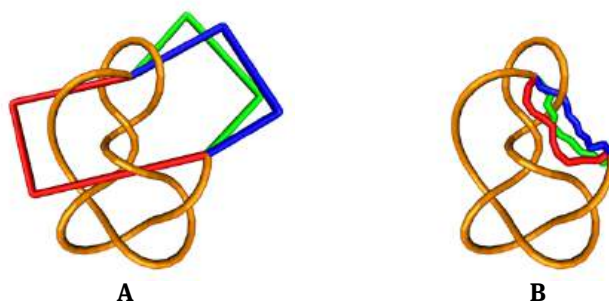


Fig. 9.7: The closed chains coming from stochastic closure techniques associated with a subchain of the 98-edge KnotPlot $+7_5$ knot [36]. For (A) uniform and (B) random arc, we show three examples of closures from the technique.

ing a pair of directions which is chosen “to minimize the risk of knot modifications or disentanglements”. In [20], the authors use a combination of random closures (seemingly from $S^2 \times S^2$), the center of mass technique, and a deterministic technique. In the deterministic approach, they extend interior endpoints to the exterior by adding length to the chain as if it is “repelled from the protein surface, until they reach an outer sphere”. We guide the reader to these articles for more details.

9.2.4 Topology of knotted arcs

Because the standard deformations used to define closed knot equivalence can turn any embedded open arc into a straight line segment, different techniques are required to study knotted arcs topologically. Classically, this topological description has been in the form of the so-called ball-arc pair [8]. The ball-arc pair consists of an open (non self-intersecting) chain c and an oriented topological ball B (which need not be “round”) with the property that the endpoints of the chain lie on the surface of the ball and the interior of the chain lies in the interior of the topological ball. Then two ball-arc pairs (B_1, c_1) and (B_2, c_2) are equivalent if there exists an orientation-preserving homeomorphism between B_1 and B_2 which takes c_1 to c_2 . In [39], the authors use the ball-arc pair to define when an open lattice chain forms a “tight knot”, a concept which was critical in the proof that long lattice knots are almost always knotted.

For a given open chain, the assignment of a topological ball (to create a ball-arc pair) is equivalent to the assignment of a corresponding set of closures for the chain, all of which have the same knot type. To see this, suppose we are given a ball-arc pair (B, c) . If we join the two endpoints of c via simple arcs lying on the boundary of B , we will obtain closed knots with a common knot type. These simple arcs are the corresponding set of closures.

Conversely, the assignment of a closing arc to a given open chain is equivalent to the assignment of a corresponding set of topological balls, namely the set of balls whose 1) interiors contain the interior of the chain, 2) boundaries contain the endpoints of the chain, and 3) boundaries contain the closing arc. The ball-arc pairs from this set of balls with the given chain are all equivalent as ball-arc pairs.

For any open arc and any knot type, one can construct a topological ball so that the closed knot obtained via a simple arc along the boundary of the ball results in the given knot type. Similarly, for any given open arc and any knot type, one can construct a closed knot of the given knot type which contains the arc as a subchain. All of the techniques presented in this section essentially restrict the set of closures, and thus, the set of ball-arc pairs from which one is sampling in a particular fashion. It may be possible to define the knotting in open chains by restricting to a given set of balls for the ball-arc pair instead of working at the level of closures, although it is not clear how one might do so. Still, while much of the modern literature on open knotting does

not specifically mention the ball-arc pair, it is important to acknowledge the historical topological contributions and note the connections to current work.

9.3 Visualizing knotting in open chains using the knotting fingerprint

In this section, we introduce the knotting fingerprint as a means of visualizing the knotting in open subchains of an open or closed chain. The knotting fingerprint can be used with any of the above techniques. The basic idea is to generate an image matrix, where each subchain is associated with a unique cell in the matrix, and use colors in the cells to communicate the knotting within the subchain. The first such fingerprints were seen in [20], in their work to find slipknots within proteins. For open chains, the fingerprint takes the form of a triangle, where the position along the x -axis indicates the starting vertex index of the given subchain and the position along the y -axis indicates the ending vertex index. Since the ending vertex index must be greater than the starting index, the portion of the matrix above the main diagonal (i.e. the diagonal from upper left to lower right) is empty and the diagonal indicates the behavior of zero-length subchains. The entire chain is associated with the sector in the lower left corner. The triangular fingerprint also has been used in [18, 38] and seems to be accepted as the standard. An example of the image matrix for the DehI protein (PDB code 3bjx), the protein forming the most complicated knot yet seen ($+6_1$), is shown in Figure 9.8 using the uniform closure technique. The knot type of a subchain is the knot type appearing with the highest percentage in the closures of the subchain and the opacity of the coloring within the cells is that percentage. The ordering of the key (from top to bottom) is by the length of subchain at which the knot type first appears. We discuss the coloring of the cells more below where we compare the images coming from the different approaches.

To illustrate different ways of visualizing the knotting of subchains, in Figures 9.9, 9.11, and 9.12, we analyze a relatively nice configuration, the KnotPlot $+7_5$ [36], an image of which appears in the right picture of Figure 9.9. This knot was chosen because the resulting pictures have some, but not too much, complexity.

For closed chains, using the square image matrix is a little awkward (see the left image of Figure 9.9). The problem is that there is a subchain from every vertex to every other vertex, so this is like $S^1 \times S^1$ (which, topologically, is a torus) or $S^1 \times [0, l]$ (where l is the length, which gives us a cylinder). In [46], the authors adapt the open chain approach for closed chains: the x -axis indicates the starting vertex index and the y -axis indicates the ending vertex index. As such, the blank part of the matrix for open chains gets filled in with subchains that include the vertex with index zero. This approach is perfectly reasonable, although the matrices become hard to read due to periodic nature of $S^1 \times S^1$. In particular, just below the diagonal, one has length

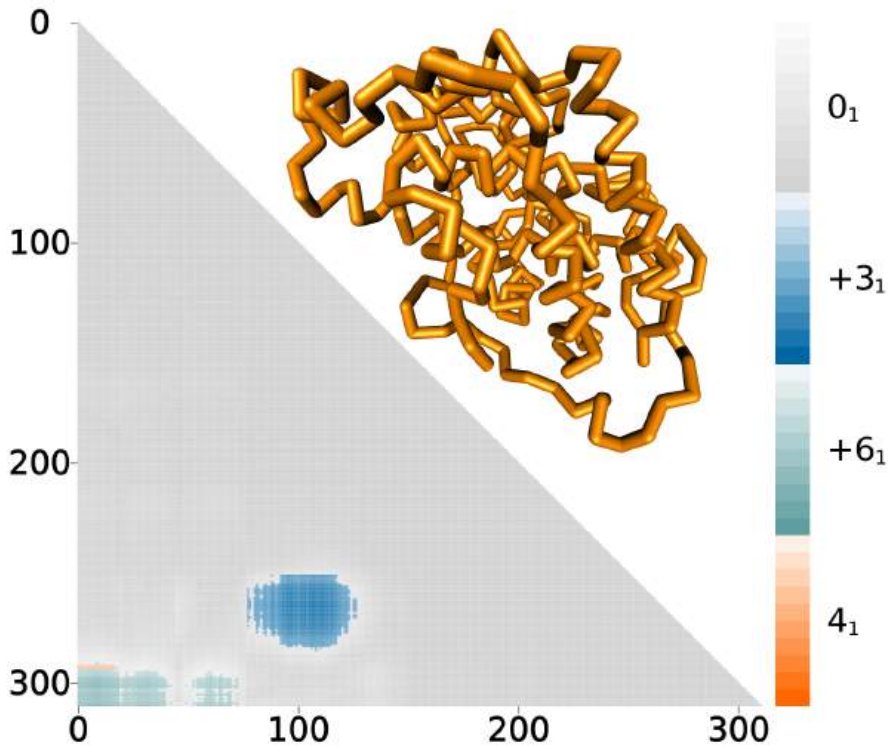


Fig. 9.8: The image matrix for the Dehl protein (PDB code 3bjx) using the uniform closure technique.

one subchains, but right above the diagonal, one has the whole chain minus an edge. Furthermore, there is a vertical wrapping of the image from bottom to top. One could “fix” the diagonal issue by having the x -axis indicate the starting vertex index and the y -axis indicate the length of the subchain (see the right image of Figure 9.9). There would still be a wrapping from left to right (or one could show the image on a cylinder), but this approach would be easier on the eye. This approach has not been employed, seemingly due to the desire to be consistent with the original triangular matrix used for proteins in [20]. The mathematically correct way to show the image would be on a torus or cylinder, although 3D images are not easy to see or manipulate and it is not clear that there would be any unique insight from putting the image on a torus or cylinder.

In Figure 9.10, we show the disk matrix [17, 32] for an 81-edge KnotPlot $+5_2$ knot. For the sake of the disk matrix, the polygon is an oriented based list of n coordinates in 3D space, i.e. we specify a first vertex (the base vertex) and a second vertex (which uniquely determines the orientation) to the list. The disk matrix is most easily described in terms of polar coordinates: the polar θ value indicates the midpoint of the

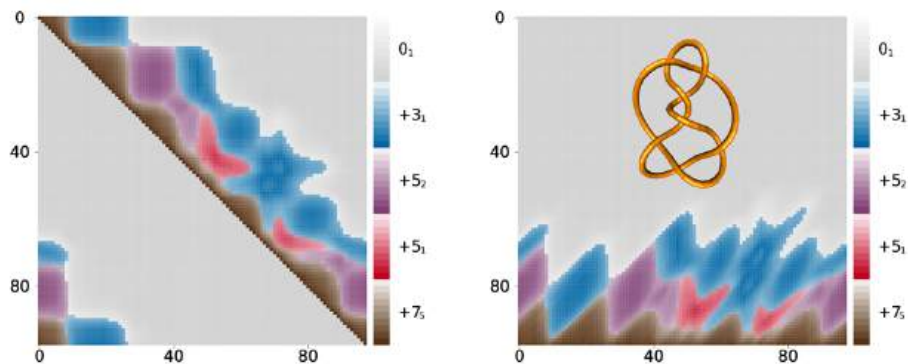


Fig. 9.9: Two square image matrices for the 98-edge KnotPlot $+7_5$ knot [36] (configuration shown on the right image) that communicate the knotting within its subchains using the uniform closures technique. On the left, the x -axis indicates the starting vertex index and the y -axis indicates the ending vertex index. On the right, the x -axis indicates the starting vertex index and the y -axis indicates the number of edges in the subchain.

subchain and the r value indicates the length. The disk matrix has a sort of brick pattern since open chains with odd numbers of vertices have a vertex midpoint, but open chains with even numbers of vertices have a midpoint which is between two vertices. Adjacent to the center of the disk are all length one subchains, which are too short to be knotted, and so are all colored as the unknot, 0_1 , in light gray. As we move out from the center, we are seeing the knot type of longer and longer subchains, until the last stripe which contains subchains with n vertices and $n - 1$ edges. Let us fix a particular vertex v of the configuration. One can add an edge on both sides of v to obtain a subchain with two edges, do this again to obtain a subchain with four edges, etc. All of these subchains have a common midpoint, namely v . In Figure 9.10, the polar sectors associated with these subchains have been colored black. The orange and green sectors correspond to the subchains that have v as a starting and ending vertex, respectively. Since we can add one edge on either end as opposed to when we add edges at both ends, the orange and green sectors appear at every radius whereas the black sectors appear at every other discretized radius. In the next section, we will see the orange and green spiral patterns appear frequently. We compute the knot types only for the subchains ending at vertex points. One could subdivide the edges to get a finer image disk matrix and this would provide more detail. Note that changing the orientation of the knot would have the effect of flipping the disk matrix across the x -axis and changing the identification of the base vertex would have the effect of rotating the matrix.

For the single closure techniques, one has a single “knot type” for any given subchain. For the stochastic techniques, the “knot type” of a subchain is a probability distribution. In [17, 32], and even our work with open chains [18, 38], we assign the

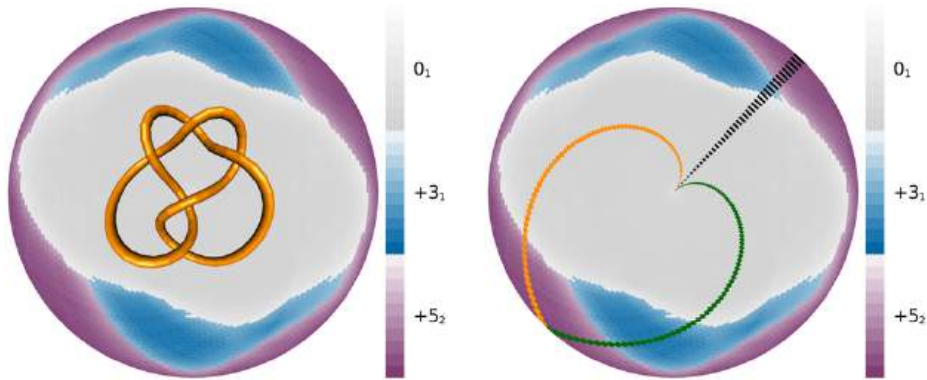


Fig. 9.10: The disk matrix for the 81-edge KnotPlot $+5_2$ knot [36] using the uniform closures technique. On the left, the disk matrix is shown with the configuration superimposed. On the right, the black sectors correspond to subchains with a fixed vertex midpoint. The orange and green sectors correspond to subchains with the same vertex as the starting and ending vertices, respectively. The polar θ value indicates the midpoint of the corresponding subchain and the polar r indicates the length of the corresponding subchain. Each sector is colored by the knot type which appears most frequently when closed via the uniform closure technique, with the opacity of the coloring set to the percentage of closures with that knot type. The key to the right of each disk matrix shows the color associated with each knot type that appears in the matrix along with a swatch of that color at opacities 10%, 20%, ..., 100%.

knot type of an open chain to be the knot type appearing most often in the distribution of closure knot types. Clearly, we are losing information here, although, as mentioned above, for the great majority of open chains there is one knot type that appears in over 50% of the closures. To give a sense for the dominance of the knot type, we set the opacity of the color within each cell to be the percentage of closures that are the given knot type (swatches of the colors at opacities 10%, 20%, ..., 100% appear in the key to the right of each disk matrix). Thus, light coloring indicates a low percentage and dark coloring indicates a high percentage for the dominant knot type seen amongst the closures. Like with the open chain triangular/square matrices, the key on the right is ordered by the length of the subchain at which the knot type first appears.

KnotPlot $+7_5$ disk matrices for the single and stochastic closure techniques are shown in Figures 9.11 and 9.12, respectively. Because this configuration is relatively spread out and smooth, we would expect a higher level of agreement between the different techniques than for less well-behaved configurations. By eye, the images look similar in some ways and different in others. The direct closure technique is the outlier here, although it might be more consistent with the other techniques than one would expect, especially when compared to the disk matrices for the stochastic techniques. The disk matrices for the center of mass and minimal interference techniques share features and the uniform and random arc closure disks share features, although many of the features differ between the two pairs. The single closure techniques tend

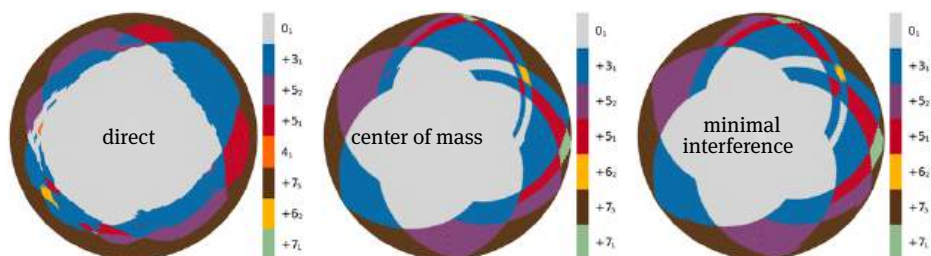


Fig. 9.11: Disk matrices for the 98-edge KnotPlot $+7_5$ knot [36] using single closure techniques: direct closure (left), center of mass (middle), and minimal interference (right).

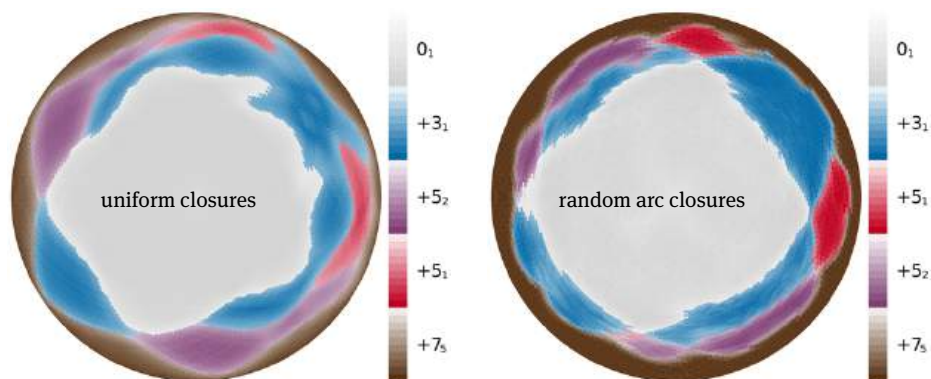


Fig. 9.12: Disk matrices for the 98-edge KnotPlot $+7_5$ knot [36] using stochastic closure techniques: uniform closures (left) and random arc closures (right).

to introduce more subchain knot types than the stochastic techniques, as is seen in this example where $+6_2$ and $+7_1$ appear for the single closure techniques but not the stochastic techniques. Furthermore, the single closure techniques tend to have sharper features than the stochastic techniques.

9.4 Features of knotting fingerprints, knotted cores, and crossing changes

In this section, we show some examples of disk matrices, using the uniform closure technique, for several random knot configurations. We make some observations about the relationship between the configurations and their disk matrices. This section is by no means comprehensive, but rather we hope that it whets the reader's appetite for understanding how open knotting manifests itself in closed chains.

We begin with a highly localized knot in Figure 9.13. We know that the -3_1 knot is highly localized because the knotting appears near the center of the disk and persists at longer lengths (i.e. near the edge of the disk). In general, when the subchains have sufficient length to change the knot type, we see cusp-like widow's peaks appearing with a new color. Note that there are some cells on the edge of the disk matrix that are gray, indicating that the corresponding 149-edge subchains are unknotted. In this case, this behavior occurs because the polygon passes very close to itself at the borders of the teal region of the configuration (the minimal knotted region). The removal of one edge here is enough to see the subchain as unknotted. If one were to subdivide the polygon's edges, the gray regions would not occur on the border, so this behavior is an artifact of the discretization of the disk matrix.



Fig. 9.13: The disk matrix for a highly localized 150-edge -3_1 knot, with an image of the configuration superimposed over the disk matrix. The teal region of the configuration is the shortest subchain that is knotted.

In Figure 9.14, we see a disk matrix that has two clear peaks. This behavior is not unusual and we often seen disk matrices with multiple widow's peaks. In this configuration, there are two simple loops (on the left in a gray colored arc and on the right in a green colored arc) and a core red arc region. To form the -3_1 , the subchain must pass fully through the red middle region and one of the green or gray regions. The two peaks correspond to the red+green and red+gray subchains. So, in general, a peak

occurs when the arc has completed some, but not all, of the knotting, and an unknotted arc wanders away before coming back to complete the knotting. In fact, we see the same behavior in Figure 9.13 on the very left portion of the disk matrix. However, in Figure 9.13, the knotting is so tight in this area of the polygon that this wandering region is very short.

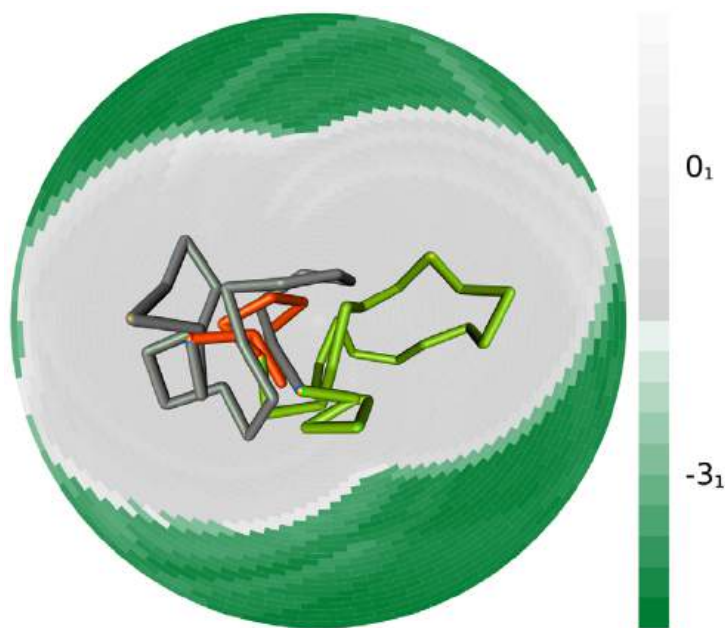


Fig. 9.14: The disk matrix for a 50-edge -3_1 knot with double peaks.

Typically, we see a lot of different knot types appear in the disk matrix. Some of these knot types are intermediate entanglement in subchains needed to create the more complicated knotting of the entire knot (which will be discussed in more detail below). However, oftentimes the knot types are not “contributing” to the knotting of the whole curve. The simplest such example is the case of a slipknot, i.e. an unknotted closed curve that contains knotted subchains. A knotted subchain within an unknotted closed curve was termed an *ephemeral knot* by Millett [29]. For the sake of our discussion, we consider ephemeral knots to be knotted subchains whose knotting is not a key feature to the knotting of the entire closed knot.

In Figure 9.15, we see an example of a configuration with ephemeral knots [29]. It is difficult to see anything within the configuration, but we include a picture for the sake of completeness. There are (at least) three green streaks that are ephemeral -3_1 knots. Note that these green streaks can be broken for a number of cells, forming an unknot or some other type of knot in the mean time. When the streak is broken, it is

unclear whether to mentally join them or to call them new ephemeral knots (although we leave that discussion for the next paragraph). In this example, the -3_1 slipknots have only one cell depth (although in Figure 9.16, we see more robustness along these lines), so the removal of one vertex on one of the ends is enough to unknot the arc. On the other hand, adding vertices to the other end keeps the knotting for a number of vertices.

In addition, there is one very short (lasting only for a single edge) $+3_1$ slipknot with a black arrow pointing to it in the image. The subchain corresponding to this cell is a $+3_1$ knot, but it seems safe to assume that the knotting of this subchain is not a key feature to the knotting of the base $+3_1$ knot (although that could be debated). In some cases, like Figure 9.13, it is pretty clear where the core of the knotting is. But in many/most cases, finding the core is not so simple. Tubiana et al. [46] defined the bottom-up and top-down procedures to define the knotted core (which they called the shortest knotted arc). In the bottom-up procedure, one searches for the shortest subchain with the same knot type as the base knot. In the top-down procedure, one searches for the shortest subchain that has the knot type of the base knot and is a subchain of increasing length subchains that also have the same knot type as the base knot. In the disk matrix, this is akin to searching from the center outwards (for bottom-up) and searching for the widow's peak closest to the center (for top-down). The example of Figure 9.15 shows that the bottom-up procedure might not provide the information one expects. On the other hand, thinking of the subchain corresponding to the peak closest to the center (the one above and slightly to the right of center), the $+3_1$ becomes unknotted for some longer chains containing that subchain. One might be tempted to forgive the $+3_1$ for this slight transgression and declare that subchain as the core. Or we might forgive the $+3_1$ for having longer subchains that form the 4_1 knot since, in some sense, the 4_1 is more complicated than the $+3_1$. Ultimately, we do not have answers to these questions, but rather wish to illuminate the reader about some of the issues in defining the knotted core.

Also note that the configuration from Figure 9.15 has a number of different knot types that appear in only a few cells, which is typical for random knot configurations using any of the notions of open knotting.

Ephemeral knotting also can form after the creation of a given knot. In Figure 9.16, we have a -3_1 knot which adds an ephemeral -3_1 knot (golden region) to form a composite knot. The extra -3_1 persists nearly to the entire length of the configuration. Here the ephemeral knotting is more robust than in Figure 9.15 in that there are subchains where both endpoints can be removed and the $-3_1 \# -3_1$ knot remains.

One might label the core as the subchain associated with the green cell closest to the center. However, note that there is a streak of gray cells signifying unknotted chains. As mentioned above, one might forgive this transgression as it only has depth one. Still, if we forgive the depth one transgression, then why not depth two, or three, or more? Certainly the composition of -3_1 with the additional -3_1 would seem to be a forgivable offense, as well as the smattering of -8_{19} subchains.

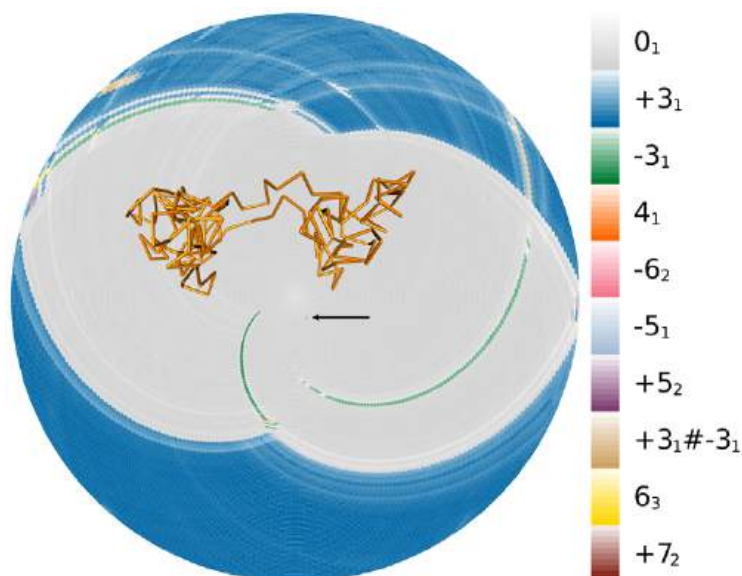


Fig. 9.15: The disk matrix for a 150-edge $+3_1$ knot with ephemeral knotting.

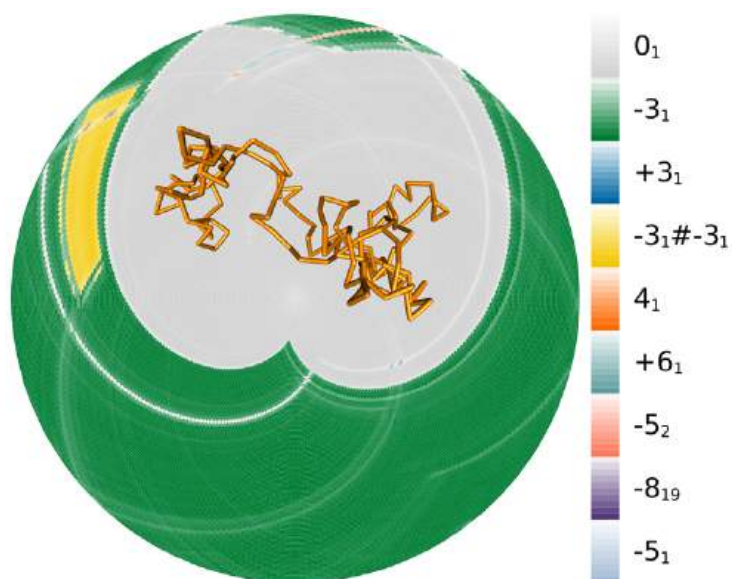


Fig. 9.16: The disk matrix for a 150-edge -3_1 knot with ephemeral knotting in the form of knot composition.

One way of “adding” ephemeral knotting is in the form of knot composition, as is seen in Figure 9.16. However, one can also add knotting by introducing additional

entanglement in the form of a more complicated knot. In Figure 9.17, we have a $+3_1$ which becomes additionally tangled to form a $+5_2$ knot. This additional entanglement eventually disappears, leaving only the $+3_1$. In [17, 32], we introduced the idea of subknotting (here $+3_1$ would be considered a *subknot* of $+5_1$). We might also consider $+5_1$ to be a *superknot* of $+3_1$. In fact, one might consider creating a family tree of knotting using the idea of subknots and superknots. Not surprisingly, the creation of such a family tree is not clear cut, although we guide the interested reader to [17, 32] for some discussion along these lines.

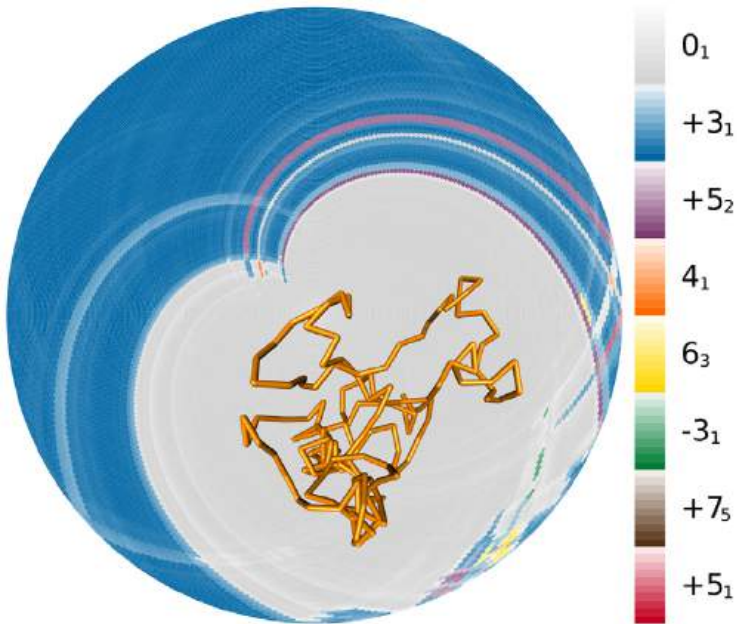


Fig. 9.17: The disk matrix for a 150-edge $+3_1$ knot with ephemeral knotting in the form of more complicated knotting.

The idea of a family tree (or, similarly, a graph connecting certain knots) has been explored, most notably with regards to crossing changes (see e.g. [10, 11, 13]). In [32], we showed that subknotting is similar to, but not the same as, crossing changes performed on a projection of a configuration. The knot type $+5_2$ has unknotting number one, meaning that it is possible to perform one crossing change and transition from an unknot to a $+5_2$ knot, or vice versa. In Figure 9.18, we see a $+5_2$ knot that is created by transitioning straight from an unknot. The $+5_2$ knot also can be created by making a $+3_1$ knot and then adding complexity, as can be seen in Figure 9.19.

Figure 9.20 has a number of interesting features. First, we see the creation of the $+5_2$ via $+3_1$ transitioning to $+5_1$, which then transitions to $+7_3$ (getting more compli-



Fig. 9.18: The disk matrix for a 150-edge $+5_2$ knot whose subchains transition straight from an unknot to a $+5_2$ knot.



Fig. 9.19: The disk matrix for a 150-edge $+5_2$ knot whose subchains form $+3_1$ knots before transitioning to the $+5_2$ knot.

cated) and then down to the $+5_2$. Second, we see an ephemeral 4_1 . This 4_1 is likely not as ephemeral as it might appear since, despite the unknotted padding between the regions, the 4_1 likely creates twisting that is later needed in the transition to the $+5_2$.

Finally, we see transitions from the unknot straight to $+5_2$, and also passing through intermediate $+3_1$ and 4_1 states.



Fig. 9.20: The disk matrix for a 50-edge $+5_2$ knot with a complex transition from unknot to $+5_2$.

9.5 Conclusions

We have presented the definitions of the commonly used notions of open knotting, including some discussion of desirable and undesirable properties of each. We introduced different matrix representations of the knotting within subchains of open and closed chains. Finally, we showed some examples of disk matrices for random knots and highlighted the relationships between some features of the disk matrices and the configurations.

We believe that the uniform closure method provides the most desirable properties for analyzing knotting in open chains. The main disadvantage is computation speed, and, indeed, with respect to speed, uniform closure is at a significant disadvantage to the single closure techniques. Furthermore, we believe that the disk matrix provides an intuitive means for the viewer to glean information about the knotting within a closed chain.

While researchers might not agree on which notion of open knotting to use, we hope the discussion here will inform future decisions.

Acknowledgment: This material is based upon work supported by the National Science Foundation Grant No. 1418869. The authors thank the referee, Kenneth Millett, and Andrzej Stasiak for their constructive comments and suggestions.

Bibliography

- [1] Bob Allanson. Martin's polyhedra. <http://members.ozemail.com.au/~llan/mpol.html>, 2013.
- [2] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helga Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28:235–242, 2000. <http://www.rcsb.org>.
- [3] Daniel Bölinger, Joanna I. Sułkowska, Hsiao-Ping Hsu, Leonid A. Mirny, Mehran Kardar, Jose N. Onuchic, and Peter Virnau. A stevedore's protein knot. *PLoS Comput. Biol.*, 6(4):e1000731, 2010.
- [4] Gregory Buck and Jeremy Orloff. A simple energy function for knots. *Topology Appl.*, 61(3):205–214, 1995.
- [5] Roman V. Buniy, Jason Cantarella, Thomas W. Kephart, and Eric J. Rawdon. The spectrum of tightly knotted flux tubes in qcd. *Journal of Physics: Conference Series*, 544(1):012025, 2014.
- [6] Roman V. Buniy and Thomas W. Kephart. A model of glueballs. *Phys. Lett. B*, B576:127–134, 2003.
- [7] Roman V. Buniy and Thomas W. Kephart. Glueballs and the universal energy spectrum of tight knots and links. *Int. J. Mod. Phys. A*, 20:1252–1259, 2005.
- [8] Gerhard Burde, Heiner Zieschang, and Michael Heusener. *Knots*, volume 5 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, extended edition, 2014.
- [9] Jason Cantarella, Bertrand Duplantier, Clayton Shonkwiler, and Erica Uehara. A fast direct sampling algorithm for equilateral closed polygons. arXiv:1510.02466, 2015.
- [10] Isabel K. Darcy and De Witt Sumners. Rational tangle distances on knots and links. *Math. Proc. Cambridge Philos. Soc.*, 128(3):497–510, 2000.
- [11] Isabel Dazey Darcy and De Witt Sumners. A strand passage metric for topoisomerase action. In *KNOTS '96 (Tokyo)*, pages 267–278. World Sci. Publishing, Singapore, 1997.
- [12] Patrícia F. N. Faísca. Knotted proteins: A tangled tale of structural biology. *Comput. Struct. Biotechnol. J.*, 13:459–468, 2015.
- [13] Alessandro Flammini, Amos Maritan, and Andrzej Stasiak. Simulations of action

- of DNA topoisomerases to investigate boundaries and shapes of spaces of knots. *Biophys. J.*, 87(5):2968–2975, 2004.
- [14] Michael H. Freedman, Zheng-Xu He, and Zhenghan Wang. Möbius energy of knots and unknots. *Ann. of Math. (2)*, 139(1):1–50, 1994.
 - [15] Carl Friedrich Gauss. Zur mathematischen Theorie der electrodynamischen Wirkungen. *Werke: Fünfter Band*, pages 605–630, 1867.
 - [16] Oscar Gonzalez and John H. Maddocks. Global curvature, thickness, and the ideal shapes of knots. *Proc. Natl. Acad. Sci. USA*, 96(9):4769–4773, 1999.
 - [17] David A. B. Hyde, Joshua Henrich, Eric J. Rawdon, and Kenneth C. Millett. Knotting fingerprints resolve knot complexity and knotting pathways in ideal knots. *J. Phys.: Condens. Matter*, 27:354112, 2015.
 - [18] Michal Jamroz, Wanda Niemyska, Eric J. Rawdon, Andrzej Stasiak, Kenneth C. Millett, Piotr Sulkowski, and Joanna I. Sulkowska. Knotprot: a database of proteins with knots and slipknots. *Nucleic Acids Research*, 43(D1):D306–D314, 2015.
 - [19] Vsevolod Katritch, Wilma K. Olson, Alexander Vologodskii, Jacques Dubochet, and Andrzej Stasiak. Tightness of random knotting. *Physical Review E*, 61:5545–5549, 2000.
 - [20] Neil P. King, Eric O. Yeates, and Todd O. Yeates. Identification of rare slipknots in proteins and their implications for stability and folding. *J. Mol. Biol.*, 373:153–166, 2007.
 - [21] Kleanthes Koniaris and Murugappan Muthukumar. Self-entanglement in ring polymers. *J. Chem. Phys.*, 95(4):2873–2881, 1991.
 - [22] Robert B. Kusner and John M. Sullivan. Möbius energies for knots and links, surfaces and submanifolds. In *Geometric topology (Athens, GA, 1993)*, pages 570–604. Amer. Math. Soc., Providence, RI, 1997.
 - [23] Richard A. Litherland, Jonathan Simon, Oguz Durumeric, and Eric Rawdon. Thickness of knots. *Topology Appl.*, 91(3):233–244, 1999.
 - [24] Rhonald C. Lua and Alexander Y. Grosberg. Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput. Biol.*, 2(5):351–357, 2006.
 - [25] Marc L. Mansfield. Are there knots in proteins? *Nat. Struct. Biol.*, 1:213–214, 1994.
 - [26] Marc L. Mansfield. Fit to be tied. *Nat. Struct. Biol.*, 4:166–167, 1997.
 - [27] Boris Marcone, Enzo Orlandini, Attilio L. Stella, and Francesco Zonta. Size of knots in ring polymers. *Phys. Rev. E*, 75(4):041105, 2007.
 - [28] Kenneth Millett, Akos Dobay, and Andrzej Stasiak. Linear random knots and their scaling behavior. *Macromolecules*, 38(2):601–606, 2005.
 - [29] Kenneth C. Millett. The length scale of 3-space knots, ephemeral knots, and slipknots in random walks. *Prog. Theor. Phys. Suppl.*, 191:182–191, 2011.
 - [30] Kenneth C. Millett and Benjamin M. Sheldon. Tying down open knots: A statistical method of identifying open knots with applications to proteins. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 203–217. World Sci. Publishing, Singapore, 2005.
 - [31] Jun O’Hara. Energy of a knot. *Topology*, 30(2):241–247, 1991.

- [32] Eric J. Rawdon, Kenneth C. Millett, and Andrzej Stasiak. Subknots in ideal knots, random knots, and knotted proteins. *Sci. Rep.*, 5:8928, 2015.
- [33] Kurt Reidemeister. Elementare begründung der knotentheorie. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 5(1):24–32, 1927.
- [34] Jane S. Richardson. β -sheet topology and the relatedness of proteins. *Nature*, 268:495–500, 1977.
- [35] Valentin V. Rybenkov, Nicholas R. Cozarelli, and Alexander V. Vologoskii. Probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl. Acad. Sci. USA*, 90:5307–5311, 1993.
- [36] Robert G. Scharein. KnotPlot. <http://www.knotplot.com>, 1998. Program for drawing, visualizing, manipulating, and energy minimizing knots.
- [37] Stanley Y. Shaw and James C. Wang. Knotting of a DNA chain during ring closure. *Science*, 260(5107):533–536, 1993.
- [38] Joanna I. Sulkowska, Eric J. Rawdon, Kenneth C. Millett, Jose N. Onuchic, and Andrzej Stasiak. Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Natl. Acad. Sci. USA*, 109(26):E1715–E1723, 2012.
- [39] De Witt Sumners and Stuart G. Whittington. Knots in self-avoiding walks. *J. Phys. A*, 21(7):1689–1694, 1988.
- [40] Peter Guthrie Tait. On knots. *Trans. Roy. Soc. Edin.*, 28:273–317, 1877.
- [41] Peter Guthrie Tait. On knots, Part II. *Trans. Roy. Soc. Edin.*, 32:327–342, 1884.
- [42] Peter Guthrie Tait. On knots, Part III. *Trans. Roy. Soc. Edin.*, 32:493–506, 1885.
- [43] Fusao Takusagawa, , and Shigehiro Kamitori. A real knot in protein. *J. Am. Chem. Soc.*, 118(37):8945–8946, 1996.
- [44] William R. Taylor. A deeply knotted protein and how it might fold. *Nature*, 406:916–919, 2000.
- [45] William Thomson. On vortex atoms. *P. Roy. Soc. Edin.*, VI:94–105, 1867.
- [46] Luca Tubiana, Enzo Orlandini, and Cristian Micheletti. Probing the entanglement and locating knots in ring polymers: A comparative study of different arc closure schemes. *Prog. Theor. Phys. Suppl.*, 191:192–204, 2011.
- [47] Laura Zirbel and Kenneth C. Millett. Characteristics of shape and knotting in ideal rings. *Journal of Physics A: Mathematical and Theoretical*, 45(22):225001, 2012.

Yuanan Diao, Claus Ernst*, Uta Ziegler, and Eric J. Rawdon

The Knot Spectrum of Random Knot Spaces

Abstract: It is well known that knots exist in natural systems. For example, in the case of (mutant) bacteriophage P4, DNA molecules packed inside the bacteriophage head are considered to be circular since the two sticky ends of the DNA are close to each other. The DNAs extracted from the capsid, without separating the two ends, can preserve the topology of the (circular) DNAs, and hence are well-defined knots. Furthermore, knots formed within such systems are often varied and different knots occur with different probabilities. Such information can be important in biology. Mathematically, we may view (and model) such a biological system as (by) a random knot space and attempt to obtain information about the system via mathematical analysis and numerical simulation. The question here is to find the probability that a randomly (and uniformly) chosen knot from this space is of a particular knot type. This is equivalent to finding the distribution of all knot types within this random knot space (called the knot spectrum in an earlier paper by the authors). In this paper, we examine the behavior of the knot spectrums for knots up to 10 crossings. Using random polygons of various lengths under different confinement conditions as the random knot spaces (model biological systems), we demonstrate that the relative spectrums of the knots, when divided into groups by their crossing numbers, remain surprisingly robust as these knot spaces vary. For a given knot type \mathcal{K} , we let $P_{\mathcal{K}}(L, R)$ be the probability that an equilateral random polygon of length L in a confinement sphere of radius R has knot type \mathcal{K} . We give a model for the family of functions $P_{\mathcal{K}}(L, R)$ and show that our model function fits the random polygon data we generated. For a fixed crossing number Cr , $3 \leq Cr \leq 10$, let S_{Cr} be the subspace consisting of random polygons which form knots that have crossing number Cr . We study the relative distribution of all the different knot types within S_{Cr} and illustrate how this distribution changes if we keep the length L fixed (or the confinement radius R fixed) and vary the confinement radius R (or the length L). We observe that this distribution is quite robust and remains essentially unchanged under length and confinement radius variation, especially if one concentrates on subfamilies such as alternating prime knots, non-alternating prime knots, or composite knots.



Yuanan Diao, Department of Mathematics and Statistics, University of North Carolina Charlotte, Charlotte, NC 28223, USA, E-mail: ydiao@uncc.edu

***Corresponding author: Claus Ernst**, Department of Mathematics, Western Kentucky University, Bowling Green, KY 42101, USA, E-mail: claus.ernst@wku.edu

Uta Ziegler, School of Engineering and Applied Sciences, Western Kentucky University, Bowling Green, KY 42101, USA, E-mail: uta.ziegler@wku.edu

Eric J. Rawdon, Department of Mathematics, University of St. Thomas, St. Paul, MN 55105, USA, E-mail: ejrawdon@stthomas.edu

<https://doi.org/10.1515/9783110571493-009>

 Open Access.  © 2018 Yuanan Diao et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

Keywords: DNA packing; topology of circular DNA; random polygons; DNA knots; knot spectrum of random polygons

Mathematics Subject Classification 2010: 57M25, 92B99

10.1 Introduction

In the world of pure mathematics, or more specifically knot theory, one typically considers the space of knots as the entire collection of all knots, with each distinct knot type as an element of this discrete space. It is commonly agreed that some knots are in general “more complicated” than other knots, although there are many different ways to define knot complexity and sometimes a knot more complicated with one complexity measure may be simpler in terms of a different complexity measure. For example, some knots with high crossing numbers have small bridge numbers and some knots with relatively small crossing numbers have high bridge numbers. One could imagine that in an ideal knot space, each knot type is no more probable than any other knot type. However, this is not the case for knots occurring in the world of nature. For example, if one is to tie a knot with a rope of some fixed length, then one can only tie finitely many different types of knots ([4, 10, 11, 25]). Furthermore, if one is to tie a knot with a rope in a random fashion, then one will inevitably reach the conclusion that some knots are “easier” to tie than other knots (meaning some knots will be tied more often than other knots if the experiment is repeated). Some important biological problems concerning subjects such as DNA packing are also related to this topic. In DNA research, a relatively simple virus called bacteriophage is commonly used to study the DNA packing mechanism. This virus keeps its genome in a spherical protein container called a capsid. Although the packing of the DNA inside the capsids cannot be directly observed, DNA extracted from the capsids without being broken retains its topological information and this information is used as a probe in studying the DNA packing mechanism [2]. It is reported that many different DNA knots form within the capsids of bacteriophages, with certain knot types appearing with much higher frequencies [2].

Mathematically, a biological system such as the DNA knots within the bacteriophage head is really just a “knot space” whose elements are geometric closed curves with a certain (continuous) probability distribution. If one is to randomly and uniformly sample a knot from this space, the probability that one gets a particular knot type is the same as the percentage of knots of this knot type within the space. The distribution of the knot types within the space is called the *knot spectrum* in [12, 16]. In a sense, the knot spectrum (of a knot space) measures how easily a knot type can be realized by a randomly chosen knot in that space and provides a knot complexity measure. While it is quite conceivable that specific knot spaces can be constructed to favor a particular type of knots, we are interested in the following question: Are there

knots that are intrinsically easier to form in most random knot spaces without obvious topological biases? For example, in the space of random equilateral polygons with a fixed length, it is known that composite knots have higher frequencies than the prime knots when the polygons are long (in fact the probability of getting a prime knot goes to zero as the length of the polygon goes to infinity [9]).

In this paper, we study the above problem by examining the behavior of the knot spectrum for knots up to 10 crossings. Using random polygons of various lengths under different confinement conditions as the random knot spaces (model biological systems), we demonstrate that the relative spectrums of the knots, when divided into groups by their crossing numbers, remain surprisingly robust as these knot spaces vary. For a given knot type \mathcal{K} , we let $P_{\mathcal{K}}(L, R)$ be the probability that an equilateral random polygon of length L in a confinement sphere of radius R has knot type \mathcal{K} . We give a model for the family of functions $P_{\mathcal{K}}(L, R)$ and show that our model function fits data we generated from a large sample of random polygons. Furthermore, for a fixed crossing number Cr ($3 \leq Cr \leq 10$) we consider the knots space S_{Cr} of random polygons that form knots with Cr crossings. We observe how the frequency distribution of knots in S_{Cr} changes under various lengths and confinement conditions. One of the main findings in this article is that this distribution is quite robust and remains essentially unchanged under length and confinement radius variation, especially if one concentrates on subfamilies such as alternating prime knots, non-alternating prime knots, or composite knots. The support of the previous statement is given by the data we collected. The details of the data collection will be explained in Section 10.3. Even though our data set is quite large, it does not allow us to support any statements on the individual knot type distributions for knots with more than 10 crossings since there are very few samples per knot type for crossing numbers above 10. Furthermore, already for nine and 10 crossings, the natural sampling error is large enough to only weakly support any conjecture about the independence of knot type distributions on polygon length and confinement radius. As a final remark, we need to acknowledge that we do not see any way to prove any statement on knot distributions. At this point we cannot even prove much simpler statements like: The probability that a random polygon is knotted is larger if the polygon is confined when compared to an unconfined random polygon of the same length.

This article is organized as follows: In Section 10.2 we introduce some basic terminology in knot theory. In Section 10.3 we outline the size of our random polygon sample and how the knot types in our sample were determined. In Sections 10.4 and 10.5 we explain the choice of a model function for $P_{\mathcal{K}}(L, R)$ and show how well our data fits the model. Next, in Sections 10.6 and 10.7 we study how the confinement radius R and the polygon length L affect the distribution of the different knot types in S_{Cr} . Finally, in Section 10.8 we summarize our findings.

10.2 Basic mathematical background in knot theory

For the convenience of our reader, we outline and discuss briefly a few topological concepts that are most relevant to this paper. For a more detailed exposition, please refer to a standard text on knot theory such as [1, 6, 21, 26].

A *knot* K is a simple closed curve in \mathbb{R}^3 . Here we assume that such a curve is a piece-wise smooth curve (this includes a space polygon without self-intersections). Two knots are considered topologically equivalent if one can be continuously deformed, together with the entire \mathbb{R}^3 space surrounding it and without being broken or causing self-intersection in the process, to the other. The class of all knots equivalent to a knot is called a *knot type*. The knot type that contains the unit circle is called the *trivial knot* (type). For a fixed knot K , a *regular projection* of K is a projection of K onto a plane such that no more than two segments of K cross at the same point in the projection. At each intersection point of a regular projection of K , it is usually marked which strand is over and which strand is under. These intersection points are called *crossings*. A regular projection with this over/under information marked at the crossings is also called a *knot diagram*. The minimum number of crossings among all possible knot diagrams of knots with the same knot type as K is called the *crossing number* of K (and is usually denoted by $Cr(K)$). A knot diagram is *alternating* if at the crossings under and over alternate as one travels along the knot projection. A knot type is *alternating* if it has an alternating diagram and is *non-alternating* if it does not have any alternating diagram. A knot is called a *composite knot* if it is realized by connecting two nontrivial knots as shown in Figure 10.1. If a knot is not a composite knot, then it is a *prime knot*.

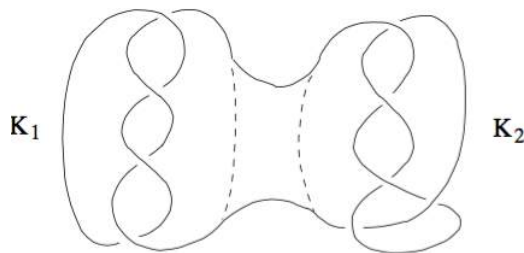


Fig. 10.1: A composite knot from two non-trivial knots K_1 and K_2 .

One fundamental problem in knot theory is to determine the knot type of a given knot. The most common and powerful tools for this purpose are various knot polynomials. The well-known knot polynomials include the Alexander polynomial, the Jones polynomial, and the HOMFLYPT polynomial. These polynomials can be com-

puted from any knot diagram and remain unchanged when computed from a different knot diagram of the same knot type. In this paper, the authors rely on the HOMFLYPT polynomial and the knot table for knots up to 16 crossings for knot identification. For the definition of the HOMFLYPT polynomial and its properties, one may refer to [24]. It is important to note that the knot polynomials are not sufficient to distinguish all knot types and the tabulation of knots remains a difficult question in general. Furthermore, only prime knots up to 16 crossings have been completely tabulated [20].

10.3 Spaces of random knots, knot sampling and knot identification

In a series of papers, three authors of this paper have developed algorithms for several models to generate equilateral random polygons that are confined inside a sphere of fixed radius [13, 14, 15, 17]. The model presented in [14] is the one chosen for the study presented in this paper. The model can be described as follows: Consider equilateral random polygons that are “rooted” at the origin and assume that there is an algorithm that samples such objects with uniform probability. Now consider a confinement sphere S_R of radius $R \geq 1$ with its center at the origin. We keep those randomly generated equilateral polygons that are contained in the confinement sphere S_R . Note that the algorithm used in [14] to generate polygons in confinement is not based on a direct accept-reject method (since such a method is extremely inefficient). Instead, it uses conditional probability density functions that can be explicitly formulated to guide the generation process. Each polygon is generated one edge at a time and there is no rejection involved. Furthermore, this algorithm generates polygons that are totally independent of each other so no de-correlation is necessary. Interested readers please refer to [14] for a detailed description of this algorithm. There is no biological or other reason for the polygons to be rooted at the center. It is rather a choice for simplicity: as it turns out, equilateral random polygons defined this way are much easier to generate due to the symmetry of the confining sphere (relative to the root) imposed on the equilateral random polygons. In order to obtain statistically significant results, it is necessary to use large samples and the simplicity of this model allows the computations to be feasible. We want to point out to the reader that, recently, a different and very efficient method to generate such confined polygons (also centered at the origin) was developed, see [7].

Our knot space consists of all random equilateral polygons of a given length and a given radius of confinement. By varying these two parameters, we obtain a family of knot spaces. Notice that the differences among these knot spaces are imposed by different geometric conditions and no explicit topological constraints exist. The data set used for our analysis in this paper comes from two sources.

The larger portion of the data is new and we describe how it is obtained here. To trace the effect of polygon length on knotting, we fixed the confinement radius at $R = 3$ and used a range of lengths of the random polygons from 10 to 90 in increments of 10. We used a maximal polygon length of 90 steps and a confinement radius $R = 3$ to ensure that, even for the longest sample, we can still identify the knot type for most polygons. This means that many of the knots that are generated must have at most 16 crossings so they are in the current knot table. From our past experience [12], we know that if we choose a confinement radius that is smaller than $R = 3$, then we would not be able to identify the knot types for the longer lengths. To trace the effect of the confinement radius, we fixed the polygon length at 30 steps and used radii of confinement ranging from $R = 1$ to $R = 4.5$. In more detail, between $R = 1$ and $R = 3$ the confinement radius increases with an increment of $1/10$, while from $R = 3$ to $R = 4.5$ the confinement radius increases with an increment of $1/2$. The reason for the wider spacing of the larger radii is that the confinement effect diminishes rather quickly for the polygons within the length-range of this study. The polygon length of 30 is chosen so that sufficiently many different knot types populate our knot spaces and that we are able to identify the knot type for most polygons. For each knot space (with a fixed confinement radius and a given polygon length), we sampled 100,000 different polygons. In total, this sample space consists of 32 sets, each containing 100,000 polygons (24 samples of length 30 with varying confinement and 9 samples of varying length with confinement $R = 3$), yielding a total of 3,200,000 random polygons (since one sample is in both sets).

A smaller portion of the data (1,640,000 polygons) is from one of our earlier papers [12]. We basically took the same set of radii (24 values) as described above and the same set of different length (9 values). For each combination of a radius and length 10,000 polygons were collected. However, for some of the larger lengths and smaller confinement radii, we could not identify the knot types reliably. The knots are too complex for our knot identification process to work – see the description below. Thus, some of these data sets were not used. We refer the reader to [12] for information about the exact description of the data. It is enough to know that this data set is grouped into subsets of size 10,000 for each fixed (L, R) pair, where $1 \leq R \leq 4.5$ and $10 \leq L \leq 90$. Moreover, data points that are collected from this older data set have a larger error margin since the sample size is much smaller (10,000 versus 100,000 polygons).

As a final remark, our combined data set contains almost five million polygons. In the following sections we will refer to the two data sets as the old data (the 1,640,000 polygons) and the new data (the 3,200,000 polygons).

In the following we briefly outline our procedure of knot identification for a given polygon P . The polygon P is projected onto a plane to obtain a knot diagram. Then *unraveller* [28] is used to (potentially) simplify the crossing information via a collection of simplification operations based on Reidemeister moves. The code *unraveller* produces two types of output: a DT-code which is used by *knotfind* [19] and crossing information

which is used to compute the HOMFLYPT polynomial H using a program written by Ewing and Millett [18]. We have a table of HOMFLYPT polynomials for all chiral knot types, prime and composite, with 16 or fewer crossings (under the generally accepted assumption that the crossing number of composite knots is additive). For each P , we obtain a list of chiral knot types $\{K_1, \dots, K_s\}$ from the table, all of which have H as their HOMFLYPT polynomial. Note that if this list is empty then the polygon does not represent any prime or composite knot with fewer than 17 crossings. In addition to the HOMFLYPT polynomial calculation, for each polygon we also use *knotfind* to compute the (non-chiral) knot type. We use the simplified DT-code D generated by *unraveller*, which might be the DT-code of a minimal diagram or might be the DT-code of a diagram close to the minimal diagram. Then *knotfind* takes D and creates a “canonical” DT-code D' . If D' represents a knot within the DT-code knot table (which only contains prime knot types) then that uniquely identifies that knot (up to chirality). If both methods agree then we claim that the knot type of P is identified. If the two methods disagree, that is one method produces a knot that we should be able to identify, but we cannot confirm this with the other method, then we resolve this issue as follows: Either we start over using a different projection of the polygon P or we run another diagram simplification program written by some of the authors. In this way, all such conflicts were resolved. Thus, we claim that with very high probability we identified all knots that are in the current knot table [16].

However there are still several additional issues which we address in the following paragraphs.

(i) The polygon P might represent a knot that is not prime and therefore is not in the DT-code knot table. At every stage of the simplification process using DT-codes, *knotfind* attempts to identify factors of composite knots. If there is an “obvious” connected sum then a part of the DT-code maps onto itself. If such a situation is detected, the simplification process is applied separately to each DT-code of the two factors. In the end, we obtain a collection of simplified DT-codes from which it is possible to reconstruct the original composite knot type. If the HOMFLYPT polynomial of this reconstruction agrees with the originally computed HOMFLYPT polynomial of the polygon then we identify P as the appropriate composite knot.

(ii) If the initial calculation of H indicates that P does not represent a knot in the table, then the simplified DT-code (or the set of simplified DT-codes in the case that an obvious connected sum was identified) provides an upper bound on the crossing number. In these cases, we double-checked by computing the HOMFLYPT polynomial from the simplified DT-code(s) to ensure that it matches H . We identify P as having the crossing number provided by the simplified DT-code.

We observed that the approach to knot simplification based on DT-codes is extremely reliable for the knots that are within the knot table. There is no reason to believe that the simplification of DT-codes becomes suddenly unreliable once the actual

crossing number exceeds 16. Thus, we believe that a simplified DT-code, while technically only providing an upper bound on the actual crossing number, gives a value that, in most cases, is actually the topological minimal crossing number. In particular, this is true for crossing numbers that are not far above 16 (the largest crossing number in the knot table). Therefore we report these approximated crossing numbers as if they are the actual crossing numbers. We also report a knot as prime if no composition was detected during the simplification process.

10.4 An analysis of the behavior of $P_{\mathcal{K}}$ with respect to length and radius

Let $P_{\mathcal{K}}(L, R)$ denote the probability that a randomly selected polygon in the set of equilateral random polygons of length L in a confinement sphere of a fixed radius R has knot type \mathcal{K} . In this section, we analyze the dependence of $P_{\mathcal{K}}(L, R)$ on the polygon length L (for fixed R values) and on the confinement radius R (for fixed polygon lengths L). At the end of the section we view $P_{\mathcal{K}}(L, R)$ as a function of both L and R .

10.4.1 $P_{\mathcal{K}}(L, R)$ as a function of length L for fixed R

For each knot type \mathcal{K} , it is obvious that there is a minimal length L_0 such that \mathcal{K} can be realized by an equilateral polygon of length L_0 but not by any equilateral polygon of length less than L_0 (L_0 is called the *equilateral stick number* of the knot type \mathcal{K} [22, 27]). Thus, $P_{\mathcal{K}}(L, R) = 0$ for $L < L_0$. For all the data analyzed in this study, we have $2R < L$ which means that the polygons were generated under confinement pressure (though very light pressure for the data set for $L = 10$ and $R = 4.5$). Our numerical results strongly support the following conjecture: for any fixed R , as L increases, $P_{\mathcal{K}}(L, R)$ increases to a single maximum at length $L_{\mathcal{K}}$, and then declines to zero as L approaches infinity, see the left of Figure 10.2. Notice that a similar phenomenon has been observed in [8] in the absence of confinement. We further conjecture that the rate of the decline is exponential in terms of some positive power of L . This conjecture has been proven when the polygons are not confined [9]. In that proof [9], it is shown that as L increases, the number of connected sum components in the equilateral random polygon increases with a probability that goes to one, hence the probability for the polygon to be of a given knot type goes to zero. Unfortunately, this argument may not be applicable to confined knots since a long confined random equilateral polygon may not have many connected sum components. However, it is quite plausible and intuitive that the overall knot complexity of a confined polygon increases as its length increases and thus knots with a lower complexity become less likely. This is strongly supported by our numerical results.

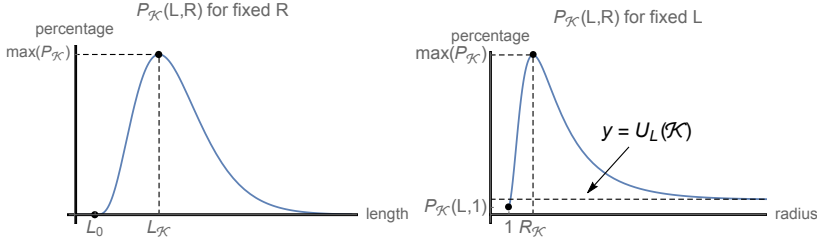


Fig. 10.2: On the left: A hypothetical curve $P_{\mathcal{K}}(L, R)$ with fixed R . On the right: A hypothetical curve $P_{\mathcal{K}}(L, R)$ with fixed L .

10.4.2 $P_{\mathcal{K}}(L, R)$ as a function of confinement radius R for fixed L

Here we assume that the fixed L value is large enough so that a knot of type \mathcal{K} can be easily formed. Since we can only model confinement radii $R \geq 1$, we assume that $P_{\mathcal{K}}(L, 1) > 0$ is a positive number. We can only speculate on the behavior of $P_{\mathcal{K}}(L, R)$ for $R < 1$. Clearly no knots can exist for $R \leq 1/2$, so $P_{\mathcal{K}}(L, R) = 0$ for $R \leq 1/2$. However, the behavior of $P_{\mathcal{K}}(L, R)$ for $1/2 < R < 1$ could be different for different knot types. Extreme confinement may favor certain knot types, so it is conceivable that, for some knot type \mathcal{K} , we have $P_{\mathcal{K}}(L, R) = 0$ for all $R \leq R_0$ for some value $R_0 > 1/2$. Moreover the value R_0 could depend on the knot type and have different values for different knot types. Since we have no data for $R < 1$, we only stipulate that $P_{\mathcal{K}}(L, 1) > 0$ is a positive number in our model. Just as in the case of dependence on length, we propose that with increasing radius $P_{\mathcal{K}}(L, R)$ rises to a single maximum at radius $R_{\mathcal{K}}$, and then declines. However, this time $P_{\mathcal{K}}(L, R)$ becomes a positive constant $U_L(\mathcal{K})$ once R becomes large enough (say $R > L/2$). This is because, for large R , the effect of confinement disappears and $P_{\mathcal{K}}(L, R)$ is simply the probability for an unconfined equilateral random polygon of length L to be of knot type \mathcal{K} . By our initial assumption on L , this probability is positive and we denote it with $U_L(\mathcal{K})$.

10.4.3 Modeling $P_{\mathcal{K}}$ as a function of length and radius.

Based on known rigorous results and our numerical results, we propose the following model function for $P_{\mathcal{K}}(L, R)$:

$$P_{\mathcal{K}}(L, R) = a \left(d + \left(\frac{L - L_0(\mathcal{K})}{R - 0.6} \right)^e \right) \exp \left(-\frac{L}{bR - c} \right),$$

where $L_0(\mathcal{K})$ is the equilateral stick number of \mathcal{K} , $L \geq L_0(\mathcal{K})$, $R \geq 1$, $R > c/b$, and $a, b, c, d, e > 0$ are positive constants depending on the knot type \mathcal{K} . The condition $R > c/b$ insures that the exponential term decreases with L . In addition we set $P_{\mathcal{K}}(L, R) = 0$

for $L < L_0(\mathcal{K})$. Notice that for a given fixed R value, $P_{\mathcal{K}}(L, R)$ is a function of L with the following properties:

1. $\lim_{L \rightarrow \infty} P_{\mathcal{K}}(L, R) = 0$; and
2. $P_{\mathcal{K}}(L, R)$ has a single maximum point.

On the other hand, for a fixed L value, $P_{\mathcal{K}}(L, R)$ is a function of R with the properties:

1. $\lim_{R \rightarrow \infty} P_{\mathcal{K}}(L, R) = ad > 0$; and
2. $P_{\mathcal{K}}(L, R)$ has a single maximum point.

10.5 Numerical results

10.5.1 The numerical analysis of $P_{\mathcal{K}}(L, R)$ based on the old data

We fitted the above model function $P_{\mathcal{K}}(L, R)$ to knot data of several knots up to seven crossings. Here we used the older data set, where each data point is based on a sample of 10,000 polygons. Three examples are shown in Table 10.1 with the corresponding fitting parameters and R^2 values:

Tab. 10.1: The parameters for the model $P_{\mathcal{K}}(L, R)$ for various knot types

parameter	3_1	5_2	6_2
a	0.000205	2.17277×10^{-7}	1.73334×10^{-9}
b	5.95573	4.28009	3.85288
c	2.2024	1.3623	1.31161
d	124.299	9736.84	335759.
e	3.43966	5.66256	7.09841
R^2	.996	.986	.982

Figure 10.3 is a plot of the fitting function $P_{3_1}(L, R)$ together with the actual data points from two different view points. Here $L_0(3_1) = 6$. The maximal z -value of the data points is ≈ 0.21 and the average value is ≈ 0.108 . If we compute the absolute value of the difference between the z -values of the data points and the best fit function, we get a maximal value of ≈ 0.026 and an average value of ≈ 0.0067 .

In Figure 10.4 we show the fitting functions for $P_{5_2}(L, R)$ and $P_{6_2}(L, R)$ together with the actual data points. Note that for both of these knots $L_0(5_2) = L_0(6_2) = 8$ [27]. The maximal z -value of the data points for 5_2 (6_2) is ≈ 0.059 (≈ 0.027) and the

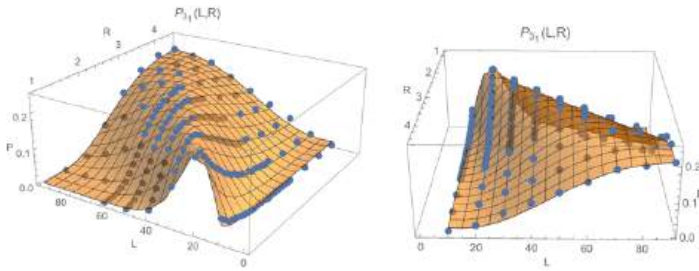


Fig. 10.3: Two different views of the surface $P_{3_1}(L, R)$ together with the actual data points. The slightly different coloring of the data points indicates if the points are above or below the surface.

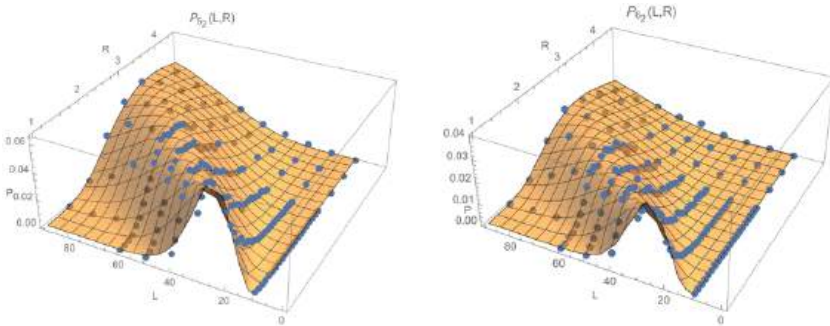


Fig. 10.4: Left: The plot of the fitting function $P_{5_2}(L, R)$ together with the actual data points; Right: The plot of the fitting function $P_{6_2}(L, R)$ together with the actual data points. The slightly different coloring of the data points indicates if the points are above or below the surface.

average value is ≈ 0.022 (≈ 0.0088). The maximum difference between the z -values of the data points for 5_2 (6_2) and the best fit function is ≈ 0.011 (≈ 0.0055) while the average difference is ≈ 0.0026 (≈ 0.0011).

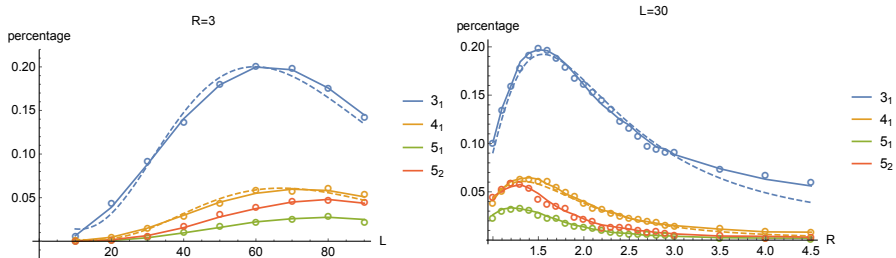


Fig. 10.5: Comparison of the fitting function $P_{\mathcal{K}}(L, R)$ obtained from the newer data and the older data for the knots 3_1 , 4_1 , 5_1 , and 5_2 . Left: Fixed R ; Right: fixed L . The solid curves are based obtained by connecting the new data points by line segments. The circles are the old data and the dashed lines represent a cross section through the surface $P_{\mathcal{K}}(L, R)$ fitted to the old data.

10.5.2 The numerical analysis of $P_{\mathcal{K}}(L, R)$ based on the new data

Overall, the newer and larger data sets did not produce any surprises in terms of the behaviors of the $P_{\mathcal{K}}(L, R)$ functions, in fact there is an amazingly good agreement. Here we show two examples. In Figure 10.5 on the left we compare the new data points with the old data points and with the slices for $R = 3$ from the surfaces $P_{\mathcal{K}}(L, R)$ obtained from the old data. The figure shows the data for the first four non-trivial knot types in the knot table. Here each round data point is based on the old data set (from a sample of 10, 000 polygons), whereas the solid lines are connecting the data points from the new data (from a sample of 100, 000 polygons). Each of the two dashed curves shows the slice for $R = 3$ of $P_{\mathcal{K}}(L, R)$ when fitted to the 164 data points of the old data for knot types 3_1 and 4_1 . Similarly, Figure 10.5 on the right shows a comparison for $L = 30$ based on the new data (solid curves) with their counterparts based on the old data (circles) and of two fitting functions $P_{\mathcal{K}}(L, R)$ (dashed curves) based on the old data. The R^2 values for the fitting functions based on the new data points are all ≥ 0.999 .

10.5.3 The location of local maxima of $P_{\mathcal{K}}(L, R)$

The maximal points of $P_{\mathcal{K}}(L, R)$ for a fixed knot type \mathcal{K} lie on a line in the (L, R) plane as shown in Figure 10.6 on the left using the example of the four knots 3_1 , 4_1 , 5_2 , and

6_2 . The points shown in the figure are computed from the model $P_{\mathcal{K}}(L, R)$ obtained from fitting the older data. The slope of these lines increases with knot complexity from about ≈ 20.41 for the trefoil knot to about ≈ 27.35 for the knot 6_2 . Furthermore, for a fixed radius R , it takes more and more length for the maximum to occur as the knot complexity increases.

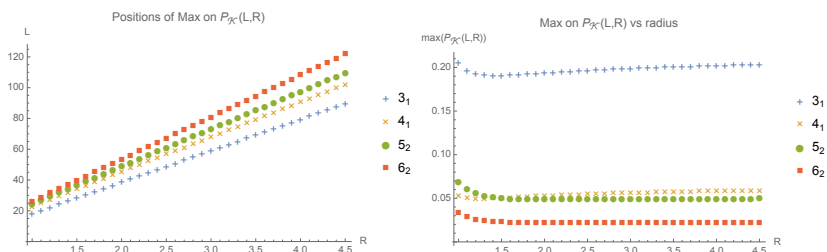


Fig. 10.6: The maxima of the fitted surfaces $P_{\mathcal{K}}(L, R)$ from the older data for the knots 3_1 , 4_1 , 5_2 , and 6_2 . Left: Dependence of the length on where the maxima occur on R . Right: Dependence of the magnitude of the maxima on R .

In Figure 10.6 on the right, we show the relationship of the maximal value of $P_{\mathcal{K}}(L, R)$ and the (L, R) plane using the same four examples of 3_1 , 4_1 , 5_2 , and 6_2 . Here we see that the height of the maxima rapidly declines as the knot complexity increases. It appears that as the radius (and length) increases, the maximal value of $P_{\mathcal{K}}(L, R)$ initially drops slightly and then remains roughly constant. In Figure 10.7 we show each of these curves in a coordinate system by itself and now it appears that the value of each maximum of $P_{\mathcal{K}}(L, R)$ slightly increases for increasing R after the initial drop.

10.6 The influence of the confinement radius on the distributions of knot types

In this section we consider only the new data, and we would like the reader to recall that these are all polygons of length 30. A recurring observation made in this section is that the relative distribution of knots in a given group of knots (e.g the group of 7-crossing alternating knots, or the 8-crossing composite knots) is remarkably steady for many different confinement radii. The last few subsections in this section illustrate this visually.

Many of the figures presented in Sections 10.6 and 10.7 contain error bars to indicate the reliability of the data. Usually the error bars are only added to some of the

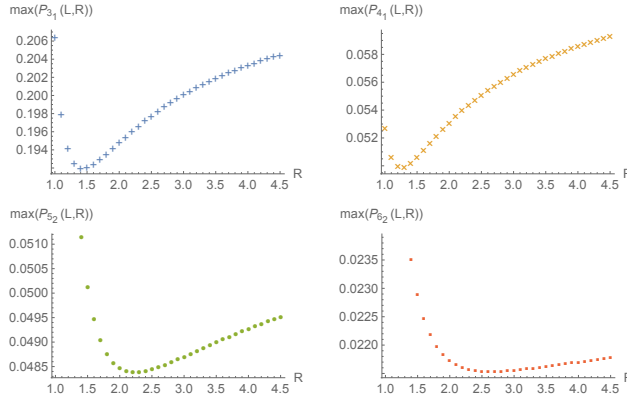


Fig. 10.7: The plot of the heights of the maxima of the fitting function $P_{\mathcal{K}}(L, R)$ depending on R for the knots 3_1 (top left), 4_1 (top right), 5_2 (bottom left) and 6_2 (bottom right).

curves of each graph so that the graph remains intelligible. The error bars were determined using the standard interval estimation for binomial proportions (see [3]). The standard confidence interval is computed as $CI_s = \hat{p} \pm z_{\alpha/2} n^{-1/2} (\hat{p}\hat{q})^{1/2}$ with $\hat{p} = \frac{X}{n}$ where n is the sample size, X is the number of successes (occurrences of knots of interest) in the sample and $\hat{q} = 1 - \hat{p}$. The value $z_{\alpha/2}$ is the $100(1 - \alpha/2)$ -th percentile of the standard normal distribution. In this article we use $\alpha = 0.05$.

10.6.1 3-, 4-, and 5-crossing knots

Our data contains 501,838 knots with three, four, or five crossings. There are 322,609 3_1 knots, 85,134 4_1 knots, 35,184 5_1 knots, and 58,911 5_2 knots. There is a maximum of 21,291 knots at radius $R = 1.0$ and a minimum of 6,951 knots at radius $R = 4.5$. We note that, for all radii, the order of the knot types (from least frequent to most frequent) is consistent, as shown in Figure 10.5 on the right.

10.6.2 6-crossing knots

Our data contains 69,080 6-crossing knots (all polygons are of length 30). There are 17,940 6_1 knots, 21,077 6_2 knots, 12,620 6_3 knots, and 17,443 $3_1\#3_1$ composite knots. There is a maximum of 7,065 knots at radius $R = 1.0$ and a minimum of 205 knots at radius $R = 4.5$. We note that we do not distinguish the square and the granny knot – instead the two different composite knots are grouped together as $3_1\#3_1$. The four functions $P_{\mathcal{K}}(30, R)$ are shown in Figure 10.8 on the left. The figure shows that the order of knot types is not the same for all radii, unlike for the smaller knot types

shown in Figure 10.5 on the right. The order of the four knot types seems to be the same for the smaller radii $1 \leq R \leq 1.5$, while it is hard to see what happens for the larger radii. In order to make this more visible, we show the relative percentage of the four different knot types, see Figure 10.8 on the right. Here, for each fixed R -value, the y -values of the four data points add up to one. This kind of normalization is done frequently in the following sections of this article. In order to make the differences of probabilities between different knot types more visible for a particular family of knots (for example knots with a fixed crossing number or knots with a particular property such as 9-crossing composite knots) we scale the actual percentages such that the sum of the scaled percentages adds up to one for all knots in this family. This new quantity is called *relative percentage* and in Figure 10.8 on the right we see this for the first time. Most noticeable in this figure is the relative decline of the composite knots as the confinement radius decreases. This is not really surprising as it has been shown in [12, 16] that tighter confinement suppresses composite knots. We also observe that the order of the alternating knots (from least frequent to most frequent) is consistent, except when the error bars become too large, and that the knot 6_3 occurs with the smallest frequency. This is due to the fact that 6_3 is achiral and achiral knots have been observed with a lower frequency than chiral knots of the same crossing number [12].

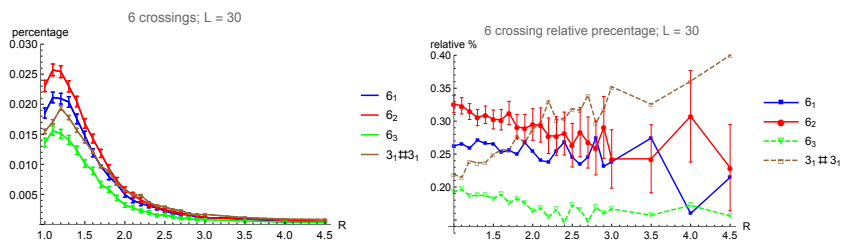


Fig. 10.8: On the left: The percentage (or actual probability) of the four different 6-crossing knots. On the right: the relative percentage distribution of the four knot types. For a fixed R , the frequencies of all data points for that R add up to one.

10.6.3 7-crossing knots

Our data contains 40,071 7-crossing knots. There are 2,587 7_1 knots, 5,125 7_2 knots, 4,763 7_3 knots, 2,835 7_4 knots, 6,497 7_5 knots, 7,665 7_6 knots, 3,717 7_7 knots, and 6,882 $3_1 \# 4_1$ composite knots. There is a maximum of 5,775 knots at radius $R = 1.1$ (there are 5,745 at radius $R = 1.0$) and a minimum of 47 knots at radius $R = 4.5$. Since the values of $P_{\mathcal{K}}(30, R)$ are quite small, we only display the relative

percentages of all 7-crossing knots in Figure 10.9. As before, we observe that the relative percentage of the composite knot $3_1 \# 4_1$ declines with decreasing confinement radius. In addition, for radii $R < 2$, the order of the relative percentage of the 7-crossing prime knots seems to be independent of the radius. The knot 7_6 is the most frequent, followed by 7_5 , 7_2 or 7_3 , 7_7 , and 7_4 or 7_1 . At this point, we have no explanation of why the data shows this ordering of the knot types. We have compared this order with the order of knot energies [23] and ropelength [5], however we found no correlation at all (see Table 10.2).

Tab. 10.2: The order of 7-crossing knots based on various measures.

knot	$P_K(30, R)$	ropelength	Möbius energy approximation
7_1	7	1	1
7_2	3 or 4	3	2
7_3	3 or 4	2	3
7_4	5 or 6	4	4
7_5	2	5	5
7_6	1	7	6
7_7	5 or 6	6	7

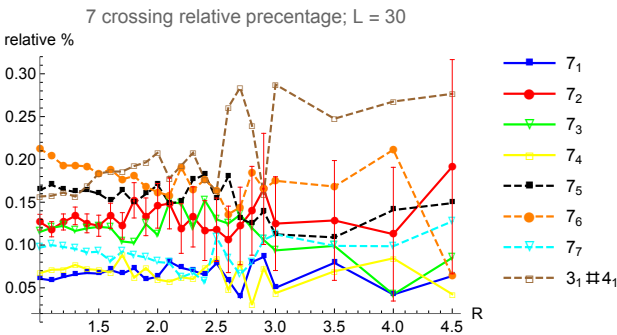


Fig. 10.9: The relative percentage of the different 7-crossing knot types on the vertical axis as a function of the confinement radius R .

10.6.4 8-crossing knots

In the following, we investigate the knot type distribution for 8-crossing knots. Our sample contains 42,711 knots with eight crossings, with a maximum of 7,301 knots

at radius $R = 1$ and a minimum of 38 knots at radius $R = 4.5$. As the number of knot types increases, the graphs become more and more complicated. There are 21 prime knot types and 3 composite knot types to consider. A figure similar to Figure 10.9 would be too crowded and so we display the information in a different way.

First, we claim that the knot type distributions are remarkably consistent regardless of the radius of confinement. In other words, while the exact value of the relative percentage is not the same, the order of the knot types (from smallest to largest relative percentage) is often the same. To support this, we no longer show a curve for each knot type versus the radius since this would be 24 curves (21 prime and 3 composite). Instead we combine the knots of 4 consecutive radii and show one curve for each group of radii reducing the number of curves to plot to six. That is we combine $R = \{1.0, 1.1, 1.2, 1.3\}$, $R = \{1.4, 1.5, 1.6, 1.7\}$, $R = \{1.8, 1.9, 2.0, 2.1\}$, $R = \{2.2, 2.3, 2.4, 2.5\}$, $R = \{2.6, 2.7, 2.8, 2.9\}$, and $R = \{3.0, 3.5, 4.0, 4.5\}$. For each of the groups of four radii, we tabulate the distribution of each of the 24 knot types by using relative percentages as before. Thus for each of the six curves the total sum of the values over all 8 crossing knot types equals to one. In effect, in Figures 10.10, 10.11, and 10.13 the information plotted vertically for each knot type is a compressed version of what Figure 10.9 shows (for 7-crossing knots) with each curve. Finally, we split the data into two groups, and display the data for the knot types on the x -axis, see Figures 10.10 and 10.11.

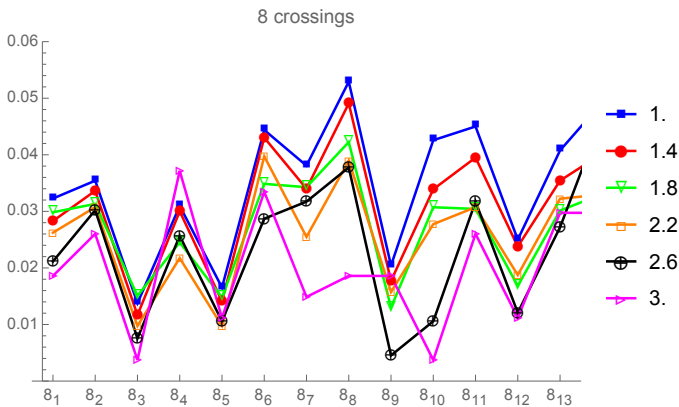


Fig. 10.10: The relative percentages of the different 8-crossing knots from 8_1 to 8_{13} grouped into six groups by the different radii. In each group, the smallest radius is shown as a plot label.

The most striking feature of Figures 10.10 and 10.11 is how much more likely the non-alternating knot types 8_{19} , 8_{20} , and 8_{21} are when compared to all other knot types (prime and composite). This property is independent of the radius of confinement.

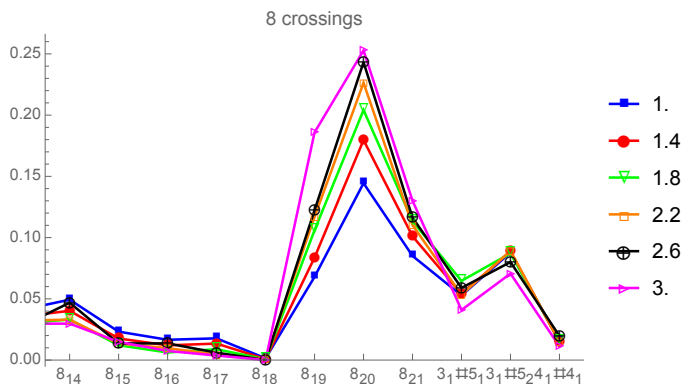


Fig. 10.11: The relative percentages of the different 8-crossing knots from 8_{13} to $4_1\#4_1$ grouped into six groups by the different radii. In each group, the smallest radius is shown as a plot label.

However, the larger confinement radii favor non-alternating knots even more than the smaller confinement radii. If one looks closely at Figure 10.11 then one can see that for the non-alternating knots 8_{19} , 8_{20} , and 8_{21} the order of the curves (from the smallest relative percentage to the largest relative percentage) is by the size of the radius. That is, the smallest group of confinement radii leads to the least relatively frequent occurrence of the knot type and the largest group of confinement radii leads to the most relative frequent occurrences of the knot types. For prime knots, the order of the curves is often the opposite, the curve representing the smallest group of confinement radii is on the top for many knots and the curve with the largest group of confinement radii is on the bottom. We suspect that this inconsistency in the order of curves maybe due the fact that our sample is not large enough, so that for knot types that have a much smaller frequency, the sample error could be large enough to create a different order. For the three composite knots $3_1\#5_1$, $3_1\#5_2$, and $4_1\#4_1$ in Figure 10.11, we notice that the curve for the largest radii is on the bottom (smallest relative percentage). However, the order of the other curves is not clear and the data points are very close to each other.

The least likely knot type is 8_{18} whose standard diagram is based on the basic Conway polyhedron 8^* , which has already been observed in [12]. More precisely, the knot 8_{18} appears only 49 times in our sample, while the next lowest occurrence is an order of magnitude larger (the knot 8_3 with 557 cases). The knot that occurs the most is 8_{20} with 6,964 cases. In order to get more insight, we look at some subsets of the knot types. Figure 10.12 on the left shows that $3_1\#5_2$ is more likely than $3_1\#5_1$, which in turn is more likely than $4_1\#4_1$. The curve for $3_1\#5_1$ has statistical error bars which are similar to those of the other two curves. This distribution of the composite types is expected since 5_2 is more likely than 5_1 and the composite $4_1\#4_1$ is achiral, and therefore less likely. The figures hint that this distribution is largely independent of

the confinement radius. Figure 10.12 on the right shows the frequencies of the three non-alternating knot types 8_{19} , 8_{20} , and 8_{21} . The figure shows that 8_{20} is more likely than the other two, which have about the same frequency, with 8_{21} being slightly more likely than 8_{19} .

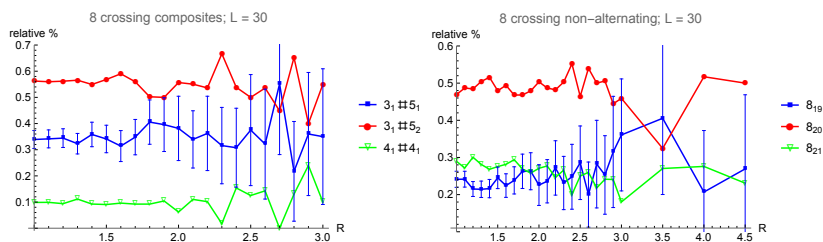


Fig. 10.12: On the left, the relative percentages of 8-crossing composites versus the confinement radius. On the right, the relative percentages of 8-crossing non-alternating prime knots versus the confinement radius.

10.6.5 9-crossing knots

In Figure 10.13, we show the 9-crossing knot spectrum in the same way as we showed the knot spectrum for the 8-crossing knots. Since there are many more knot types (49 prime and 6 composite) the knot types are not labeled on the horizontal axis. However we indicate the extent of the three knot groups: prime alternating knot types, prime non-alternating knot types, and composite knot types. Our sample contains 26,742 knots with nine crossings, with a maximum of 5,665 random knots at radius $R = 1$ and a minimum of 18 knots at radius $R = 4.5$. The total of 26,742 random knots is about 63 % of the number of 8-crossing random knots while the number of knot types is twice as large (24 versus 55 knot types). Thus as expected, the fluctuations due to the smaller random sample muddle the picture. Nevertheless, the curves (maybe with the exception of the curve for the largest radii) are still roughly parallel. Here, by roughly parallel, we mean that the six curves could be put into a relatively narrow band – it does not mean that the curves cannot intersect each other. The reader should also keep in mind that the smaller the radius the more actual knots are represented by such a curve. Thus the curve representing the largest radii has the largest statistical error, which is reflected by the fact that it often crosses through all the other curves. The smallest numbers of knots occur for 9_{40} (6 knots), 9_{41} (39 knots), 9_{35} (46 knots) and 9_{34} (60 knots). We also note that the standard diagrams of the knots 9_{40} and 9_{34} are based on the Conway basic polyhedra 9^* and 8^* respectively, see also [12]. The largest numbers of knots occur for 9_{42} (2,604 knots), 9_{44} (2,585 knots), 9_{45} (2,028 knots),

and 9_{43} (1, 799 knots). The next knot in the frequency list drops by more than 600. As in Figure 10.11, there is a large increase in the number of non-alternating knots. The same observations we made for 8-crossing knot types about the order of the different curves according to size of the confinement radius could be true. However, the picture is less clear, which could be due to the larger sample variance caused by the larger number of knot types and the smaller overall knot sample.

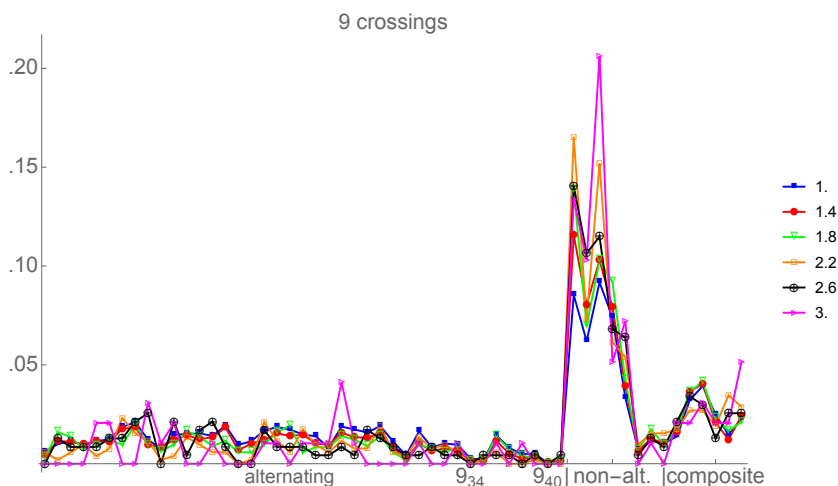


Fig. 10.13: The relative percentages of the different 9-crossing knots grouped into six groups by the different radii. In each group, the smallest radius is shown as a plot label.

In order to get more insight, we investigate some subsets of the knot types: composite knot types and non-alternating knot types. Figure 10.14 on the left shows that $3_1\#6_2$ is more likely than $3_1\#6_1$, which in turn is more likely than $3_1\#6_3$. This is as expected since 6_2 is more likely than 6_1 , which in turn is more likely than 6_3 . We also see that $4_1\#5_2$ is more likely than $4_1\#5_1$, which is as expected since 5_2 is more likely than 5_1 . The curve for $3_1\#3_1\#3_1$ has statistical error bars which are similar to the other curves. The frequency of $3_1\#3_1\#3_1$ is about the same as that of $4_1\#5_1$. This shows that this distribution is largely independent of the confinement radius. Figure 10.14 on the right shows the relative frequencies of the eight non-alternating knots 9_{42} to 9_{49} . It shows that 9_{47} , 9_{48} , and 9_{49} are much less likely than any of the others. We also note that a standard diagram of the knots 9_{47} is based on the Conway basic polyhedra 8^* .

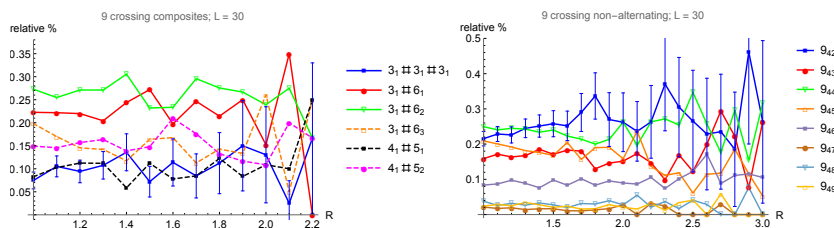


Fig. 10.14: On the left, the relative percentages of 9-crossing composites versus the confinement radius. On the right, the relative percentages of 9-crossing non-alternating prime knots versus the confinement radius.

10.6.6 10-crossing knots

In Figure 10.15 we show the relative percentages of the 10-crossing spectrum. Since there are many knot types (165 prime and 14 composite), the knots are not labeled on the x -axis. Our sample contains 24, 277 knots with 10 crossings, with a maximum of 6, 021 knots at radius $R = 1$ and a minimum of 9 knots at radius $R = 4.5$. There are about as many 10-crossing random knots as 9-crossing random knots but the number of knot types is more than three times as large. Thus, we expect the fluctuations due to the smaller random sample to be even larger than it was for 9-crossing knots. There are very few knots for the larger radii, thus it makes very little sense to use the data for the large radii. As a result, in Figure 10.15 we only show three curves, the first for radii $R = 1.0 - 1.3$ contains 18, 118 random knots, the second for radii $R = 1.4 - 1.7$ contains 4, 647 random knots, and the third for radii $R = 1.9 - 2.1$ contains 1, 004 random knots. (The fourth would only contain 310 knots and therefore does not yield a useful distribution.)

If we look at the five knots that have the lowest frequency, then four of these have a standard diagram based on Conway basic polyhedra: 10_{123} (2 knots and 10^* polyhedra), 10_{115} (3 knots and 8^* polyhedra), 10_{121} (4 knots and 9^* polyhedra), and 10_{122} (4 knots and 9^* polyhedra). In fifth place is 10_{99} (8 knots) which is not based on Conway basic polyhedra. As in the previous figures, we can see the large increase in the number of non-alternating knots. The non-alternating knots with the largest frequencies are as follows: 10_{132} (1007 knots), 10_{133} (762 knots), 10_{124} (596 knots), 10_{128} (596 knots), and 10_{137} (585 knots).

It is impossible to see in Figure 10.15 if the curves are still roughly parallel. To get a feeling of how similar the curves are, in Figure 10.16 we show a zoomed in version of the first two curves in Figure 10.15 using only the alternating knots of 10 crossings. Here we can see that there is still a lot of similarity between the two curves.

The error bars for most of the individual knot types are too large to allow definite conclusions. We illustrate this by just looking at the 10-crossing composites in Figure 10.17. The figure on the right shows the 10-crossing composite knots which include a 3_1

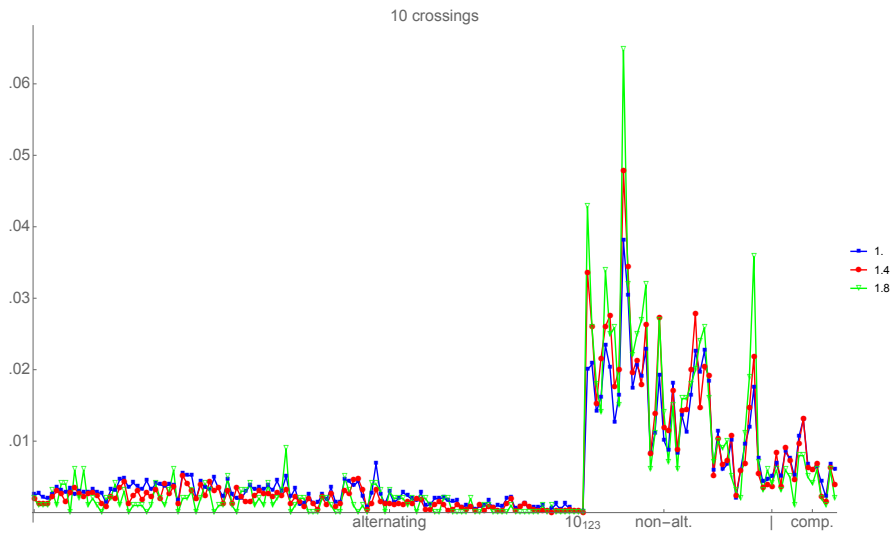


Fig. 10.15: The relative percentages of the different 10-crossing knots grouped into three groups by the different radii. In each group, the smallest radius is shown as a plot label.

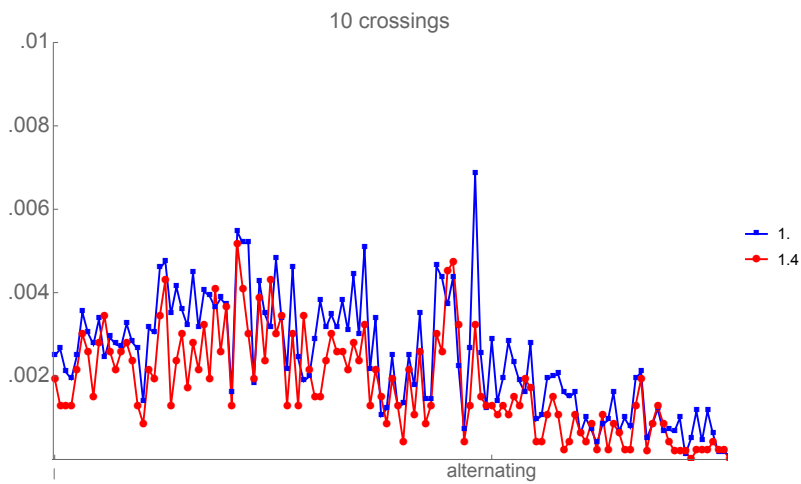


Fig. 10.16: A zoomed in version of Figure 10.15 showing only the relative percentages of the different alternating 10-crossing knots using the first two groups of the different radii.

component. Their relative percentages seem to be determined by the relative percentages of the associated 7-crossing knots, with 7_5 and 7_6 the most likely, causing $3_1 \# 7_5$ and $3_1 \# 7_6$ to be the most likely composite knots. Similarly, 7_1 and 7_4 are the least likely, causing $3_1 \# 7_1$ and $3_1 \# 7_4$ to be the least likely composite knots in the graph (see Figure 10.9 and Table 10.2). On the right part in Figure 10.17, a similar observation holds for 10-crossing composite knots which contain a 4_1 component.

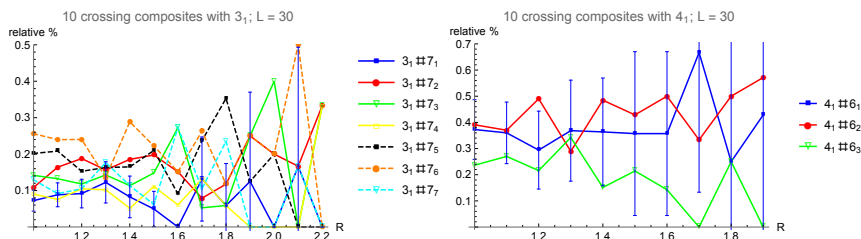


Fig. 10.17: On the left, the relative percentages of 10-crossing composites with 3_1 , and, on the right, the relative percentages of 10-crossing composites with 4_1 versus the confinement radius.

10.7 The influence of polygon length on the distributions of knot types in the presence of confinement

In this section, we discuss the dependence of the distribution of knot types on length in the presence of confinement. We want to remind the reader that we consider only the new data, and all polygons are in a sphere with confinement radius $R = 3$. This investigation is similar to that of Section 10.6. Intuitively, we expect that keeping the confinement radius fixed and increasing the length of the polygons has an analogous effect to keeping the length of polygons fixed and decreasing the confinement radius. Thus many of the results are similar to those of the last section. This includes the observations that the relative distribution of knots in a given group of knots is remarkably steady for many different confinement radii.

10.7.1 3-, 4-, and 5-crossing knots

Our data contains 184,517 knots with three, four, or five crossings. There are 117,059 3_1 knots, 31,827 4_1 knots, 13,075 5_1 knots, and 22,556 5_2 knots. There is

a maximum of 32,711 knots at length $L = 70$ and a minimum of 666 knots at length $L = 10$. We note that for all lengths the order (from most frequent to least frequent) of knot types is consistent, as shown in Figure 10.5 on the left.

10.7.2 6-crossing knots

Our data contains 33,161 6-crossing knots (all in a confinement sphere of radius $R = 3$). There are 7,351 6_1 knots, 8,311 6_2 knots, 5,107 6_3 knots, and 12,392 $3_1 \# 3_1$ composite knots. There is a maximum of 7983 knots at length $L = 90$ and a minimum of 2 knots at length $L = 10$. The four functions $P_{\mathcal{K}}(L, 3)$ are shown in Figure 10.18 on the left. The figure shows that the order of the four knot types remains unchanged, with $3_1 \# 3_1$ having the highest frequency, followed by 6_2 , 6_1 , and 6_3 with the lowest frequency. Figure 10.18 on the right shows the relative frequency of the four different knot types. Here for each fixed value L , the y -values of the four data points add up to one.

The most noticeable difference to the case in the previous section is the relative stability of the knot distribution, that is for length $L \geq 30$ the four functions in Figure 10.18 on the right look basically constant. The previous section shows that stronger confinement causes a relative decline in the number of composite knots leading to the expectation of a decline of composite knots for increased polygon lengths (and fixed confinement radius). Figure 10.18 indicates that this is not the case. It has already been observed in [12, 16] that for composite knots, the analogy between decreasing the radius of the confinement sphere for polygons and increasing the length of the polygons in a fixed confinement sphere breaks down. We also observe that the relative order of the prime knot types is the same in both cases (see Figure 10.8 for the smaller confinement radii) and that the achiral knot 6_3 occurs with the smallest frequency.

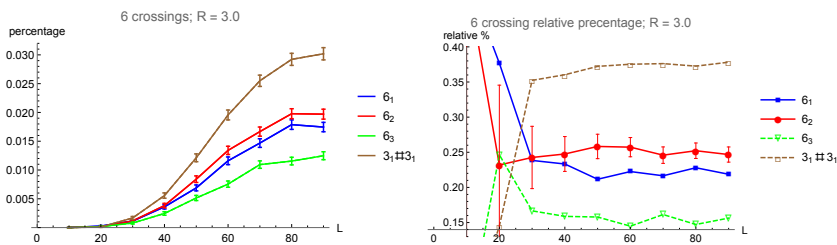


Fig. 10.18: On the left: The relative percentages of the four different 6-crossing knots versus length. On the right: the relative percentages of the four knot types versus length.

10.7.3 7-crossing knots

Figure 10.19 shows the dependence of the distribution of 7-crossing knot types in the presence of a confinement sphere of radius $R = 3$. Our sample contains 19,971 7-crossing knots. There are no 7-crossing knots of length 10 in our sample and the data starts with 15 knots of length 20, rising to 5,587 knots of length 90. The least likely knot is 7_4 which occurs 1063 times in the sample, while the composite knot $3_1\#4_1$ occurs with the largest frequency of 5,769. It appears that the relative distribution of 7-crossing knots varies only slightly with the length of the polygons for $L \geq 40$. From the data for 6-crossing knots (see Figure 10.18), one can speculate that the relative distribution of 7-crossing knots is also independent of the lengths of the polygons. There are significant similarities between this relative percentage distribution and the data for the dependence of the relative percentage of 7-crossing prime knots on the radius of confinement as shown in Figure 10.9: The curves show that 7_6 is the most frequent, followed by 7_5 . Next are 7_2 and 7_3 which are again approximately equally likely. This is followed by 7_7 and at the end we have 7_1 and 7_4 (again approximately equally likely). Just as in the case of the 6-crossing knots, the difference between the two distributions shown in Figures 10.19 and 10.9 is the relative percentage of the composite knot $3_1\#4_1$, which for small radii has a frequency similar to 7_5 (or 7_6), but for radius $R = 3$ is the most likely.

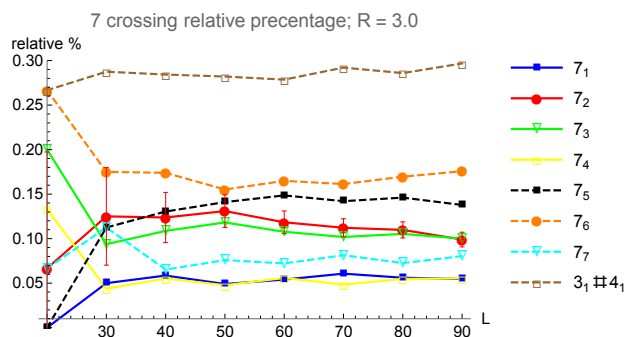


Fig. 10.19: The relative percentages of the different 7-crossing knot types versus length.

10.7.4 8-crossing knots

Next we look at 8-crossing knot types in the presence of a confinement sphere of radius $R = 3$. Our sample contains 20,921 8-crossing knots. There is a single knot of length 10, 12 knots of length 20, rising to 6,475 knots of length 90. The least likely knot is

8_{18} which occurs 19 times in the sample, while the composite knot $3_1\#5_2$ occurs with the largest frequency of 3,486. The prime knot with the largest frequency is the non-alternating knot 8_{20} , which occurs 2,521 times.

It again appears that the relative distribution of 8-crossing knots varies little with the length of the polygons once the polygons are long enough. Data supporting this claim is presented in several graphs. For Figures 10.20 and 10.21, all polygons up to length 50 are combined into one group since, for the smaller lengths, there are not many 8-crossing knots. As before we show relative percentages, that is for each curve (representing different lengths) the sum of the percentages over all knot types adds up to one. Similar to the data for the dependence of the relative percentages of 8-crossing knots on the radius of the confinement sphere, there are many more non-alternating prime knots than alternating prime knots, although there are also many more composite knots. Just as in the case of 7-crossing knots, the composite knots behave differently with increasing polygon length than they do with decreasing confinement radius. We also note that for the non-alternating prime knots, the curves are strictly ordered by length, with the shortest length leading to the largest relative frequency (on top) and the longest length leading to the lowest relative frequencies (on the bottom). The behavior for the alternating prime knots and composite knots is inconsistent but the order of curves is mostly reversed. Thus with the exception of the composite knots, this behavior is similar to the behavior of the 8-crossing knot types with a changing confinement radius.

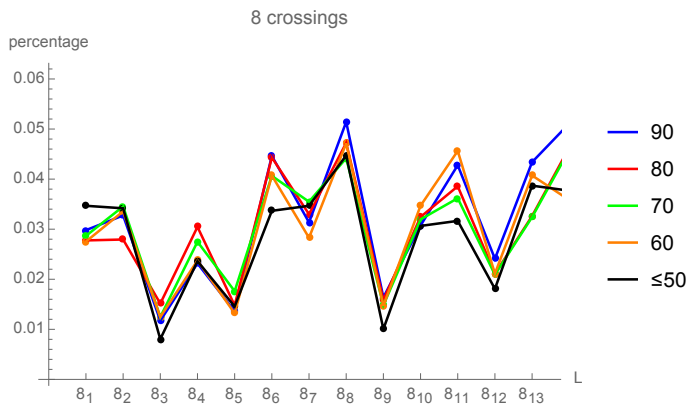


Fig. 10.20: The relative percentage of the 8-crossing knots from 8_1 to 8_{13} grouped into six groups by the different lengths.

Figure 10.22 shows the relative percentages of non-alternating knots (right) and the composite knots (left). It is remarkable how close the relative percentages for composite knots of length 90 ($R = 3$) ($\{0.32, 0.58, 0.1\}$) rounded to the nearest per-

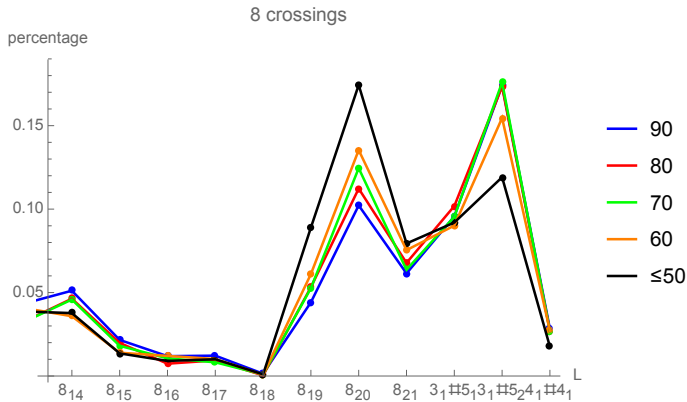


Fig. 10.21: The relative percentages of the 8-crossing knots from 8_{13} to $4_1\#4_1$ grouped into six groups by the different lengths.

cent) are to the relative percentages for composite knots at radius $R = 1$ ($L = 30$) $\{0.34, 0.56, 0.1\}$. For the non-alternating knot types, the data for polygons of length 90 is ($R = 3$) $\{0.21, 0.49, 0.30\}$ compared to ($L = 30$) $\{0.24, 0.47, 0.29\}$ for length 30 polygons in confinement of radius $R = 1$.

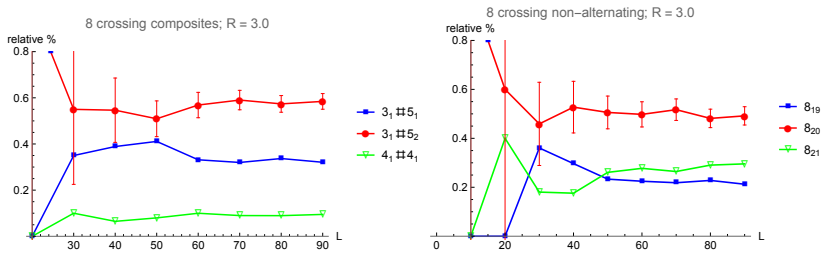


Fig. 10.22: The relative percentages of the different 8-crossing knot types on the vertical axis and its dependence on the polygon length. On the left: composite knots; on the right: non-alternating knots. The frequencies are normalized so that adding all values for a fixed length yields one.

For alternating prime knots of eight crossings, the picture is not clear since the error bars are too large to draw reliable conclusions about the frequency of all knot types. This is shown in Figure 10.23. Clearly, the knots that are the most frequent, 8_8 and 8_{14} , are also on the top of the knot type distribution for a fixed length and varying confinement radius, see Figures 10.10 and 10.11.

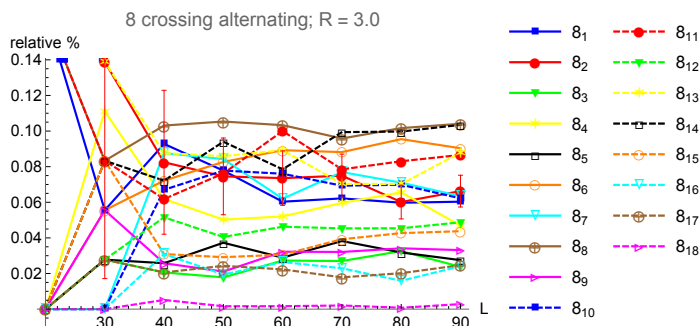


Fig. 10.23: The relative percentages of 8-crossing alternating prime knot types versus length.

10.7.5 9-crossing knots

Our sample contains 14, 406 9-crossing knots. There are no knots of length 10, 4 knots of length 20, rising to 4, 913 knots of length 90. The least likely knot is 9_{40} which occurs 6 times in the sample, while the composite knots $3_1\#6_1$ and $3_1\#6_2$ occur with the largest frequency of 1, 007 and 1, 115 respectively. The prime knot with the largest frequency is the non-alternating knot type 9_{44} , which occurs 971 times. The most frequent alternating prime knot is 9_8 , which occurs 307 times. There are 5, 925 alternating prime knots (9_1 to 9_{41}), 3, 836 non-alternating prime knots (9_{42} to 9_{49}), and 4, 645 composite knots. This shows again that the frequency of non-alternating knot types and composite knot types is larger than the frequency of alternating prime knot types. Similar to the last two subsections, the data suggests that the relative distribution of 9-crossing knots varies little with changes of the polygon length once the polygons are long enough ($L \geq 50$). Data to support this claim is presented in two graphs. Figure 10.24 shows the relative percentage of non-alternating knots (right) and of composite knots (left). It is remarkable how close the percentages for composite knots of length 90 (and radius $R = 3.0$), rounded to the nearest percent, are to the percentages for composite knots at radius $R = 1$ (and length $L = 30$) (see Table 10.3). Just like in the section about the dependence of the knot distribution on the confinement radius, the order of the relative frequencies of 9-crossing composite knots is strongly related to the relative frequencies of the 6- and 5-crossing prime knots. Similar results are obtained for the non-alternating prime knot types of length 90 in confinement of radius $R = 3$ compared with length 30 polygons in confinement of radius $R = 1$ (see Table 10.3).

Tab. 10.3: Similar relative percentages for polygons of length 90 (and radius $R = 3.0$) and polygons in confinement radius $R = 1.0$ (and length $L = 30$). On the left for composite knots and on the right for non-alternating knots.

knot	length 90 R = 3	R = 1 L = 30	knot	length 90 R = 3	R = 1 L = 30
$3_1\#3_1\#3_1$	0.17	0.07	9_{42}	0.24	0.22
$3_1\#6_1$	0.22	0.22	9_{43}	0.18	0.16
$3_1\#6_2$	0.25	0.27	9_{44}	0.24	0.25
$3_1\#6_3$	0.14	0.20	9_{45}	0.18	0.21
$4_1\#5_1$	0.08	0.08	9_{46}	0.10	0.08
$4_1\#5_2$	0.15	0.15	9_{47}	0.01	0.02
			9_{48}	0.03	0.04
			9_{49}	0.02	0.02

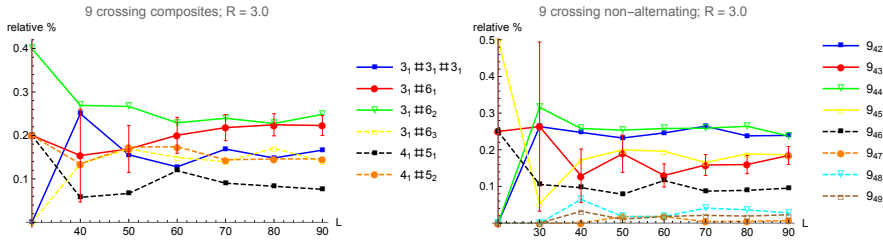


Fig. 10.24: On the left the relative percentages of different 9-crossing composite knot types versus length. On the right, the relative percentages of different 9-crossing non-alternating prime knot types versus length.

10.7.6 10-crossing knots

Our sample of knots with 10 crossings contains 12, 326 knots. There are no knots of length 10, one knot of length 20, rising to 4, 739 knots of length 90. The least likely knot is 10_{123} , which is the only knot that does not occur in the sample. A minimal standard diagram of 10_{123} is based on the 10^* basic Conway polyhedron. We note that this is the only knot type with less or equal to 10 crossings that was not observed in the old data set either [12]. This confirms the observation that knots whose minimal standard diagrams are based on basic Conway polyhedrons occur with a probability that is much lower than the probability of other knot types [12]. The composite knot $3_1\#3_1\#4_1$ occurs with the largest frequency of 457. The prime knot with the largest frequency is the non-alternating knot 10_{132} , which occurs 379 times. The most frequent alternating prime knot is 10_{20} , which occurs 81 times. There are 3, 751 alternating prime knots (10_1 to 10_{123}), 5, 520 non-alternating prime knots (10_{124} to 10_{165}), and

3,055 composite knots. This shows, just as for other crossing numbers, that the frequency of non-alternating knot types and composite knot types is larger than the frequency of alternating prime knot types. The relative distributions of 10-crossing knots in our data seems to vary more than for other graphs with the length of the polygons (once the polygons are long enough), however this is due to the large error bars. Thus, the authors claim that for a large enough sample, the statements of the previous sections would still hold: The relative distributions vary only slightly with changes in the length of the polygons. To support this, we show the non-alternating knots and the composite knots. The reader should compare Figure 10.25 with Figure 10.17 to follow the discussion in this paragraph. In the figure on the left are the composite knots with a 3_1 component. Their frequency seems to be determined by the frequency of the 7-crossing knots, with 7_5 and 7_6 the most likely knots, causing $3_1\#7_5$ and $3_1\#7_6$ to be the most likely composite knots. Similarly 7_1 and 7_4 are the least likely knots, causing $3_1\#7_1$ and $3_1\#7_4$ to be the least likely composite knots, see Figure 10.9. On the right are the results for composite knots with a 4_1 component. Here we would expect that $4_1\#6_1$ occurs with the highest frequency – but this does not always happen. However, as the error bar on the knot $4_1\#6_2$ indicates, the curves do not allow any definite conclusions.

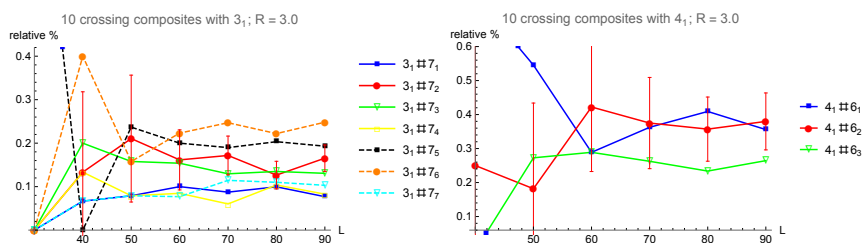


Fig. 10.25: On the left, the relative percentages of 10-crossing composites with a 3_1 -component versus length. On the right, the relative percentages of 10-crossing composites with a 4_1 -component versus length.

10.8 Conclusions

We finish this chapter with several conjectures. Here R is the radius of confinement and n is the length of a polygon P in our sample. We also want to remind the reader that the evidence collected in this article is restricted to $1 \leq R \leq 9/2$ and $10 \leq L \leq 90$. Furthermore our sample does not contain enough knots for large crossing numbers $Cr > 10$ to draw conclusions about the frequencies of individual knot types.

- The distributions of different prime knot types are virtually independent of the confinement radius and length. To be more precise, we claim that if \mathcal{K}_1 and \mathcal{K}_2 are two prime knot types in a knot space S_{Cr} (that is both knots have the same crossing number Cr) and if $P_{\mathcal{K}_1}(L_0, R_0) \ll P_{\mathcal{K}_2}(L_0, R_0)$ for some confinement radius R_0 and some length L_0 (where L_0 is big enough so that both knots can be formed) then $P_{\mathcal{K}_1}(L, R) \ll P_{\mathcal{K}_2}(L, R)$ for all confinement radii $R \geq 1$ and all lengths $L \geq L_0$. Similarly if $P_{\mathcal{K}_1}(L_0, R_0) \approx P_{\mathcal{K}_2}(L_0, R_0)$ for some confinement radius R_0 and some length L_0 (where L_0 is big enough so that both knots can be formed) then $P_{\mathcal{K}_1}(L, R) \approx P_{\mathcal{K}_2}(L, R)$ for all confinement radii $R \geq 1$ and all lengths $L \geq L_0$.

- The above property also holds if both knot types in S_{Cr} are composite knots.

- The relative probabilities of a composite knot type versus the probabilities of a prime knot type are virtually independent of the length but not independent of the confinement radius. To be more precise, we claim that if \mathcal{K}_1 is a prime knot and \mathcal{K}_2 a composite knot in a knot space S_{Cr} (that is both knots have the same crossing number Cr) and if $P_{\mathcal{K}_1}(L_0, R_0) \ll P_{\mathcal{K}_2}(L_0, R_0)$ for some confinement radius R_0 and some length L_0 (where L_0 is big enough so that both knots can be formed) then $P_{\mathcal{K}_1}(L, R) \ll P_{\mathcal{K}_2}(L, R)$ for all lengths $L \geq L_0$. Similarly if $P_{\mathcal{K}_1}(L_0, R_0) \approx P_{\mathcal{K}_2}(L_0, R_0)$ for some confinement radius R_0 and some length L_0 (where L_0 is big enough so that both knots can be formed) then $P_{\mathcal{K}_1}(L, R) \approx P_{\mathcal{K}_2}(L, R)$ for all lengths $L \geq L_0$.

- For a fixed crossing number Cr the probability that a polygon represents a knot in S_{Cr} eventually decreases in strong confinement. That is to say, provided that the length L is long enough so that \mathcal{K} can easily be formed, the function $P_{\mathcal{K}}(L, R)$ is initially an increasing function in R considering $R = 1/2$ as a starting point. For composite knots \mathcal{K} this increase is more pronounced than for prime knots \mathcal{K} .

Acknowledgment: Y. Diao was supported by Chinese NSF grant number 91324201. E. J. Rawdon was supported by USA National Science Foundation grant number 1418869.

Bibliography

- [1] Adams, C. (2004). *The Knot Book: An elementary introduction to the mathematical theory of knots*. American Mathematical Society.
- [2] Arsuaga, J., Vázquez, M., Trigueros, S., Sumners, D. W., and Roca, J. (2002). Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids. *Proc. Natl. Acad. Sci. USA*, 99:5373–5377.
- [3] Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical science*, pages 101–117.

- [4] Buck, G. and Simon, J. (1999). Thickness and crossing number of knots. *Topology Appl.*, 91(3):245–257.
- [5] Buniy, R. V., Cantarella, J., Kephart, T. W., and Rawdon, E. J. (2014). Tight knot spectrum in QCD. *Physical Review D*, 89(5):054513.
- [6] Burde, G. and Zieschang, H. (2003). *Knots*, volume 5. Walter de Gruyter.
- [7] Cantarella, J., Deguchi, T., and Shonkwiler, C. (2014). Probability theory of random polygons from the quaternionic viewpoint. *Comm. Pure Appl. Math.*, 67(10):1658–1699.
- [8] Deguchi, T. and Tsurusaki, K. (1998). Numerical application of knot invariants and universality of random knotting. In *Knot theory (Warsaw, 1995)*, volume 42 of *Banach Center Publ.*, pages 77–85. Polish Acad. Sci., Warsaw.
- [9] Diao, Y. (1995). The knotting of equilateral polygons in \mathbf{R}^3 . *J. Knot Theory Ramifications*, 4(2):189–196.
- [10] Diao, Y. and Ernst, C. (1998). The complexity of lattice knots. *Topology Appl.*, 90(1-3):1–9.
- [11] Diao, Y., Ernst, C., and Janse van Rensburg, E. J. (1997). Knot energies by ropes. *J. Knot Theory Ramifications*, 6(6):799–807.
- [12] Diao, Y., Ernst, C., Montemayor, A., Rawdon, E., and Ziegler, U. (2014). The knot spectrum of confined random equilateral polygons. *Molecular Based Mathematical Biology*, 2(1):19–33.
- [13] Diao, Y., Ernst, C., Montemayor, A., and Ziegler, U. (2011). Generating equilateral random polygons in confinement. *J. Phys. A: Math. Theor*, 44:405202.
- [14] Diao, Y., Ernst, C., Montemayor, A., and Ziegler, U. (2012a). Generating equilateral random polygons in confinement II. *J. Phys. A: Math. Theor*, 45:275203.
- [15] Diao, Y., Ernst, C., Montemayor, A., and Ziegler, U. (2012b). Generating equilateral random polygons in confinement III. *J. Phys. A: Math. Theor*, 45:465003.
- [16] Diao, Y., Ernst, C., Rawdon, E., and Ziegler, U. (2017). Relative frequencies of composite, alternating and non-alternating prime knots in random knot spaces. *Experimental Mathematics*, 6:1–18.
- [17] Diao, Y., Ernst, C., Saarinen, S., and Ziegler, U. (2015). Generating random walks and polygons with stiffness in confinement. *J. Phys. A: Math. Theor*, 48(9):095202.
- [18] Ewing, B. and Millett, K. C. (1997). Computational algorithms and the complexity of link polynomials. In *Progress in knot theory and related topics*, pages 51–68. Hermann, Paris.
- [19] Hoste, J. and Thistlethwaite, M. (1999). Knotscape. <http://www.math.utk.edu/~morwen/knotscape.html>. Program for computing topological information about knots.
- [20] Hoste, J., Thistlethwaite, M., and Weeks, J. (1998). The first 1,701,936 knots. *Math. Intelligencer*, 20(4):33–48.
- [21] Kawauchi, A. (1996). *A Survey of Knot Theory*. Birkhäuser Verlag, Basel.
- [22] Kim, H., No, S., and Oh, S. (2014). Equilateral stick number of knots. *Journal of Knot Theory and Its Ramifications*, 23(07):1460008.

- [23] Kusner, R. B. and Sullivan, J. M. (1998). Möbius-invariant knot energies. In *Ideal knots*, pages 315–352. World Sci. Publishing, Singapore.
- [24] Lickorish, W. R. (2012). *An introduction to knot theory*, volume 175. Springer Science & Business Media.
- [25] Litherland, R. A., Simon, J., Durumeric, O., and Rawdon, E. (1999). Thickness of knots. *Topology Appl.*, 91(3):233–244.
- [26] Livingston, C. (1993). *Knot theory*, volume 24. Cambridge University Press.
- [27] Rawdon, E. J. and Scharein, R. G. (2002). Upper bounds for equilateral stick numbers. In *Physical knots: knotting, linking, and folding geometric objects in \mathbb{R}^3 (Las Vegas, NV, 2001)*, volume 304 of *Contemp. Math.*, pages 55–75. Amer. Math. Soc., Providence, RI.
- [28] Thistlethwaite, M. (1990). unraveller. Code provided through private communication.

Kyle Chapman, Kenneth C. Millett, and Laura Plunkett

Sampling Spaces of Thick Polygons

Abstract: Open and closed polygons provide an attractive coarse grained model for many molecular structures. The random selection of polygons of a specified thickness is, however, an objective that has not been achievable until very recently. Here, we give an elementary description of the distinct sampling strategies that have been employed, their limitations, and the new algorithms that now allow one to randomly sample the spaces of open and closed polygons with specified thickness. We observe that the introduction of even a very modest thickness has an immediate and profound effect on the shape, the size, and the type of knot formed.

11.1 Introduction



A polygon in three dimensional Euclidean space consists of a finite sequence of straight line segments, called edges, meeting only at the endpoints of adjacent edges, called vertices. Polygons are closed if they form a loop and are, otherwise, open. Analogous to the ball and rod molecular models, polygons have provided scientifically useful coarse grained models of locally linear molecular structures. These polygons have provided productive coarse grained models of a wide range of molecules: rubber, amber, polyethylene, or other polymer melts or individual DNA chains, or proteins. In some cases, the edge lengths are allowed to vary in length according to a given distribution and, in others, they are required to have a constant length, in some cases reflecting the monomers and bonds of the structure while, in others, reflecting the homogeneity of the persistence (or Kuhn) length of the modeled macromolecule. Similarly, the bending angles between adjacent edges reflect the curvature character of the molecule and a triple of adjacent edges can be used to measure the intrinsic torsion of the chain. Another facet of the modeling challenge is to capture the thickness of the macromolecule or, similarly, the excluded volume of the modeled chains. The notion of excluded volume, introduced by Kuhn [10], was employed by Paul Flory [6] to explain that certain fundamental aspects of polymer physics are a consequence of the fact that a portion of a long polymer can not occupy the same space as another portion of the same polymer. As a result, for example, circular polymer chains can not change their topological knot type, i.e. “crossings are forbidden,” and, furthermore, their properties are distinct from those of infinitesimal mathematical polygonal rings.

Kyle Chapman, Department of Mathematics, University of Georgia, Athens, GA, 30602, USA

Kenneth C. Millett, Department of Mathematics, University of California, Santa Barbara, CA 93106, USA

Laura Plunkett, Department of Mathematics and Science, Holy Names University, Oakland, CA 94619, USA

<https://doi.org/10.1515/9783110571493-010>

 Open Access.  © 2018 Kyle Chapman et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The properties of excluded volume and thickness have proved to be a particularly challenging facet of the structure to simulate, even in a coarse grained manner, across a meaningful physical range of values.

Our focus will be on the available methods to sample the sets of open and closed polygons with which one may try to access structural information related to the thickness of the configuration. After an overview of how one may define the thickness of a polygon, we will describe these ways in sufficient detail that one may develop an algorithm to implement these methods. The first topic will be the ways in which one samples open and closed polygons without thickness constraints. These are subsequently sorted by thickness with the hope of generating a statistically robust sample for each range of thickness. Unfortunately, the rapid decay of sample size with increasing thickness limits the effectiveness of this method. We will next discuss two new methods that allow one to randomly sample collections of open and closed polygons having any desired thickness and, in conclusion, briefly describe some observations resulting from these new methods.

11.2 Classical Perspectives

The concept of *excluded volume* was introduced by Kuhn [10] and applied by Flory in the study of polymers [6, 21]. For example, the Kuhn length of polyethylene (PE) is 14 angstroms. Of course, the molecule has a cross-sectional molecular diameter as well as steric forces that give it an effective thickness that must be taken into consideration in a model. As noted, this has proved to be quite challenging due to the computational cost of a molecular model and the lack of a cruder, coarse grained, simulation method. If the effects of the excluded volume are neutralized in a given situation, described as being at the *theta point*, one ignores these interactions and employs the *ideal chain* model. These are defined as freely joined polygonal chains in which one ignores interactions between edges. We will now describe how one measures the thickness of polygonal chains, how such chains can be easily generated, and the character of the resulting thickness distribution.

11.2.1 Thickness of polygons

We employ the ideas of Rawdon [12, 20], to define the thickness of a polygon in 3-space (which Rawdon calls its injectivity radius). The underlying concept is that of a tubular neighborhood of a smooth curve consisting of the embedded family of discs of radius r at each point of the curve. The supremum of the radii of the embedded tubular neighborhoods is called the thickness of the curve. Rawdon extended this concept to polygons by taking the minimum, over all vertex angles, of a *curvature constraint*,

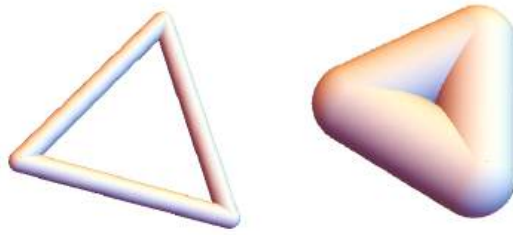


Fig. 11.1: Tubular neighborhoods of the unit equilateral triangle including one of maximal radius $\frac{\sqrt{3}}{6}$

$\frac{\tan[\frac{\pi}{6}]}{2}$, and half the minimum distance between distant edges, taken over all pairs of distant edges, the *distal distance constraint*. Tubular neighborhoods of an equilateral triangle are illustrated in Figure 11.1 and, of random walks, in Figure 11.2.

11.2.2 Self-avoiding random walks

An embedded open chain with n unit edge lengths and random vertex angles is a *self-avoiding random walk*, shown in Figure 11.2, is given by a sequence of vertices, v_i with $i = 0, 1, \dots, n$ or, equivalently, a sequence of edge vectors, e_j with $j = 1, \dots, n$. These are related by $v_0 = \{0, 0, 0\}$ and $v_i = \sum_{j=1}^i e_j$. To randomly generate an n step random walk (or n edge random open polygon) it is sufficient to individually generate the corresponding n edge vectors as follows: select $\theta = 2\pi \text{randomreal}[0, 1]$ and $\varphi = \text{ArcCos}[2 \text{randomreal}[0, 1] - 1]$ to define $e_j = (\text{Sin}(\theta) \text{Sin}(\varphi), \text{Cos}(\theta) \text{Sin}(\varphi), \text{Cos}(\varphi))$.

11.2.3 Closed polygons: fold algorithm

An embedded closed chain with n unit edge lengths and random vertex angles is a *self-avoiding polygon* constrained by the requirement that the initial and terminal vertex are equal, i.e. the closure constraint. The satisfaction of the closure constraint is the critical feature of sampling algorithms. The fold method preserves the closure by randomly selecting a pair of non-adjacent vertices to define an axis dividing the polygon into two arcs. One of these arcs is rotated through a random angle and adjoined to the complementary arc to define, following confirmation that resulting configuration is embedded by directly checking for possible edge intersections between edge segments, a new polygon, see Figure 11.3. We also show the result of 100 random folds applied to the 10 edge regular polygon. While its randomness is an open question, the fold algorithm is known to be a transitive algorithm, i.e. one capable of taking any configuration to any other configuration, whose resulting statistics, for example those of

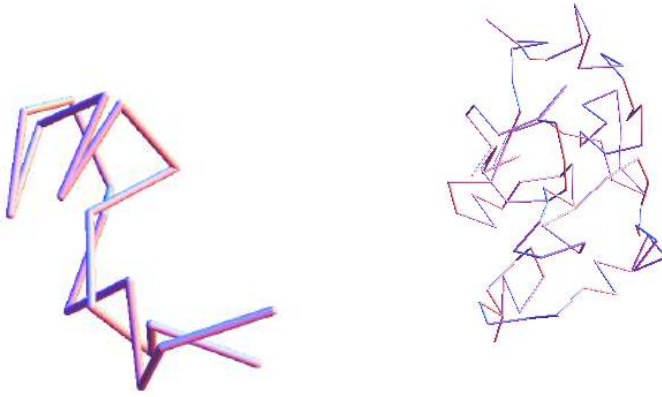


Fig. 11.2: A 20 step random walk: curvature constraint is 0.041 and the distal distance constraint is 0.071. The 100 step random walk has angle constraint of 0.054 and distal distance constraint 0.026.

observed knot types or average radius of gyration, are consistent with being random [16, 17].

11.2.4 Closed polygons: crankshaft algorithm

The crankshaft method is a computationally attractive algorithm that has also been shown to be transitive and having statistics consistent with randomness [1]. This algorithm randomly selects a pair, larger subset if desired, of edge vectors of an equilat-

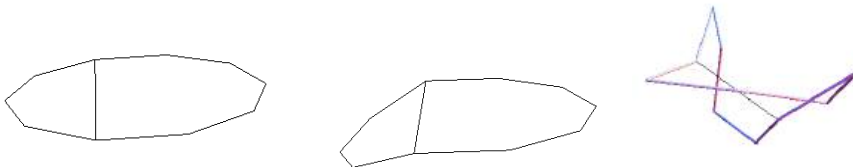


Fig. 11.3: A regular 10 edge polygon, a first fold, and one created from the regular polygon using 100 random folds along axes such as the one indicated.

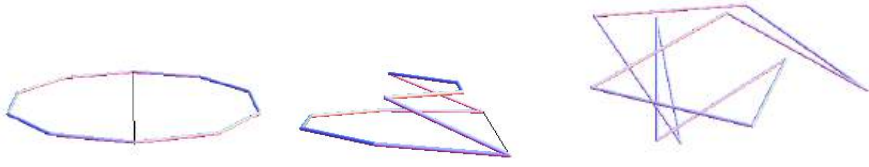


Fig. 11.4: A regular 10 edge polygon, a first crankshaft, and one created from the regular polygon using 100 random crankshaft modifications.

eral polygon and defines a segment based at the origin using this ordered sequence of vectors. Following a random rotation about the axis defined by the origin and the endpoint of the segment, one replaces the originally selected edge vector by the corresponding one in the resulting segment. Since the sum of these edge vectors has not changed, the newly defined polygon is closed. If it is embedded, the result is the new polygon, see Figure 11.4. The fact that one only has to modify a small collection of edge vectors, in our case just two, makes the calculations fast as they are explicit as shown in the following formulae.

$$e_i \mapsto \text{crank}(e_i, e_j, \theta) = \frac{e_i + e_j}{2} + \cos(\theta) \frac{e_i - e_j}{2} + \sin(\theta) \frac{e_i \times e_j}{\|e_i + e_j\|} \quad (11.2.1)$$

$$e_j \mapsto \text{crank}(e_j, e_i, \theta) = \frac{e_j + e_i}{2} + \cos(\theta) \frac{e_j - e_i}{2} + \sin(\theta) \frac{e_j \times e_i}{\|e_j + e_i\|} \quad (11.2.2)$$

11.2.5 Quaternionic Perspective

Building upon a construction by Hausmann and Knutson [8], it is possible to describe a natural measure on open and closed polygons in 3-space [2, 3]. This is based on a map from the complex Stiefel manifold of 2-frames in n -space to the space of closed n edge polygons in 3-space of total length 2 whose construction employs the Hopf map on quaternions. In the setting of equilateral n edge polygons, this approach allows one to randomly sample these polygons by jointly sampling the polyhedron determined by a complete set of secants and $n - 3$ dihedral space angles between the triangles formed by the secants and the edges of the polygon, each with the uniform measure. Data generated from this approach is consistent with the data from the random fold and crankshaft methods and, therefore, provides evidence that both the fold and crankshaft algorithms sample the space of n edge equilateral polygons randomly.

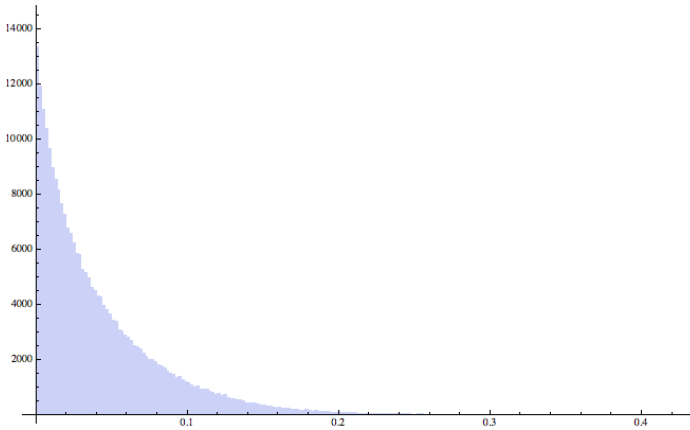


Fig. 11.5: A thickness distribution of a random sample of 250000 twenty edge equilateral polygons.

11.3 Sampling Thick Polygons

A consequence of the desire to understand the geometric and topological consequences of excluded volume on the average character on n edged open and closed polygons, i.e. thick coarse grained models of polymers, one is led to use the methods we've discussed to acquire data that gives insight into the world of thick polygons. Unfortunately, the proportion of a random sample decays so rapidly that this approach means that generating sufficient data required to give robust estimates as thickness increases, see Figure 11.5, is unreasonable. This is especially true if one wishes to study the average properties of, for example, trefoil knots. In this section, we describe two new sampling algorithms that permit one to randomly sample the sets of open and closed polygons of a specified thickness. They were first presented in the the PhD dissertations of Laura Plunkett, for open polygons, and Kyle Chapman, for closed polygons [4, 19] at University of California, Santa Barbara. They are described below.

11.3.1 Primer on Probability Theory

Much of the value of the following algorithms comes from being able to use them to prove a Markov chain arising from the moves involved can be used to determine a probability measure on various sets. In order to motivate these algorithms, we start with some of the key definitions involved.

We consider a space of states, X . We have a space of moves M , which gives a map $F : X \times M \rightarrow X$. If we choose a starting point x_0 , then a random sequence in M gives a random sequence in X . This is a Markov chain which gives a sequence of probability distributions $P^n(x_0, A)$ on our space X defined by the probability of starting at x_0

and landing in the Borel set A after n moves coming from M . The goal is to show that this sequence of probability distributions $P^n(x_0, A)$ converges to a limiting probability distribution $P(A)$.

The first step is knowing that there isn't some lower dimensional subset we are restricted to. This is given by the notion of **Forward Accessible**, which means that independent of the choice of x_0 the orbit under M has non-empty interior. While any single sequence only reaches a countable number of points, the set of all possible sequences can reach a full dimensional set.

The next step is knowing that the topology of the space is sufficiently well respected. We have the sequence of probability distributions P^n , so given any distribution on the natural numbers a , we get a new probability distribution $K_a(x, A) = \sum_{n=1}^{\infty} (a(n) \cdot P^n(x, A))$. Our Markov chain is a **T-chain** if there is some measure T on X with $T(x, A) \leq K_a(x, A)$ for all A , $T(x, X) > 0$ for all x and $T(x, A)$ lower semi-continuous in x for all A . This is a complex and difficult to show directly, but having other properties, we use Lemma 11.3.1. (See below.) A Markov chain which is forward accessible and with M having a lower semi-continuous probability density function is a T-chain. In our cases, moves from M are being chosen uniformly, so it has the required PDF, and so we have a T-chain.

Having a reachable state x_0 for our space, and we having a T-chain allows us to assert that there is an invariant measure on X , but it may not be a probability measure. For example, if we consider a random walk on the line or in the plane, the sequence P^n spreads steadily over the whole region, and the invariant measure is Lebesgue measure. The property which shows this issue does not happen and that the invariant measure P is a probability measure is **bounded in probability on average**. This says that for every ε there is a compact set C with $\overline{P^n} = \frac{1}{n} \sum_{i=1}^n P^i$ satisfying $\liminf \overline{P^n}(C) \geq 1 - \varepsilon$. Essentially this amounts to showing that the probability of going to " ∞ " is zero. We will be dealing with state spaces X which are compact, so this issue is automatically resolved, and so the invariant measure P is a probability distribution.

Finally, these can be combined to allow the use of the aperiodic ergodicity theorem to get the desired result. Letting $B(X)$ denote the Borel subsets of X ,

$$\sup_{A \in B(X)} |P^n(x, A) - P(A)| \rightarrow 0.$$

11.3.2 Open polygons: Plunkett algorithm [18]

We would like to generate random elements from the set of walks with n edges and thickness t . We will call the set of these walks $\mathcal{W}(n, t)$. To sample this space, we will start with a walk in the space, $W \in \mathcal{W}(n, t)$, and modify it, while ensuring we always have a thickness of at least t , until the resulting walk, W' , is independent of the original.

Our method is inspired by the pivot method, a Monte Carlo method proposed by Lal, and later implemented and proven to be ergodic by Madras and Sokal. The pivot method generates self-avoiding walks on the lattice by applying successive lattice symmetries at randomly selected vertices along the walk and accepting the new walk if it was not self-intersecting [13, 11]. Analogously, we perform reflections at vertices to portions of the walk.

- A single reflection will consist of selecting a random vertex and a random plane through that vertex. One portion of the walk will be reflected through the plane, and the other will remain fixed. If the new walk is still in $\mathcal{W}(n, t)$, we accept the move and continue. If it is not, we return to the walk before our reflection and try again with a new vertex and a new plane.
- A double reflection will consist of selecting two random vertices and a random plane through each vertex. The reflections are performed successively, and if the end result is a walk in $\mathcal{W}(n, t)$, we accept the change and continue. If it is not, we return to the walk before our reflections and try again.
- The acceptance probability of a reflection move varies with the thickness and length of walk. For example, for random walks of length 1000 and thickness of 1.00, the acceptance rate is 25.59% [18].

After a sufficient number of single and double reflections, we have a new walk, W' , independent of the starting walk W . What remains is to show that this method samples the space $\mathcal{W}(n, t)$ ergodically [14]. Informally, this means that the Monte Carlo Markov Chain gives a sampling method that converges to the probability measure on, in this case, a continuous state space. The first requirement in showing that our method is ergodic is to show it is transitive. A transitive Monte Carlo method is capable of sampling (or visiting) every state in the state space. For our method, proving transitivity would require showing we can go from any walk, W_0 , to any other, W_1 , through a finite series of single and double reflections with each intermediate walk in $\mathcal{W}(n, t)$. Because these reflections are reversible, this is equivalent to showing that we can get from any configuration, W_0 , to the straight configuration, W_s , a straight rod of length n .

So suppose we start with a walk $W_0 \in \mathcal{W}(n, t)$. We find a path to the straight configuration, W_s , by the following steps:

- we will perform single reflections on the outside of the convex hull of our walk until the ends of the walk determine the diameter of the convex hull, then
- we will use double reflections to remove critical points (maxima and minima) from the walk, and lastly
- we will perform more single reflections to straighten the configuration until we arrive at W_s .

If we have a walk, W_i , then we may perform any reflection through a vertex on the surface of the convex hull by a plane incident with only the surface of the convex hull. No reflection of this type will decrease the thickness of the resulting walk, as all pairs of edges either remain in the same position relative to each other, or the distance between them is sufficiently increased so as to never be less than $2t$. We will begin by reflecting in this way until both ends of the walk are on the surface of the convex hull and, further, that the two ends determine the diameter of the convex hull, as in Figure 11.6.

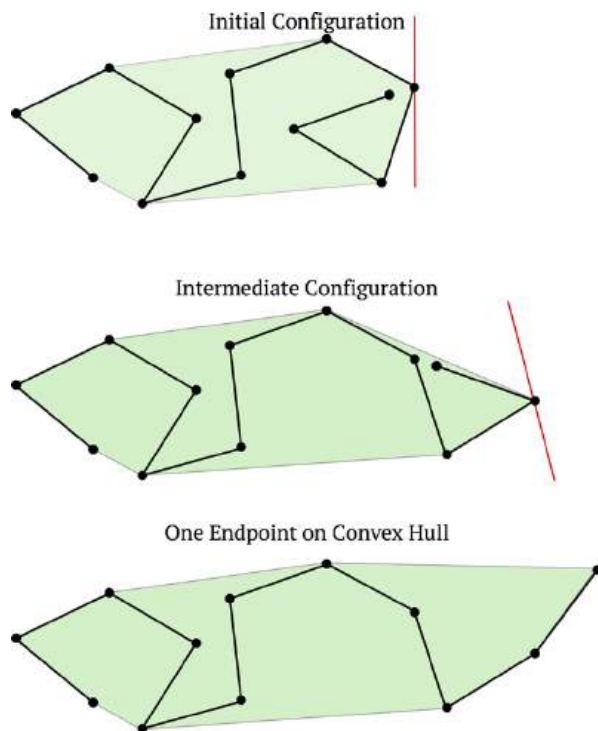


Fig. 11.6: Here we have three successive reflections through planes incident with points v_i on the convex hull. The end result is that the two ends of the chain have been moved to the surface of the convex hull, and determine the diameter of the convex hull. While the represented case is two dimensional, the application is in three dimensions.

Now we will remove pairs of maxima and minima (relative to the diameter of the convex hull, d^* , from the last step) through a series of double reflections. If our walk W_i has no critical points, we may move on to the last step, the straightening. If there are critical points, each critical point (with respect to this diameter of the convex hull), is either at a vertex v_i or an edge s_i . There is a plane through the critical point and perpendicular to the diameter, which we will call P_i .

We will select our maximum and minimum very carefully. First, We select the highest local maximum, that is, the local maximum m_i such that the distance between its plane P_i and the plane P_n containing the terminal and highest point is minimized, as in Figure 11.7. Between this maximum m_i and the end of the walk v_n there is at least one local minimum. We will select the local minimum n_j , $i < j < n$, such that the distance between the plane containing n_j , P_j , and the plane containing v_i , P_i , is greatest, as in Figure 11.7.

First, we will reflect the terminal end of the walk from m_i to v_n through the plane P_i . Then, we will reflect the terminal end of our modified chain, from \hat{n}_j to \hat{v}_n through the plane \hat{P}_j , the reflection of P_j . As in Figure 11.7, we have eliminated a pair of critical points. We now have three subsections of the walk to consider: from v_0 to m_i , m_i to \hat{n}_j and \hat{n}_j to \hat{v}_n . Between v_0 to m_i and m_i to \hat{n}_j , we have performed a reflection similar to part 1, and no thickness reducing interactions have been introduced. Similarly for the interaction between m_i to \hat{n}_j and \hat{n}_j to \hat{v}_n . For the interaction between the first and third segments, from v_0 to m_i and \hat{n}_j to \hat{v}_n , the third segment has been translated further away from the first segment. This increased distance is enough to verify that the thickness of the resulting walk has not been decreased.

Lastly, we will straighten this configuration by performing single reflections. Assume we have straightened the segments from v_{j+1} to v_n and we wish to straighten the next edge s_j between v_j and v_{j+1} . As in Figure 11.8, we perform a reflection so that the straight segment from v_{j+1} to v_n is in line with the edge s_j . We proceed until we have the straight configuration.

To prove that this method is ergodic, in addition to being transitive, we need the following lemmas about Markov chains.

Lemma 11.3.1. The Markov Chain F being forward accessible and the state space X having a probability density function implies that F is a T -chain.

These are modifications of the arguments in Meyn and Tweedie [14]: Propositions 7.1.5 and 6.2.4.

Lemma 11.3.2. A T -chain with a reachable state is bounded in probability on average if and only if it is positive Harris recurrent.

This is proposition 18.3.2 in Meyn and Tweedie [14].

Lemma 11.3.3. F is aperiodic.

Sketch of Proof Seeking a contradiction, suppose there exist C_i , subsets of the state space X or in this case $\mathcal{W}(n, t)$, such that $i \in \mathbb{Z}_n$, the cyclic group of order n , and they are a collection of disjoint non-empty closed sets with the probability of going from C_i to C_{i+1} is one. Let $w \in C_i$. In our context, if we apply a pair of reflections at the second

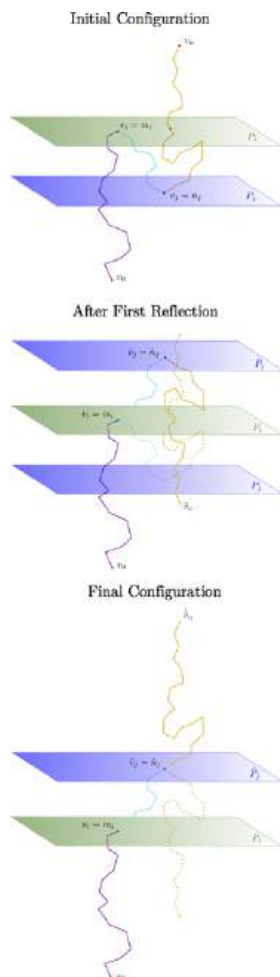


Fig. 11.7: Here we have a typical double reflection move for part two of the proof. First, we identify the local maximum, m_i , and local minimum, m_j such that m_i is the highest maximum and n_j is the lowest minimum, relative to the axis selected at the beginning of this step. The first reflection through the plane P_i reflects the segments between m_i and v_n . The reflection through \hat{P}_j , the reflected image of P_j , reflects the segments between \hat{n}_j and \hat{v}_n . This gives us our new configuration, with two fewer critical points and with the same (or greater) thickness.

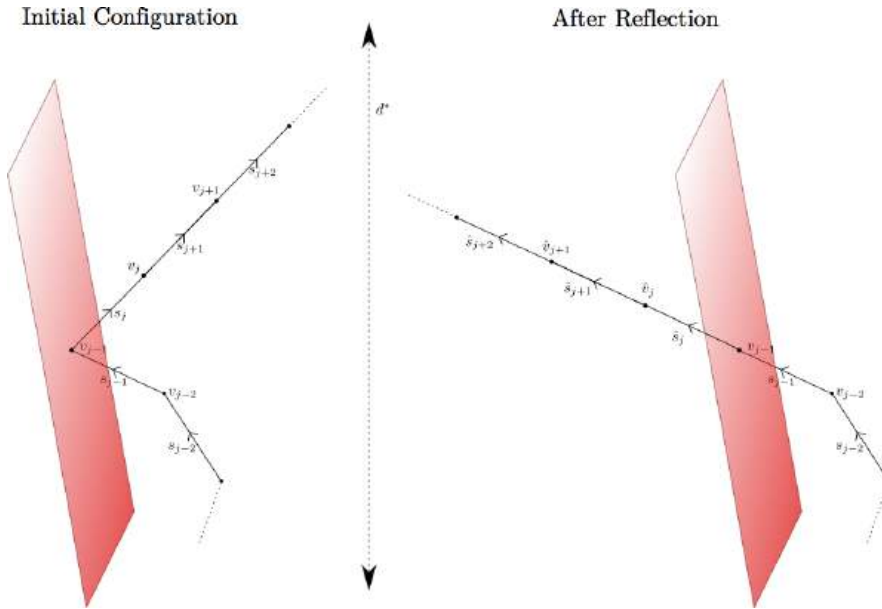


Fig. 11.8: In the previous step, we eliminated pairs of maxima and minima relative to d^* , the diameter of the convex hull in the first step. Now, via a single reflection reflection, we straighten the segments s_{j+1}, \dots, s_n to be in line with s_j .

to last vertex through planes containing the last edge, then w is left fixed by this double reflection. Therefore there is a double reflection move r with $r(w) = w \notin C_{i+1}$. Since C_{i+1} is closed, there is an open neighborhood of w which is not in C_{i+1} so there is a positive probability of landing in this open neighborhood of w . This contradicts the requirements of periodicity. Thus, F is aperiodic. \square

Theorem 11.3.1. The Markov chain on walks in $\mathcal{W}(n, t)$, as described above, is ergodic.

Proof. From our work showing transitivity, we have that for any walk there is a sequence of finitely many single and double reflections which can transform this walk to the straight walk, and each single and double reflection in this sequence is on the interior of a smooth section of F . Then, by concatenating this sequence with the reverse of another such sequence, we can connect any two walks on the interior of $\mathcal{W}(n, t)$ with a sequence of reflections on the interior of a smooth section of F . Therefore F creates a forward accessible Markov chain. Next, since our noise parameter M has a probability density function, namely a constant, we can combine this with forward accessibility to get that the Markov chain is a T-chain. We get that any sequence of probability distributions is tight, so in particular our Markov chain F is bounded in probability on average because our state space is compact. Because there is a sequence of reflections ensur-

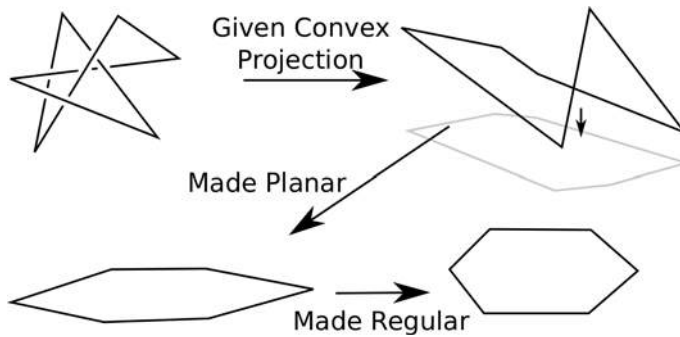


Fig. 11.9: An example of the three aspects being applied to a hexagon

ing that every walk can be modified to be within any neighborhood of the straight walk with positive probability, we conclude that F has a reachable state. Therefore F , being a T -chain with a reachable state which is bounded in probability on average, must be positive Harris recurrent. Finally, since there is a fixed point at the interior of a smooth section, F cannot be periodic and so the aperiodic ergodicity theorem [14] tells us that F is ergodic. \square

11.3.3 Closed polygons: Chapman algorithm

We work in the space $\mathcal{R}(n, t)$ of polygonal rings with n unit length edges and a thickness of greater than or equal to t . We show that any knot $K \in \mathcal{R}(n, t)$ has a sequence of polygonal folds which brings it to the regular planar polygon P_n , without having the thickness be less than t . Note that this algorithm unknots the polygon and, therefore, may not preserve the topological knot type of the polygon. The overall structure for making these choices has 3 aspects. The first aspect is being able to expand the knot until its orthogonal projection to the $x-y$ plane is a convex projection, without changing the number of minimal height vertices, with respect to the orthogonal projection. The second aspect is being able to take a knot with convex projection and increase the number of minimal height vertices, with respect to the orthogonal projection. These two actions are combined and alternated so as to allow one to make the knot planar and convex. The third aspect is to take a planar and convex polygon and make it regular.

A key piece for the first aspect, making the projection convex, is the fact that a reflection move across a plane which does not intersect the interior of the convex hull, does not increase the thickness. This is a special example of the fact that if no two vertices are made closer together, then the thickness is not increased. For making the projection convex, we utilize the argument of the Grünbaum-Zaks theorem [7]. To properly state this theorem we use need a weakened notion of convexity called exposed which

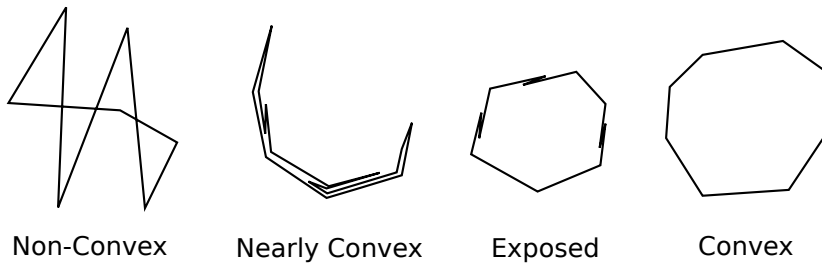


Fig. 11.10: Examples of the types of convexity. Overlapping edges are separated for clarity.

is shown in Figure 11.10. While a polygon in the plane is convex if it is an embedding onto the boundary of the convex hull, it is exposed if the image is the boundary of the convex hull and the pre-image of each vertex is connected. This means that the polygon might double back over an edge, but still cycles around the convex hull. The Grünbaum-Zaks theorem says that any polygon can be made exposed using a finite sequence of reflections across edges of the convex hull.

The Grünbaum-Zaks theorem allows us to make the projection exposed, so we need to make the projection convex. For that, we consider an edge of the convex hull which is not hit injectively. Taking the pre-image of that edge we get a strip, with an arc of the knot moving from one side of the strip to the other, as in Figure 11.11. Looking at the convex hull of this arc we can find a sub arc which has end points which share a bottom edge of that convex hull. This sub arc can be pushed out slightly using a rotation about the line labeled l through a small angle, which can be done with a pair of reflections. This makes the projection more injective and so alternating with Grünbaum-Zaks theorem completes the first aspect.

The second aspect is much simpler. Since the knot is not planar, we can find a sub arc which is above the minimum height except at the endpoints which are minimal height vertices. This arc can be rotated down until another vertex has reached minimal height, as in Figure 11.12.

The final aspect uses a special move which consists of six reflections. Within the planar convex polygon, we choose an inscribed quadrilateral, which has two opposite corners which correspond to vertices at which the interior angles are smaller than the regular angle, and the complementary two corners correspond to vertices at which the interior angles are larger than regular the regular angle. Focusing on the inscribed quadrilateral, two reflections can give a new planar quadrilateral with different angles. This leaves four flaps in the form of the arcs connecting the vertices of the inscribed quadrilateral, and so an additional four reflections will make the whole knot planar. Intermediate value theorem allows this collection of six reflections to be chosen so one additional angle is regular.

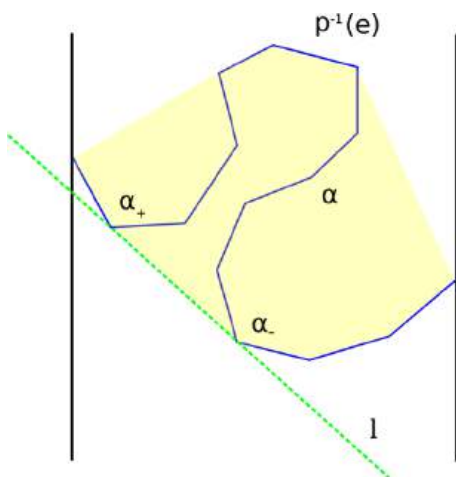


Fig. 11.11: An example of the pre-image of an edge of the convex hull. α denotes a vertex which is strictly above another point in the arc. α_{\pm} denote the vertices in either direction which are on the bottom of the convex hull of the arc. l is the boundary line of this convex hull shared by α_{\pm} .

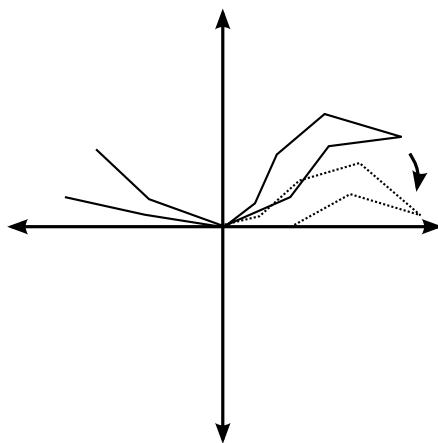


Fig. 11.12: An example of a rotation which increases the number of minimum height vertices, shown from a projection which is parallel to the axis of rotation.

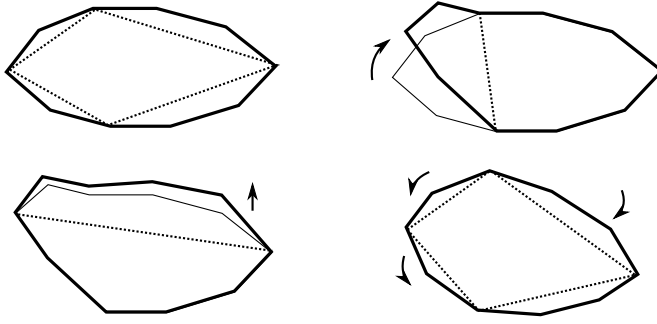


Fig. 11.13: An example of applying the six reflection move to a decagon. The initial diagram has four distinguished vertices connected by dotted lines.

We now let F denote the Markov chain arising from applying sequences of 6 reflection moves at a time to a closed polygon x in the state space $X = \mathcal{R}(n, t)$ of n edge rings with thickness at least t .

Lemma 11.3.4. The Markov chain F is ergodic

The algorithm we described shows that any point x can be brought to the regular planar polygon with a finite sequence of moves, so by using this and its reverse, we can connect any two points in the state space. This means that we can utilize the same argument for $\mathcal{R}(n, t)$ as for $\mathcal{W}(n, t)$. The algorithm shows forward accessibility and a reachable point. This lets us claim that it is a T -chain. Since it is a compact state space we get an invariant probability measure. Finally, since there exists moves which take a point to itself, the Markov chain is aperiodic, so the aperiodic ergodicity theorem applies.

11.4 Discussion and Conclusions

Considering open polygons, we generated a data set consisting of 5000 samples for each of the following pairs of lengths and thicknesses:

- lengths, N : 100 to 1000 in steps of 100.
- thicknesses: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1 (Note, these correspond to minimal interior angles of about 0° , 23° , 44° , 62° , 77° , 90° , 100° , 109° , 116° , 122° , and 127° .)

Even for the thickest and longest of these samples, the acceptance rate for a proposed reflection or double reflection was above 25%. From this data we examined

squared radius of gyration, squared end to end distance, and for the length 300 samples, knotting in the open walks [15].

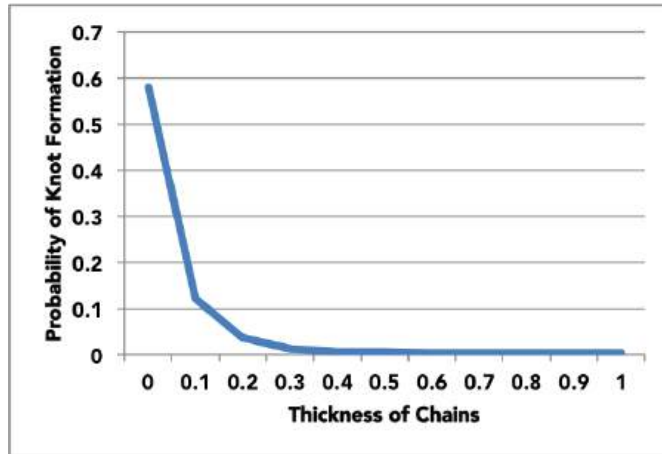


Fig. 11.14: Knot probability as a function of thickness for self avoiding walks of length 300. For each 0.1 thickness is increased, the probability of knot formation is cut by 50% or more, up to $t = 0.5$. For $t > 0.5$, the probability of knot formation is constant.

As we see in Figure 11.14, the probability of knot formation for walks with length 300 is quite high without the introduction of thickness, with more than half of samples having a dominant, nontrivial knot type, i.e. a specific knot type appearing more than half the time. For each 0.1 (to be compared to the length of each segment, 1,) that the thickness is increased, the probability of knot formation is cut in half until the thickness is at 0.5 after which this probability is constant. We conclude that the introduction of thickness has an immediate and profound effect on knot formation, and we expect that further studies will confirm that this effect persists to longer length scales and in both thicker and thinner walks.

We also analyzed the growth exponent for the thicknesses from 0.0 to 1.0 by finding the average radius of gyration squared, RG_n^2 , for each data set, and performing a linear regression with vertical offsets on the log of the data. Assuming that $RG_n^2 \propto N^\nu$, this allowed us to solve for ν as a function of thickness.

For ideal chains it has been shown that $RG_N^2 \propto N$, where N is the number of segments. It is easy to verify that for $t = \infty$ we have the straight configuration and $RG_N^2 \propto N^2$ [22]. Excluded volume has been classically characterized by the estimate $\nu = 1.2$ [5]. In wet experiments and numerical simulation for good solvent polymers, ν has been estimated between 1.1 and 1.2, with $2(0.588) = 1.18$ in walks on the simple, cubic lattice [5]. We can see that there is an immediate impact of thickness: the scal-

ing exponent increases to the $[1.14, 1.17]$ interval almost immediately, as predicted by Vologodskii's simulations of very thin walks and rings [9].

For comparison, we will refer to ν_t as the scaling exponent for the self avoiding walk with thickness t , and μ_t as the scaling exponent for the family of walks with only a bending constraint for adjacent edges, and no restrictions equivalent to excluded volume.

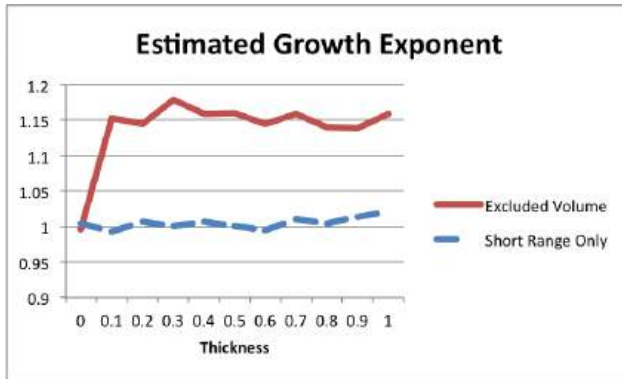


Fig. 11.15: We analyzed how the squared radius of gyration scaled as a function of N . Assuming that $RG_N^2 \propto N^\nu$, we solved for ν and plotted it as a function of thickness for self avoiding walks (red, solid). The exponent, as expected for the ideal case, was close to 1 for thickness 0.0 and increased to 1.16 where it was fairly stable (± 0.01) for thicknesses $t \in [0.2, 1.0]$.

For those walks without long range or excluded volume interaction the growth exponent is, as expected, close to 1 as these walks should scale like ideal chains. The thick walks behave very differently. The growth exponents for families of walks with very small thicknesses immediately jumps to 1.16 where it was fairly stable (± 0.01) for thicknesses $t \in [0.2, 1.0]$. Therefore the effect of excluded volume is as immediate and profound on growth exponent as it is for knot formation: even the introduction of very modest thicknesses has a profound effect on shape, size and knotting.

Bibliography

- [1] Sotero Alvarado, Jorge Alberto Calvo, and Kenneth C. Millett. The generation of random equilateral polygons. *J. Stat. Phys.*, 143(1):102–138, 2011. Supplementary material available online.
- [2] Jason Cantarella, Tetsuo Deguchi, and Clayton Shonkwiler. Probability theory of random polygons from the quaternionic viewpoint. *Comm. Pure Appl. Math.*, 67(10):1658–1699, 2014.

- [3] Jason Cantarella, Alexander Y. Grosberg, Robert Kusner, and Clayton Shonkwiler. The expected total curvature of random polygons. *Amer. J. Math.*, 137(2):411–438, 2015.
- [4] Kyle Leland Chapman. *An Ergodic Algorithm for Generating Random Knots with a Thickness Constraint*. PhD thesis, University of California, Santa Barbara, 2015.
- [5] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Clarendon Press, Oxford, 1986.
- [6] Paul J Flory. *Principles of polymer chemistry*. Cornell University Press, 1953.
- [7] Branko Grünbaum and Joseph Zaks. Convexification of polygons by flips and by flipturns. *Discrete Math.*, 241(1-3):333–342, 2001. Selected papers in honor of Helge Tverberg.
- [8] Jean-Claude Hausmann and Allen Knutson. Polygon spaces and Grassmannians. *Enseign. Math. (2)*, 43(1-2):173–198, 1997.
- [9] K.V. Klenin, A.V. Vologodskii, V.V. Anshelevich, A.M. Dykhne, and M.D. Frank-Kamenetskii. Effect of excluded volume on topological properties of circular DNA. *Journal of Biomolecular Structure and Dynamics*, 5(6):1173–1185, 1988.
- [10] Werner Kuhn. über die gestalt fadenförmiger moleküle in lösungen. *Kolloid-Zeitschrift*, 68(1):2–15, 1934.
- [11] M. Lal. Monte Carlo computer simulation of chain molecules. *Molecular physics*, 17(1):57–64, 1969.
- [12] R. A. Litherland, J. Simon, O. Durumeric, and E. Rawdon. Thickness of knots. *Topology Appl.*, 91(3):233–244, 1999.
- [13] N. Madras and A.D. Sokal. The pivot algorithm: a highly efficient Monte Carlo method for the self-avoiding walk. *Journal of Statistical Physics*, 50(1):109–186, 1988.
- [14] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [15] Kenneth Millett, Akos Dobay, and Andrzej Stasiak. Linear random knots and their scaling behavior. *Macromolecules*, 38(2):601–606, 2005.
- [16] Kenneth C. Millett. Knotting of regular polygons in 3-space. *J. Knot Theory Ramifications*, 3(3):263–278, 1994. Random knotting and linking (Vancouver, BC, 1993).
- [17] Kenneth C. Millett. Monte Carlo explorations of polygonal knot spaces. In *Knots in Hellas '98 (Delphi)*, volume 24 of *Ser. Knots Everything*, pages 306–334. World Sci. Publ., River Edge, NJ, 2000.
- [18] Laura Plunkett (nee Zirbel) and Kyle Chapman. Off-lattice random walks with excluded volume: a new method of generation, proof of ergodicity and numerical results. *Journal of Physics A: Mathematical and Theoretical*, 49(13):135203, 2016.
- [19] Laura Plunkett. *An Analysis of Shape, Scaling and Knotting in Polymer Models With and Without Excluded Volume*. PhD thesis, University of California, Santa Barbara, 2013.
- [20] Eric J. Rawdon. Approximating smooth thickness. *J. Knot Theory Ramifications*, 9(1):113–145, 2000.

- [21] M. Rubinstein and R. Colby. *Polymer Physics*. Oxford University Press, 2003.
- [22] L. Zirbel and K.C. Millett. Characteristics of shape and knotting in ideal rings. *Journal of Physics A: Mathematical and Theoretical*, 45(22):225001, 2012.

Eugene L. Starostin and Gert H.M. van der Heijden

Equilibria of elastic cable knots and links

Abstract: We present a theory for equilibria of geometrically exact braids made of two thin, uniform, homogeneous, isotropic, initially-straight, inextensible and unshearable elastic rods of circular cross-section. We formulate a second-order variational problem for an action functional whose Euler–Lagrange equations, partly in Euler–Poincaré form, yield a compact system of ODEs for which we define boundary-value problems for braids closed into knots or links. The purpose of the chapter is to present a pathway of deformations leading to braids with a knotted axis, thereby offering a way to systematically compute elastic cable knots and links. A representative bifurcation diagram and selected numerical solutions illustrate our approach.

Keywords: elastic knots and links, cable knots, equilibria, variational problem, bifurcation

MSC: 74K10, 74G60, 65L10


12.1 Introduction

Contact problems in the theory of elasticity have seen a surge of interest recently. They are challenging because of the one-sided constraints they introduce. Determining the topology and geometry of the contact set given the boundary conditions is probably the hardest part of any contact problem. Although progress has been made in some of the 1D continua problems [8], [5], [13], we are still far from a full understanding of the solutions of the Euler–Lagrange equations for general contact problems. The theory of elastic braids in [12] deals with a special type of (self-)contact problems of elastic rods involving only equality constraints.

In this work we focus our attention on closed braids whose axes make a torus knot, i.e., so-called cable knots or links [2]. Such knots or links form naturally in (virtual) experiments with elastic torus knots/links if they are made to buckle under the insertion of twist and then allowed to go through a (multiple) self-intersection (the simplest example being the passage through a figure-8 with a single self-intersection). We do not intend to present in this chapter a global classification of all equilibria of knotted 2-braids; rather we explain a particular scenario of how they appear in a typical bifurcation diagram as a control parameter (such as the twisting moment) is varied. We make no assumptions on either the topology or the geometry of the contact set ex-

Eugene L. Starostin, Gert H.M. van der Heijden, Department of Civil, Environmental & Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom, E-mails: e.starostin@ucl.ac.uk, g.heijden@ucl.ac.uk

<https://doi.org/10.1515/9783110571493-011>

Open Access.  © 2018 Eugene L. Starostin and Gert H.M. van der Heijden, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

cept that it is closed on itself. We obtain the shape of the contact curve as part of the solution.

In Section 2 we first recall the essential equations of the elastic 2-braid theory [12] specialised to the present case of $(n, 2)$ -cable knots/links. Then, in Section 3, we formulate boundary conditions that smoothly seal the ends of the rods so that a knot or a link is formed and discuss a bifurcation diagram and numerical solutions before closing this study, in Section 4, with some concluding remarks.

12.2 Theory of elastic braids made of two equidistant strands

12.2.1 Equidistant curves, reference frames and strains

Consider a pair of smooth curves $\mathbf{r}_1(s) \in \mathbb{R}^3$, $s \in [0, L_1]$, $\mathbf{r}_2(\sigma) \in \mathbb{R}^3$, $\sigma \in [0, L_2]$, both parametrised by arclength, that serve as centrelines of two inextensible and unshearable elastic rods of length $L := L_1$ and L_2 , respectively. We denote by $\mathbf{t}_1(s) = \frac{d\mathbf{r}_1(s)}{ds}$ and $\mathbf{t}_2(\sigma) = \frac{d\mathbf{r}_2(\sigma)}{d\sigma}$ the unit tangents to each of the curves. We define the point-to-point squared distance function $D_2(s, \sigma) = \boldsymbol{\rho}^2(s, \sigma)$, where $\boldsymbol{\rho}(s, \sigma) = \mathbf{r}_2(\sigma) - \mathbf{r}_1(s)$ is a chord vector connecting the two curves. We assume that there is a one-to-one mapping $[0, L_1] \leftrightarrow [0, L_2] : s \leftrightarrow \sigma(s)$ between the two curves such that D_2 is bi-critical at corresponding points, i.e., $\left. \frac{\partial D_2}{\partial s}(s, \sigma) \right|_{\sigma=\sigma(s)} = -2\boldsymbol{\rho}(s, \sigma(s)) \cdot \mathbf{t}_1(s) = 0$ and $\left. \frac{\partial D_2}{\partial \sigma}(s, \sigma) \right|_{\sigma=\sigma(s)} = 2\boldsymbol{\rho}(s, \sigma(s)) \cdot \mathbf{t}_2(\sigma(s)) = 0$. Then $D_2(s, \sigma(s))$ is constant, say equal to Δ^2 , and the two curves are said to be at constant distance Δ . Moreover the chords $\boldsymbol{\rho}(s, \sigma(s))$ are orthogonal to the curve tangents at both ends. Because of the one-to-one mapping there exists a common parametrisation for both curves. Taking arclength along the first curve as the common parameter, we can introduce the unit chord vector $\mathbf{d}_1(s) = \frac{1}{\Delta}\boldsymbol{\rho}(s, \sigma(s)) = \frac{1}{\Delta}[\mathbf{r}_2(\sigma(s)) - \mathbf{r}_1(s)]$. We shall henceforth write expressions like $\mathbf{r}_2(s)$ instead of $\mathbf{r}_2(\sigma(s))$.

Along each curve we define two moving frames as follows (Fig. 12.1). We define $\mathbf{u}_1 := \mathbf{t}_1 \times \mathbf{d}_1$ so that the vectors \mathbf{t}_1 , \mathbf{d}_1 and \mathbf{u}_1 form an orthonormal frame that we call the first braid frame. We also define an orthonormal material frame on the first rod: $\{\mathbf{t}_1, \mathbf{d}_0, \mathbf{v}_1\}$ (\mathbf{d}_0 pointing in an arbitrary normal direction). The two frames differ by a rotation about \mathbf{t}_1 through an angle ξ_1 measured from \mathbf{d}_0 to \mathbf{d}_1 . Correspondingly, we define two moving orthonormal frames on the second curve \mathbf{r}_2 . The first one (called the second braid frame) is made up of the vectors \mathbf{t}_2 , \mathbf{d}_1 and $\mathbf{u}_2 := \mathbf{t}_2 \times \mathbf{d}_1$, while the second one (called the second material frame) is $\{\mathbf{t}_2, \mathbf{d}_2, \mathbf{v}_2\}$. They differ by a rotation about the tangent vector \mathbf{t}_2 through an angle ξ_2 , measured from \mathbf{d}_1 to \mathbf{d}_2 .

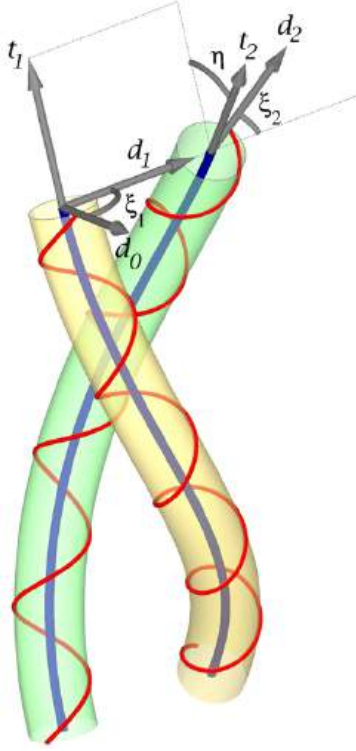


Fig. 12.1: A braid made of two rods. Centrelines are drawn as thick blue curves while red helices show how the material twists. For clarity the rods are shown with smaller diameter.

After choosing a coordinate system we may identify the orientations of the above four reference frames with elements of the group of orthogonal 3×3 matrices:

$$R_{ij}(s) := (\mathbf{t}_j(s), \mathbf{d}_i(s), \boldsymbol{\beta}_{ij}(s)) \in SO(3),$$

$$\boldsymbol{\beta}_{ij}(s) = \mathbf{t}_j(s) \times \mathbf{d}_i(s) = i(2 - i)\mathbf{u}_j(s) + |i - 1|\mathbf{v}_j(s),$$

where $i \in \{0, 1, 2\}$, $j \in \{1, 2\}$, $ij \neq 02, 21$. These define four skew-symmetric 3×3 matrices in the Lie algebra $\mathfrak{so}(3)$ as follows:

$$\hat{\omega} = R_{11}^T R'_{11}, \quad \hat{\Omega} = R_{12}^T R'_{12}, \quad \hat{\tilde{\omega}} = R_{01}^T R'_{01}, \quad \hat{\tilde{\Omega}} = R_{22}^T R'_{22}, \quad (12.2.1)$$

where we have introduced the ‘hat’ isomorphism between skew-symmetric matrices

$$\hat{w} = \begin{pmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{pmatrix} \text{ in } \mathfrak{so}(3) \text{ and axial (rotation) vectors } w = (w_1, w_2, w_3)^T \text{ in}$$

\mathbb{R}^3 .^{12.1} Here and in the following a prime denotes differentiation with respect to s . Thus we have defined four axial vectors $\omega = (\omega_1, \omega_2, \omega_3)^\top$, $\Omega = (\Omega_1, \Omega_2, \Omega_3)^\top$, $\tilde{\omega} = (\tilde{\omega}_1, \tilde{\omega}_2, \tilde{\omega}_3)^\top$ and $\tilde{\Omega} = (\tilde{\Omega}_1, \tilde{\Omega}_2, \tilde{\Omega}_3)^\top$, which are the rotation vectors of, respectively, the braid frames $\{\mathbf{t}_1, \mathbf{d}_1, \mathbf{u}_1\}$, $\{\mathbf{t}_2, \mathbf{d}_1, \mathbf{u}_2\}$ and the material frames $\{\mathbf{t}_1, \mathbf{d}_0, \mathbf{v}_1\}$ and $\{\mathbf{t}_2, \mathbf{d}_2, \mathbf{v}_2\}$.

The orthonormal frames form a sequence under consecutive rotations about \mathbf{t}_1 , \mathbf{d}_1 and \mathbf{t}_2 . Thus

$$R_{11} = R_{01}R_1(\xi_1), \quad R_{12} = R_{11}R_2(\eta), \quad R_{22} = R_{12}R_1(\xi_2), \quad (12.2.2)$$

where

$$R_1(\xi_i) = \exp(\xi_i \hat{e}_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \xi_i & -\sin \xi_i \\ 0 & \sin \xi_i & \cos \xi_i \end{pmatrix} \in SO(3), \quad i = 1, 2, \quad e_1 = (1, 0, 0)^\top,$$

$$R_2(\eta) = \exp(\eta \hat{e}_2) = \begin{pmatrix} \cos \eta & 0 & \sin \eta \\ 0 & 1 & 0 \\ -\sin \eta & 0 & \cos \eta \end{pmatrix} \in SO(3), \quad e_2 = (0, 1, 0)^\top,$$

and η is the angle, about \mathbf{d}_1 , from the first tangent, \mathbf{t}_1 , to the second, \mathbf{t}_2 (see Fig. 12.1). From Eqs. (12.2.1), (12.2.2) it follows that the rotation vectors of the material and braid frames are related as

$$\hat{\omega} = R_1^\top(\xi_1)\hat{\tilde{\omega}}R_1(\xi_1) + R_1^\top(\xi_1)R_1'(\xi_1), \quad (12.2.3)$$

$$\hat{\Omega} = R_2^\top(\eta)\hat{\omega}R_2(\eta) + R_2^\top(\eta)R_2'(\eta), \quad (12.2.4)$$

$$\hat{\tilde{\Omega}} = R_1^\top(\xi_2)\hat{\Omega}R_1(\xi_2) + R_1^\top(\xi_2)R_1'(\xi_2). \quad (12.2.5)$$

Owing to the inextensibility condition $|\frac{d\mathbf{r}_2}{d\sigma}| = 1$, the arclength parameter σ along the second centreline satisfies

$$\sigma' = \sqrt{(\Delta\omega_1)^2 + (\Delta\omega_3 - 1)^2}. \quad (12.2.6)$$

We express the angle η as a function of the components of ω as

$$\sin \eta = -\Delta\omega_1/\sigma', \quad \cos \eta = (1 - \Delta\omega_3)/\sigma', \quad (12.2.7)$$

and hence

$$\omega_1 = \left(\omega_3 - \frac{1}{\Delta} \right) \tan \eta. \quad (12.2.8)$$

12.1 Throughout we adopt the notation that for any vector $\mathbf{v} \in \mathbb{R}^3$ the sans-serif symbol \mathbf{v} denotes the triple of components $(v_1, v_2, v_3)^\top = (\mathbf{v} \cdot \mathbf{t}_1, \mathbf{v} \cdot \mathbf{d}_1, \mathbf{v} \cdot \mathbf{u}_1)^\top$ in the first braid frame. An exception is however made for sans-serif rotation vectors, which are always triples of components in their corresponding frames.

Eqs. (12.2.3)–(12.2.5) allow us to express the strand strains $(\tilde{\omega}_1, \tilde{\omega}_2, \tilde{\omega}_3, \tilde{\Omega}_1, \tilde{\Omega}_2, \tilde{\Omega}_3)$ in terms of the variables $(\omega_1, \omega_2, \omega_3, \xi_1, \xi_2)$, which we shall call the braid strains. In much of the following it will be convenient to use Eq. (12.2.8) to eliminate ω_1 in favour of the angle η , which also has an intuitive physical meaning as braid angle. With either choice of variables we have an unconstrained description in which, by construction, the constant-distance constraint is automatically satisfied.

We note that if $\omega_1 = 0$ and $\omega_3 \neq 1/\Delta$, then $\eta = 0$ and the two strands are parallel (this condition does not restrict in any way the shape of the first centreline). When $\omega_3 = 1/\Delta$, and $\omega_1 \neq 0$, the second strand becomes orthogonal to the first one ($\eta = \pm\pi/2$). If both $\omega_1 = 0$ and $\omega_3 = 1/\Delta$ then $\sigma' = 0$, i.e., the induced parametrisation of the second centreline is singular. The angle η is then not defined. We need to rule out this case.

For a non-selfintersecting braid, tubes of radius $\Delta/2$ around each of the centrelines must not overlap. Locally this means that points connected by the bicritical chords must be closest-approach points of the centrelines. It may be shown that this requires $\eta \in (-\pi/2, \pi/2)$ [12].

12.2.2 Equations for the standard 2-braid

For the elastic energy we make the usual assumption of frame indifference, i.e., the energy is invariant under Euclidean motions. It will then depend only on the strains (and possibly arclength) and not, for instance, on the centreline \mathbf{r}_1 . We have already assumed both rods in the braid to be inextensible and unshearable, so there is no elastic energy on account of stretches. Apart from this restriction we allow for arbitrary hyperelastic rods [1] and write

$$U_1 = \int_0^{L_1} f_1(\tilde{\omega}) \, ds \quad \text{and} \quad U_2 = \int_0^{L_2} f_2(\tilde{\Omega}(\sigma)) \, d\sigma = \int_0^{L_1} f_2(\tilde{\Omega}/\sigma') \sigma' \, ds \quad (12.2.9)$$

for the elastic strain energy of the first and second rod, respectively, where f_1 and f_2 are the strand strain energy densities, and the explicit argument of $\tilde{\Omega}$ in the σ integral emphasises that $\tilde{\Omega}$ is to be regarded as a function of arclength σ of the second rod in this integral.

Assuming fixed ends of the first strand, we add $-\mathbf{F} \cdot [\mathbf{r}_1(L) - \mathbf{r}_1(0)] = -\int_0^L \mathbf{F} \cdot \mathbf{t}_1 \, ds$ as an isoperimetric constraint expression with \mathbf{F} a (constant) Lagrange multiplier (to become the internal force in the braid). Also, by inextensibility of the second rod we require the length of the second strand

$$L_2 = \int_0^{L_2} d\sigma = \int_0^L \sigma' \, ds =: \int_0^L f_\sigma(\omega_1, \omega_3) \, ds \quad (12.2.10)$$

to be constant, giving a second isoperimetric constraint.

We can express the arguments of the integrands in Eqs. (12.2.9) and (12.2.10) in terms of the braid strains (ω, ξ_1, ξ_2) and their derivatives by using Eqs. (12.2.3)–(12.2.5) and (12.2.6) and thus formulate a second-order variational problem for the reduced functional $l: 2\mathfrak{so}(3) \times \mathbb{R}^3 \times TS^1 \times TS^1 \rightarrow \mathbb{R}$, $l(\omega, \omega', F, \xi_1, \xi_1', \xi_2, \xi_2') = f_1(\tilde{\omega}) + f_2(\tilde{\Omega}(\sigma)) + h\sigma' - F \cdot t_1$,

$$\delta \int_0^L l(\omega, \omega', F, \xi_1, \xi_1', \xi_2, \xi_2') ds = 0, \quad (12.2.11)$$

for variations that keep the end positions and orientations of the rods fixed; h is a (constant) Lagrange multiplier.

The Euler–Lagrange equations are derived partly in Euler–Poincaré form and comprise

- (a) balance equations for the components of the overall braid force $F = (F_1, F_2, F_3)^\top$ and braid moment $M = (M_1, M_2, M_3)^\top$ expressed in the first braid frame [12]

$$F' + \omega \times F = 0, \quad (12.2.12)$$

$$M' + \omega \times M + t_1 \times F = 0, \quad (12.2.13)$$

- (b) the ‘constitutive’ equations

$$M = \mathcal{E}_\omega(l), \quad (12.2.14)$$

where $\mathcal{E}_\zeta(k) := \frac{\partial k}{\partial \zeta} - \left(\frac{\partial k}{\partial \zeta} \right)'$ is the Euler–Lagrange operator for the variable ζ ,

- (c) the phase equations for the twist angles of the rods

$$\mathcal{E}_{\xi_i}(l) = 0, \quad i = 1, 2. \quad (12.2.15)$$

Equations (12.2.12) and (12.2.13) are the moving-frame (advected) versions of the familiar balance equations $F' = \mathbf{0}$, $M' + r_1' \times F = \mathbf{0}$ expressed in an inertial frame. It follows that $|F|$ and $F \cdot M$ are first integrals.

Equations (12.2.15) can be written as a set of four first-order equations by introducing new variables (the strand torques) T_i , $i = 1, 2$:

$$T_i = \frac{\partial l}{\partial \xi_i'}, \quad T_i' = \frac{\partial l}{\partial \xi_i}. \quad (12.2.16)$$

If we eliminate the variable ω_1 in favour of η by using $\eta = \eta(\omega_1, \omega_3)$ given in Eq. (12.2.7), the reduced density l takes the form

$$l = g(\omega_2, \omega_3, \eta, \eta', F_1, \xi_1, \xi_1', \xi_2, \xi_2'), \quad (12.2.17)$$

while Eq. (12.2.14) transforms into [11]

$$M_1 = \frac{\partial \eta}{\partial \omega_1} \varepsilon_\eta(g) = \frac{\Delta \cos^2 \eta}{\Delta \omega_3 - 1} \left[\frac{\partial g}{\partial \eta} - \frac{d}{ds} \left(\frac{\partial g}{\partial \eta'} \right) \right], \quad (12.2.18)$$

$$M_2 = \frac{\partial g}{\partial \omega_2}, \quad (12.2.19)$$

$$M_3 = \frac{\partial g}{\partial \omega_3} + \frac{\partial \eta}{\partial \omega_3} \varepsilon_\eta(g) = \frac{\partial g}{\partial \omega_3} - \frac{\Delta \sin \eta \cos \eta}{\Delta \omega_3 - 1} \left[\frac{\partial g}{\partial \eta} - \frac{d}{ds} \left(\frac{\partial g}{\partial \eta'} \right) \right] \quad (12.2.20)$$

(cf. the equations for an elastic strip reduced to the centreline in [10]).

We now specialise our theory to the case where the two strands are linearly elastic, uniform, isotropic, initially-straight rods with bending stiffness B_1 and B_2 and torsional stiffness C_1 and C_2 , respectively. We allow the strands to be intrinsically twisted with twist rates ω_{01} and ω_{02} . We call this special case the standard 2-braid. The density g in Eq. (12.2.17) then takes the form

$$g = g_e(\omega_2, \omega_3, \eta, \eta', \xi'_1, \xi'_2) + h(1 - \Delta \omega_3)/\cos \eta - F_1, \quad (12.2.21)$$

with the elastic energy density $g_e (= f_1 + f_2)$, the sum of bending and torsional energy densities of the two strands, given by (using Eqs. (12.2.3)–(12.2.5))

$$\begin{aligned} g_e &= \frac{1}{2} \left[B_1(\tilde{\omega}_2^2 + \tilde{\omega}_3^2) + \frac{B_2}{\sigma'}(\tilde{\Omega}_2^2 + \tilde{\Omega}_3^2) + C_1(\tilde{\omega}_1 - \omega_{01})^2 + \frac{C_2}{\sigma'}(\tilde{\Omega}_1 - \sigma' \omega_{02})^2 \right] = \\ &= \frac{1}{2} \left\{ B_1(\omega_2^2 + \omega_3^2) + B_2 \left[\frac{\cos \eta}{1 - \Delta \omega_3} (\omega_2 + \eta')^2 + \frac{(\omega_3 - \sin^2 \eta / \Delta)^2}{(1 - \Delta \omega_3) \cos \eta} \right] + \right. \\ &\quad \left. + C_1 \left[\left(\omega_3 - \frac{1}{\Delta} \right) \tan \eta - \xi'_1 - \omega_{01} \right]^2 + \right. \\ &\quad \left. + C_2 \left[\frac{\cos \eta}{1 - \Delta \omega_3} \left(\frac{\sin \eta}{\Delta} - \xi'_2 \right)^2 + 2\omega_{02} \left(\frac{\sin \eta}{\Delta} - \xi'_2 \right) + \omega_{02}^2 \frac{1 - \Delta \omega_3}{\cos \eta} \right] \right\}. \quad (12.2.22) \end{aligned}$$

This gives a system of 13 ODEs in $(F_1, F_2, F_3, M_1, M_2, M_3, \eta, \omega_2, \omega_3, \xi_1, T_1, \xi_2, T_2)$: Eqs. (12.2.12), (12.2.13), (12.2.16), (12.2.18), (12.2.19) and (12.2.20). The last three in fact contain two algebraic equations because g does not depend on ω'_2 and ω'_3 , but these can be turned into ODEs by differentiation (and the algebraic equations used to solve for η' and h). Also note that since g does not depend on the phase angles ξ_1 and ξ_2 , the strand torques T_i are constants.

12.2.3 Kinematics equations

Reconstruction of the centreline of the first strand requires solving for the tangent \mathbf{t}_1 and integrating this to get \mathbf{r}_1 . We choose a parametrisation of the first braid frame $\{\mathbf{t}_1, \mathbf{d}_1, \mathbf{u}_1\}$ in terms of four Euler parameters (or quaternions) $\mathbf{q} = (q_0, q_1, q_2, q_3)$ subject

to the normalisation condition $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$ [6] and write

$$\begin{aligned} \mathbf{t}_1 &= \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 \\ 2q_1q_2 + 2q_0q_3 \\ 2q_1q_3 - 2q_0q_2 \end{pmatrix}, & \mathbf{d}_1 &= \begin{pmatrix} 2q_1q_2 - 2q_0q_3 \\ q_0^2 - q_1^2 + q_2^2 - q_3^2 \\ 2q_2q_3 + 2q_0q_1 \end{pmatrix}, \\ & & \mathbf{u}_1 &= \begin{pmatrix} 2q_1q_3 + 2q_0q_2 \\ 2q_2q_3 - 2q_0q_1 \\ q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix}. \end{aligned}$$

These Euler parameters, unlike Euler angles, give a singularity-free description of arbitrary rotations in space and are therefore convenient for numerical computations. The kinematics equations are obtained by differentiating the above vectors and using Eq. (12.2.1):

$$\frac{d}{ds} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & -\omega_1 & -\omega_2 & -\omega_3 \\ \omega_1 & 0 & \omega_3 & -\omega_2 \\ \omega_2 & -\omega_3 & 0 & \omega_1 \\ \omega_3 & \omega_2 & -\omega_1 & 0 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}, \quad (12.2.23)$$

where we recall from Eq. (12.2.8) that $\omega_1 = (\omega_3 - \frac{1}{\Delta}) \tan \eta$. It is straightforward to verify that $\mathbf{q} \cdot \mathbf{q}$ is a first integral of this system of equations. To find the centreline \mathbf{r}_1 we solve Eq. (12.2.23) in conjunction with the equation $\mathbf{r}'_1 = \mathbf{t}_1$. The second centreline is then given by $\mathbf{r}_2 = \mathbf{r}_1 + \Delta \mathbf{d}_1$.

12.3 Numerical solution

In this section we present results obtained by numerically solving boundary-value problems for the standard 2-braid equations developed in Section 2 using the continuation and bifurcation code AUTO [3]. AUTO solves ODEs using collocation and uses pseudo-arclength continuation to advance solutions as a control (or bifurcation) parameter is incremented. The code is also able to detect pitchfork or other bifurcation points along a solution branch where this branch intersects another solution branch and to switch to and compute this second branch. In this work we choose the strand torque as a bifurcation parameter.

Here we consider only the closed braid problem where the two rods form a torus knot or link. We define a linking number to distinguish between topologically different types of such 2-braid solutions by following Fuller's extension of the concept of a linking number to a (not necessarily closed) ribbon [4]. By a ribbon is here meant a curve with a field of normal vectors of constant length. First note that the contact curve of our knot $\frac{1}{2}[\mathbf{r}_1(s) + \mathbf{r}_2(s)] = \mathbf{r}_1(s) + \frac{\Delta}{2}\mathbf{d}_1(s)$, $s \in [0, L]$, is a (smoothly) closed curve and the director $\frac{\Delta}{2}\mathbf{d}_1(s)$ is an orthogonal vector defined for all $s \in [0, L]$. It therefore forms a non-closed ribbon. The linking number of this ribbon is defined as $\mathcal{L}k = \mathcal{T}w + \mathcal{W}r$,

where $\mathcal{W}r$ is the writhing number of the contact curve and $\mathcal{T}w = \frac{1}{2\pi} \int_0^L tw(s) ds$ is the total twist (the twisting number) of the braid with twist rate tw . To have a closed single centreline, $\mathcal{L}k$ has to be semi-integer, $\mathcal{L}k = n/2$, $n \in \mathbb{Z}$. Alternatively, if we think of the ribbon as extended in both $\pm \frac{1}{2}\mathbf{d}_1$ directions, we may say that, for our knots, this ribbon is one-sided (like a Möbius strip [10]) and its edge (given by the centreline of the rod) is a single knotted curve.

We start from an analytical solution (two linked equidistant circular rings) and perform continuations mimicking the cutting, twisting and resealing of the rings to obtain different torus knots (and links) [12] and eventually, after torsionally buckling these knots/links, cable knots/links. The braid axis (contact curve) of this starting solution is an unknot. We compute solutions with different linking numbers for which we derive boundary conditions that ensure smooth closure of the braid into a knot (for semi-integer linking number of the braid) or a link (for integer linking numbers).

In addition to the 13-dimensional system of ODEs derived in Section 12.2.2 we solve the 7 kinematics equations for \mathbf{q} and \mathbf{r}_1 of Section 12.2.3, so as to be able to apply displacement boundary conditions and plot solution shapes, and Eq. (12.2.6) to compute the second arclength. Integration of this last equation allows us to impose the total-length constraint $L_{\text{tot}} := L + L_2 = \text{const.}$, the natural constraint to consider for knot solutions. Since $\mathbf{q} \cdot \mathbf{q}$ is an integral of the kinematics equations, the quaternion normalisation condition $|\mathbf{q}| = 1$ can simply be imposed by choosing boundary conditions for the q_i that satisfy it.

All numerical runs are for elastic parameters $B_1 = B_2 = 1$, $C_1/B_1 = C_2/B_2 = 1/(1 + \nu)$, where Poisson's ratio $\nu = 0.3$, and in all cases we set $\omega_{01} = \omega_{02} = 0$. Given a solution of the 21D system of ODEs we can compute other properties such as the normal contact pressure p or the curvature κ_c of the contact curve (see details in [12]). Here we make the assumption of frictionless hardcore contact in which the only distributed force acting between the two contacting strands is the normal pressure directed along the chord vector \mathbf{d}_1 . First integrals are monitored throughout to keep track of the numerical accuracy of solutions.

12.3.1 Torus knots

12.3.1.1 Boundary conditions

To compute knot solutions of total length L_{tot} we apply the following boundary conditions

$$\begin{array}{ll}
x_1(0) = 0, & (12.3.1) \\
y_1(0) = 0, & (12.3.2) \\
z_1(0) = 0, & (12.3.3) \\
q_0(0) = 0, & (12.3.4) \\
q_1(0) = -1/\sqrt{2}, & (12.3.5) \\
q_2(0) = -1/\sqrt{2}, & (12.3.6) \\
q_3(0) = 0, & (12.3.7) \\
\xi_1(0) = 0, & (12.3.8) \\
\xi_2(0) = 0, & (12.3.9) \\
T_1(0) = t_{10}, & (12.3.10) \\
T_2(0) = t_{20}, & (12.3.11) \\
\sigma(0) = 0, & (12.3.12)
\end{array}
\qquad
\begin{array}{ll}
x_1(L) - x_1(0) = \Delta, & (12.3.13) \\
y_1(L) = 0, & (12.3.14) \\
z_1(L) = 0, & (12.3.15) \\
q_0(L) = q_3(L), & (12.3.16) \\
q_1(L) = -q_2(L), & (12.3.17) \\
4q_0(L)q_1(L) = \sin \eta(L), & (12.3.18) \\
\eta(L) = \eta(0), & (12.3.19) \\
[M_2(0) - B_1\omega_2(0)]/B_2 = -\omega_2(L), & (12.3.20) \\
(1 - \Delta\omega_3(0))/\cos \eta(0) = \cos \eta(L)/(1 - \Delta\omega_3(L)), & (12.3.21) \\
\sigma(L) = L_2, & (12.3.22) \\
L + L_2 = L_{\text{tot}}, & (12.3.23)
\end{array}$$

where t_{10} and t_{20} are constants. Since for a knot the two strands are part of a single closed rod, it only makes good sense to take $t_{20} = -t_{10}$ and also $B_2 = B_1$, $C_2 = C_1$.

Conditions Eqs. (12.3.1)–(12.3.7), (12.3.13)–(12.3.21) ensure smooth closure of the braid into a doubly-covered ring, smooth here meaning continuity up to second derivatives of \mathbf{r}_i , i.e., curvatures. Conditions Eqs. (12.3.13)–(12.3.15) place the end of the first centreline \mathbf{r}_1 at the beginning of the second centreline \mathbf{r}_2 , both strands thus forming a single closed rod. Conditions Eqs. (12.3.16) and (12.3.17) imply $\mathbf{d}_1(L) = -\mathbf{d}_1(0)$. Condition Eq. (12.3.18) guarantees that $\mathbf{t}_1(L)$ makes an angle $\eta(L)$ with $\mathbf{t}_1(0)$ of the same strand. Together with condition Eq. (12.3.19) this ensures matching of the tangents: the tangent at $s = L$ of the first strand aligns with the tangent at $s = 0$ of the second strand, and vice versa. Conditions Eqs. (12.3.19)–(12.3.21) are equivalent to $\Omega_1(0)/\sigma'(0) = \omega_1(L)$, $\Omega_2(0)/\sigma'(0) = -\omega_2(L)$ and $\Omega_3(0)/\sigma'(0) = -\omega_3(L)$, which ensure that the curvatures at the end of the first strand match the curvatures at the beginning of the second strand.

The total number of boundary conditions is 23 and the three free parameters required for solution branches are L , L_2 and t_{10} . Note that the material frames of these knot solutions will in general not be closed. This is not different from similar studies of closed single rods (see, e.g., [9, 7]). Material closure requires $\xi_1(0) + \xi_2(L) = \pi \pmod{2\pi}$ and can be achieved by inserting the right amount of twist by adjusting t_{10} ($= -t_{20}$).

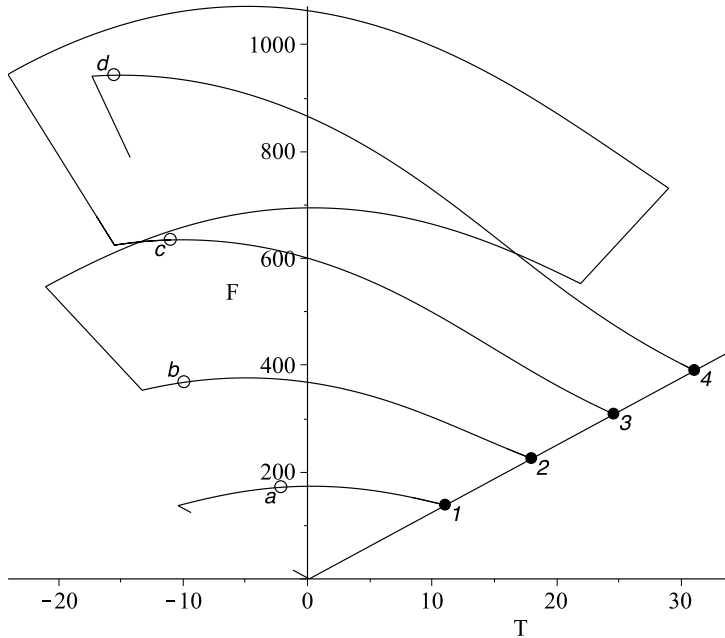


Fig. 12.2: Partial bifurcation diagram for the $(7, 2)$ torus knot with the first four bifurcation points (marked with solid black circles) and bifurcating branches. $\Delta = 0.0051364907419$, $L_{\text{tot}} = 1.98889$. Empty circles mark particular cable knots whose shapes are shown separately in the following figures.

12.3.1.2 Examples of cable knots

Cable knots [2] can be obtained by inserting twist into the torus knots computed using the above boundary conditions thereby forcing them to buckle (bifurcate). The bifurcation diagram in Fig. 12.2 shows how cable knots appear along bifurcating solution branches. The figure presents a (T, F) diagram, where $T := t_{10} = T_1 = -T_2$ is the common strand torque and $F = |\mathbf{F}|$ is the magnitude of the constant braid force as a characteristic solution measure. The diagram is for the (right-handed) $(7, 2)$ torus knot. The trivial branch, representing unbuckled knots with unknotted braid axis, makes an infinite v-shaped curve (only part of this branch, marked 1 – 2 – 3 – 4, is shown in Fig. 12.2). The solutions along the v-branch are D_7 -symmetric with the central axis aligned with the force vector. We shall refer to them as ‘flat’ solutions because in many ways they behave like twisted planar circular isotropic rods (which have linear F - T relationship for branches of planar equilibria [7]). This is despite the fact that neither the contact curve of the braid nor the centreline of the rod is a plane curve.

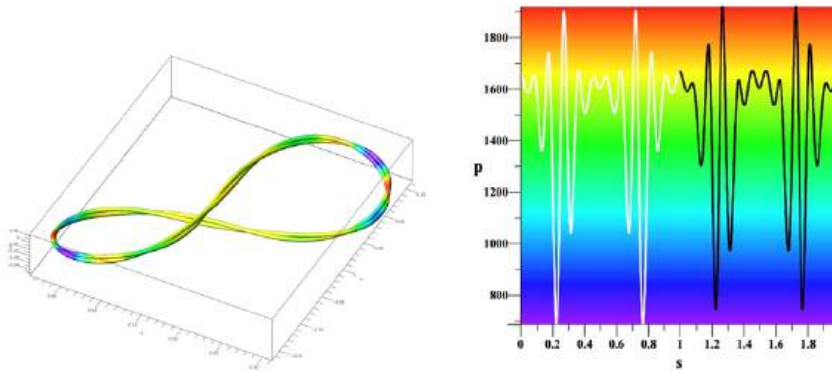


Fig. 12.3: Figure-8-shaped (11, 2) torus knot. $T = -2.10752$, $F = 171.50131$ ((a) in Fig. 12.2). $\Delta = 0.0051364907419$, $L_{\text{tot}} = 1.98889$.

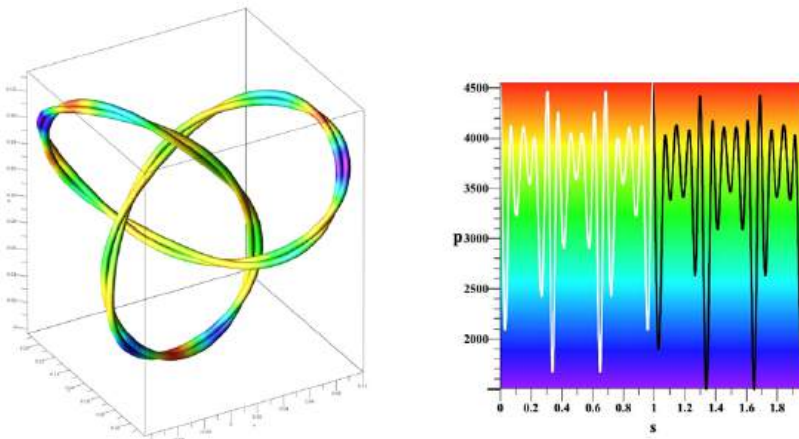


Fig. 12.4: Buckled right-handed cable knot formed by knotting a right-handed (13, 2) torus knot into a right-handed trefoil. $T = -9.89105$, $F = 368.075$ ((b) in Fig. 12.2). $\Delta = 0.0051364907419$, $L_{\text{tot}} = 1.98889$.

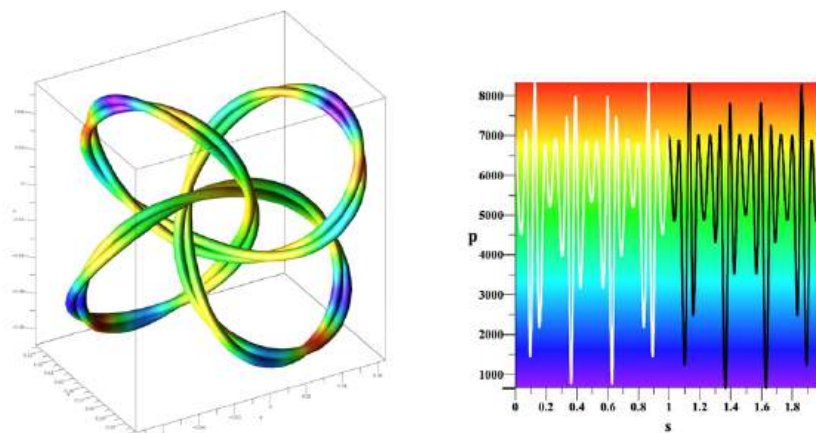


Fig. 12.5: Buckled right-handed cable knot formed by knotting a right-handed $(7, 2)$ torus knot into a right-handed $(4, 3)$ torus knot. $T = -10.97932$, $F = 634.5698$ ((c) in Fig. 12.2). $\Delta = 0.0051364907419$, $L_{\text{tot}} = 1.98889$.

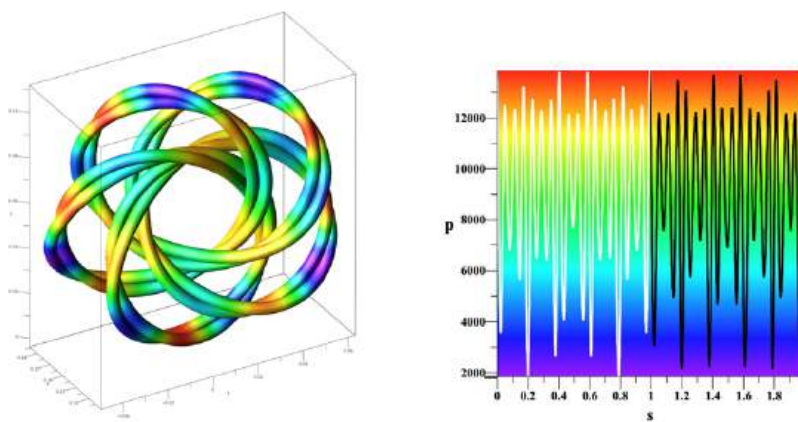


Fig. 12.6: Buckled right-handed cable knot formed by knotting a right-handed $(7, 2)$ torus knot into a right-handed $(5, 4)$ torus knot. $T = -15.51471$, $F = 942.989$ ((d) in Fig. 12.2). $\Delta = 0.0051364907419$, $L_{\text{tot}} = 1.98889$.

There exist critical torque levels T , corresponding to bifurcation points 1, 2, 3, 4, ... in the diagram, where new branches of more warped (buckled) solutions begin. This sequence of buckling modes is again similar to the behaviour of a single closed elastic rod (cf. Fig. 3 in [7]). Figure 12.2 shows branches for the first four bifurcating modes, with the bifurcation points marked by solid black circles. As we move further away from the right side of the v-branch, shapes become more warped and at some point begin to self-overlap. The braid axis exhibits a multiple self-intersection depending on the mode, becoming knotted for modes 2 and higher. Continuing along the branch we come to an interval where the shape becomes free of self-overlapping, see points a , b , c and d in Fig. 12.2 (this may not happen if the rod is too thick). These solutions are cable knots, examples of which are shown in Figs. 12.3–12.6.

The braid axis of the first mode shape in Fig. 12.3, going through a single figure-8-type self-intersection, remains unknotted; we present it here only for completeness of the picture. We see that the right-handed $(7, 2)$ torus knot becomes a $(11, 2)$ torus knot after self-crossing. The linking number of the shape shown is $3.5 + 2 = 5.5$. Each of the curves corresponding to the various modes ends at another v-branch (only parts of which are shown in Fig. 12.2). These represent multi-covered ‘flat’ solutions also having their own bifurcation points. As they are self-overlapping for modes higher than one, we do not explore them further.

All solutions in Figs. 12.3–12.6 have been computed by continuation from the $(7, 2)$ torus knot and have the same total length $L_{\text{tot}} (= 1.98889)$. (The shape in Fig. 12.6 has a small overlapping domain because the chosen value of Δ is just too large.) In particular, the trefoil cable in Fig. 12.4 is obtained by continuation of a right-handed $(7, 2)$ torus knot that after a triple self-crossing becomes a $(13, 2)$ torus knot tied into a right-handed trefoil. Thus, the torus knot (inside the tubular neighbourhood of the companion trefoil knot) has linking number $3.5 + 3 = 6.5$. The corresponding linking numbers for the cable knots in Figs. 12.5 and 12.6 are 7.5 and 8.5, respectively.

The set of all knot solutions breaks down into two components. A disjoint bifurcation diagram, the mirror image of the present diagram under reflection in the F -axis, consists of all solutions connected to the left-handed $(7, 2)$ torus knot (hence, the left-handed trefoil).

The contact pressure profile is drawn in white for $s \in [0, L]$ and in black for $s \in [L, L_{\text{tot}}]$ in Figs. 12.3–12.6. This allows one to see the differences for both contact-ing strands. The profile differences are larger for thicker rods. We observe that for all solutions shown the contact pressure is non-negative and increases with the mode.

12.3.2 Torus links

12.3.2.1 Boundary conditions

To compute link solutions we apply the following boundary conditions:

$$x_1(0) = 0, \quad (12.3.24)$$

$$y_1(0) = 0, \quad (12.3.25)$$

$$z_1(0) = 0, \quad (12.3.26)$$

$$q_0(0) = 0, \quad (12.3.27)$$

$$q_1(0) = -1/\sqrt{2}, \quad (12.3.28)$$

$$q_2(0) = -1/\sqrt{2}, \quad (12.3.29)$$

$$q_3(0) = 0, \quad (12.3.30)$$

$$\xi_1(0) = 0, \quad (12.3.31)$$

$$\xi_2(0) = 0, \quad (12.3.32)$$

$$T_1(0) = t_{10}, \quad (12.3.33)$$

$$T_2(0) = t_{20}, \quad (12.3.34)$$

$$\sigma(0) = 0, \quad (12.3.35)$$

$$x_1(L) = x_1(0), \quad (12.3.36)$$

$$y_1(L) = y_1(0), \quad (12.3.37)$$

$$z_1(L) = z_1(0), \quad (12.3.38)$$

$$q_0(L) = q_3(L), \quad (12.3.39)$$

$$q_1(L) = q_2(L), \quad (12.3.40)$$

$$q_0(L) + q_1(L) = -\sin(\Phi/2 - \pi/4), \quad (12.3.41)$$

$$\omega_2(0) \text{ fixed}, \quad (12.3.42)$$

$$\omega_2(L) = \omega_2(0) \cos \Phi + \omega_3(0) \sin \Phi, \quad (12.3.43)$$

$$\sigma(L) = L_2, \quad (12.3.44)$$

where t_{10} and t_{20} are constants and $\omega_2(0)$ is fixed at whatever value the starting solution has.

The q_i conditions at $s = 0$ imply that \mathbf{t}_1 is aligned with the y axis, \mathbf{d}_1 is aligned with the x axis and \mathbf{u}_1 is aligned with the $-z$ axis. Conditions Eqs. (12.3.39) and (12.3.40) ensure alignment of the tangents $\mathbf{t}_1(0)$ and $\mathbf{t}_1(L)$. Φ is the angle between $\mathbf{d}_1(0)$ and $\mathbf{d}_1(L)$ in the plane normal to these tangents. Closed links have $\Phi = \pm 2\pi n$, where n is a non-negative integer. By using parameter continuation in Φ solutions with different linking number $\mathcal{L}k = \pm n$ may be obtained, positive linking numbers corresponding to right-handed and negative linking numbers corresponding to left-handed links.

This gives a total of 21 conditions for a 21-dimensional system of equations. To compute branches of solutions for a given link we may vary one of the parameters, for example t_{10} or Δ . For equal twisting of the two strands (in the same direction), t_{20} may be taken equal to $-t_{10}$. Note that closure of the material frames of these link solutions requires $\xi_i(L) = \xi_i(0) \pmod{2\pi}$, $i = 1, 2$, which will not in general be satisfied.

12.3.2.2 Examples of cable links

Similar to how we computed cable knots, we first compute a trivial branch of ‘flat’ torus links with unknotted contact curve. Then, following the branches of post-buckled solutions of different modes, we observe knotting of the braid axis by multiple self-intersection, thus forming what we call cable links. Figs. 12.7–12.9 present, respectively, mode-2, -3 and -4 cable links with equal-length components ($L_1 = L_2$). The

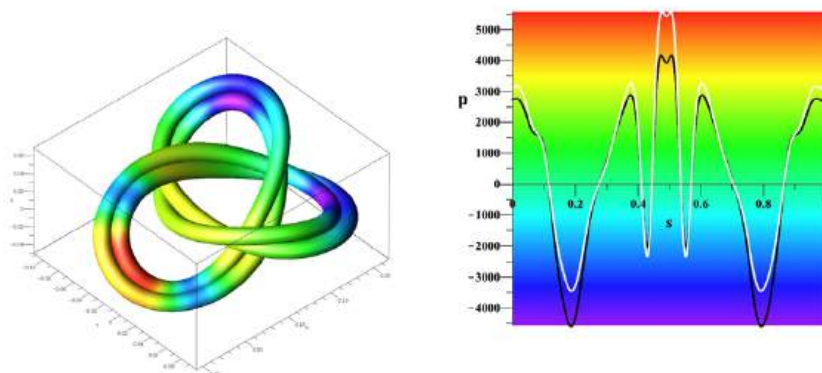


Fig. 12.7: Buckled right-handed cable link formed by knotting a left-handed $(6, 2)$ torus link into a right-handed $(3, 2)$ torus knot. $T_1 = -T_2 = -11.46623$, $F = 364.061$, $\Delta = 0.016016913268$, $L_1 = L_2 = 1$.

shape in Fig. 12.7 is obtained by taking a left-handed ‘flat’ torus link $(6, 2)$ and following the second bifurcating branch. After self-intersection, the axis of the braid becomes a right-handed trefoil (a companion knot). As can be verified directly from the image in Fig. 12.7, the two strands are not linked inside a tube imagined around the knotted braid axis. The linking number of this link equals 3. A similar procedure applied to a ‘flat’ torus link emanating from a third mode bifurcation, results in a cable of a right-handed $(4, 3)$ torus knot (Fig. 12.8). The two components form a 1-link inside an imaginary tube around the knotted axis of the braid. The linking number of the link is 9. The fourth mode bifurcating branch of the initial ‘flat’ torus link $(10, 2)$ leads to a cable of a right-handed $(5, 4)$ torus knot (Fig. 12.9) and the two strands are not linked inside the knotted tube. The linking number of this link is 15.

In all Figs. 12.7–12.9, the contact pressure profiles are displayed as functions of arclength s along the first strand, in white for the pressure on the first strand and in black for the pressure on the second strand. Note that for all three examples the contact pressure changes sign, which means that the physical realisation would require a mechanism (e.g., glue or a clamp) to keep the strands together.

12.4 Concluding remarks

We have shown how the simplest cable knots and links emerge as equilibrium configurations of closed braids made of two equidistant elastic strands. Cable links present an obvious generalisation of cable knots when there is more than one component. A

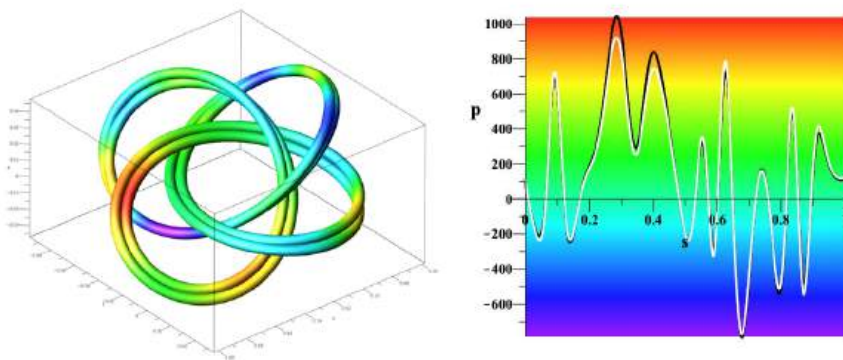


Fig. 12.8: A cable of a right-handed (4, 3) torus knot. $T_1 = -T_2 = -14.49330$, $F = 633.545$, $\Delta = 0.006$, $L_1 = L_2 = 1$.

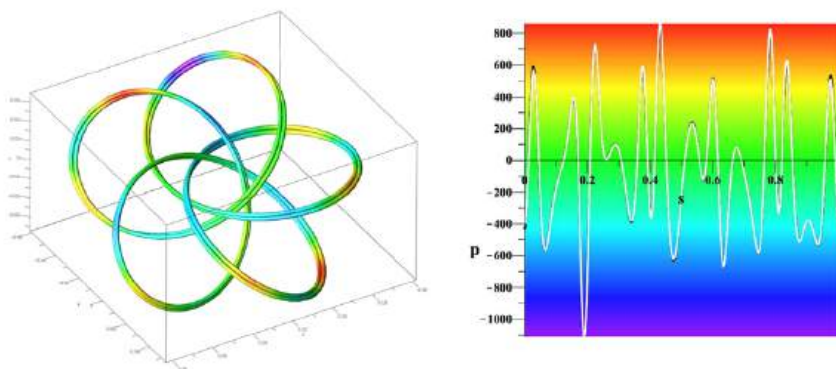


Fig. 12.9: Buckled right-handed cable link formed by knotting a left-handed (10, 2) torus link into a right-handed (5, 4) torus knot. $T_1 = -T_2 = -14.63829$, $F = 946.310$, $\Delta = 0.002$, $L_1 = L_2 = 1$.

further natural extension of this mechanical problem would be to increase the number of strands in the braid, which requires the development of a multi-strand elastic braid theory. Although this calls for a more sophisticated technique to account for equidistant constraints on multiple strands, we do not expect a significant difference in qualitative geometrical properties of mechanical equilibrium solutions for thin elastic rods.

Bibliography

- [1] S. S. Antman. *Nonlinear Problems of Elasticity*. Springer, 2nd edition, 2005.
- [2] G. Burde and H. Zieschang. *Knots*. de Gruyter, Berlin, 2nd edition, 2002.
- [3] Doedel, E. et al. AUTO: Software for continuation and bifurcation problems in ordinary differential equations. User's manual, Concordia University, Montreal, Canada, 2007.
- [4] F. B. Fuller. Decomposition of the linking number of a closed ribbon: A problem from molecular biology. *Proceedings of the National Academy of Sciences of the United States of America*, 75(8):3557–3561, 1978.
- [5] O. Gonzalez, J. H. Maddocks, F. Schuricht, and H. von der Mosel. Global curvature and self-contact of nonlinearly elastic curves and rods. *Calculus of Variations and Partial Differential Equations*, 14:29–68, 2002.
- [6] A. J. Hanson. Quaternion Frenet frames: making optimal tubes and ribbons from curves. Technical report TR407, Indiana Univ., Dept. of Computer Science, 1994.
- [7] K. A. Hoffman. Stability results for constrained calculus of variations problems: an analysis of the twisted elastic loop. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 461(2057):1357–1381, 2005.
- [8] F. Schuricht and H. von der Mosel. Euler-Lagrange equations for nonlinearly elastic rods with self-contact. *Archive for Rational Mechanics and Analysis*, 168:35–82, 2003.
- [9] E. L. Starostin. Three-dimensional shapes of looped DNA. *Meccanica*, 31:235–271, 1996.
- [10] E. L. Starostin and G. H. M. van der Heijden. The shape of a Möbius strip. *Nature Materials*, 6(8):563–567, 2007.
- [11] E. L. Starostin and G. H. M. van der Heijden. Force and moment balance equations for geometric variational problems on curves. *Physical Review E*, 79(6):066602, Jun 2009.
- [12] E. L. Starostin and G. H. M. van der Heijden. Theory of equilibria of elastic 2-braids with interstrand interaction. *Journal of the Mechanics and Physics of Solids*, 64:83–132, 2014.
- [13] G. H. M. van der Heijden, M. A. Peletier, and R. Planqué. Self-contact for rods on cylinders. *Archive for Rational Mechanics and Analysis*, 182:471–511, 2006.

Renzo L. Ricca* and Francesca Maggioni

Groundstate energy spectra of knots and links: magnetic versus bending energy

Abstract: In this paper we review recent results on the groundstate energy spectra of magnetic knots and links and compare these results with new results on bending energy of tight knots and links obtained by using RIDGERUNNER data on curvature. Remarkable similarities between the two systems are found. Comparative analysis between magnetic and bending energy at groundstate energy level shows that information based on bending energy provides a very good proxy for magnetic end-states.

Keywords: Knots, links, energy spectrum, constrained minimization, magnetic relaxation, bending energy, topological fluid mechanics.

MSC: 57M25, 76M30, 76B47, 76W99

13.1 Introduction

In recent years much progress has been done in applications of knot theory to mathematical physics, from classical to quantum field theory, and in physical and biological sciences as well. This has led mathematical research to explore new territories at the cross-road of several, different disciplines. One interesting problem here is the study of minimum energy states of physical knots and links subject to continuous relaxation (through diffeomorphisms) of some energy functional. Magnetic relaxation of knots and links embedded in an ideal fluid provides indeed a prototype example of minimization useful to explore and understand similar features present, for instance, in elastic systems. Here we present a brief review of results on the groundstate energy spectra of magnetic knots and links (published in *J. Phys. A: Math. & Theor.* [20]) that sheds light on similar aspects when we consider elastic, rather than magnetic knots and links. Indeed proof of how good this analogy can be is given by comparing, as we do here, those results with new results based on bending energy estimates of tight knots and links obtained by using data readily available from RIDGERUNNER [1].

The material is presented as follows. In Sec. 13.2 we consider magnetic knots and links as tubular embeddings in ideal magnetohydrodynamics and introduce basic definitions. The prototype problem of magnetic relaxation under constraints is discussed in Sec. 13.3. A solution to the problem of magnetic relaxation of (zero-framed) knots

*Corresponding author: **Renzo L. Ricca**, Department of Mathematics and Applications, U. Milano-Bicocca, Via Cozzi 55, 20125 Milano, Italy & Beijing–Dublin International College, Beijing U. Technology, 100 Pingleyuan, Beijing 100124, PR China, E-mail: renzo.ricca@unimib.it

Francesca Maggioni, Department of Management, Economics & Quantitative Methods, U. Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy, E-mail: francesca.maggioni@unibg.it

<https://doi.org/10.1515/9783110571493-012>

Open Access.  © 2018 Renzo L. Ricca and Francesca Maggioni, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

and links is presented in Sec. 13.4. By using RIDGERUNNER data the groundstate magnetic energy spectra of the first 250 prime knots and 130 prime links are presented (Sec. 13.5). Similar spectra for the bending energy of tight knots and links are presented in Sec. 13.6. A comparative analysis of these results is made in Sec. 13.7. Finally, conclusions are drawn in Sec. 13.8.

13.2 Magnetic knots and links in ideal conditions

We consider magnetic knots and links in an ideal, incompressible, perfectly conducting fluid in S^3 (i.e. $\mathbb{R}^3 \cup \{\infty\}$, simply connected). Let $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ be the fluid velocity, smooth function of the position vector \mathbf{x} and time t , with $\nabla \cdot \mathbf{u} = 0$ in S^3 and $\mathbf{u} = 0$ at infinity. The magnetic field $\mathbf{B} = \mathbf{B}(\mathbf{x}, t)$ is frozen in the fluid and has finite energy, that is

$$\mathbf{B} \in \{\nabla \cdot \mathbf{B} = 0, \partial_t \mathbf{B} = \nabla \times (\mathbf{u} \times \mathbf{B}), L_2\text{-norm}\}. \quad (13.2.1)$$

A magnetic knot is a magnetic flux tube prescribed by the knot type \mathcal{K} and the magnetic field \mathbf{B} , defined on a *regular* tubular support $\mathcal{T}(\mathcal{K})$ centered on \mathcal{K} . We assume \mathcal{K} to be a C^3 -smooth, closed loop (i.e. a submanifold of S^3 homeomorphic to S^1), simple (i.e. non-self-intersecting) and parametrized by arc-length s . The tube $\mathcal{T} = \mathcal{K} \otimes \mathcal{S}$, given by the cartesian product of \mathcal{K} and the circular disk \mathcal{S} , is centered on the knot, whose total length is $L = L(\mathcal{K})$ (hence $s \in [0, L]$), local radius of curvature $\rho > 0$, and cross-sectional area $A = \pi R^2$ of radius $R > 0$.

Since the magnetic knot is a physical tube, it is useful to introduce the volume $V(\mathcal{T})$, the magnetic flux Φ and the magnetic energy M . The total volume is given by $V = V(\mathcal{T}) = \pi R^2 L$, with tubular boundary $\partial\mathcal{T}$ a magnetic surface, i.e.

$$\text{supp}(\mathbf{B}) := \mathcal{T}(\mathcal{K}), \quad \mathbf{B} \cdot \hat{\nu}_\perp = 0 \text{ on } \partial\mathcal{T}, \quad (13.2.2)$$

where $\hat{\nu}_\perp$ is a unit normal to $\partial\mathcal{T}$. The existence and regularity of non-self-intersecting nested tori, support of the magnetic field inside \mathcal{T} , is guaranteed by the tubular neighborhood theorem [22], provided $\rho \geq R$ all along \mathcal{K} . The magnetic flux Φ is defined by

$$\Phi = \int_{\mathcal{S}} \mathbf{B} \cdot \hat{\nu} \, d^2 \mathbf{x}, \quad (13.2.3)$$

where now $\hat{\nu}$ is the unit normal to \mathcal{S} ; the magnetic energy M is given by

$$M = \frac{1}{2} \int_{V(\mathcal{T})} \|\mathbf{B}\|^2 \, d^3 \mathbf{x}. \quad (13.2.4)$$

13.3 The prototype problem

For a magnetic knot, whose field is confined to a single tube of *signature* (V, Φ) , the combined action of magnetic stresses and Lorentz force induces the field lines to shrink like elastic bands, by shortening the knot, while conserving volume and flux [12]. Magnetic energy gets gradually converted into kinetic energy, and eventually dissipated by viscosity or other dissipative effects, if present. As the relaxation progresses, the average cross-section increases proportionately, and the tubular knot becomes thicker and tighter, until knot topology prevents any further adjustment: the final state is ultimately reached when the relaxation comes to a complete stop (see figure 13.1). During this process the knot is also gradually deformed by the action of a signature-preserving flow (through diffeomorphisms), that governs the relaxation from the initial configuration. Since the tight configuration of the end-state resembles that of an *ideal* knot of platonic features [23], magnetic relaxation provides physical mechanism to investigate optimal geometric properties of ideal knots.

Let (r, ϑ_R, s) denote an *orthogonal*, curvilinear coordinate system centered on \mathcal{K} (see [11]); $r \in [0, R]$ and $\vartheta_R \in [0, 2\pi]$ are the radial and azimuthal coordinates in the cross-sectional plane of \mathcal{S} , with origin O at $s = 0$ and $\vartheta_R = 0$ given by the direction of the principal normal to \mathcal{K} at O . The metric is orthogonal, with scale factors $h_r = 1$, $h_{\vartheta_R} = r^2$, $h_s = 1 - cr \cos \vartheta$, where $c = c(s)$ is curvature,

$$\vartheta = \vartheta(\vartheta_R, s) = \vartheta_R - \int_0^s \tau(\bar{s}) d\bar{s}, \quad (13.3.1)$$

and $\tau = \tau(s)$ torsion. Orthogonality is ensured by eq. (13.3.1), which provides the necessary correction to the standard azimuthal angle by the torsion contribution (see details in [11], Sec. 3). The results presented here are derived by using this metric.

The magnetic field \mathbf{B} may be decomposed into meridian and longitudinal components, that is

$$\mathbf{B} = (0, B_{\vartheta_R}(r), B_s(r)), \quad (13.3.2)$$

and in general we assume that the longitudinal field is far greater than the meridian field, i.e. $B_s \gg B_{\vartheta_R}$. This is consistent with the usual definition of twisted flux tube, whose field lines wind around the knot axis in the longitudinal direction. By using the solenoidal condition $\nabla \cdot \mathbf{B} = 0$, the magnetic field can be expressed in terms of poloidal (meridian) and toroidal (longitudinal) fluxes Φ_P and Φ_T , i.e.

$$\mathbf{B} = \left(0, \frac{1}{L} \frac{d\Phi_P}{dr}, \frac{1}{2\pi r} \frac{d\Phi_T}{dr}\right) + \left(0, \frac{\partial \tilde{\psi}}{\partial s}, -\frac{\partial \tilde{\psi}}{\partial \vartheta_R}\right), \quad (13.3.3)$$

where the total field is given by the sum of an average field plus a fluctuating field with zero net flux, in terms of the flux function $\tilde{\psi} = \tilde{\psi}(r, \vartheta_R, s)$. The twist $h = \Phi_P/\Phi_T$ of the



Fig. 13.1: Ideal relaxation of a magnetic trefoil knot and a Hopf link.

field lines provides the magnetic field *framing* given by $(2\pi)^{-1}$ times the turns of twist required to generate poloidal field from toroidal field, starting from $\Phi_p = 0$.

According to the process described above, knot topology dictates a lower bound on the relaxation of magnetic energy M , which must be bounded from below by a minimum $M_{\min} > 0$, that on dimensional grounds is given by (see [13])

$$M_{\min} = m(h)\Phi^2 V^{-1/3}, \quad (13.3.4)$$

where $m(h)$ is a positive, dimensionless function of the internal twist h . Of particular interest is the value of h for which $m(h)$ is minimal (m_{\min}). A fundamental problem is this:

Problem 1 ([14]). Determine m_{\min} for knots of minimum crossing number 3, 4, 5, ...

If c_{\min} denotes the *topological crossing number* of the knot and $h = 0$ (a condition referred to as *zero-framing*), one can prove the following result:

Theorem 1 ([16]). Let \mathcal{K} be a zero-framed magnetic knot with signature $\{V, \Phi\}$. We have $m(0) = (2/\pi)^{1/3} c_{\min}$; hence

$$M_{\min} = \left(\frac{2}{\pi}\right)^{1/3} c_{\min} \Phi^2 V^{-1/3}. \quad (13.3.5)$$

For a signature-preserving flow, eq. (13.3.5) establishes a correspondence between minimum energy levels and topology. However, $M_{\min} \propto c_{\min}$ is a rather loose result. From a direct inspection of the knot table (see, for instance, the standard tabulation in [21]) with the exception of the trefoil and the 4-crossing knot, for all other values of $c_{\min} > 4$ there are several distinct knot types for each given c_{\min} , and the number gets exponentially large for increasing values of c_{\min} (2 for $c_{\min} = 5$, 3 for $c_{\min} = 6$, 7 for $c_{\min} = 7$, 21 for $c_{\min} = 8$, and so on). Hence, the question is to determine whether a one-to-one relationship between energy minima and different knot types of same c_{\min} may exist or not.

13.4 Relaxation of magnetic knots and constrained minima

To explore this problem let us consider the relaxation of a magnetic flux tube in some generality. Let $V_r = \pi r^2 L$ be the partial volume of the tubular neighborhood of radius r ; the ratio of the partial to total volume is given by $V_r/V(\mathcal{T}) = (r/R)^2$. Now, let $f(r/R)$ be a monotonically increasing function of r/R ; for example $f(r/R) = (r/R)^\gamma$, with $\gamma > 0$; $\gamma = 2$ defines the *standard* flux tube. A detailed analysis of the relaxation of magnetic flux tube with twist is done in [11]. By using the orthogonal, curvilinear system (r, ϑ_R, s) and the magnetic field decomposition given by (13.3.3), standard minimization of (13.2.4) is carried out and under the periodicity of ϑ_R and s , subject to these assumptions:

- (i) $\{V, \Phi\}$ is invariant;
- (ii) the circular cross-section is independent of s ;
- (iii) $\tilde{\psi}$ is independent of s ;
- (iv) the knot length is independent of h .

We have:

Theorem 2 ([11]). Let \mathcal{K} be an essential magnetic knot with signature $\{V, \Phi\}$ and magnetic field given by (13.3.3). Constrained minimization of magnetic energy yields

$$M^* = \left(\frac{\gamma^2 L^{*2}}{8(\gamma - 1)V} + \frac{\gamma \pi h^2}{2L^*} \right) \Phi^2, \quad (13.4.1)$$

where L^* is the minimal tube length of the tight knot.

For a standard flux tube ($\gamma = 2$), (13.4.1) reduces to

$$M^* = \left(\frac{L^{*2}}{2V} + \frac{\pi h^2}{L^*} \right) \Phi^2. \quad (13.4.2)$$

This result is equivalent to the eq. (4.2) of [7] (coefficients left undetermined). Note that because of the constraints, for any given knot family we have that $\langle M^* \rangle_{c_{\min}} \geq M_{\min}$, where angular brackets denote averaging over the number of knots of the same c_{\min} family.

In order to investigate the relation between energy and knot topology, let us refer to standard flux tubes; it is useful to rewrite eq. (13.4.2) in terms of *ropelength*, a useful measure of knot complexity [4]: this is defined by $\lambda = L^*/R^*$, where L^* is the minimal length and R^* the maximal cross-sectional radius of the knot in tight configuration. In the case of the unknot, the least possible value of λ (say λ_0) is that given by the tight torus; hence $\lambda \geq \lambda_0 = 2\pi$. By using $V = \pi R^{*2} L^* = \text{cst.}$, and after some straightforward algebra, we have

$$M^* = \left(\frac{\lambda^{4/3}}{2\pi^{2/3}} + \frac{\pi^{4/3} h^2}{\lambda^{2/3}} \right) \Phi^2 V^{-1/3}. \quad (13.4.3)$$

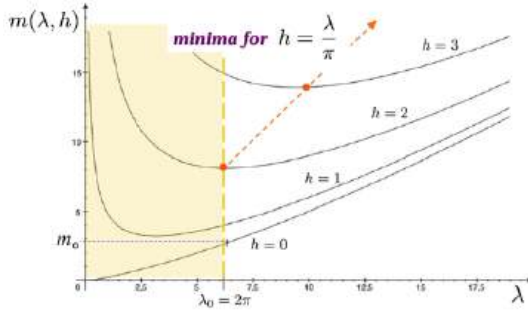


Fig. 13.2: Influence of twist h on the energy function $m(\lambda, h)$, plotted against ropelength λ , according to eq. (13.4.4). The absolute minimum is given by the tight torus, for which $\lambda = \lambda_0 = 2\pi$ and $m_0 \approx 2.70$ (from [20]).

By comparing (13.3.4) and (13.4.3), and under the above assumptions, we can state that

$$m(\lambda, h) = \frac{\lambda^{4/3}}{2\pi^{2/3}} + \frac{\pi^{4/3}h^2}{\lambda^{2/3}}, \quad (13.4.4)$$

showing the explicit dependence of minimum energy on ropelength and framing

13.5 Groundstate magnetic energy spectra

Let us first investigate the minima $m_{\min} = m_{\min}(h)$ by plotting (13.4.4) against λ for $h = 0, 1, 2, 3, \dots$ (see figure 13.2). The absolute minimum m_0 corresponds to the zero-framed unknot (tight torus), given by $h = 0$ and $\lambda = \lambda_0 = 2\pi$: $m_0 = (2\pi^2)^{1/3} \approx 2.70$. The groundstate energy of zero-framed flux tubes provides the absolute minimum energy level; $m(h)$ remains a monotonic increasing function of λ for $h \leq 2$: at $\lambda_0 = 2\pi$ we have $m(h = 1) = 4.05$ and $m(h = 2) = 8.11$. For $h \geq 2$ the energy minima are attained at $h = \lambda/\pi$; thus, by substituting the optimal value $\lambda = \pi h$ in (13.4.4), we have

$$m_{\min}(h) = \frac{3}{2}\pi^{2/3}h^{4/3} \quad (h \geq 2). \quad (13.5.1)$$

For $h > 2$ (and $\lambda \geq \lambda_0$) the functional dependence of $m(h)$ on λ ceases to be monotonic. The same $h^{4/3}$ power-law of eq. (13.5.1) was also found by [7] (p. 206, eq. 4.15, by scaling arguments).

The minimum energy spectra of the first prime knots and links is determined by setting $h = 0$ in (13.4.4) and by using ropelength data (λ_K) of each knot/link type K obtained by RIDGERUNNER, a tightening algorithm developed by [1]. A particularly simple expression is obtained by normalizing $m(\lambda_K, 0)$ with respect to the minimum energy

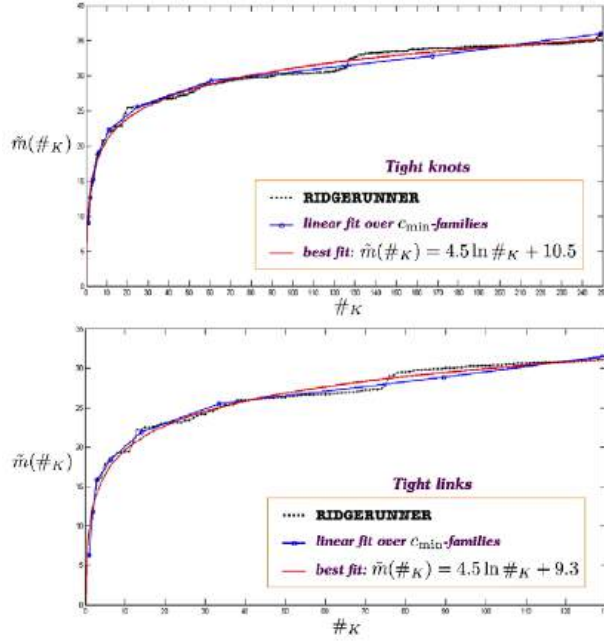


Fig. 13.3: Magnetic energy spectrum $\tilde{m} = \tilde{m}(\#_K)$ of tight knots (top) and tight links (bottom) plotted against the knot/link number $\#_K$, given by the position of the knot/link K listed according to increasing value of ropelength $\lambda_K = \lambda(\#_K)$ (from [20]).

value m_o of the tight torus; thus, we have

$$\tilde{m}(K) = \frac{m(\lambda_K, 0)}{m_o} = \left(\frac{\lambda_K}{2\pi} \right)^{4/3}, \quad (13.5.2)$$

that gives a one-to-one relationship between minimum energy level and knot ropelength. Since the relation $\lambda_K = \lambda(K)$ is not known analytically, it must be reconstructed from numerical data. We take $\lambda_K = \lambda(\#_K)$, where $\#_K$ denotes the position of the knot/link K listed according to increasing values of ropelength given by RIDGERUNNER. Hence, instead of tabulating energy levels as function of the knot/link position given by standard knot tabulation, by taking $\lambda_K = \lambda(\#_K)$ we plot $\tilde{m} = \tilde{m}(\#_K)$, according to increasing ropelength data. The energy spectra are shown in figure 13.3 for the first 250 prime knots up to 10 crossings (top diagram) and 130 prime links up to 9 crossings (bottom diagram). Remarkably, magnetic energy levels of knot and link types seem to follow an almost identical logarithmic law given by the best fit curve shown in the plots.

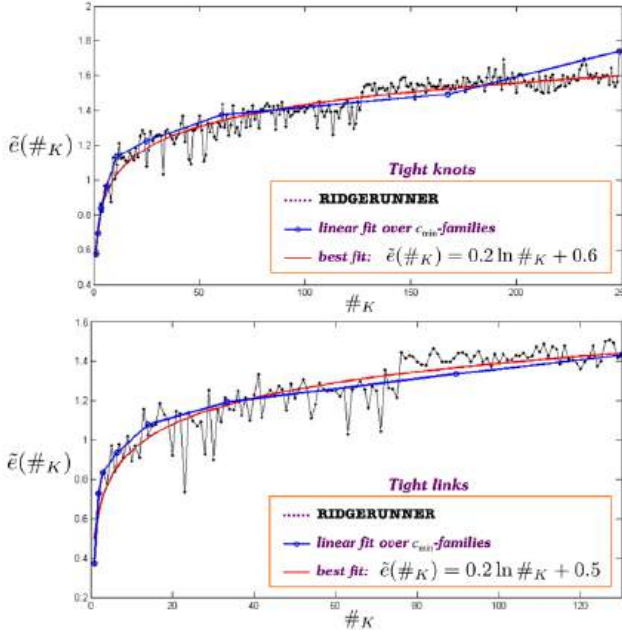


Fig. 13.4: Bending energy spectrum $\tilde{e} = \tilde{e}(\#_K)$ of tight knots (top) and links (bottom) plotted against the knot/link number $\#_K$, given by the position of the knot/link K listed according to increasing values of ropelength $\lambda_K = \lambda(\#_K)$.

13.6 Bending energy spectra

It is interesting to compare groundstate magnetic energy spectra with the corresponding bending energy spectra obtained by considering bending energy in place of magnetic energy. Since magnetic relaxation is driven by the Lorentz force, that is mainly a curvature force, computation of the elastic energy due solely to curvature (bending energy) provides an interesting comparison. Bending energy is defined by

$$E_b = \frac{1}{2} \oint_{\mathcal{C}} K_b [c(s)]^2 ds, \quad (13.6.1)$$

where K_b is bending rigidity and $c(s)$ is local curvature. By normalizing this quantity with respect to the reference value $E_o = \pi K_b / R^* = K_b 2^{1/3} \pi^{5/3}$ of the tight torus, we have the normalized bending energy given by

$$\tilde{e} = \frac{E_b}{E_o} = \frac{\oint_{\mathcal{C}} [c(s)]^2 ds}{2^{4/3} \pi^{5/3}}. \quad (13.6.2)$$

By using curvature data of tight knots obtained by RIDGERUNNER, we can easily plot the energy spectrum $\tilde{e} = \tilde{e}(\#_K)$ according to increasing ropelength data. The corresponding energy spectra are shown in figure 13.4. Remarkably, and similarly to the magnetic

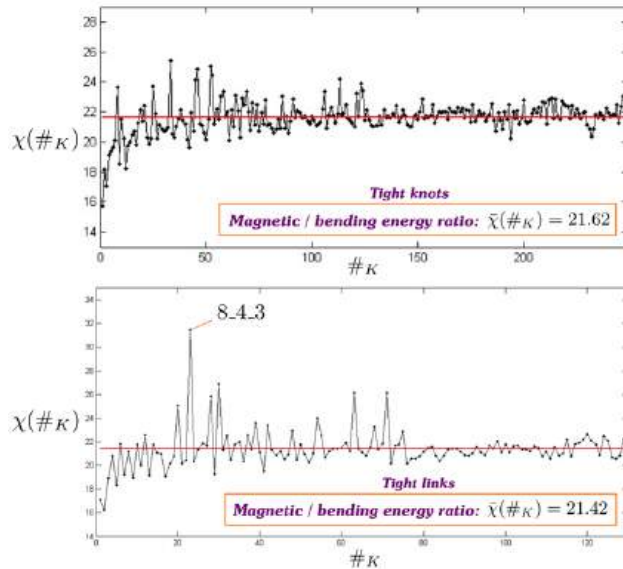


Fig. 13.5: Ratio of magnetic to bending energy $\chi = \tilde{m}/\tilde{e}$ of tight knots (top) and tight links (bottom) plotted against the knot/link number $\#_K$, given by the position of the knot/link K listed according to increasing value of ropelength $\lambda_K = \lambda(\#_K)$.

case, bending energy levels of knot and link types seem to follow an almost identical logarithmic law given by the best fit curve shown in the plots.

13.7 Magnetic energy versus bending energy

Let us compare the energy values by taking the ratio of magnetic to bending energy $\chi = \tilde{m}/\tilde{e}$ and plot for comparison $\chi = \chi(\#_K)$ according to increasing ropelength data. Results are presented in figure 13.5. As we see, with a few exceptions, ratios tend to level up around a constant value that for sufficiently large c_{\min} for tight knots is given by $\bar{\chi} = 21.62$ and for tight links is given by $\bar{\chi} = 21.42$. Thus, on average, for complex topologies we see that information on bending energy is proportional to magnetic energy. This is actually in good agreement with expectations: since magnetic relaxation is driven by the Lorentz force, which to a first approximation is a curvature force, it is not so surprising to discover that relaxed magnetic states are, on average, proportional to bending energy end-states of tight knots and links. Departure from average values is greater at lower c_{\min} , where, probably, the constraints imposed by the assumptions affect the most the results.

Let us now re-examine a common feature of the plots of figures 13.3-13.4. The curves dotted by circles results from a linear fit made over each c_{\min} -family, while

the continuous curve is the best-fit interpolation over all available data. To the first decimal digit, we find that best-fit interpolations follow an almost identical logarithmic law, given by

$$\tilde{m}(\#_K) = a \ln \#_K + b, \quad (13.7.1)$$

with $a = 4.5$, $b = 10.5$ for magnetic knots and $a = 4.5$, $b = 9.3$ for magnetic links, and similarly

$$\tilde{e}(\#_K) = a \ln \#_K + b, \quad (13.7.2)$$

with $a = 0.2$, $b = 0.6$ for elastic knots and $a = 0.2$, $b = 0.5$ for elastic links. These unexpected results are quite remarkable and call for some justification.

Ropelength is certainly an increasing function of topological complexity (given by c_{\min}), also because an increasing number of crossings implies an increasing minimal length necessary to tie a flux tube into a knot or a link. Results on ropelength bounds [4, 6, 8, 9] show that

$$O(c_{\min}^{3/4}) \leq \lambda_K \leq O(c_{\min} \ln^5 c_{\min}), \quad (13.7.3)$$

where $O(\cdot)$ denotes order of magnitude. From (13.5.2) we have that $\tilde{m}(\#_K) \propto [\lambda(\#_K)]^{4/3}$; by combining this with (13.7.2), we have

$$[\lambda(\#_K)]^{4/3} \propto a \ln \#_K + b. \quad (13.7.4)$$

Now, if we assume that the number of knots grows exponentially with c_{\min} (a plausible assumption), then $\#_K \sim C^{c_{\min}}$ for some constant C . Hence, by (13.7.4) we have $[\lambda(\#_K)]^{4/3} \propto c_{\min}$, or

$$\lambda(\#_K) \propto c_{\min}^{3/4}, \quad (13.7.5)$$

a result that, if not true in full generality, is certainly in good agreement with the lower estimate given by (13.7.3). Furthermore, let us set (for simplicity) $V = \Phi = 1$ in (13.3.5), and define

$$\overline{m}(c_{\min}) \equiv \frac{M_{\min}}{m_o} = \frac{1}{\pi} c_{\min}. \quad (13.7.6)$$

We can then relate (13.3.5) to (13.5.2), and write

$$\langle \tilde{m}(K) \rangle_{c_{\min}} \geq \overline{m}(c_{\min}) = \frac{1}{\pi} c_{\min}, \quad (13.7.7)$$

since for any given K $\tilde{m}(K)$ can be further decreased to its actual minimum by relaxing the constraints (i)-(iv) of Theorem 2. Similar considerations apply to $\tilde{e}(K)$. By writing (13.5.2) in terms of $\#_K$ and substituting this latter into the above equation, we have

$$\langle \lambda(\#_K) \rangle_{c_{\min}} \geq 2\pi^{1/4} c_{\min}^{3/4}, \quad (13.7.8)$$

that gives a new relation between ropelength, averaged over each c_{\min} -family, and c_{\min} . Note that the coefficient $2\pi^{1/4} \approx 2.66$ is independent of the knot family, and this result, in good agreement with (13.7.3), is still one of the best analytical results valid for any c_{\min} to date (see, for instance, [4]).

13.8 Conclusions

By using analytical results for the constrained minimum energy of magnetic knots obtained in [11], we have established a general functional relationship between minimum energy levels of knots and links and internal twist h , given by an $h^{4/3}$ -power law. In the case of standard flux tubes our result is in good agreement with an earlier result by [7] obtained by a scaling argument. By using ropelength data obtained by the RIDGERUNNER tightening algorithm developed by [1] we have computed the groundstate energy spectra of the first 250 prime knots and 130 prime links; we have shown that the two spectra follow an almost identical logarithmic law. We have then extracted data on curvature and by computing the bending energy we have compared magnetic and bending energy spectra, finding a remarkable proportionality between end-states. By assuming that the number of knot types grows exponentially with the topological crossing number c_{\min} , we have shown that this generic behavior can be justified by a general relationship between ropelength and crossing number, that is independent of the number of components (knots or links). Moreover, by considering ropelength averaged over a given knot family, we have established a new relation between this averaged ropelength and $c_{\min}^{3/4}$, valid for knots/links of *any* c_{\min} . However, as recent analytical work shows [9], these results cannot be considered fully general and further improvements are expected. In the context of magnetic relaxation, corrections are expected to come from finer realization of the analytical constraints (for instance, by allowing the cross-section to adapt to optimal shape) and from further improvements of the tightening procedure. In any case, our results demonstrate the great potential of magnetic energy methods to investigate and establish new relationships between energy contents and topological properties of complex systems. Moreover, by using curvature information of tight knots and links we can also estimate optimal properties of 3D-packing and global geometry. These results can find useful applications in many disparate fields, from the study of structural complexity of physical and biological filamentary systems [10, 18, 5], to applications in plasma physics and solar physics [15, 19]. They may also provide new insight into the ongoing search for fundamental aspects in the mass-energy relations of high-energy theoretical physics [3, 2].

Acknowledgment: R.L.R. wishes to express his gratitude to the Fonds National Suisse (FNS) for financial support.

Bibliography

- [1] Ashton, T., Cantarella, J., Piatek M. & Rawdon, E. (2011) Knot tightening by constrained gradient descent. *Experim. Math.* 20 57–90.

- [2] Buniy, R. V., Cantarella, J., Kephart, T. W. & Rawdon, E. J. (2014) Tight knot spectrum in QCD. *Phys. Rev. D* 89, 054513.
- [3] Buniy, R. V. & Kephart, T. W. (2005) Glueballs and the universal energy spectrum of tight knots and links. *Int. J. Mod. Phys. A* 20 1252–1259.
- [4] Buck, G. & Simon, J. (1999) Thickness and crossing number of knots. *Topol. Appl.* 91 245–257.
- [5] Buck, G. & Simon, J. (2012) The spectrum of filament entanglement complexity and an entanglement phase transition, *Proc. Roy. Soc. A* 468 4024–4040.
- [6] Cantarella, J., Kusner, R. B. & Sullivan, J. M. (2002) On the minimum ropelength of knots and links. *Invent Math* 150 257–286.
- [7] Chui, A. Y. K. & Moffatt, H. K. (1992) Minimum energy magnetic fields with toroidal topology. In *Topological Aspects of the Dynamics of Fluids and Plasmas*, (ed. H.K. Moffatt *et al.*), pp. 195–218. Kluwer Acad. Pubs., Dordrecht.
- [8] Diao, Y. (2003) The lower bounds of the lengths of thick knots. *J. Knot Theory Its Ram.* 12 1–16.
- [9] Diao, Y., Ernst, C., Por, A. & Ziegler, U. (2016) The ropelengths of knots are almost linear in terms of their crossing numbers. Preprint at arXiv: 0912.3282 (accessed on 20 April, 2018).
- [10] Kauffman, L. H., Editor (1995) *Knots and Applications*. Series of Knots and Everything 6. World Scientific, Singapore.
- [11] Maggioni, F. & Ricca, R. L. (2009) On the groundstate energy of tight knots. *Proc. R. Soc. A* 465 2761–2783.
- [12] Moffatt, H. K. (1985) Magnetostatic equilibria and analogous Euler flows of arbitrarily complex topology. Part I. Fundamentals. *J Fluid Mech* 159 359–378.
- [13] Moffatt, H. K. (1990) The energy spectrum of knots and links. *Nature* 347 367–369.
- [14] Moffatt, H. K. (2001) Some remarks on topological fluid mechanics. In *An Introduction to the Geometry and Topology of Fluid Flows*, (ed. R.L. Ricca), p. 3. Kluwer Acad. Pubs., Dordrecht.
- [15] Moffatt, H. K., Bajer, K. & Kimura, Y., Editors (2013) *Topological Fluid Dynamics: Theory and Applications*. Elsevier, Dordrecht.
- [16] Ricca, R. L. (2008) Topology bounds energy of knots and links. *Proc. R. Soc. A* 464 293–300.
- [17] Ricca, R. L., Editor (2009) *Lectures on Topological Fluid Mechanics*. Springer-Verlag, Heidelberg.
- [18] Ricca, R. L. (2009) Detecting structural complexity: from visiometrics to genomics and brain research. In *Mathknow* (ed. M. Emmer and A. Quarteroni), pp. 167–181. Springer-Verlag, Berlin.
- [19] Ricca, R. L. (2013) New energy and helicity lower bounds for knotted and braided magnetic fields. *Geophys. Astrophys. Fluid Dyn.* 107 385–402.
- [20] Ricca, R. L. & Maggioni, F. (2014) On the groundstate energy spectrum of magnetic knots and links. *J. Phys. A: Math. & Theor.* 47, 205501.
- [21] Rolfsen, D. (1976) *Knots and Links*. Publish or Perish, Berkeley.

- [22] Spivak, M. (1979) *A comprehensive introduction to differential geometry*. Vol. 1. Publish or Perish, Houston.
- [23] Stasiak, A., Katritch, V. & Kauffman, L. H. Editors, (1998) *Ideal Knots*. Series on Knots and Everything 19. World Scientific, Singapore.