Physics Reports 530 (2013) 257–290

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

Menger curvature as a knot energy

Paweł Strzelecki [a,*], Heiko von der Mosel [b]

[a] Institute of Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warsaw, Poland
[b] Institut für Mathematik, RWTH Aachen University, Templergraben 55, D-52062 Aachen, Germany

ARTICLE INFO

Article history:
Accepted 13 May 2013
Available online 20 May 2013
editor: Randall Kamien

Keywords:
Menger curvature
Knot energies

ABSTRACT

Motivated by the suggestions of Gonzalez and Maddocks, and Banavar et al. to use geometrically defined curvature energies to model self-avoidance phenomena for strands and sheets we give a self-contained account, aimed at non-experts, on the state of art of the mathematics behind these energies. The basic building block, serving as a multipoint potential, is the circumradius of three points on a curve. The energies we study are defined as averages of negative powers of that radius over all possible triples of points along the curve (or via a mixture of averaging and maximization). For a suitable range of exponents, above the scale invariant case, we establish self-avoidance and regularizing effects and discuss various applications in geometric knot theory, as well as generalizations to surfaces and higher-dimensional submanifolds.

Contents

* Corresponding author. Tel.: +48 225544212.
E-mail addresses: pawelst@mimuw.edu.pl (P. Strzelecki), heiko@instmath.rwth-aachen.de (H. von der Mosel).

0370-1573/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physrep.2013.05.003

1. Introduction

Mathematically, a knot is simply a continuous embedding of a circle into three-dimensional space, where we do not distinguish between one such closed knotted curve and its deformations through space avoiding any kind of self-intersections or cutting and gluing. Such admissible deformations, avoiding also the "pull-tight phenomenon" illustrated in Fig. 11, are called ambient isotopies and belong to the basic tool set of knot theory. Knots occur in diverse branches of modern physics or biology. Numerous microscopic and macroscopic examples of this particular connection between Mathematics and the Sciences come to mind, from the interplay between knot theory and statistical or quantum physics, described e.g. in Louis H. Kauffman's essay [1], and microscopic defect lines in chiral nematic liquid-crystal colloids forming various knots and links,[1] to knotted field lines in hydrodynamics and optics, and to polymer chains and knotted structures in DNA. A close interaction between ingenious experiments and hard theory, drawing heavily from several branches of mathematics, including the calculus of variations, nonlinear and geometric analysis, and topology, is present in most of that research.

On the mathematical side, modern knot theory produces numerous sophisticated *knot invariants*, which can be viewed as mappings from the complicated space of all knots to some simpler space, for example to the real numbers or to polynomials or groups. The term *invariant* means that if two knotted curves are in the same knot class, i.e., if they are ambient isotopic, then their values of a knot invariant are also the same. At present, none of the known invariants permits to determine algorithmically the knot type of every given knot, or to decide whether two seemingly different embeddings of a circle do indeed represent the same knot.

All this leads to a quest for knot recognition methods and to the desire to deform a given knot to a fairly simple, *optimal* or *model* shape. One of the possible approaches is to simulate the physical movement of knots under the influence of some sort of self-repelling potential. One of the most famous tools used for serious experiments with pictures of different knot conformations is Robert Scharein's KnotPlot program.[2] The mathematical concept behind such intuitively appealing simulations of ambient isotopies is that of a knot energy, proposed by Shinji Fukuhara [3], and later made more precise and investigated in depth by various authors, see e.g. [4–7]. Our main aim will be to describe mathematically the properties of one such energy representing a whole family of geometrically defined self-avoidance energies, that is particularly interesting (at least for us) because of its links both to modelling physical and biological objects and to deep advances in abstract harmonic analysis, geometric measure theory and related branches of mathematics.

For the purposes of this introduction, the reader is invited to think of a knot energy as a functional defined on the space of all knotted curves in $\mathbb{R}^3$ that assigns some real number to *each conformation* of every knot, in such a way that several natural conditions are satisfied. First, distinct knot types should be separated by infinitely high energy walls. Then the gradient flow, following the path of steepest descent in the energy landscape, will be confined to a single knot type. Secondly, it would be desirable to know that a bound on the energy value restricts both the set of available knot types and the geometry of their particular representatives: those that are particularly (or unnecessarily) complicated should correspond to high energy values. Thirdly, ideally, the energy should distinguish different knot types, e.g. through distinct minimal energy values on different knot types.

Numerical modelling of curves that avoid self-interpenetration leads to several challenges. For example, the fact that a curve has no self-intersections cannot be deduced just from qualitative *local* properties of its parametrization, such as local curvature, no matter how much is known about them. In addition, the general problem that a numerical gradient flow gets trapped in a local minimum or even in some stable but non-minimal critical point, seems to gain even more relevance in the presence of these nonlocal (and highly nonlinear) interactions between different strands of a curve. Kauffman [8] points out some of the caveats present behind numerical experiments with knots.

Oscar Gonzalez and John H. Maddocks [9], in search for ideal shapes of knots within the context of DNA knotting, proposed the concept of a *global radius of curvature* which can be used to give a characterization of *thickness*.[3] To define the global

[1] This has actually been achieved experimentally, by using laser tweezers as a micromanipulation tool to produce all knot and link types of up to six crossings, cf. Tkalec et al. [2].

[2] As the authors – both of them mathematicians by education, employment and experience – have learned during illuminating contacts with members of the physics' community at the Kavli Institute of Theoretical Physics in summer 2012, KnotPlot is widely used by physicists, to manipulate knotted curves in order to recognize their knot types, and simply to produce illustrations.

[3] The term "thickness", or *ropelength* as the quotient of length and thickness, comes from the following plain language statement of a variational problem: suppose you have a fixed length of string or rope and want to tie a given knot; how thick can this rope be so that your task is still possible?

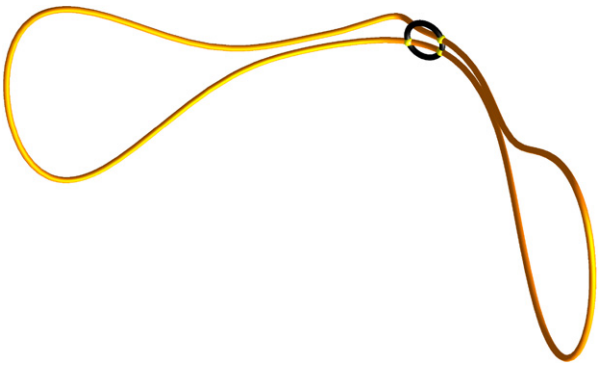


Fig. 1. Two adjacent almost parallel strands of the curve cause $1/R$ to be high at many locations while the classic local curvature remains moderate.

radius of curvature, one considers all triples of distinct points x, y, z on a curve γ in $\mathbb{R}^3$, and for each of them computes the circumradius $R(x, y, z)$, i.e. the radius of the unique circle passing through x, y, z (— this circle degenerating to an infinite straight line if the three points are collinear). Keeping one of the points fixed and varying the remaining two along the curve, one computes the global radius of curvature as

$$\varrho_G[\gamma](x) := \inf_{\substack{y,z\in\gamma \\ z\neq x\neq y\neq z}} R(x, y, z). \tag{1}$$

The inverse, $1/\varrho_G[\gamma](x)$, is referred to as the *global curvature* of γ at x. The thickness $\triangle[\gamma]$ of γ is defined as the infimum of $\varrho_G[\gamma](x)$ over all points x on the curve. It is clear that ϱ_G takes into account both the local curvature of γ and its global properties: it is easy to see that if two nearly straight strands of γ run close to each other, more or less parallelly, then global curvature must be large for many of the points $x \in \gamma$ while the classic local curvature is small at those points; see Fig. 1. If, on the other hand, y and z tend to x along the curve γ, then $1/R(x, y, z)$ tends to the local curvature of γ at x as long as γ is sufficiently smooth.

In Section 2, we describe in more detail the relations between thickness and classical curvature, and the contrast between thickness and self-repulsion. Let us just say here that the intuitive suggestion in the final section of [9] that 'circumradius and global radius curvature … lead to families of integral knot energies that do not require explicit regularization or mollification' has been one of the starting points for a research program that we have followed over nearly a decade, reaching a much better understanding of these energies. One of them, the *integral Menger curvature*

$$\mathscr{M}_p(\gamma) := \int_\gamma \int_\gamma \int_\gamma \frac{d\mathscr{H}^1(x)\, d\mathscr{H}^1(y)\, d\mathscr{H}^1(z)}{R(x, y, z)^p}, \tag{2}$$

where $d\mathscr{H}^1$ denotes the integration with respect to the one-dimensional Hausdorff measure, i.e. the arclength, will serve as a role model in the present paper: we shall explain several properties of related knot energies, and the geometric reasons behind those properties, using the example of $\mathscr{M}_p$.

To give the reader a glimpse of the smoothing properties of $\mathscr{M}_p$, let us invoke the following analogy. Imagine a closed, possibly nonsmooth and possibly self-intersecting curve of length 1 in a dark room; you cannot see the curve but a scanning device can measure the radii $R(x, y, z)$ for lots of randomly selected triples of points and supply you with the statistics of the inverses $1/R$. Then, using a Monte Carlo procedure, you might be able to compute a reasonable approximation of $\mathscr{M}_p(\gamma)$ for various exponents p. Miraculously, if that integral converges for $p > 3$, then you do know the following. First, the curve must be free of self-intersections; it also has no corners or cusps: the tangent vector is defined everywhere and is continuous.[4] Secondly, the value of the $\mathscr{M}_p$-energy explicitly defines the length scale r_0 below which the curve is nearly straight and does not bend too much, just like a rather stiff necklace with a certain number of (conical) beads of fixed size. The knotting – if any – must happen beyond that scale; no little knots on γ can be seen if you scan the balls of radius r_0 or smaller. All this is linked to the control of geometry of the curve which is strong enough to restrict the number of possible knot types the curve might form. You could obtain a crude yet explicit estimate of that number. You could also estimate the number of sticks needed to build a polygonal model of the knot the curve forms, and give an explicit bound for the average crossing number of γ, i.e. for the average number of crossings that are seen in a projection of γ onto a plane. The limit as $p \to \infty$ of $\mathscr{M}_p(\gamma)^{1/p}$ gives $1/\triangle[\gamma]$, the inverse of thickness, or the ropelength of the curve.

It is still an open problem to prove rigorously that the gradient flow of $\mathscr{M}_p$ does exist. However, recent extensive numerical simulations of that flow by Tobias Hermes [11] indicate that unknotted curves – including some not entirely trivial

[4] In particular, finite integral Menger curvature $\mathscr{M}_p$, $p > 3$, of a loop γ implies that this loop cannot be topologically *wild* (which, a priori, could happen if γ were just rectifiable and simple) and must be *tame*, cf. Freedman et al. [10, Section 4] where a similar phenomenon is described for another knot energy, the *Möbius energy* $\mathscr{E}_{\text{Möb}}$, which we discuss briefly in Section 3.3.3.

unknots – untangle and flow to round circles (see figures in Section 5). A far reaching dream is that a profound analytic understanding of the gradient flow for $\mathscr{M}_p$ would help to explore in detail its presumably very complex energy landscape over knot space. In addition, sending p to infinity this might help to define and investigate analytically a gradient flow for the nonsmooth limit energy ropelength, for which Jason Cantarella's and Eric Rawdon's algorithm RIDGERUNNER provides fascinating numerical results [12].

The paper is organized as follows. In Section 2, we explain the regularizing and self-avoidance effects of $\mathscr{M}_p$, and the geometry behind those effects. In Section 3, we show that those properties of $\mathscr{M}_p$ can immediately be translated to simple features of the knot energy landscape: banning the pull-tight phenomenon, existence of energy minimizers in each knot class, bounds on the number of knot classes under each energy level etc. We also compare the properties of integral Menger curvature to those of several other knot energies, including a repulsive potential introduced by Jun O'Hara [13], also known as *Möbius energy*, which can be viewed as a regularization of self-repulsion via electrostatic forces. In Section 4 we investigate in more detail how Menger curvature controls the geometry of a knot: we obtain the stick number estimates and explain why two curves of bounded $\mathscr{M}_p$-energy that are sufficiently close to each other in space simply as point sets, i.e. close w.r.t. the so-called Hausdorff distance, do represent the same knot. (Do bear in mind that the closedness itself is by no means sufficient here: one of the curves might wind around the other one, forming lots of extra little knots). Finally, in Section 5 we discuss some of the open problems, in particular the regularity and shape of minimizers, and the existence of the flow, and mention several generalizations of such geometric curvature energies to surfaces and higher dimensional submanifolds.

At the end of this introduction, to avoid the misleading impression that integral Menger curvature is related first and foremost to knots and simulations of the physical models and their ambient isotopies, let us digress and mention a deep, purely mathematical link between Menger curvature and complex analysis. In the late XIXth century Paul Painlevé has been studying the problem of *removable singularities of bounded analytic functions*: suppose you have a compact set K in the complex plane $\mathbb{C}$; under which circumstances can all bounded analytic functions $f : \mathbb{C} \setminus K \to \mathbb{C}$ be extended to analytic functions defined on all of $\mathbb{C}$? In other words and in light of the classic Liouville theorem, what are the necessary and sufficient conditions on K implying that all bounded analytic functions $f : \mathbb{C} \setminus K \to \mathbb{C}$ are constant? Such sets have been termed *removable for bounded analytic functions*. Every student learns in a basic course on analytic functions that isolated point singularities are removable. It is a bit more complicated, but still on the level of exercises for a graduate course in complex analysis, to see that (a) each compact set K with $\mathscr{H}^1(K) = 0$ is removable, (b) no continuous arc of non-zero length is removable. In the 1960's Anatoli G. Vitushkin conjectured that a compact set K with $0 < \mathscr{H}^1(K) < \infty$ is removable if and only if it is *purely unrectifiable*, i.e. if its projections to almost every straight line have zero length. (Examples of such sets include certain analogues of the Cantor set in the plane.)

After more than three decades, Vitushkin's conjecture has been proved due to mutual efforts and discoveries of several mathematicians, including Mark S. Melnikov, Xavier Tolsa, Joan Verdera, Pertti Mattila, Guy David and others. The story is now well-documented in research papers and surveys, see e.g. [14,15] or [16]. The gist is that to define analytic functions by means of their boundary values one uses the Cauchy integral formula for a curve (or a set) K; this leads to a question for what sets K this formula defines a linear operator with good properties, like boundedness on L^2 etc. Two of the key steps in the solution of Vitushkin's conjecture were the discovery by Melnikov and Verdera that the L^2-norm of the Cauchy integral along a curve γ is intimately related to the integral Menger curvature $\mathscr{M}_2(\gamma)$ of that curve, and the contribution by Jean-Christophe Léger who proved that finiteness of the integral Menger curvature $\mathscr{M}_2(E)$ of a one-dimensional Borel set $E \subset \mathbb{C}$, i.e. the condition

$$\mathscr{M}_2(E) = \iiint_{E\times E\times E} \frac{1}{R^2(x,y,z)}\, d\mathscr{H}^1(x)\, d\mathscr{H}^1(y)\, d\mathscr{H}^1(z) < \infty,$$

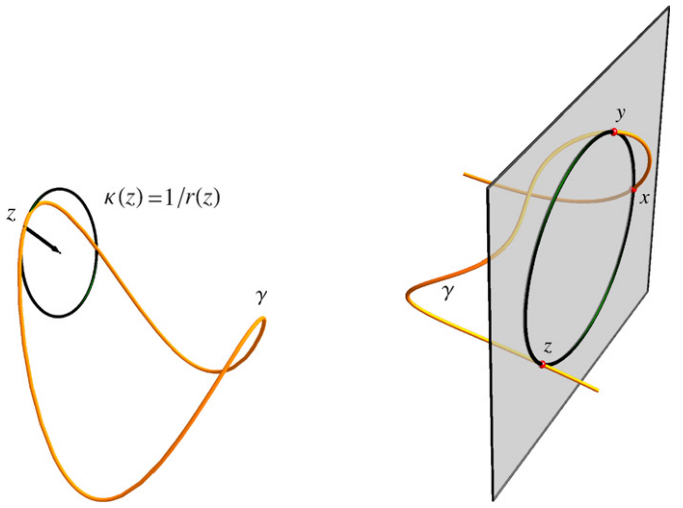
implies that E is rectifiable. This means that E is contained – up to a negligible subset of one-dimensional measure zero – in a union of countably many C^1-curves.

If the reader is a physicist, then he or she should bear in mind that to a randomly selected mathematician the name 'Menger curvature' might only ring the (complex analytic) bells hinted at in the previous paragraph. However, as many other mathematical tools, integral Menger curvature serves more than just one purpose.

2. Self-avoidance and regularizing effects

Since we are interested in different conformations of knots, i.e., specific curves in $\mathbb{R}^3$, and the plain term 'curve' is ambiguous even inside mathematics, let us make it more precise for our purposes. Everywhere below, we consider the class $\mathcal{C}$ of all *closed and rectifiable* curves $\gamma \subset \mathbb{R}^3$ whose length, i.e., one-dimensional Hausdorff measure $\mathscr{H}^1(\gamma)$, is equal to 1. Moreover, for technical reasons we assume that all curves in $\mathcal{C}$ contain a fixed point, say the origin in $\mathbb{R}^3$, and that all loops in $\mathcal{C}$ are parametrized by arclength defined on the interval $[0, 1]$, that is, $\gamma : [0, 1] \to \mathbb{R}^3$ is Lipschitz continuous with $|\gamma'| = 1$ almost everywhere[5] and $\gamma(0) = \gamma(1)$. The curves in $\mathcal{C}$ will sometimes be referred to as *(unit) loops*. If γ is injective on $[0, 1)$, then we say that γ is *simple*.

[5] By Rademacher's theorem, the tangent $\gamma'(t)$ is defined for almost every parameter $t \in [0, 1]$, since γ is Lipschitz.



(a) Local curvature $\kappa(z)$ obtained as the inverse of the local radius of curvature $r(z)$ of the osculating circle of γ at z.

(b) Menger curvature $1/R(x, y, z)$ depends on the triple of points on γ.

Fig. 2. Local curvature $\kappa(z)$ vs. Menger curvature of x, y, z.

In other words, the class $\mathcal{C}$ contains – up to scaling, which might be necessary to fix the length – about every planar curve that one can physically draw, and about every curve in $\mathbb{R}^3$ that one can imagine. Corners, cusps and self-intersections are *a priori* allowed.

We shall often use the name γ both for the parametrization itself and for the *image* $\gamma([0, 1]) \subset \mathbb{R}^3$, hoping that this does not create too much ambiguity, and using a clear, explicit distinction between the two whenever necessary.

2.1. Between thickness and classical curvature

The notion of classic local curvature at a point $z = \gamma(s)$ for a curve $\gamma \in \mathcal{C}$ requires more smoothness: if γ happens to be twice continuously differentiable (in mathematical terms we write $\gamma \in C^2$), then the local curvature at z is given by $\kappa(z) := |\gamma''(s)|$. A more geometric way of determining local curvature at the point $z \in \gamma$ is to search for the best local approximation of the curve γ near z by circles. The so-called *osculating circle* will do the job: it is tangent to γ at z, approximates γ nicely up to second order locally near z, and its radius $r(z)$, the *local radius of curvature* at z, equals the inverse local curvature; see Fig. 2(a).

In contrast to this local function, that depends on single curve points only, Karl Menger [17] considered the circumradius $R(x, y, z)$ of three curve points $x, y, z \in \gamma$ with the knowledge that the coalescent limit of $R(x, y, z)$ as x and y tend to z coincides with the local radius of curvature $r(z)$ if γ is sufficiently smooth. Besides that, Menger was 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 the circumradius Menger indeed intended to develop a purely metric geometry in contrast to classic differential geometry.[6] Motivated by computational issues in the modelling of DNA, Gonzalez and Maddocks reconsidered in [9] the circumradius function, but with a focus on capturing global features of the curve by searching for the minimal circumradius that one can find upon varying the two points y, z along γ, to obtain the global radius of curvature $\varrho_G[\gamma](x)$ at x (see (1)), or even thickness

$$\triangle[\gamma] = \inf_{x \neq y \neq z \neq x} R(x, y, z) \tag{3}$$

by varying all three points along the curve. Let us add to Menger's two insights at this point a third simple observation: in contrast to local curvature neither the circumradius, nor the global radius of curvature or thickness require any smoothness of the curve. Each of them is well-defined on the class $\mathcal{C}$ of unit loops. From the perspective of the calculus of variations this is crucial: very often one is forced to enlarge classic function spaces to more general spaces, for instance to so-called *Sobolev spaces* where derivatives exist only in a weak integral sense, in order to actually *find* minimizers of given energies. Only afterwards one can try to prove higher regularity of the minimizer. Likewise for optimization problems involving geometric

[6] Metric geometry in the sense of Menger is not part of the present survey, for further reading in that direction see, e.g. the treatise of Leonard M. Blumenthal and Menger [18], in particular Chapter 10.

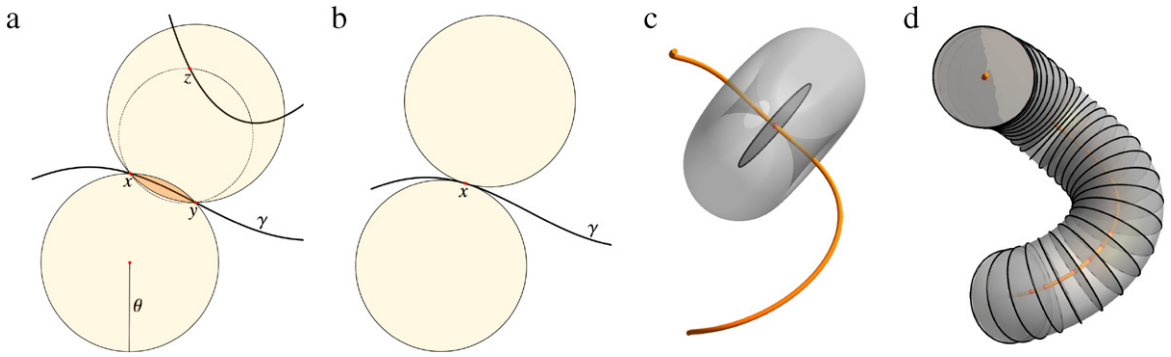


Fig. 3. Thick curves are $C^{1,1}$-manifolds. (a) A cross section of all balls of radius θ containing x and y on their boundaries. Any curve point z in the lightly shaded region would lead to a circumradius $R(x, y, z) < \theta$. (b) The limiting position of the balls in (a) as y approaches x along γ. (c) A horn torus whose cross section is shown in (b); the curve does not penetrate its interior. (d) The tubular neighbourhood of γ is formed by disjoint discs of radius θ, centred on γ.

objects: the larger the class of objects, the more likely it is that the mathematical search for minimizing configurations turns out successful.

In view of all these facts one is tempted to say that the circumradius, and minimization over selections of its arguments, might lead to geometrically defined energies that capture both local properties (like local curvature) and global behaviour of curves. This interplay between local and global control is reflected in the following theorem [19,20], which has served as the basis for analytic investigations on ideal configurations of knots and links.

Theorem 2.1 (*Thick Curves are $C^{1,1}$-Manifolds*)**.** *Unit loops $\gamma \in \mathcal{C}$ with a positive thickness $\triangle[\gamma]$ are embedded and continuously differentiable with a Lipschitz continuous tangent vector. Moreover, the curvature of γ is defined and bounded almost everywhere by $1/\triangle[\gamma]$.*

So, in mathematical terms, one finds that thick curves are, in fact, one-dimensional $C^{1,1}$-submanifolds of Euclidean 3-space, where the notation $C^{1,1}$ reflects the regularity of the tangent vector γ': it exists everywhere and is Lipschitz continuous with the estimate

$$|\gamma'(s) - \gamma'(t)| \leq \frac{1}{\triangle[\gamma]}|s - t| \quad \text{for all } s, t \in [0, 1], \tag{4}$$

where here, and in the following, $|s - t|$ denotes the *intrinsic distance* of the points $\gamma(s)$ and $\gamma(t)$ along the curve, that is, the length of the shortest subarc of γ connecting $\gamma(s)$ and $\gamma(t)$. Since Lipschitz continuous functions are – according to a classic theorem by Hans A. Rademacher – differentiable almost everywhere, one finds that local curvature $\kappa(s) = |\gamma''(s)|$ is defined and bounded by ropelength $1/\triangle[\gamma]$ for almost every parameter $s \in [0, 1]$. In other words, the energy ropelength controls local curvature and guarantees an embedding.

The geometric essence of the proof of this theorem is the following, which will at the end also justify the use of the word "thickness" for $\triangle[\gamma]$. Assume that $\theta := \triangle[\gamma]$ is positive and, for simplicity, that $\gamma \subset \mathbb{R}^3$ is embedded,[7] and consider two distinct points $x = \gamma(s)$ and $y = \gamma(t)$ of differentiability sufficiently close to each other. (By Rademacher's theorem one has many choices for these points, since almost every curve point is a point where the tangent vector exists, since γ is Lipschitz continuous.) Then we look at all three-dimensional open balls of radius θ that contain x and y in their respective boundary sphere. It turns out that γ intersects the union of all these balls only in their intersection, which forms a lens-shaped region; see Fig. 3(a). Indeed, any point $z \in \gamma$ contained in that union but not in the lens would lead to a triple $x, y, z \in \gamma$ with circumradius $R(x, y, z) < \theta$ contradicting the very definition of thickness (see (3)). Having trapped the curve γ locally in such lens-shaped regions immediately confines the tangents at x and y to the smallest double-cone with axis through $x - y$ and containing the lens, which leads to inequality (4). This uniform estimate can be readily extended to all pairs of parameters, since the parameters of differentiability form a dense set in $[0, 1]$.

In addition, if y tends to x along the curve, then this union of balls tends to a degenerate torus, an open horn torus tangent to γ at x, which contains by the previous argument no curve point; see Fig. 3(b). So, the curve is equipped with a collar of such horn tori, see Fig. 3(c), which implies the existence of a tubular neighbourhood in which the next-point projection onto the curve is uniquely defined. In other words, this tubular neighbourhood consists of the disjoint union of open planar discs of uniform radius θ, each of which is centred at and normal to the curve; see Fig. 3(d). This surrounding tube serves as an exact *excluded volume constraint*, and this is a consequence of finite ropelength $1/\triangle[\gamma]$.

There are numerous variational applications of Theorem 2.1, the most prominent of which is the existence of *ideal knots and links*, i.e., minimizers of ropelength within given knot or link classes [19–21]; see Fig. 4.

[7] If one rewrites the definition of thickness $\triangle[\gamma]$ in terms of an appropriate maximization over arclength parameters, one immediately observes that the arclength parametrization is injective for positive thickness; see [20, Definition 1 & Lemma 1].

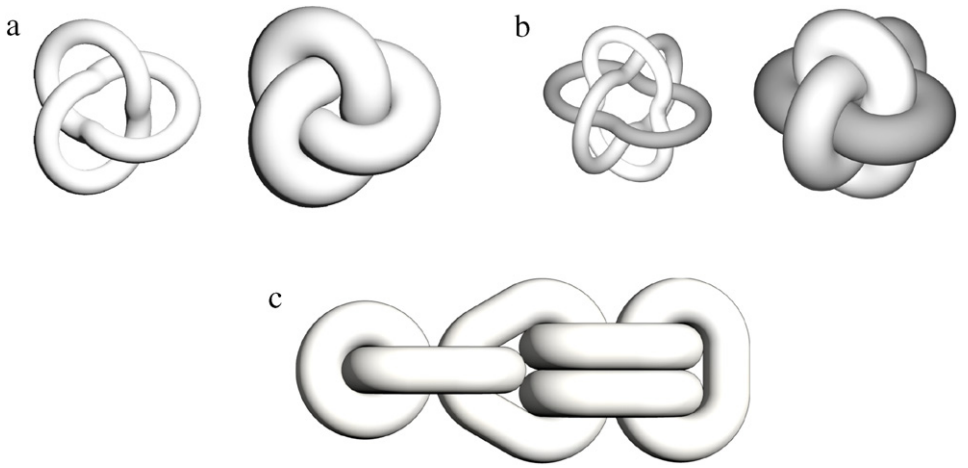


Fig. 4. (a) At present only numerical approximations of the ideal trefoil are known [22,23,12]. (b) Under natural symmetry assumptions the ideal Borromean rings are one of the most complex analytically known ideal shapes [24] (cf. [25]). (c) The depicted ideal link consists of six components and is a member of a whole family. For instance, the configuration stays ideal when rotating the loop on the left out of the drawing plane [19].
Source: Reprinted with permission from [25].

Fig. 5. Longest ropes for various prescribed thickness parameters. All curves are visualized as tubes of a fixed radius, which coincides with the actual prescribed thickness value only for the last curve in (c).

Alternatively, one can deal with elastic rods of prescribed thickness minimizing nonlinear elastic energies such as in [20]. Or one can investigate various *packing problems* for long and slender objects with a prescribed minimal thickness. As a particularly beautiful albeit mathematically idealized example, let us mention the search for the *longest rope on the unit sphere*, which boils down to the maximization of length among all (closed or open) curves on the unit sphere that have a prescribed minimal thickness. For an infinite number of given thickness values one can explicitly construct the solutions and prove their uniqueness up to rigid rotations [26,27]; see Fig. 5 and the animation for the construction in [28].

The draw back of thickness as a steric constraint and of its energy counterpart ropelength is that, as a functional acting on the space of curves, it is nonsmooth because of the pointwise maximizations involved. This has serious consequences for regularity considerations: it is highly nontrivial to apply appropriate tools from nonsmooth analysis to derive necessary conditions for minimality or criticality, e.g. Euler–Lagrange equations; see the investigations of ropelength criticality in [29,25,30,31], or for nonlinearly elastic rods in [32].

Consequently, one is naturally lead to the question if one can relax this energy functional by replacing one or several maximizations by an averaging process to increase smoothness of the functional. On the other hand, upon taking average values or integrating instead of taking pointwise maxima of $1/R(x, y, z)$ one might loose a lot of control over the regularity and the shape of curves. How much geometric information one actually has to give up in this relaxation and how much control is still present is the topic of the next sections.

2.2. Shape control. Examples

Contrary to thickness, integral Menger curvature (2) does not directly control the parametrization of a curve $\gamma\colon [0, 1] \to \mathbb{R}^3$; it does control *the shape of the image* of γ. Here is a trivial example. A doubly covered circle γ_0 of length $\frac{1}{2}$ would have finite $\mathcal{M}_p$-energy for all p, since the circumradius R is simply constant for all triples of pairwise distinct points on γ. Nevertheless, this information alone is not enough to conclude that the image of γ_0 has no multiply covered arcs. This is one of the reasons behind the definition of $\mathcal{C}$: the requirement that $\mathcal{H}^1(\gamma) = 1$ be equal to the length of the parameter domain does exclude multiple covering of whole arcs. It is not really restrictive, as we shall explain shortly.

It is clear that for $p = 3$ integral Menger curvature $\mathscr{M}_3$ is scale invariant. If one scales a given curve by a factor $\lambda > 0$, then the length, i.e., the measure in each of the three integrals in (2), scales also by λ whereas the integrand $1/R^p$ scales by λ^{-p}, i.e. the triple integral (2) remains unchanged for $p = 3$.

How do corners or cusps influence the value of integral Menger curvature? As a first step, one tries to compute the $\mathscr{M}_p$-energy for a polygonal line. Assume first we deal with the scale invariant $p = 3$. Take a polygon which consists just of two different segments with one common endpoint, say $\gamma_1 = [0, x] \cup [0, y]$ for two points $x, y \in \mathbb{R}^3$ that are not on the same straight line through the origin, so that γ_1 has a true corner at 0. Subdividing each of the segments into smaller disjoint pieces, scaled down geometrically, and then expanding the triple integral $\mathscr{M}_p$ into a series and dropping some of the terms (the non-diagonal ones), we see that

$$\mathscr{M}_p(\gamma_1) \geq \mathscr{M}_p([x/2, x] \cup [y/2, y]) + \cdots + \mathscr{M}_p([x/2^{n+1}, x/2^n] \cup [y/2^{n+1}, y/2^n]) + \cdots. \tag{5}$$

However, for $p = 3$ all the terms of the series on the right-hand side are equal due to scale invariance of $\mathscr{M}_3$, and they are non-zero, so that $\mathscr{M}_3(\gamma_1)$ is infinite. The same reasoning works for every polygonal line γ with at least one true corner, since one may simply neglect all parts of γ but the two segments forming the corner, say $\gamma_1 \subset \gamma$, and estimate $\mathscr{M}_3(\gamma) \geq \mathscr{M}_3(\gamma_1)$, which diverges as seen above. For $p > 3$ this blow-up of energy is even more drastic: the terms of the series in (5) grow to infinity. For $p < 3$ this argument fails; Sebastian Scholtes [33] shows all polygons have finite $\mathscr{M}_p$-energy if and only if $p < 3$.

A more technical variant of the above reasoning would show that $\mathscr{M}_p(\gamma)$ is infinite for each curve $\gamma\colon [0, 1] \to \mathbb{R}^3$ which is piecewise smooth, except at corners where the tangent vector has *different* one-sided limits. One would also guess that a cusp – a point where two arcs of a piecewise smooth curve γ meet tangentially – should also lead to the blow-up of integral Menger curvature, even more drastically than a corner, since the integrand blows up even faster in the neighbourhood of a cusp. Thus, one would expect that curves with $\mathscr{M}_p$ finite for some $p \geq 3$ have no self-intersections (each self-intersection would produce a corner or a cusp). This is indeed the case, as the following topological result proven in [34] shows.

Theorem 2.2. *All unit loops $\gamma \in \mathcal{C}$ with $\mathscr{M}_p(\gamma) < \infty$ for some $p \geq 3$ are homeomorphic to a circle.*

We shall explain the mechanism of the proof of Theorem 2.2 in the next subsection, but let us point out here that arbitrary closed curves in arclength parametrization with finite $\mathscr{M}_p$-energy but not necessarily in the class $\mathcal{C}$ do not quite have to behave like that. Indeed, integral Menger curvature does *not* penalize any parametrization that multiply traces out certain parts of the curve — as long as the image looks nice. The effect of finite energy then is that this image is either homeomorphic to a circle or to a closed segment, which obviously is a manifold with nonempty boundary. To see that the latter might indeed happen, recall from the beginning of this subsection the doubly-covered semicircle, a smooth one-dimensional submanifold homeomorphic to a closed segment.

2.3. Regularization and the geometry behind it

For $p > 3$, finiteness of $\mathscr{M}_p(\gamma)$ implies much more about the regularity of γ', and the reader should compare this result to Theorem 2.1 where integral Menger curvature is replaced by ropelength $1/\triangle[\gamma]$.

Theorem 2.3. *If $p > 3$ and $\gamma \in \mathcal{C}$ satisfies $\mathscr{M}_p(\gamma) \leq E < \infty$, then γ' is defined everywhere and satisfies the uniform estimate*

$$|\gamma'(t) - \gamma'(s)| \leq C(p) \left(\int_s^t \int_s^t \int_s^t \frac{1}{R^p} \right)^{1/p} |t - s|^{1-\frac{3}{p}}, \quad s < t, \tag{6}$$

whenever t and s are close enough, i.e. $|t - s| < \delta(p)E^{-1/(p-3)}$. The two constants $\delta(p)$ and $C(p)$ depend only on p.

Moreover, this is *as good as it gets*: for each Hölder exponent $\alpha > 1 - \frac{3}{p}$ there is a simple curve with $\mathscr{M}_p(\gamma) < \infty$ for which γ' fails to satisfy the Hölder estimate $|\gamma'(s) - \gamma'(t)| \lesssim |t - s|^\alpha$. Specific examples of such curves are given in Marta Szumańska's Ph.D. Thesis [35]; see also [36]. We know in fact that such behaviour is typical, since more recent work of Simon Blatt and Sławomir Kolasiński [37,38] characterizes finite energy curves as exactly those embedded curves that belong to certain fractional Sobolev spaces, which embed into the classic function space $C^{1,1-(3/p)}$ but *not* into any better $C^{1,\alpha}$. For those readers who are more familiar with the standard (nonfractional) Sobolev spaces, we mention an analogy[8]: one can interpret the integrand $1/R$ as a very weak form of discrete curvature, so that one may compare $(\mathscr{M}_p(\gamma))^{1/p}$ to the L^p-norm of second derivatives of (a sufficiently smooth curve) γ on a three-dimensional domain because of the three one-dimensional integrations in $\mathscr{M}_p$. If this L^p-norm is finite for some $p > 3$ then the classic Morrey–Sobolev embedding theorem [39, Theorem 5.4] implies that γ is indeed of class $C^{1,1-(3/p)}$. So, with this close analogy to function spaces in mind, one may view Theorem 2.3 as a *geometric variant of the Morrey–Sobolev embedding theorem.* However, a word of warning is appropriate: In contrast to ropelength, integral Menger curvature $\mathscr{M}_p$ does not control the classical curvature of γ. It can happen that $\mathscr{M}_p$ of γ is finite for some $p > 3$ yet the classical curvature of γ is *nowhere* defined.

[8] This is how Theorem 2.3 was discovered, in fact.

Let us also mention another way to interpret inequality (6). The term $|t-s|^{-3/p}$ corresponds to *averaging* of the integral; thus, (6) can be rewritten as

$$|\gamma'(t)-\gamma'(s)| \le C(p)\left(\frac{1}{\operatorname{meas}\left([s,t]^3\right)}\iiint_{[s,t]^3}\frac{1}{R^p}\right)^{1/p}|t-s|. \tag{7}$$

In other words, if $\mathscr{M}_p(\gamma)$ is finite for some $p>3$, then – in the arclength parametrization – the oscillation of the unit tangent vector γ' is controlled by the increments of length, up to a factor which depends only on the average value of the integral Menger curvature of that piece of the curve which is relevant, since the integration in (7) is performed only along the cube $[s,t]^3$ in the domain of all triples of parameters corresponding to the arc from $\gamma(s)$ to $\gamma(t)$. So, the proof of Theorem 2.3 is somewhat *semi-local*: we take into account the fact that R is a multipoint function, which secures embeddedness of the curve, but for the energy estimates leading to (7) we use only the energy contribution of a fairly small portion of the curve. We will see later in Section 4 that one can use larger parts of the curve to control specific knot invariants by means of finite integral Menger curvature.

Even if the statement of Theorem 2.3 is purely analytic, the proof and the ideas behind it are again geometric as was the case, albeit in a much simpler way, for the proof of Theorem 2.1. Let us explain some of the ideas now: during this explanation we shall encounter several consequences of finite energy that pave the way towards more sophisticated knot-theoretic properties of integral Menger curvature presented in Sections 3 and 4.

2.3.1. Beta numbers and their decay

The first step of the proof of Theorem 2.3 is to see that a curve of finite energy is (locally) confined in relatively thin and narrow tubes (for $p>3$ these tubes become thinner and thinner when scaling down). From deep mathematical work of Peter Jones [40] (see also the far reaching extensions in the monograph of Guy David and Stephen Semmes [41]) in harmonic analysis we import the technical notion of *beta numbers* defined as

$$\beta_\gamma(x,d) := \inf\left\{\sup_{y\in\gamma\cap B(x,d)}\frac{\operatorname{dist}(y,G)}{d}\colon G \text{ is a straight line through } x\right\} \quad \text{for } x\in\gamma \text{ and } d>0. \tag{8}$$

In plain words, $\beta_\gamma(x,d)$ measures how thin the *thinnest* cylinder is that contains the portion of γ in a given ball $B(x,d)$ of radius d centred at $x\in\gamma$. Dividing by d makes it dimension free: we just want to know what is the ratio of the radius to the height of that cylinder.

It turns out that control of the energy value $\mathscr{M}_p(\gamma)$ balances the scale below which the beta numbers, i.e., the widths of these cylinders, are well-controlled.

Lemma 2.4. *Fix $p\ge 3$. Let $\mathscr{M}_p(\gamma)$ be finite. There exists a constant $c_0=c_0(p)>0$ such that if $\epsilon<0.001$ and $d<\operatorname{diam}\gamma$ satisfies the* balance condition

$$\epsilon^{6+p}d^{3-p} \ge c_0(p)\mathscr{M}_p(\gamma), \tag{9}$$

then

$$\beta_\gamma(x,d)\le\epsilon \quad \text{for each } x\in\gamma.$$

Applying this lemma for ϵ and d such that the balance condition holds with an equality sign, we obtain

$$\beta(x,d)\le\epsilon=\left(c_0(p)\mathscr{M}_p(\gamma)d^{p-3}\right)^{\frac{1}{6+p}}\lesssim d^\kappa \quad \text{with } \kappa=(p-3)/(p+6).$$

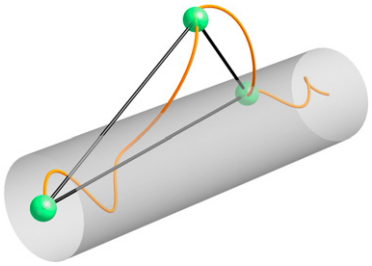
One proves Lemma 2.4 by contradiction, checking that if the narrowest tube containing γ in a given ball were too thick, than the $\mathscr{M}_p$-energy of γ would be too large. Indeed, if we had $\beta(x,d)>\epsilon$ for some $d<\operatorname{diam}\gamma$, then – see Fig. 6(a) and (c) – we would find three points $x,y,z\in\gamma\cap B(x,d)$ forming a triangle with base d and height larger than ϵd. For these three points, by elementary geometry,

$$\frac{1}{R(x,y,z)}=\frac{2\cdot\text{height}}{|x-z|\cdot|y-z|}\gtrsim\frac{\epsilon d}{d^2}=\frac{\epsilon}{d}. \tag{10}$$

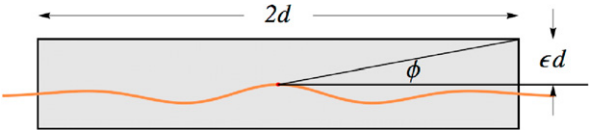
It is clear that the same estimate holds up to an absolute constant when we replace x,y,z by their sufficiently close respective neighbours, staying in the balls of radius $\epsilon^2 d\ll\epsilon d\ll d$ centred around x,y,z. Thus, estimating the total energy $\mathscr{M}_p(\gamma)$ by the portion coming from three little arcs near x, y and z – obviously, each of them of length at least $\epsilon^2 d$ – we obtain (see Fig. 6(c))

$$\mathscr{M}_p(\gamma)>\iiint_{\text{three little arcs}}\frac{1}{R^p}\gtrsim(\epsilon^2 d)^3\cdot\left(\frac{\epsilon}{d}\right)^p=\epsilon^{6+p}d^{3-p}, \tag{11}$$

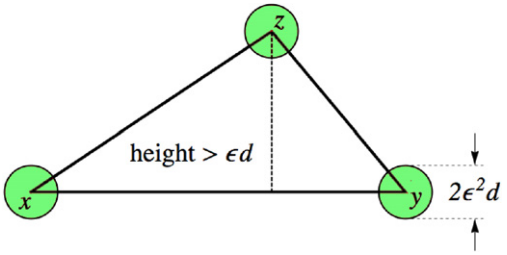
a contradiction to the balance condition.



(a) If $\epsilon < \beta_\gamma(x, d)$, then (any) tube of length $2d$ and radius ϵd is too narrow to contain $\gamma \cap B(x, d)$.



(b) For ϵ sufficiently small the angle at which the base of the tube in (a) is seen from the centre of the tube is (roughly) ϵ, as $\tan\phi = \epsilon d/d = \epsilon$.



(c) Slight perturbations of the vertices do not change $1/R$ too much, so all triangles based on three arbitrary points within small balls of radius $\epsilon^2 d$ about x, y, and z, contribute similarly to the energy.

Fig. 6. Beta numbers and the proof of Lemma 2.4.

Remark. In the scale invariant case $p = 3$ we have $p + 6 = 9$, $p - 3 = 0$. Lemma 2.4 yields then the following:

$$\sup_{x\in\gamma} \beta_\gamma(x, d) \lesssim \omega(d), \quad d \le \operatorname{diam}\gamma, \tag{12}$$

where

$$\omega(d) := \sup\left(\int_{A_1}\int_{A_2}\int_{A_3} \frac{1}{R^3}\right)^{1/9}, \tag{13}$$

the supremum being taken over all triples of subsets $A_1, A_2, A_3 \subset \gamma$ with $\mathscr{H}^1(A_i) < d$. In particular,

$$\sup_{x\in\gamma} \beta_\gamma(x, d) \to 0 \quad \text{as } d \to 0.$$

This is enough to prove Theorem 2.2, via an iterative analysis of beta numbers at small scales, see [42, Thm. 1.4] for details.

2.3.2. Scaling down: tilting tubes and double cones

As explained above, the condition $\mathscr{M}_p(\gamma) < \infty$ for $p > 3$ yields

$$\beta_\gamma(x, d) \lesssim d^\kappa \quad \text{for } x \in \gamma \text{ and } \kappa = \frac{p-3}{p+6} \in (0, 1). \tag{14}$$

Thus, if d goes to 0 geometrically, then $\beta_\gamma(x, d)$ does the same. This observation allows to iterate Lemma 2.4 and learn more about the geometry of curves with finite energy. This is the second step of proof of Theorem 2.3.

Let us fix a point $x \in \gamma$ and a number $d < \operatorname{diam}\gamma/2$. There is another point $y = y_0$ of γ on the surface of the sphere $\partial B(x, d)$. Now, follow the curve from y_0 towards x and define y_n, $n = 1, 2, \ldots$, as the points where γ hits the sphere $\partial B(x, d/2^n)$ for the first time. Write T_n to denote the narrowest tube which contains $\gamma \cap B(x, d/2^{n-1})$ and has the line through x and y_n as its axis of rotation. Then, by (14), T_n has the ratio of its radius to height at most proportional to $(d/2^{n-1})^\kappa$. Inequality (14) specifies how much the tubes $T_1, T_2, \ldots$ can tilt as the curve approaches x. Since the axis of T_{n+1} is determined by the point $y_n \in T_n$, and for small angles ϕ we have $\phi \approx \tan\phi$, the maximum total tilt angle is controlled, up to a constant depending only on p and the energy of γ, by

$$d^\kappa + \left(\frac{d}{2}\right)^\kappa + \left(\frac{d}{4}\right)^\kappa + \cdots \lesssim d^\kappa$$

(we simply sum a geometric series). This is a small number if one starts the iteration in a sufficiently small scale d.

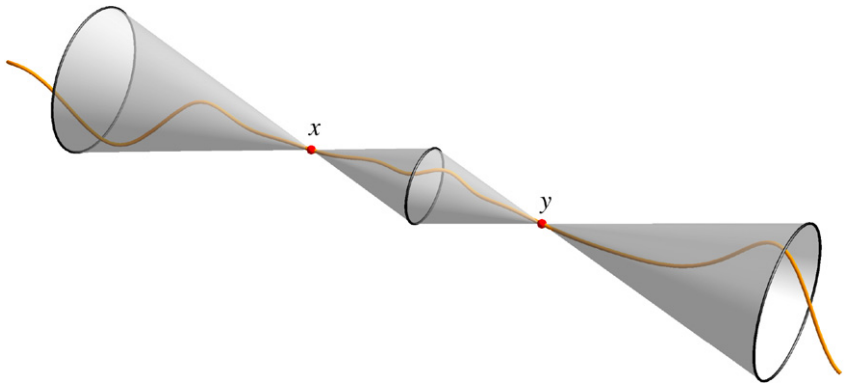


Fig. 7. The (d_0, φ)-diamond property: at small scales, the curve is trapped in a conical region and does not meander back and forth: each cross section of the cones contains exactly one point of the curve.

In fact, since the tilting tubes T_n that contain pieces of γ in $B(x, d/2^n)$ get infinitely thin as $n \to \infty$, one can deduce that $\gamma \cap B(x, d)$ is contained in a double cone with vertex at x, the axis of rotation determined by any point $y \in \partial B(x, d/2)$ and the opening angle proportional to d^κ. Replacing the roles of x and y, one checks that the curve at small scales must be trapped in the intersection of two double cones with vertices at $x, y \in \gamma$, common axis of rotation given by the line through x and y, and the opening angle $\approx |x-y|^\kappa$, see Fig. 7.

Even more is true. Since the situation persists at all scales smaller than a threshold value determined by the energy, one can check that in fact in each double cone the intersection of the curve and each plane perpendicular to the cone axis can consist of only one point. Thus, the curve does not meander back and forth. Since this property is the key to several other geometric, analytic and topological consequences, it merits a name.

For $x \neq y \in \mathbb{R}^3$ and $\varphi \in (0, \frac{\pi}{2})$ we denote by $C_\varphi(x; y)$ the double cone whose vertex is at the point x, with cone axis passing through y, and with opening angle φ, or in mathematical terms,

$$C_\varphi(x; y) := \left\{ z \in \mathbb{R}^3 \setminus \{x\} : \exists\, t \neq 0 \text{ such that } \sphericalangle(t(z-x), y-x) < \frac{\varphi}{2} \right\} \cup \{x\}.$$

Definition 2.5 (*Diamond Property*)**.** We say that a curve $\gamma \in \mathcal{C}$ has the *diamond property at scale* d_0 *and with angle* $\varphi \in (0, \pi/2)$, in short *the* (d_0, φ)*-diamond property*, if and only if for each couple of points $x, y \in \gamma$ with $|x-y| = d \leq d_0$ two conditions are satisfied: we have

$$\gamma \cap B_{2d}(x) \cap B_{2d}(y) \subset C_\varphi(x; y) \cap C_\varphi(y; x) \tag{15}$$

(cf. Fig. 7), and moreover each plane $a + (x-y)^\perp$, where $a \in B_{2d}(x) \cap B_{2d}(y)$, contains exactly one point of $\gamma \cap B_{2d}(x) \cap B_{2d}(y)$.

Using this language, one easily translates the geometric considerations above to the following.

Proposition 2.6 (*Energy Bounds Imply the Diamond Property*)**.** *Let $\gamma \in \mathcal{C}$ and $0 < E < \infty$. Assume that $\mathcal{M}_p(\gamma) \leq E$ for some $p > 3$. Then, there exist constants $\delta = \delta(p) \in (0, 1)$ and $c(p) < \infty$ (both depending only on p) such that γ has the (d_0, φ)-diamond property for each couple of numbers (d_0, φ) satisfying*

$$d_0 \leq \delta(p) E^{-1/(p-3)}, \qquad \varphi \geq c(p) E^{1/(p+6)} d_0^\kappa, \tag{16}$$

where $\kappa = (p-3)/(p+6)$.

Note that condition (16) means that one can choose the 'trapping cones' in Fig. 7 with angle $\varphi \approx |x-y|^\kappa$. Take two points $x = \gamma(t)$ and $y = \gamma'(s)$ where γ' exists. Since γ', existing a.e. by the classic theorem of Rademacher, is a unit vector for an arclength parametrization, the oscillation $|\gamma'(t) - \gamma'(s)|$ is controlled by the opening angle of the cones. Using this observation, one checks that γ' satisfies

$$|\gamma'(t) - \gamma'(s)| \lesssim |t-s|^\kappa \tag{17}$$

for all t, s in a set of full measure in $[0, 1]$. Such a set is necessarily dense. Thus, γ' can be uniquely extended to a function which still satisfies the same Hölder estimate. Elementary real analysis shows that this unique extension coincides with the derivative of γ not just *almost everywhere* but in fact everywhere. Notice the difference to the proof of Theorem 2.1, where we could use a bound on ropelength: There we were able to trap the curve in lens-shaped regions to obtain (4), here we used cruder trapping cones to establish the uniform estimate (17) which is much weaker than (4).

In order to finish the proof of Theorem 2.3, one basically has to improve now the Hölder exponent of γ' from $\kappa = (p-3)/(p+6)$ to the (optimal) $\alpha = (p-3)/p$. This is the third and last step of the proof. Here is a word of informal explanation. Suppose that a curve is just $C^{1,\alpha}$ for $\alpha = 1 - 3/p$ and not any smoother, say γ locally looks like the graph

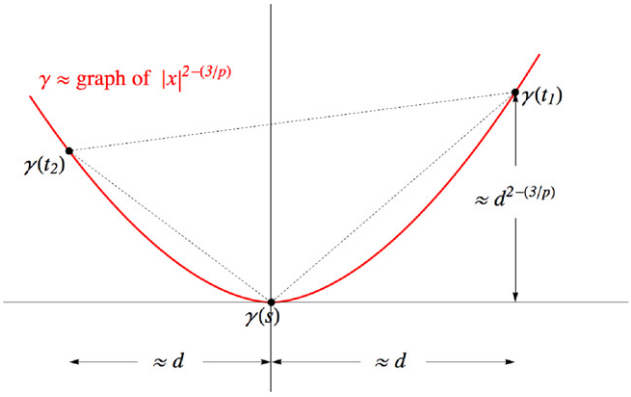


Fig. 8. The location of typical triples on the nonsmooth graph of the function $x \mapsto |x|^{2-(3/p)}$.

of $x \mapsto |x|^{2-(3/p)}$ near zero. We would then expect that typical points $\gamma(t_i)$ $(i = 1, 2)$ with $|t_1 - t_2| \approx d$ can be located roughly at the distance $d^{1+\alpha}$ from the tangent line at $\gamma(s)$ when $|\gamma(s) - \gamma(t_i)| \approx d \approx |s - t_i| \approx d$. But then, again typically, $1/R(\gamma(t_1), \gamma(t_2), \gamma(s))$ would not exceed a constant multiple of $d^{1+\alpha}/d^2 = d^{-3/p}$, just by elementary geometry and formula (10); see Fig. 8.

As we know nothing about the existence of γ'', there are no a priori upper bounds for $1/R$ that we might use, even locally. However, expecting $C^{1,\alpha}$ to be the optimal classic function space for finite energy curves, it is illustrative to look at the sets of 'bad points' where the model bound $1/R \lesssim d^{-3/p}$ is violated. A measure theoretic argument, mathematically known as *slicing*, indeed, shows that there are 'not too many' such bad points at all scales, and this is enough to conclude, since one can use the many 'good points' to determine that γ does not deviate too much from its secants. We refer the reader to [34] for the details.

2.3.3. Necklaces of disjoint cones

Leaving Theorem 2.3 and local properties of γ' aside, let us still assume $p > 3$ and stick to the situation depicted in Fig. 7 in order to see that it has some *global* consequences. Obviously, one can imagine a sequence of such small double cones – trapping the neighbouring short arcs of γ – positioned along the whole curve, with vertices evenly spaced, at sufficiently small distances, proportional to $d \propto E^{-1/(p-3)}$, where E is some constant larger than the energy $\mathscr{M}_p(\gamma)$. Think now about three of the neighbouring vertices: they do determine two double cones with a common tip. The (d_0, φ)-diamond property implies that the angle between the axes of these two double cones must be small. To see that, look back at Fig. 7 and note that if we add a third point $z \in \gamma$ with $|z - y| \le |x - y|$ to the right of y, then z must be inside the double cone $C_\varphi(y; x)$ with vertex at y and axis given $v = x - y$. Thus, the angle between the axes of two neighbouring cones, i.e. between the vectors $z - y$ and $y - x$ determined by the three consecutive vertices of the cones, is at most $\frac{\varphi}{2}$. (The assumption $|z - y| \le |y - x|$ is not restrictive at all, as we may always relabel the points and call them z, y, x instead of x, y, z). Thus, going from one cone to another one, the curve cannot make sharp turns. This is why – despite the *lack of control* of the local curvature, since the curves of finite $\mathscr{M}_p$-energy do not have to be C^2 – integral Menger curvature does yield the means to control how much the curve bends.

There is more to it. If the vertices $x_1, x_2, \ldots, x_N, x_{N+1} = x_1$ of the cones are evenly spaced along the curve, at distances $|x_i - x_{i+1}| \equiv d \propto E^{-1/(p-3)}$, then each ball $B_d(x_i)$ contains *only* the arcs of γ coming from the two double cones with common vertex at x_i, see Fig. 9. The arcs contained in all the other double cones but these two must not enter $B_d(x_i)$. A relatively simple argument that we are going to skip in order to avoid technicalities implies that all such double cones along the curve must have disjoint interiors, as depicted on Fig. 9.

Let us give, however, a precise statement of that property, since we will refer to it in the sequel. Here is the necessary notation. For $x \ne y \in \mathbb{R}^3$ we denote the closed halfspace

$$H^+(x; y) := \{z \in \mathbb{R}^3 : \langle z - x, y - x\rangle \ge 0\} \tag{18}$$

(y is contained in the interior of $H^+(x; y)$, x is on its boundary, and the boundary plane is perpendicular to $x - y$). 'Double cones' with fixed opening angles $\frac{1}{4}$ are denoted by

$$K(x, y) := C_{1/4}(x; y) \cap C_{1/4}(y; x) \cap H^+(x; y) \cap H^+(y; x). \tag{19}$$

Then, the following holds.

Lemma 2.7 (*Necklace of Disjoint Double Cones*)**.** *Suppose that $\gamma \in \mathcal{C}$ is simple and has the $(d_0, \frac{1}{4})$-diamond property. If $0 = t_1 < \cdots < t_N < 1 = t_{N+1}$ and $x_i = \gamma(t_i)$ are such that $|x_{i+1} - x_i| \le d_0$, then the open double cones*

$$K_i = \operatorname{int} K(x_i, x_{i+1}) \quad \text{and} \quad K_j = \operatorname{int} K(x_j, x_{j+1})$$

are disjoint whenever $i \ne j \pmod N$. Moreover, the vectors $v_i = x_{i+1} - x_i$ satisfy $\sphericalangle(v_{i+1}, v_i) < 1/8$.

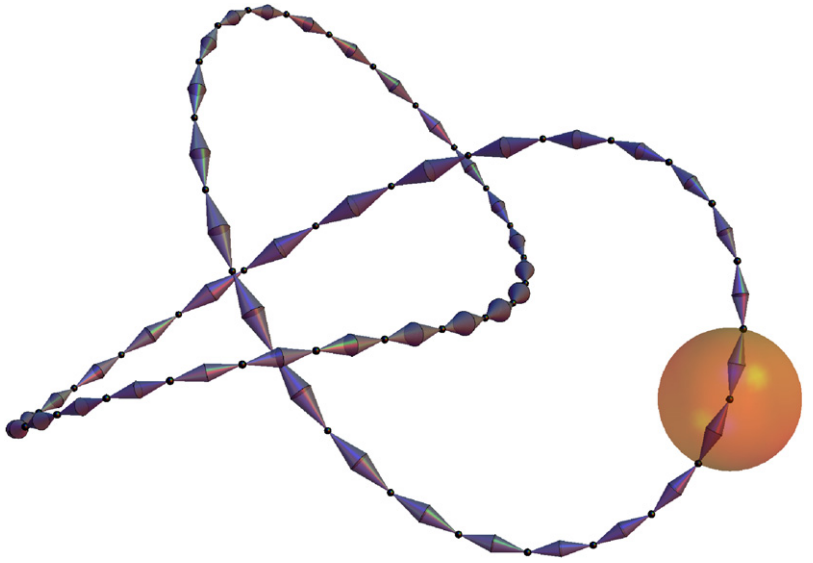


Fig. 9. The intuitive meaning of the "Necklace Lemma 2.7": small double cones with vertices along the curve have pairwise disjoint interiors. Moreover, different strands of the necklace stay well away from each other.

Remark 2.8. The number $1/4$ in the lemma has been chosen just for the sake of simplicity. The result holds in fact for any angle $\varphi \leq \frac{1}{4}$, with $\frac{1}{8}$ replaced by $\varphi/2$.

One can view the Necklace Lemma as a sort of 'weak excluded volume constraint' of curves γ with $\mathscr{M}_p(\gamma) < \infty$. It is not an exact excluded volume constraint as described in Section 2.1 for curves with finite ropelength. Here, we have no uniform tube consisting of the disjoint union of uniformly sized normal discs, and we have no unique next-point projection onto the curve in a neighbourhood of curves with finite integral Menger curvature. Already Herbert Federer [43] has shown that such an exact excluded volume constraint, in Federer's terminology *positive reach*, is equivalent to $C^{1,1}$-smoothness, which generally we do not have for curves with bounded integral Menger curvature. Nevertheless, this weaker form of excluded volume by the necklace of disjoint double cones can be used to derive crude but explicit bounds on the average number of crossings and on the so-called stick number in terms of integral Menger curvature. We shall return to that point in Sections 4.1 and 4.2.

3. Applications in geometric knot theory

Let us now specify in more detail what is meant by a *knot energy*. We follow here the definition of O'Hara, cf. [44, Def. 1.1].

3.1. Being charge: the definition of a knot energy

The crucial requirement of the definition is that you are not allowed to change the knot class if the energy stays bounded; each knot class is surrounded by infinite energy barriers as visualized in Fig. 10. Mathematically, a functional $\mathscr{E} : \mathcal{C} \to [-\infty, \infty]$ that is finite on all simple smooth loops $\gamma \in \mathcal{C}$ with the property that $\mathscr{E}(\gamma_i)$ tends to $+\infty$ as $i \to \infty$ on any sequence of simple loops $\gamma_i \in \mathcal{C}$ that converges uniformly to a limit curve with at least one self-intersection, is called *self-repulsive* or *charge*. If $\mathscr{E}$ is self-repulsive *and* bounded from below, it is called a *knot energy*.

Now, integral Menger curvature $\mathscr{M}_p$ is certainly bounded from below, and it is finite on simple smooth loops since then $1/R$ is bounded. (A priori and according to (10), $1/R$ might blow up locally, when all the points coalesce, but recall that if the curve is at least C^2, then $1/R(x, y, z)$ tends to the local curvature $\kappa(z)$ as $x, y \to z$, and local curvature is bounded for C^2-curves!)

To see that $\mathscr{M}_p$ is charge, we employ *reductio ad absurdum*.[9] Assume that a sequence of simple unit loops $\{\gamma_i\} \subset \mathcal{C}$ with uniformly bounded energy $\mathscr{M}_p(\gamma_i) \leq E < \infty$, converges uniformly, that is, in the supremum norm to $\tilde{\gamma}$ which is not a simple loop. All the γ_i have the diamond property in the same scale d_0, dictated by E, and form a bounded subset in the space $C^{1,\alpha}([0, 1])$, $\alpha = (p - 3)/p$, of C^1 curves having their derivatives Hölder continuous with exponent α. It is then a simple exercise in analysis to use Theorems 2.2 and 2.3 to prove the following.

[9] In G.H. Hardy's words: *It is a far finer gambit than any chess gambit: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game.*

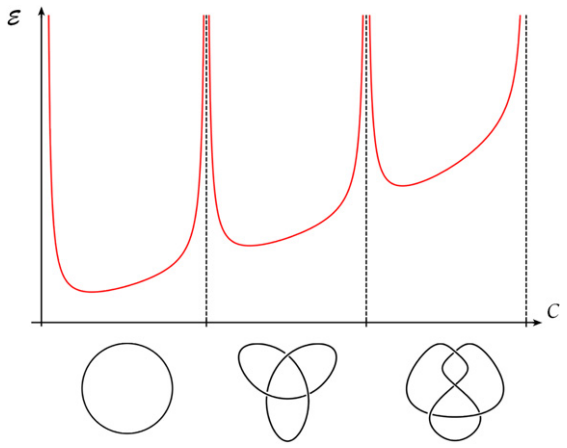


Fig. 10. The concept of a knot energy $\mathscr{E} : \mathcal{C} \to [-\infty, \infty]$: infinitely high walls separate the loops representing different knots.

Lemma 3.1 (*Quantitative Self-Avoidance*)**.** *There exists a number $\delta = \delta(E) > 0$, such that for all curves $\gamma \in \mathcal{C}$ with $\mathscr{M}_p(\gamma) < E$ we have*

$$|\gamma(s) - \gamma(t)| > \min\left\{\delta, \frac{|s-t|}{2}\right\}. \tag{20}$$

(Recall that $|s-t|$ denotes the intrinsic distance between $\gamma(s)$ and $\gamma(t)$ on the curve. One can check that the statement is satisfied with some number $\delta \propto E^{-1/(p-3)}$.) Intuitively, the lemma ascertains that (a) locally, in a scale solely dictated by the energy, γ is nearly straight, (b) bounds on the energy prevent distant strands of the curve from being too close to each other.

Now, as $\tilde{\gamma} = \lim \gamma_i$ is not simple, there exist two parameters $s \neq t \in [0, 1)$ such that $\tilde{\gamma}(s) = \tilde{\gamma}(t)$. For sufficiently large i we have

$$\min\left\{\delta, \frac{|s-t|}{2}\right\} > |\gamma_i(s) - \tilde{\gamma}(s)| + |\gamma_i(t) - \tilde{\gamma}(t)|,$$

as the left-hand side is positive and the right-hand side goes to zero as $i \to \infty$. However, since $\tilde{\gamma}(s) = \tilde{\gamma}(t)$, by the triangle inequality the right-hand side exceeds $|\gamma_i(s) - \gamma_i(t)|$. This contradicts the quantitative self-avoidance of γ_i, cf. Lemma 3.1.

So, $\mathscr{M}_p$ is indeed charge, and therefore it is a knot energy.

3.2. First impressions of the energy landscape

Let us now investigate a few other knot-theoretic properties of $\mathscr{M}_p$ that can be easily obtained from Theorem 2.3 and the geometric machinery described in Section 2.

3.2.1. Pull-tight phenomenon

For some knot energies $\mathscr{E}$ (we shall come to the examples later) knots can pull tight in a convergent sequence of loops for which $\mathscr{E}$ stays uniformly bounded from above. This *pull-tight phenomenon* is characterized by the presence of nontrivially knotted arcs $A_i \subset \gamma_i$ of a fixed knot type, each A_i being a fragment of a loop γ_i in a sequence $(\gamma_i) \subset \mathcal{C}$, with the additional property

$$A_i \subset B_i \equiv B(x_i, r_i) \subset \mathbb{R}^3, \quad \text{such that } r_i \to 0 \text{ as } i \to \infty; \tag{21}$$

see [7, Definition 1.3] and Fig. 11. In principle this phenomenon could be the cause why minimizing sequences for $\mathscr{E}$ on a fixed knot class converge to a limit that is embedded itself *but* in a different knot class. Think of it as a particularly nasty way of tunnelling from one knot class to another: you deform the curve so that it does not cross itself, and in the limit you have again a simple curve which, alas, represents a different knot since a small, topologically nontrivial, part of the original knot has been pulled tight to single point.

It should be clear by now that the pull tight phenomenon cannot happen for sequences of curves with uniformly bounded integral Menger curvature. If $\mathscr{M}_p(\gamma) < E$, and we look at any of the balls $B(x, d)$ centred on γ, with radius $d \propto E^{-1/(p-3)}$, then – due to the diamond property and Proposition 2.6 – the arc of γ contained in $B(x, d)$ is nearly straight, and has only one common point with each cross-section of the small double cone determined by two of its endpoints. So, unlike in (21), knotting cannot happen at small length scales.

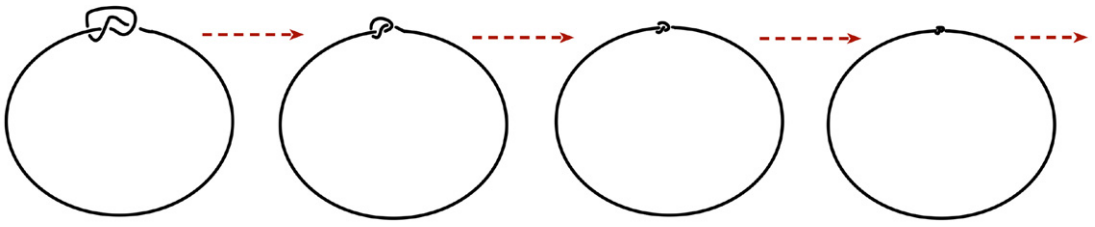


Fig. 11. The pull-tight phenomenon.

Thus, for the integral Menger curvature $\mathscr{M}_p$ with $p > 3$, the energy walls between different knot classes are indeed infinitely high and the 'pull-tight' tunnelling is forbidden. Any knot energy $\mathscr{E}$ with that more specific property, namely, such that $\mathscr{E}(\gamma_i)$ tends to $+\infty$ on every sequence $\{\gamma_i\} \subset \mathcal{C}$ with a pull-tight phenomenon, is called *tight*. In particular, $\mathscr{M}_p$ is tight for $p > 3$.

3.2.2. Existence of minimizers in all knot classes

A knot energy $\mathscr{E}$ is *minimizable*[10] if in each knot class there is at least one representative in $\mathcal{C}$ minimizing $\mathscr{E}$ within this knot class.

For functionals defined on infinite dimensional spaces, the standard way to prove existence of minimizers is to mix two ingredients: compactness of the energy sub-levels in an appropriate topology, and continuity or lower semicontinuity of the function in that topology. (In some cases, it is a delicate art to decide which topology is appropriate: ideally, it should have many compact sets and allow for many functions to be lower semicontinuous. An improvement in one direction is often a sacrifice in the other.) Luckily, for $\mathscr{M}_p$ considered on all simple unit loops, the situation is rather straightforward — thanks to our uniform estimate in Theorem 2.3.

Let us fix a knot class $[K]$. To minimize $\mathscr{M}_p$ with $p > 3$ on a given knot class $[K]$ within $\mathcal{C}$, note first that by rescaling a smooth and regular representative of $[K]$ to length one and reparametrizing by arclength, we find a representative of $[K]$ in $\mathcal{C}$. In particular, there certainly exists a minimizing sequence $\{\gamma_i\} \subset \mathcal{C}$ with $\gamma_i \in [K]$ for all $i \in \mathbb{N}$, such that

$$\lim_{i\to\infty} \mathscr{M}_p(\gamma_i) = \inf_{\mathcal{C}\cap[K]} \mathscr{M}_p.$$

The right-hand side is finite since $\mathscr{M}_p$ is nonnegative. Therefore, there is a constant E which serves as a common upper bound for all the energy values, $\mathscr{M}_p(\gamma_i) \le E$ for all $i \in \mathbb{N}$. Now, Theorem 2.3 yields the crucial uniform bound

$$\|\gamma_i\|_{C^{1,\alpha}([0,1],\mathbb{R}^3)} \le C(p, E), \qquad \alpha = 1 - \frac{3}{p}.$$

Thus, the sequences γ_i and γ_i' are equicontinuous, and by the elementary compactness theorem of Arzela–Ascoli we can extract a subsequence $\{\gamma_{i_k}\} \subset \{\gamma_i\}$ such that γ_{i_k} converges to γ in C^1, so that in particular $|\gamma'| \equiv 1$.

Since we already know that all curves with finite integral Menger curvature are simple, and that $\mathscr{M}_p$ is charge, the limit curve γ is injective. Hence

$$\mathscr{H}^1(\gamma) = \int_0^1 |\gamma'(s)|\, ds = \lim_{k\to\infty} \int_0^1 |\gamma_{i_k}'(s)|\, ds = \lim_{k\to\infty} \mathscr{H}^1(\gamma_{i_k}) = 1$$

because of the continuity of the curve length with respect to C^1 convergence. Therefore the limit curve γ is a simple loop in $\mathcal{C} \cap C^1([0, 1], \mathbb{R}^3)$. To conclude, one would now like to estimate

$$\mathscr{M}_p(\gamma) \le \liminf_{k\to\infty} \mathscr{M}_p(\gamma_{i_k}) = \inf_{\mathcal{C}\cap[K]} \mathscr{M}_p.$$

Indeed, there is nothing dangerous here: the circumradius $R(\cdot, \cdot, \cdot)$ is clearly continuous at all triples of pairwise distinct non-collinear points in $\mathbb{R}^3$, so that, for a sequence of γ_{i_k} converging in C^1 to γ, the inequality above follows from Fatou's lemma of Lebesgue integration theory.

3.2.3. Finite number of knot classes under each energy level

At this stage, one can quickly see that there are only finitely many distinct knot types under each fixed energy level $\mathscr{M}_p(\gamma) < E$. We shall prove that now using *reductio ad absurdum* again. An alternate proof, using the stick number bounds, or explicitly bounding the average crossing number, shall be discussed later on.

So, let us assume the contrary: for some E and some $p > 3$, there are infinitely many pairwise distinct knot types $[K_i]$ with representatives $\gamma_i \in \mathcal{C}$ such that $\mathscr{M}_p(\gamma_i) \le E$ for all $i \in \mathbb{N}$. Again, as in the last subsection, we invoke Theorem 2.3 and

[10] O'Hara calls this property *minimizer producing*; see [7, Definition 1.2].

Arzela–Ascoli to extract a subsequence $\gamma_{i_k} \to \gamma$ in the C^1-topology. Since we already know that $\mathscr{M}_p$ is lower semicontinuous with respect to the C^1-convergence, the inequality $\mathscr{M}_p(\gamma) \le E$ follows. In addition, by Lemma 3.1 no self-intersection or multiply covered arc can emerge in the limit, and γ is also a simple loop in $\mathcal{C}$.

To reach a contradiction, let us recall that isotopy type is stable under C^1-convergence. In the C^2-category one finds this result, e.g., in Morris Hirsch's book [45, Chapter 8], whereas the only published proofs in C^1 we are aware of, are in the papers by Philipp Reiter [46] and by Simon Blatt in higher dimensions [47]. Anyway, $[\gamma] = [\gamma_{i_k}]$ for all sufficiently large indices $k \in \mathbb{N}$, contradicting the assumption that all γ_{i_k} represent pairwise different knot types.

A knot energy $\mathscr{E}$ which has only finitely many knot types under each energy level is called *strong*. To summarize the contents of Section 3.2, we state the following.

Theorem 3.2. *For each $p > 3$, $\mathscr{M}_p$ is a charge, minimizable, tight, and strong knot energy.*

It should be clear that the Hölder regularity of γ' obtained in Theorem 2.3 is the crucial analytic tool here: this is why sequences of curves under each energy level must contain subsequences that converge not just uniformly but in the more restrictive C^1 topology which preserves length and the isotopy type of simple loops. Without Theorem 2.3 we would be lost.

There are, however, simple knot-theoretic questions concerning $\mathscr{M}_p$ that we cannot answer yet. What do the local and global minimizers look like? Can $\mathscr{M}_p$ distinguish a knot from an unknot? Both questions are imprecise; we take them up – in comparison to some other knot energies – in the next section.

3.3. Comparison to other knot energies

Once you have a knot energy $\mathscr{E}$, it would be desirable to know that it can – at the very least – distinguish a knot from an unknot, and that global minima of the energy have a particularly symmetric shape, making the unknottedness obvious. We refer to the following two properties of knot energies:

(i) A knot energy $\mathscr{E}$ *distinguishes the unknot* or is called *unknot-detecting* if the infimum of $\mathscr{E}$ over the trivial knots (the "unknots") in $\mathcal{C}$ is strictly less than the infimum of $\mathscr{E}$ over the nontrivial knots in $\mathcal{C}$.
(ii) A knot energy $\mathscr{E}$ is called *basic* if the round circle (of length one) is the unique minimizer of $\mathscr{E}$ in $\mathcal{C}$.

We do not know whether integral Menger curvature $\mathscr{M}_p$, $p > 3$, has these properties. The main difficulty we have faced when trying to obtain an answer (which we conjecture to be positive in both cases: we are tempted to believe that $\mathscr{M}_p$ is basic and does detect the unknots, and there is some numerical evidence for that which we will discuss in Section 5) is the fact that the integrand $1/R^p$ in $\mathscr{M}_p$ as a multipoint function is very nonlocal: Already the slightest perturbation of a small arc of a curve affects $1/R$ *globally*. This makes any comparison argument rather difficult.

The same questions – of being basic or unknot-detecting – have been much easier to decide for other knot energies, including several other relatives of global curvature and ropelength, and the Möbius invariant energy of O'Hara, the so-called *Möbius energy*, studied also by Michael H. Freedman, Zheng-Xu He, and Zhenghan Wang [10].

3.3.1. Knot energies interpolating between ropelength and integral Menger curvature

The global radius of curvature $\varrho_G[\gamma](x)$, defined in (1), is obtained by the infimization of $R(x, y, z)$ with respect to all points $y, z \in \gamma \setminus \{x\}$. Obviously, we have $R \ge \varrho_G[\gamma] \ge \triangle[\gamma]$. There is one more natural intermediate radius, namely

$$\varrho[\gamma](x, y) := \inf_{\substack{z \in \gamma \\ z \ne x \ne y \ne z}} R(x, y, z). \tag{22}$$

Repeated integrations over inverse powers of all these radii with respect to the remaining variables lead to other *Menger curvature energies*, as already suggested by Gonzalez and Maddocks in [9, Section 6],

$$\mathscr{I}_p(\gamma) := \int_\gamma \int_\gamma \frac{d\mathscr{H}^1(x) d\mathscr{H}^1(y)}{\varrho[\gamma](x, y)^p}, \tag{23}$$

and

$$\mathscr{U}_p(\gamma) := \int_\gamma \frac{d\mathscr{H}^1(x)}{\varrho_G[\gamma](x)^p}, \tag{24}$$

where the integration is taken with respect to the one-dimensional Hausdorff-measure $\mathscr{H}^1$. Since

$$R(x, y, z) \ge \varrho[\gamma](x, y) \ge \varrho_G[\gamma](x) \ge \triangle[\gamma],$$

the energy values on a fixed loop $\gamma \in \mathcal{C}$ are ordered as

$$\mathscr{M}_p^{1/p}(\gamma) \le \mathscr{I}_p^{1/p}(\gamma) \le \mathscr{U}_p^{1/p}(\gamma) \le \frac{1}{\triangle[\gamma]} \quad \text{for all } p \ge 1 \tag{25}$$

with the limits

$$\lim_{p\to\infty} \mathscr{M}_p^{1/p}(\gamma) = \lim_{p\to\infty} \mathscr{I}_p^{1/p}(\gamma) = \lim_{p\to\infty} \mathscr{U}_p^{1/p}(\gamma) = \frac{1}{\triangle[\gamma]}, \tag{26}$$

and each of the sequences $\{\mathscr{M}_p^{1/p}(\gamma)\}$, $\{\mathscr{I}_p^{1/p}(\gamma)\}$, $\{\mathscr{U}_p^{1/p}(\gamma)\}$ is nondecreasing as $p \to \infty$ on a fixed loop $\gamma \in \mathcal{C}$. Allowing higher order contact of circles (or spheres) to a given loop $\gamma \in \mathcal{C}$ one defines various other radii as discussed in detail in [48]. A particular example is the *tangent-point* radius

$$r_{\mathrm{tp}}[\gamma](x, y) \tag{27}$$

defined as the radius of the unique circle through $x, y \in \gamma$ that is tangent to γ at the point x. For loops $\gamma \in \mathcal{C}$ this is a correct definition for almost every $x \in \gamma$. This leads to the corresponding *tangent-point* and *symmetrized tangent-point* energy (as mentioned in [9, Section 6])

$$\mathscr{E}_p(\gamma) := \int_\gamma \int_\gamma \frac{d\mathscr{H}^1(x)d\mathscr{H}^1(y)}{r_{\mathrm{tp}}[\gamma](x, y)^p}, \qquad \mathscr{E}_p^{\mathrm{sym}}(\gamma) := \int_\gamma \int_\gamma \frac{d\mathscr{H}^1(x)d\mathscr{H}^1(y)}{\Big(r_{\mathrm{tp}}[\gamma](x, y)r_{\mathrm{tp}}[\gamma](y, x)\Big)^{p/2}}, \tag{28}$$

to complement the list of Menger curvature energies on $\mathcal{C}$. As the tangent-point radius $r_{\mathrm{tp}}[\gamma](x, y)$ is certainly not smaller[11] than the double radius $\varrho[\gamma](x, y)$ defined by (22), we can complement the order of integrals in (25) by the following inequalities

$$(\mathscr{E}_p^{\mathrm{sym}})^{1/p}(\gamma) \le \mathscr{E}_p^{1/p}(\gamma) \le \mathscr{I}_p^{1/p}(\gamma) \le \mathscr{U}_p^{1/p}(\gamma) \le \frac{1}{\triangle[\gamma]} \quad \text{for all } p \ge 1\,; \tag{29}$$

the leftmost inequality follows easily from $ab \le (a^2 + b^2)/2$.

The self-avoidance and regularization properties, described for the integral Menger curvature $\mathscr{M}_p$ in Section 2, persist for the curvature-related energies mentioned above. The overall scheme of reasoning that one uses to check this is, basically, pretty similar to the one described in Section 2.3; one has to adjust numerous technical details and we shall not dwell too much on that. Let us just mention two notable differences.

The first one, rather obvious, is that the scale invariant exponents are different.

For all the energies that involve double integration, i.e. for $\mathscr{I}_p$, $\mathscr{E}_p$ and $\mathscr{E}_p^{\mathrm{sym}}$, the scale invariant exponent is $p = 2$. The analogue of Theorem 2.2 holds: one can prove that if any of these double integral energies of a Lipschitz curve γ is finite for some $p \ge 2$, then the image of γ is a one-dimensional topological manifold. If $\mathscr{I}_p$, $\mathscr{E}_p$ or $\mathscr{E}_p^{\mathrm{sym}}$ is finite for a loop $\gamma \in \mathcal{C}$ and $p > 2$, then γ' exists everywhere and is Hölder continuous with exponent $\alpha_2 = 1 - \frac{2}{p}$; see [49,42,50]. Counterparts of the (d_0, φ)-diamond property, cf. Proposition 2.6, do also hold, for distances d_0 and angles φ satisfying an appropriate variant of the balance condition (16), namely

$$d_0 \lesssim E^{-1/(p-2)}, \qquad \varphi \gtrsim E^{1/(p+4)} d_0^{\kappa_2}, \quad \text{where } \kappa_2 = (p-2)/(p+4),$$

and where E denotes the upper bound for a given knot energy. The reader who wishes to think just in terms of pictures is again invited to visualize the necklace of double cones from Fig. 9. Quantitatively, the relation between the energy bound E and the size and proportion of the cones is (formally) different. Qualitatively, the picture is the same for $\mathscr{I}_p$, $\mathscr{E}_p$ and $\mathscr{E}_p^{\mathrm{sym}}$: the energy value specifies a scale d_0 below which there is no knotting, the curve is nearly straight. Moreover, the energy controls the bending (though it does not control the second derivative at all) at small and intermediate length scales. Thus, the following is true.

Theorem 3.3. *Assume $p > 2$. Each of the three energies $\mathscr{I}_p$, $\mathscr{E}_p$ and $\mathscr{E}_p^{\mathrm{sym}}$ is charge, minimizable, tight and strong.*

For $\mathscr{U}_p$, i.e., for the integral of p-th power of the global curvature $\kappa_G[\gamma](x) = 1/\varrho_G[\gamma](x)$, the scale invariant exponent is $p = 1$. Here, one can show, cf. [51], that all curves $\gamma \in \mathcal{C}$ are simple and have continuous tangents, also for $p = 1$. It turns out that the integrability of global curvature prevents self-intersections[12] and yields good control of the oscillations of the tangent vector, via the uniform estimate

$$|\gamma'(s) - \gamma'(t)| \le \int_s^t \frac{1}{\varrho_G(\gamma(\tau))}\, d\tau \le |s - t|^{1-\frac{1}{p}} \left(\int_s^t \frac{1}{\varrho_G(\gamma(\tau))^p}\, d\tau\right)^{1/p} \le |s - t|^{1-\frac{1}{p}}\, \mathscr{U}_p(\gamma)^{1/p}.$$

This is a source of one contrast between $\mathscr{U}_p$ and other energies discussed above: due to the left-most inequality, $\gamma'\colon [0, 1] \to \mathbb{R}^3$ is an absolutely continuous function, therefore the second derivative γ'' exists almost everywhere on $[0, 1]$. Thus, finiteness of $\mathscr{U}_p$-energy implies that the local curvature of γ is defined almost everywhere, and is dominated by the global curvature $\kappa_G = 1/\varrho_G$. However, assuming $p = 1$ is not enough for compactness arguments similar to those that we have earlier described for $\mathscr{M}_p$. For $p > 1$ the following holds.

[11] One deals with a limit $z \to x$ in the definition of $r_{\mathrm{tp}}(x, y)$ instead of the infimum over all z in the definition of $\varrho[\gamma](x, y)$.

[12] The reader might try and prove it by hand; if the curve has a double point, then the integral $\int \kappa_G$ diverges at least as fast as $\int (1/x)$ near 0.

Theorem 3.4. *Assume $p > 1$. The $\mathscr{U}_p$-energy is charge, minimizable, tight and strong.*

The second important difference between the integral Menger curvature $\mathscr{M}_p$ and all the other energies related to the circumradius R via a mixture of maximizations and integrations is that in all the cases different from $\mathscr{M}_p$ it is easier – due to maximization which carries strict global pointwise information and not just the averaged one – to control the behaviour of $\mathscr{U}_p$, $\mathscr{I}_p$, $\mathscr{E}_p$ and $\mathscr{E}_p^{\mathrm{sym}}$ on *some* curves. As a consequence, the following results hold [52].

Theorem 3.5. *Assume that $p > 2$. The energies $\mathscr{I}_p$, $\mathscr{E}_p$ and $\mathscr{E}_p^{\mathrm{sym}}$ are basic.*

Theorem 3.6. *Assume that $p \geq 1$. The energy $\mathscr{U}_p$ is basic and unknot-detecting.*

The main ingredient behind Theorem 3.5 is the relation between $\mathscr{E}_2^{\mathrm{sym}}$ and the average crossing number; we shall come back to that in Section 4. For $\mathscr{U}_p$, being basic is related to a simple isoperimetric inequality, see [51, Section 3]. To see why $\mathscr{U}_p$ does detect the unknots, we need to compare it with yet another well-known energy, the total curvature.

3.3.2. The total curvature

To see that $\mathscr{U}_p$ does detect the unknots, one has to quote a celebrated result in classic differential geometry: the Farý–Milnor theorem [53,54] which ascertains that for a nontrivially knotted curve the *total curvature*, defined as the integral of the absolute value of curvature along the curve,

$$TK(\gamma) := \int_0^1 |\kappa(s)|\, ds, \quad \gamma \in \mathcal{C},$$

must be at least 4π, whereas for the unknots the absolute minimum of $\int |\kappa|$ is equal to 2π. Since we already know that $1/\varrho_G$ dominates the local curvature, it is easy to conclude that

$$4\pi \leq \int_\gamma |\kappa|\, ds \leq \mathscr{U}_1(\gamma) \leq \mathscr{U}_p(\gamma)^{1/p} \quad \text{for each nontrivially knotted } \gamma \in \mathcal{C},\ p \geq 1,$$

whereas, since $\mathscr{U}_p$ is basic,

$$\inf\left\{\mathscr{U}_p(\gamma)^{1/p} \colon \gamma \in \mathcal{C} \text{ is unknotted}\right\} = \mathscr{U}_p(\text{round circle})^{1/p} = 2\pi.$$

Thus, $\mathscr{U}_p$ is unknot-detecting because it is minimized (only) by round circles and it dominates the total curvature which is unknot-detecting by Farý–Milnor theorem.

Notice that total curvature itself is *not* a reasonable knot energy: being unknot-detecting is its only property from the list we have discussed so far. Since the curvature κ is determined *locally* by the parametrization of the curve, *TK* does not even detect self-intersections. On the class of simple loops in $\mathcal{C}$ it is neither charge, nor strong, nor minimizable. It certainly is not basic, as it only measures the amount of turning: each convex curve γ in the plane has the total curvature of 2π according to a well-known theorem of Werner Fenchel.

3.3.3. The Möbius energy. A summary

One of the first knot energies studied in detail was the *Möbius energy*, introduced by O'Hara in [13],

$$\mathscr{E}_{\mathrm{Möb}}(\gamma) := \int_0^1 \int_0^1 \left\{ \frac{1}{|\gamma(s) - \gamma(t)|^2} - \frac{1}{|s-t|^2} \right\} ds\, dt \qquad \text{for } \gamma \in \mathcal{C}, \tag{30}$$

which is nonnegative, since the intrinsic distance $|s-t|$ of the two curve points $\gamma(s)$, $\gamma(t)$ always dominates the extrinsic Euclidean distance $|\gamma(s)-\gamma(t)|$. (The term $1/|\gamma(s)-\gamma(t)|^2$ alone, modelling the situation where each two curve points repel each other with a force analogous to the electrostatic one, would produce a divergent integral, hence the regularization[13]). That this energy is indeed self-repulsive is proven in [44, Theorem 1.1] and [10, Lemma 1.2]. However, O'Hara observed in [44, Theorem 3.1] that the Möbius energy is not tight. This is related to the property behind the name of this energy: $\mathscr{E}_{\mathrm{Möb}}$ is Möbius invariant, i.e. invariant under all transformations of the curve which belong to the Möbius group, generated by scalings, translations *and* inversions of $\mathbb{R}^3$ with respect to spheres.

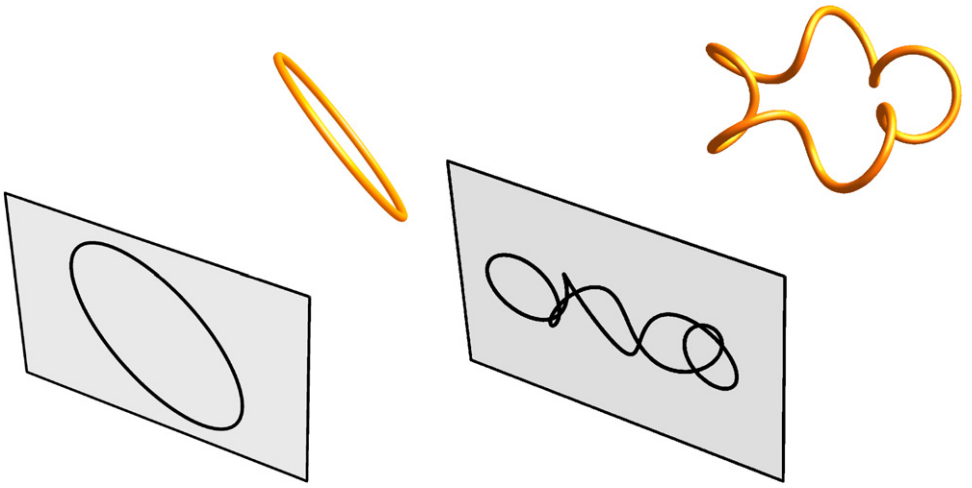
Freedman, He and Wang [10] used this invariance to establish the existence of $\mathscr{E}_{\mathrm{Möb}}$-minimizing knots but only restricted to *prime knot classes*. They have shown that $\frac{1}{2\pi}\mathscr{E}_{\mathrm{Möb}}(\gamma)$ dominates the minimal number of crossings of the knot that is represented by γ; this implies that $\mathscr{E}_{\mathrm{Möb}}$ is strong. Finally, they have also demonstrated that loops minimizing $\mathscr{E}_{\mathrm{Möb}}$ must be of class $C^{1,1}$ (= have Lipschitz continuous derivative). Later on, He [56] improved this, demonstrating – via heavy analytic methods applied to the gradient flow of $\mathscr{E}_{\mathrm{Möb}}$ – that all loops minimizing $\mathscr{E}_{\mathrm{Möb}}$ are in fact infinitely smooth. Simon Blatt, Philipp Reiter, and Armin Schikorra have very recently shown that even all critical points of $\mathscr{E}_{\mathrm{Möb}}$ are C^∞-smooth [57].

The properties of the energies $\mathscr{E}\colon \mathcal{C} \to \mathbb{R}$ discussed so far are summarized in a compact way in Table 1. All the presented properties concerning ropelength $1/\triangle$ can be derived from the corresponding properties of $\mathscr{U}_p$, but they were established long before Menger curvatures were systematically studied; see, e.g., [5,58,59,19,20,60].

[13] There is certainly some ambiguity of what kind of regularization one should work with for repulsive potentials, and this ambiguity is one of the central arguments of Jayanth R. Banavar et al. [55] to propose serious analytic (and numerical) research for global curvature and integral Menger curvature.

Table 1
The comparison of integral Menger curvature (and other energies related to the global radius of curvature) to ropelength $1/\triangle$, Möbius energy of O'Hara and total curvature.

Is the energy:	$\mathcal{M}_{p>3}$	$\mathcal{I}_{p>2}$	$\mathcal{U}_{p>1}$	$\mathcal{E}_{p>2}$	$\mathcal{E}^{\mathrm{sym}}_{p>2}$	$1/\triangle$	$\mathcal{E}_{\mathrm{Möb}}$	*TK*
Charge	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Minimizable	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Tight	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Strong	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Unknot-detecting	?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic	?	?	Yes	?	?	Yes	Yes	No



(a) A round circle always projects to an ellipse. (b) A projection of an unknot *can* have many crossings.

Fig. 12. The average crossing number is not a topological invariant.

4. Controlling knot invariants

In this section, we explain in more detail how integral Menger curvature is related to other quantities, considered in knot classic theory and geometric knot theory.

4.1. Average crossing number estimates

The average crossing number of a knotted loop $\gamma \in \mathcal{C}$ is defined as follows. For each unit vector $v \in \mathbb{S}^2$ one projects γ orthogonally onto the two-dimensional plane $P_v = (v)^\perp$ perpendicular to v, and counts the number $n(\gamma, v)$ of crossings of the planar curve obtained in P_v. (See Fig. 12.) The average crossing number acn γ is defined simply as the average of $n(\gamma, v)$ over all directions $v \in \mathbb{S}^2$, i.e.

$$\operatorname{acn}(\gamma) = \frac{1}{4\pi} \int_{\mathbb{S}^2} n(\gamma, v)\, d\mathcal{H}^2(v). \tag{31}$$

As Freedman, He and Wang explain in [10, Section 3], there is another – often more handy – formula for $\operatorname{acn}(\gamma)$. Namely, one has

$$\operatorname{acn}(\gamma) = \frac{1}{4\pi} \iint_{[0,1]\times[0,1]} \frac{\left|\det(\gamma'(s), \gamma'(t), \gamma(s) - \gamma(t))\right|}{|\gamma(s) - \gamma(t)|^3}\, ds\, dt \quad \text{for any } \gamma \in \mathcal{C}. \tag{32}$$

To see that both formulae define the same quantity, identify $[0, 1]$ with the circle $\mathbb{S} = \mathbb{R}/\mathbb{Z}$ of length one, and take the map

$$F\colon \mathbb{S} \times \mathbb{S} \setminus \{(t, t)\colon t \in [0, 1]\} \to \mathbb{S}^2, \qquad F(s, t) = \frac{\gamma(s) - \gamma(t)}{|\gamma(s) - \gamma(t)|}.$$

A computation shows that the absolute value of the Jacobian determinant of this map is equal to the integrand in (32). Therefore, the right-hand side of (32) is just the area covered by F on the unit sphere $\mathbb{S}^2$ – counting (unsigned) multiplicities – divided by 4π, i.e. by the area of $\mathbb{S}^2$. On the other hand, for a fixed vector $v \in \mathbb{S}^2$, the number $n(\gamma, v)$ of crossings of the projection of γ onto $P_v = (v)^\perp$ is equal to the number of points in the preimage $F^{-1}(\{v\})$. Thus, according to the area formula [61, Theorem 3.2.3] one computes in (31) and (32) in fact the same area and then divides by 4π.

Please note that the average crossing number is *not* a topological invariant of a knot. A flat circle γ_0 has $\operatorname{acn}(\gamma_0) = 0$; a complicated unknot can have the average crossing number as high as one wishes. The *crossing number* of a knot is the minimum of $n(\gamma, v)$ over all simple loops γ representing that knot and over all directions $v \in \mathbb{S}^2$. For simple loops $\gamma \in \mathcal{C}$, the average crossing number obviously dominates the crossing number of the knot-type represented by γ, and the crossing number is a knot invariant.

To see that bounds on integral Menger curvature $\mathscr{M}_p$, or in fact on any of the energies $\mathscr{I}_p$, $\mathscr{E}_p$ and $\mathscr{E}_p^{\mathrm{sym}}$ discussed in the previous section, imply rough but direct bounds on the crossing number, we shall discuss a crude estimate of the average crossing number for curves that have the diamond property. Here is a technical statement.

Proposition 4.1. *Let $\gamma \in \mathcal{C}$. Assume that there exists d_1 such that for each $d \le d_1$ the curve γ satisfies the $\big(d, \varphi(d)\big)$-diamond property, where $\varphi(d) = Ad^\alpha$ for some $\alpha \in (\frac{1}{2}, 1]$ and $\varphi(d_1) \le \frac{1}{4}$. Then the average crossing number of the curve is finite and there exist two absolute constants c_1 and c_2 such that*

$$\operatorname{acn}(\gamma) < \frac{A^2 c_1}{2\alpha - 1} d_1^{2\alpha - 1} + c_2 d_1^{-\frac{4}{3}}. \tag{33}$$

Before translating this estimate into an inequality, relating the energy bounds directly to the average crossing number, let us mention that the general idea of proof of (33) is analogous to Gregory Buck and Jonathan Simon's papers [5, Cor. 4.1] and [58, Cor. 2.1]. We split the integral expressing the average crossing number into two parts; one of them, the local contribution, can be controlled using the local smoothness properties of the curve; the other one takes into account the interactions of distant portions of the curve. These long-range interactions have not really been taken into account to prove our key-estimate (6) in Theorem 2.3, but the diamond property as a weak excluded volume constraint can be used here. Indeed, Lemma 2.7 and the necklaces of double cones provides an excluded volume that prevents tight stuffing of many strands in nested thin spherical shells. Such stuffing would contribute a lot to the average crossing number.

More precisely, let us split

$$\operatorname{acn}(\gamma) = I_{\mathrm{close}} + I_{\mathrm{distant}}$$

where

$$I_{\mathrm{close}} = \frac{1}{4\pi} \int_{\mathbb{S}^1} \int_{\{t \in \mathbb{S}^1 \,:\, |s-t| \le d_1\}} \frac{\big|\det(\gamma'(s), \gamma'(t), \gamma(s) - \gamma(t))\big|}{|\gamma(s) - \gamma(t)|^3}\, dt\, ds$$

and

$$I_{\mathrm{distant}} = \frac{1}{4\pi} \int_{\mathbb{S}^1} \int_{\{t \in \mathbb{S}^1 \,:\, |s-t| > d_1\}} \frac{\big|\det(\gamma'(s), \gamma'(t), \gamma(s) - \gamma(t))\big|}{|\gamma(s) - \gamma(t)|^3}\, dt\, ds.$$

It is easy to estimate the local term I_{close}. Since the curve γ satisfies the (d, φ)-diamond property, the arc of γ between $\gamma(s)$ and $\gamma(t)$ is trapped in the double cone with vertices at $\gamma(s)$ and $\gamma(t)$ and the opening angle $\varphi = A|\gamma(s) - \gamma(t)|^\alpha$. Thus, the three unit vectors $w_1 = \gamma'(s)$, $w_2 = \gamma'(t)$ and $w_3 = (\gamma(s) - \gamma(t))/|\gamma(s) - \gamma(t)|$ belong to the same cone with opening angle φ, and therefore, using the geometric interpretation of the determinant as the volume of the parallelepiped spanned by three vectors, one quickly obtains

$$|\det(\gamma'(s), \gamma'(t), \gamma(s) - \gamma(t))| \le |\gamma(s) - \gamma(t)| \sin^2 \varphi \le A^2 |\gamma(s) - \gamma(t)|^{1+2\alpha}$$

where, by assumption, we can use $\varphi = C|\gamma(s) - \gamma(t)|^\alpha$. Thus, for $|s - t| \le d_1$, when the curve points $\gamma(s)$ and $\gamma(t)$ are at a distance at most d_1, the integrand in (32) is at most

$$A^2 |\gamma(s) - \gamma(t)|^{2\alpha - 2}.$$

Since at small distances γ is nearly straight, we have in fact $|\gamma(s) - \gamma(t)| \approx |s - t|$ in this regime, and Therefore,

$$I_{\mathrm{close}} \lesssim \int_{\mathbb{S}^1} \int_{s - d_1}^{s + d_1} |s - t|^{2\alpha - 2} dt\, ds = 2 \frac{d_1^{2\alpha - 1}}{2\alpha - 1}.$$

(Here, to obtain convergence, the assumption $\alpha > \frac{1}{2}$ on the local Hölder exponent of the derivative is necessary). This estimate of I_{close} corresponds to the first term on the right hand side of (33).

Here is the rough idea how to the estimate the integral I_{distant}. We split this integral into the terms so that in each term the distance $|\gamma(s) - \gamma(t)|$ is roughly constant, equal to a fixed multiple of d_1, and then use a brute force estimate of the determinant, yielding

$$\begin{aligned} I_{\mathrm{distant}} &= \frac{1}{4\pi} \sum_{k=1}^{N} \int_{\mathbb{S}^1} \int_{\{t \in \mathbb{S}^1 : |\gamma(s) - \gamma(t)| \approx k \cdot d_1\}} \frac{|\det(\gamma'(s), \gamma'(t), \gamma(s) - \gamma(t))|}{|\gamma(s) - \gamma(t)|^3} dt\, ds \\ &\lesssim \sum_{k=1}^{N} \int_{\mathbb{S}_1} \int_{\{t \in \mathbb{S}^1 : |\gamma(s) - \gamma(t)| \approx k \cdot d_1\}} |\gamma(s) - \gamma(t)|^{-2}\, dt\, ds. \end{aligned}$$

To see how large the above sum is, we do three things. First, we assume the worst case scenario: the curve is packed as densely as possible around each of its points $\gamma(s)$. (Then, the integrand $|\gamma(s)-\gamma(t)|^{-2}$ is large on sets that are as large as possible.) Second, we fix a necklace of double cones along γ, with vertices x_i at distances between $d_1/2$ and d_1, and *fixed* opening angles equal to $\frac{1}{4}$. The length of γ in a spherical shell $S_k = \{t \in \mathbb{S}^1 : |\gamma(s)-\gamma(t)| \approx k \cdot d_1\}$ having the fixed point $\gamma(s)$ as its centre and the radii, say, $(k-1)d_1$ and $(k+1)d_1$ is, roughly, proportional to $N_k \cdot d_1$ where N_k is the number of double cones of the necklace falling into S_k. Third, since the cones in the necklace are disjoint, the sum of their volumes – which is proportional to the total length of the axes times d_1^2, i.e. to $N_k d_1^3$ – cannot exceed the volume of the shell which is proportional to $(kd_1)^2 \cdot d_1$. Comparing these two estimates, we conclude that

$$\text{length of } \gamma \cap S_k \approx N_k \cdot d_1 \le k^2\, d_1. \tag{34}$$

Plugging this inequality into the sum that dominates the integral I_{distant} and noting that in each term of that sum $|\gamma(s)-\gamma(t)|^{-2} \approx (kd_1)^{-2}$, we obtain

$$I_{\text{distant}} \lesssim \sum_{k=1}^{N} \int_{\mathbb{S}_1} \int_{\{t\in\mathbb{S}^1 : |\gamma(s)-\gamma(t)|\approx k\cdot d_1\}} |\gamma(s)-\gamma(t)|^{-2}\, dt\, ds \lesssim \frac{1}{d_1}\cdot N. \tag{35}$$

The last step is to estimate the number N of the terms in the sum we just dealt with. The worst case scenario (of dense packing) we assumed means that the length of γ in S_k in (34) is in fact *proportional* to $k^2 d_1$. Then, computing the total length of the loop gives

$$\mathscr{H}^1(\gamma) = 1 \lesssim \sum_{k=1}^{N} k^2 d_1 \lesssim N^3 d_1,$$

which means that the worst scenario of dense packing occurs for $N \propto d_1^{-1/3}$, and (35) translates to

$$I_{\text{distant}} \lesssim \frac{1}{d_1}\cdot N \lesssim d_1^{-4/3},$$

up to an absolute constant. This is the second term of the inequality (33).

Using the quantitative relation between the energy bounds for $\mathscr{M}_p$ and the sizes of the double cones, stated earlier in Proposition 2.6, one can express the constants d_1 and A from Proposition 4.1 as

$$d_1 = \delta(p) E^{-\beta}, \qquad A = c(p) E^{\alpha\beta},$$

where $\beta = 1/(p-3)$ and $\alpha = (p-3)/(p+6)$. Inserting the above quantities into formula (33), and next using the elementary inequality $E^\beta \le 1 + E^{4\beta/3}$, we obtain the following direct corollary.

Corollary 4.2. *Let $\gamma \in \mathcal{C}$ and $0 < E < \infty$. If $\mathscr{M}_p(\gamma) < E$ for some $p > 12$ then there exist constants $c_1(p)$ and $c_2(p)$, such that*

$$acn(\gamma) < c_1(p) + c_2(p) E^{\frac{4}{3(p-3)}}. \tag{36}$$

Notice that the requirement $p > 12$ guarantees that the assumption $\alpha > 1/2$ in Proposition 4.1 is satisfied, which was necessary to control the local contribution I_{close} of the average crossing number. In addition, one can compare inequality (36) with the corresponding average crossing number estimate in terms of ropelength of Buck and Simon [5, Corollary 4.1] and [58, Corollary 2.1]:

$$\operatorname{acn}(\gamma) \le \frac{11}{4\pi}\cdot\left(\frac{1}{\triangle[\gamma]}\right)^{4/3}. \tag{37}$$

Although our constants $c_1(p)$ and $c_2(p)$ in (36) are quite large (in comparison to the $11/(4\pi)$ in (37)) we see the characteristic power $4/3$ for the energy in both estimates: as $p \to \infty$ in (36), the energy term behaves like $E^{4/(3p)}$, which bounds $(\mathscr{M}_p(\gamma))^{4/(3p)}$, and the latter converges to $(1/\triangle[\gamma])^{4/3}$ as $p \to \infty$. Examples of infinite families of curves such that

$$\operatorname{acn}(\gamma) \approx \left(\frac{1}{\triangle[\gamma]}\right)^{4/3}$$

are given e.g. by Jason Cantarella, Rob Kusner and John Sullivan in [62]. (It is not difficult to imagine one: fix a solid rotational torus T with both radii $\approx 1/n$, and consider a tightly packed curve γ in T which represents the $(n, n-1)$-torus knot; since a cross section of T has the area $\approx 1/n^2$ and contains n equal cross sections of the thick tube centred on γ, the thickness of γ must be $\approx \sqrt{1/(n^2\cdot n)} = 1/n^{3/2}$. Thus, the right hand side of (37) is roughly n^2. The left hand side is also $\approx n^2$, as the crossing number of the $(n, n-1)$-torus knot is $n(n-2) \approx n^2$.)

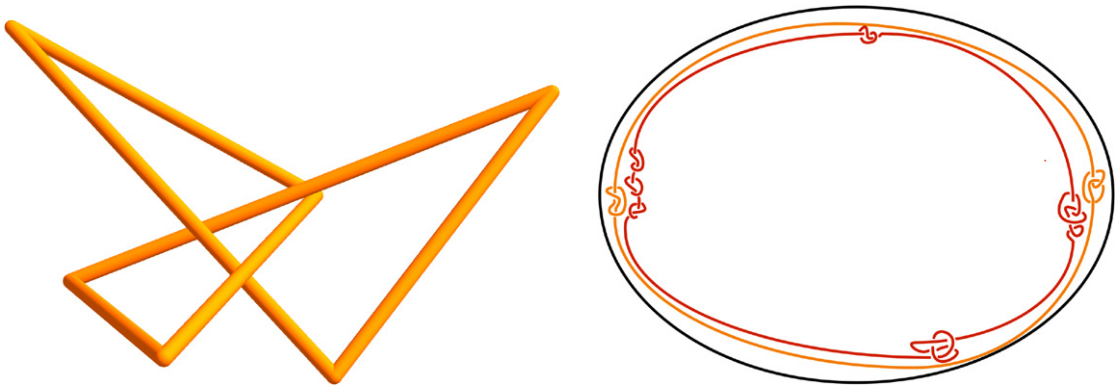


Fig. 13. Left: The stick number of the trefoil knot equals six. Right: Hausdorff distance alone does not distinguish the knot type.

4.2. Stick numbers

The *stick number* $\mathrm{seg}[K]$ of a knot-type $[K]$ is the minimal number of segments (="sticks") needed to construct a polygonal representative of $[K]$. It is clear that three sticks never suffice to tie a polygonal knot, since the resulting triangle is a planar (and therefore unknotted) curve. Looking at possible planar projections of a polygonal knot with very few segments one can actually show that one needs at least six sticks to tie a nontrivial polygonal knot; see the nice argument in [63, Section 1.6] and Fig. 13(left).

For a very complex knot-type it is supposedly very difficult to determine the stick number, but if that knot-type can be represented below certain energy levels for an energy that controls some features of the curve, then one might hope to at least bound the stick number from above. That this is indeed possible for ropelength, because of its control of local curvature and the exact excluded volume constraint, was shown by Richard A. Litherland, Jonathan K. Simon, Oguz C. Durumeric, and Eric J. Rawdon in [59]:

$$\mathrm{seg}[K] \leq \left\lfloor \frac{1}{\pi} \cdot \frac{1}{\triangle[\gamma]} \right\rfloor + 1, \tag{38}$$

where $\lfloor a \rfloor$ denotes the largest integer below the real number a. This estimate implies a bound on the number of knot-types representable below given values for ropelength, since the stick number is strongly related to the crossing number, and explicit bounds can be derived as in [10, Section 3].

Integral Menger curvature, on the other hand, does not control local curvature, but due to the diamond property (see Definition 2.5) that controls the amount of local bending and serves as a weak excluded volume constraint, one can hope that stick numbers may be estimated. That this is indeed the case follows from the following result.

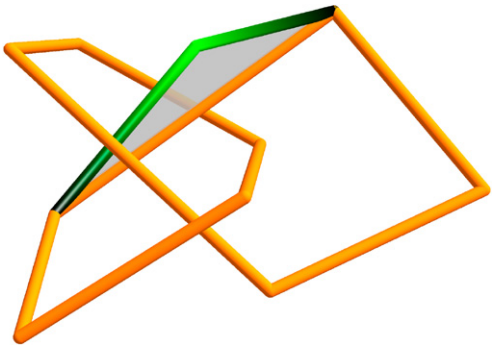
Theorem 4.3. *A unit loop $\gamma \in \mathcal{C}$ with $\mathcal{M}_p(\gamma) \leq E < \infty$ for some $p > 3$ represents the same knot-type as any inscribed polygon of which the edge-length is bounded by $\delta_1(p)E^{1/(3-p)}$, where $\delta_1(p) \in (0, 1)$ is an absolute constant depending only on p.*

Since the maximal number of edges is proportional to the inverse of the maximal edge-length one immediately deduces the desired stick number bound, which in turn can be used to explicitly bound the number of knot-types by integral Menger curvature.

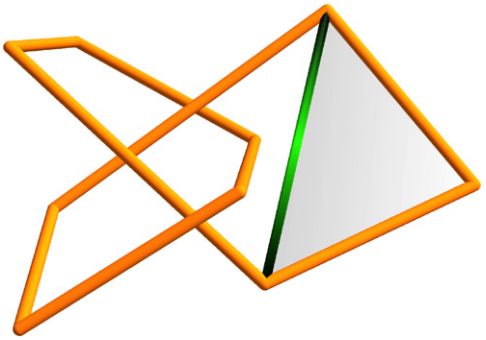
Corollary 4.4 (*Stick Number*)**.** *If the knot-type $[K]$ possesses a representative $\gamma \in \mathcal{C}$ with $\mathcal{M}_p(\gamma) \leq E < \infty$ for some $p > 3$, then*

$$\mathrm{seg}[K] \leq \left\lfloor \frac{E^{1/(p-3)}}{\delta_1(p)} \right\rfloor + 1. \tag{39}$$

For the proof of Theorem 4.3 we use Lemma 2.7 and the diamond property to construct explicitly a deformation of 3-space that maps the curve to such an inscribed polygon with vertices $x_i \in \gamma$ without ever leaving the knot-class, that is, we construct an ambient isotopy between γ and the inscribed polygon. To that end cover γ with a necklace of double cones $K(x_i, x_{i+1})$ that, by virtue of Lemma 2.7, have pairwise disjoint interiors, since for each i the polygonal edge connecting x_i and x_{i+1} is sufficiently short by assumption. The desired isotopy is constant off the union of the $K(x_i, x_{i+1})$, and in each double cone it maps each two-dimensional cross section $D_i(z) := K_i \cap (z + v_i^{\perp})$, containing $z \in [x_i, x_{i+1}]$ and perpendicular to $v_i := x_{i+1} - x_i$, homeomorphically to itself, keeping the boundary of $D_i(z)$ fixed and moving the point $\gamma(s) \in D_i(z)$ along a straight segment until it hits the axis of the cone.



(a) A Δ-move replaces an edge of a polygonal knot by the union of two new edges.



(b) A Δ^{-1}-move is an inverse operation: it replaces two edges by one.

Fig. 14. Δ and Δ^{-1}-moves. In both cases, the (full) triangle formed by the new and old edges cannot be intersected by the other edges of the knot.

A surprising second consequence of Theorem 4.3, besides the stick number bound in Corollary 4.4, is that Hausdorff-distance can be used for finite energy curves to determine neighbourhoods in the space $\mathcal{C}$ of unit loops where only one knot-type is present. (This would even work for loops without fixed length as long as their $\mathcal{M}_p$-energy remains bounded.) Hausdorff-distance $\mathrm{dist}_{\mathscr{H}}$ is a means to measure the distance between sets, say $X, Y \subset \mathbb{R}^3$, and is defined as

$$\mathrm{dist}_{\mathscr{H}}(X, Y) := \inf\{\epsilon > 0 : X \subset B_\epsilon(Y) \text{ and } Y \subset B_\epsilon(X)\},$$

where we used the notation $B_\epsilon(X) := \{z \in \mathbb{R}^3 : \mathrm{dist}(z, X) < \epsilon\}$ for the ϵ-neighbourhood of a set $X \subset \mathbb{R}^3$.

In general, Hausdorff-distance is by no means a reasonable tool to separate distinct knot-types: For any given $\epsilon > 0$ and any embedded curve $\gamma_0 \in \mathcal{C}$ one finds infinitely many other unit loops $\gamma_n \in \mathcal{C}$ with $\mathrm{dist}_{\mathscr{H}}(\gamma_0, \gamma_n) < \epsilon$ representing mutually distinct knot-types $[K_n]$; see Fig. 13(right). However, finite integral Menger curvature introduces via the diamond property so much rigidity that the following result is true.

Theorem 4.5 (*Isotopy by Hausdorff-Distance*)**.** *Any two unit loops $\gamma_1, \gamma_2 \in \mathcal{C}$ with $\mathcal{M}_p(\gamma_i) \le E < \infty$ for $i = 1, 2$ and some $p > 3$, are of the same knot-type, as long as*

$$\mathrm{dist}_{\mathscr{H}}(\gamma_1, \gamma_2) \le \delta_2(p) E^{1/(3-p)}, \tag{40}$$

where $\delta_2(p)$ is a universal constant depending just on p.

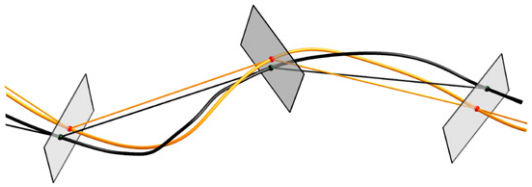
The idea of proof for this isotopy result is as follows. We know by Proposition 2.6 that both curves enjoy the diamond property beyond the same scale $d = d(E, p) := \delta(p) E^{1/(3-p)}$ depending only on E and p (see (16)). Choosing $\delta_2(p) := 0.001 \cdot \delta(p)$ will do the job. Indeed, by Theorem 4.3, we know that every polygon P inscribed in γ_1, with vertices $x_i = \gamma_1(t_i)$ for equidistantly spaced and pairwise sufficiently close parameters t_i, is of the same knot-type as γ_1. The goal is, to find a second polygon Q inscribed in the second curve γ_2 with the same number of edges, such that, on the one hand, this polygon Q is of the same knot-type as γ_2, which can be guaranteed by sufficiently short edges, and, on the other hand, such that one can see by elementary topological operations on the polygons, that also the polygons P and Q are of the same knot-type. Then one concludes that γ_1 and γ_2 represent the same knot-type.

In order to construct such a polygon Q one introduces planes Π_i through x_i and orthogonal to the tangent $\gamma_1'(t_i)$. Then one can show by means of the quantitative self-avoidance estimate in Lemma 3.1 that the planes Π_i and Π_{i+1} bound *disjoint* small tubular regions $B_\epsilon(\alpha_i)$ about the subarc $\alpha_i := \gamma_1([t_i, t_{i+1}])$ connecting x_i and x_{i+1} on γ_1. Here one can choose $\epsilon = 20\,\mathrm{dist}_H(\gamma_1, \gamma_2)$. In a second step, one finds points $y_i \in \Pi_i \cap \gamma_2$ at distance, say $\epsilon/10$ from x_i for each i, by means of the diamond property of γ_1, and the fact that $\epsilon/10 = 2\,\mathrm{dist}_H(\gamma_1, \gamma_2)$. Crucial for this is the fact that the diamond property controls the amount of bending of γ_1. These points $y_i \in \gamma_2$ can be shown to produce a polygon Q (inscribed in γ_2) with sufficiently small edge length such that Q and γ_2 are automatically of the same knot-type according to Theorem 4.3. Finally, it remains to be shown, that P and Q are also of the same knot-type.

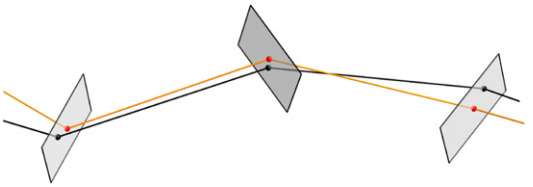
To this end, one can use so-called Δ, *and* Δ^{-1}*-moves*[14] from classic knot theory; see Fig. 14.

The first Δ-move within the first tubular region is to replace the edge $[x_1, x_2]$ by the union of two new edges, namely $[x_1, y_1]$ and $[y_1, x_2]$. The second Δ-move replaces the edge $[y_1, x_2]$ by the union of $[y_1, y_2]$ and $[y_2, x_2]$. Next, one performs one Δ^{-1}- and one Δ-move by trading first the union of $[y_2, x_2]$ and $[x_2, x_3]$ for $[y_2, x_3]$, and then replacing $[y_2, x_3]$ by the union of $[y_2, y_3]$ and $[y_3, x_3]$; see Fig. 15(c)–(f).

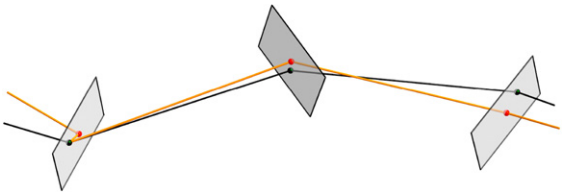
[14] These are not the famous *Reidemeister moves* as, e.g., carefully explained in [64, Chapter 1], but they do conserve the *combinatorial equivalence* and hence the knot-type of polygons.



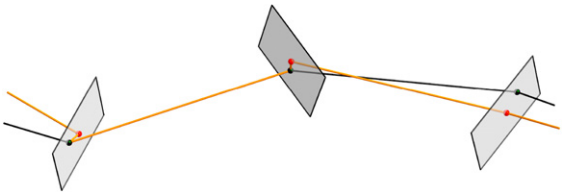
(a) Two curves γ_1, γ_2 and their inscribed polygons P, Q.



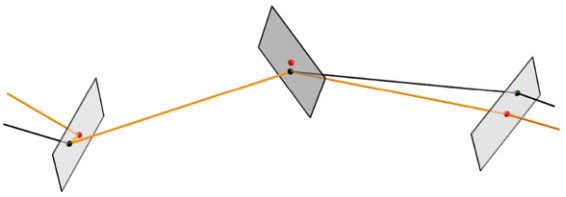
(b) Since P and Q are isotopic to γ_1 and γ_2, it is enough to establish their combinatorial equivalence.



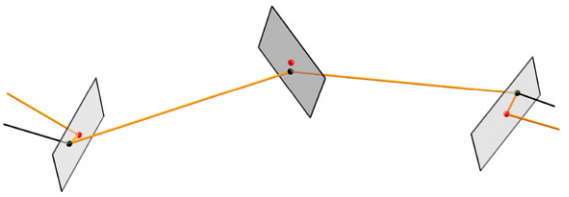
(c) Combinatorial equivalence of P and Q, step 1.



(d) Combinatorial equivalence of P and Q, step 2.



(e) Combinatorial equivalence of P and Q, step 3.



(f) Combinatorial equivalence of P and Q, step 4.

Fig. 15. The proof of Theorem 4.5.

This way, we proceed all the way around the curve constantly interchanging Δ^{-1}- and Δ-moves. Only for the last step we perform two consecutive Δ^{-1}-moves to replace the last segments, say $[y_N, x_N]$ and $[x_N, x_1]$, by $[y_N, x_1]$, and then $[y_N, x_1]$ and the first constructed segment $[x_1, y_1]$ by $[y_N, y_1]$, to finally obtain the polygon Q to conclude the proof. One could argue that Δ-moves in one tubular region surrounding one arc could also affect other parts of the polygons by producing new (and dangerous) intersections of distant polygonal edges, but such effects are excluded by the choice of ϵ which is directly related to the small Hausdorff-distance we assumed from the beginning on.

4.3. Packing problems

Among the variational applications of ropelength we mentioned in Section 2.1 packing problems, like finding the longest rope on the unit sphere. Thickness prevents a high degree of bending, and it serves as an excluded volume constraint, so one may ask for the best way to pack as much rope into a three-dimensional container. Existence theory via the calculus of variations [20,21] tells us that there is a solution, but – apart from the unique explicit longest ropes on spheres established in [26,27] – nobody was able so far to analytically describe their actual shape yet. A rather rough account on the ability of ropelength to prevent a high degree of compaction is the following inequality proven by Buck and Simon in [5, Theorem 2] for any smooth embedded closed curve γ of length L contained in a closed ball of radius r:

$$\frac{L}{\triangle[\gamma]} \geq \sqrt{\frac{3}{32}} \cdot \frac{L}{r^{3/2}}.$$

So, if the ball as a container becomes smaller and smaller as its radius r tends to zero, the ropelength of any loop of length L contained in that ball necessarily blows up like $r^{-3/2}$. However, this estimate is not sharp enough to prove, e.g., that any great circle, i.e., equator on the closed ball of radius $L/(2\pi)$ is the unique ropelength minimizer among all loops of length L contained in that ball. A direct geometric argument using [20, Lemma 3] shows, on the other hand, that ropelength is basic, which in particular implies such a result. The same is true for the interpolating energy $\mathscr{U}_p$ where one integration is involved; see [51, Lemma 7]. Relaxing $\mathscr{U}_p$ to the double integral $\mathscr{I}_p$ (see (23)) leads to an energy for which it is still open whether or not it is basic. But the packing problem described above has an affirmative solution as the following result shows; see [52].

Theorem 4.6 (*Optimal Packing in a Ball*)**.** *Among all unit loops in $\mathcal{C}$ that are contained in a fixed closed ball of radius $1/(2\pi)$, any great circle on that ball "uniquely" minimizes $\mathscr{I}_p$ for all $p \geq 2$.*

The mechanism behind the proof is a powerful geometric argument that we refer to as the "sweeping technique" the essence of which is described in the following statement regarding the integrand $\varrho[\gamma](x, y)$ of $\mathscr{I}_p$ defined in (22).

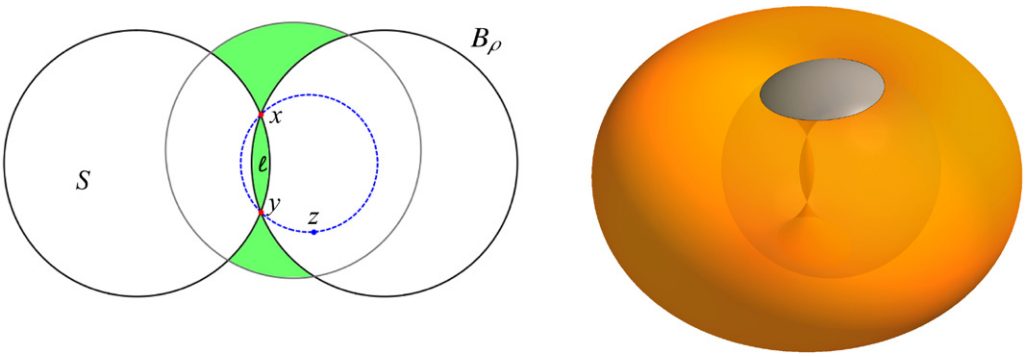


Fig. 16. The sweeping technique. Left: if $x, y \in \gamma$ and $\rho = \varrho[\gamma](x, y) > |x - y|/2$, then $\gamma \cap S = \emptyset$ for a large region $S = S(x, y)$; hence γ is confined to the shaded zones. The solid circle in the middle depicts the ball of radius $1/2\pi$ that contains γ. Right: a three-dimensional view of the sweep-out region S, bounded by a self-intersecting torus of revolution. The ball containing γ is partially hidden in that torus.

Assume that there are two distinct points $x, y \in \gamma$ with

$$\rho := \varrho[\gamma](x, y) > \frac{|x - y|}{2}, \tag{41}$$

then no point of γ is contained in the "sweep-out region"

$$S(x, y) := \bigcup_{x,y \in \partial B_\rho} B_\rho \setminus \ell(x, y), \tag{42}$$

which consists of union of all balls or radius ρ containing the (fixed) points x and y in their boundary ∂B_ρ minus the lens-shaped closure of their common intersection

$$\ell(x, y) := \overline{\bigcap_{x,y \in \partial B_\rho} B_\rho};$$

see Fig. 16.

If, moreover, the points x, y do not uniquely realize the diameter of the curve γ, i.e., the largest possible Euclidean distance of point pairs on γ, then one can easily show that the curve cannot be contained within the lens $\ell(x, y)$.[15]

Observe that the sweep-out region $S(x, y)$ in (42) can be quite voluminous if $\rho = \varrho[\gamma](x, y)$ is fairly large. In particular, if $\varrho[\gamma](x, y)$ were strictly greater than the radius $1/(2\pi)$ of the confining ball, then there would be simply no space for the curve γ left within that ball; see Fig. 16. This observation can be turned into an upper bound on $\varrho[\gamma](\xi, \eta)$ for almost all pairs $\xi, \eta \in \gamma$:

$$\varrho[\gamma](\xi, \eta) \le \frac{1}{2\pi},$$

which immediately implies the energy inequality

$$\mathscr{I}_p^{1/p}(\gamma) \ge \mathscr{I}_2^{1/2}(\gamma) \ge 2\pi = \mathscr{I}_2^{1/2}(\text{great circle}).$$

A similar argument shows also that this inequality is actually strict, if γ is not one of the great circles of the confining ball; for details see the proof of [52, Theorem 3.2].

The sweeping technique just described also leads to the nontrivial lower bound on the $\mathscr{I}_p$-energy of any unit loop $\gamma \in \mathcal{C}$,

$$\mathscr{I}_p^{1/p}(\gamma) \ge \mathscr{I}_2^{1/2}(\gamma) \ge \min\left\{2 + \pi, \frac{2}{\operatorname{diam}\gamma}\right\} \ge 4, \tag{43}$$

which states that one needs at least an $\mathscr{I}_2$-energy level of 16 to close a curve of unit length. This is vaguely reminiscent of the lower bound 2π for total curvature to close a curve according to Fenchel's theorem, only that we do not know if the bound in (43) is sharp — probably not, since we strongly believe that $\mathscr{I}_p$ is basic for all $p \ge 2$, which would give $2\pi = \mathscr{I}_p^{1/p}(\textit{circle})$ as the sharp lower bound. Since the sweeping technique relies on the one maximization remained in the definition of $\mathscr{I}_p$, we do not have a corresponding nontrivial lower bound for the $\mathscr{M}_p$-energy. One can, however, turn the arguments in the proof of the beta number estimate in Lemma 2.4 into a lower bound for the $\mathscr{M}_p$-energy for any loop, but the resulting constant would be less explicit and much smaller than the right-hand side of (43). In other words, one also needs a positive amount of $\mathscr{M}_p$-energy to close a curve, but we are far from knowing sharp bounds here. Again, numerical evidence of the simulations of Hermes [11] strongly suggests that the circle of radius $1/(2\pi)$ is the unique minimizer of $\mathscr{M}_p$ for all $p \ge 3$, so that we expect 2π as the sharp lower bound for $(\mathscr{M}_p)^{1/p}$ on the class $\mathcal{C}$ of unit loops.

[15] For $p > 2$ this would also follow from C^1-smoothness of γ with finite $\mathscr{I}_p$-energy; see [49], since γ confined in the lens $\ell(x, y)$ would lead to sharp abrupt turns of the tangent γ' in the tip points x and y of the lens.

Another consequence of the sweeping technique is the following rigidity result for curves of constant (pointwise) Menger curvatures [52, Corollary 3.5].

Theorem 4.7 (*Rigidity*). *If $\gamma \in \mathcal{C}$ satisfies either $R(x, y, z) = R_0$, or $\varrho[\gamma](x, y) = R_0$, or $\rho_G[\gamma](x) = R_0$ for some $0 < R_0 < \infty$, then $R_0 = 1/(2\pi)$, and γ is a circle of radius $1/(2\pi)$.*

Notice that such a result for classic local curvature is simply not true: there are infinitely many unit loops of constant local curvature. One can even construct arbitrary C^2-knots of constant curvature such as in [65].

5. Higher dimensions and open problems

5.1. Shapes of energy minimizers: numerical evidence and some conjectures

Problem 5.1. Assume $p \geq 3$. Is the integral Menger curvature $\mathcal{M}_p$ basic?

Even simpler versions of that question are open: it is unknown whether the round circle is the unique minimizer of $\mathcal{M}_p$ restricted to the class of *planar* unit loops. As we have indicated earlier, the main source of difficulties is the nonlocal character of $\mathcal{M}_p$: even small changes of the curve affect the integrand, $1/R$, at *all* triples of points, making all comparison arguments difficult.

Problem 5.2. Does the integral Menger curvature $\mathcal{M}_p$, $p \geq 3$, detect the unknot?

Here, we only have a partial answer: the integral Menger curvature $\mathcal{M}_p$ detects the unknot for all $p \geq p_0$, where p_0 is a finite number. Since this statement is proved by contradiction, using the convergence of $\mathcal{M}_p(\gamma)^{1/p}$ to ropelength of γ as $p \to \infty$, we have no explicit estimate of p_0. Nevertheless, we are tempted to conjecture that the answers to both questions, formulated in Problems 5.1 and 5.2, are positive. Numerical evidence, gathered by Tobias Hermes [11] (see the next subsection for more details), supports that view very strongly. It is also rigorously proved in [11] that the circle is a *critical* point of $\mathcal{M}_p$.

We do not know much about the integral Menger curvature for curves in the very interesting, scale-invariant case $p = 3$. If $\mathcal{M}_3(\gamma) < \infty$, then the loop γ is free from self-intersections (cf. Theorem 2.2). However, it is not clear whether γ is differentiable everywhere! In fact, γ cannot be a polygon (cf. Section 2.2) but we do not know how to exclude e.g. the possibility that γ spirals in a neighbourhood of a point x so that the tangent at $y \in \gamma$ has no one-sided limits as $y \to x$.

Problem 5.3. Let $p = 3$ and suppose that a loop $\gamma \in \mathcal{C}$ satisfies $\mathcal{M}_3(\gamma) < \infty$. Is γ differentiable everywhere? If yes, is the (unit) tangent continuous? How to obtain any compactness estimates, and fill in the entries of Table 1 corresponding to $\mathcal{M}_3$?

The question concerning the *regularity of minimizers* of integral Menger curvature $\mathcal{M}_p$ is wide open. Up to now, the problem of regularity of minimizers of knot energies has been successfully overcome only in a few cases. Minimizers of O'Hara's Möbius invariant energy $\mathcal{E}_{\text{Möb}}$, see (30), could be shown to be C^∞-smooth [10,56]; see also Reiter [66,67] and the recent work of Blatt, Reiter and Schikorra on critical points of $\mathcal{E}_{\text{Möb}}$ [57]. The initial gain in regularity up to $C^{1,1}$ in [10] heavily relied on this Möbius invariance, a property not shared by $\mathcal{M}_p$ or its relatives $\mathcal{U}_p$ and $\mathcal{I}_p$. The ropelength minimizing *links* constructed by Cantarella, Kusner, and Sullivan in [19] show that $C^{1,1}$ is indeed the optimal regularity in given link classes. $C^{1,1}$-regularity of ropelength minimizers might in fact be optimal in general: for ideal knots one observes numerically jumps in local curvature.

However, even for $\mathcal{U}_p$ with $1 < p < \infty$ the situation is unclear.

Problem 5.4. Are the minimizers of $\mathcal{M}_p$ and related knot energies of class $C^{1,1}$? C^∞? Does the optimal regularity for $\mathcal{M}_p$-minimizers depend explicitly on the parameter p, and does that lead to any conclusions about the still open optimal regularity of ideal knots as $p \to \infty$?

Uniqueness of ideal knots is not to be expected as exhibited by the examples of whole continuous families of ideal links in [19]; see also Fig. 4. For integral Menger curvature the situation might be different, but also this is an open question.

Problem 5.5. Study the uniqueness (and nonuniqueness) of local minima of $\mathcal{M}_p$ in various knot classes. In particular: does $\mathcal{M}_p$, restricted to the class $\mathcal{C}_{\text{unknot}}$ of all unknots in $\mathcal{C}$, have multiple local minima?

5.2. Energy landscape and flow

5.2.1. Second thoughts about the energy landscape of $\mathcal{M}_p$

The energy landscape of integral Menger curvature $\mathcal{M}_p$, and of the other geometric curvature energies interpolating between $\mathcal{M}_p$ and ropelength, on the space $\mathcal{C}$ of unit loops is largely unknown. We do not know how many local or absolute minimizers it has in particular knot classes, not to speak of the possible distribution of critical points. Even if we dismiss the restriction of a given knot type we have no rigorous proof that the circle (of length one) is the unique minimizer, although heuristic arguments and numerical evidence suggest exactly that, which would make $\mathcal{M}_p$ basic.

Theorem 5.6 (*Infinitely Many Local $\mathscr{M}_p$-Minima*)**.** *The integral Menger curvature $\mathscr{M}_p$ possesses infinitely many local minima in $\mathcal{C}$ for $p > 3$.*

According to Theorem 3.2, $\mathscr{M}_p$ is minimizable for $p > 3$, so in each prescribed knot class $[K]$ we find at least one absolute minimizer γ_K representing $[K]$. Combining this with Theorem 4.5 each such minimizer γ_K is a local minimum for $\mathscr{M}_p$ among *all* unit loops in $\mathcal{C}$ (without restrictions on the knot type). Indeed, any loop $\alpha \in \mathcal{C}$ with $\mathscr{M}_p(\alpha) < \mathscr{M}_p(\gamma_K) =: E$ and with Hausdorff distance from γ_K less than the constant in (40) would be of the same knot type $[K]$, contradicting $\mathscr{M}_p(\gamma_K) \le \mathscr{M}_p(\alpha)$. So, in that neighbourhood, there is simply no other unit loop with smaller integral Menger curvature. Since there are infinitely many different knot classes we thus obtain infinitely many local minimizers for $\mathscr{M}_p$ in the unrestricted class $\mathcal{C}$ of unit loops.

Motivated by finite-dimensional calculus one is tempted to say, that we have found infinitely many $\mathscr{M}_p$-critical points, but for a functional on an infinite dimensional domain, like $\mathcal{C}$, the notion of a critical point needs to be defined with care. In his Ph.D.-thesis Hermes [11] derived a formula for the first variation of integral Menger curvature,

$$\delta\mathscr{M}_p(\gamma, h) := \frac{d}{d\epsilon}\mathscr{M}_p(\gamma + \epsilon h)\bigg|_{\epsilon=0},$$

in a mathematically rigorous way, for exactly those curves $\gamma \in \mathcal{C}$ of which the energy is finite, and for variations h such that for all sufficiently small ϵ also the perturbation $\gamma + \epsilon h$ has finite energy.[16] The term "perturbation" is justified by the fact that the loops $\gamma + \epsilon h$ tend to γ as $\epsilon \to 0$, in Hausdorff-distance. But in general the perturbation is not of unit length, so that $\gamma + \epsilon h \notin \mathcal{C}$ for $\epsilon \neq 0$. Consequently, the (infinitely many) local minimizers γ_K obtained above, cannot be compared directly to the perturbed curves $\gamma_K + \epsilon h$. Hermes considered the scale invariant version $\mathscr{S}_p$ of integral Menger curvature instead,

$$\mathscr{S}_p(\gamma) := \frac{\mathscr{M}_p^{1/p}(\gamma)}{\mathrm{length}(\gamma)^{(3-p)/p}}, \tag{44}$$

for continuous closed curves γ of arbitrary finite length. It turns out that the $\mathscr{M}_p$-minimizers $\gamma_K \in [K] \cap \mathcal{C}$ also minimize the rescaled functional $\mathscr{S}_p$ in the class of continuous closed curves of finite length parametrized on the interval $[0, 1]$. Since all perturbations $\gamma_K + \epsilon h$ are of that class, we find therefore infinitely many critical points γ_K of the rescaled energy $\mathscr{S}_p$. Alternatively, one can use Hermes' formula for the first variation $\delta\mathscr{M}_p(\gamma, h)$ of integral Menger curvature to derive an Euler–Lagrange equation for each local minimizer γ_K of $\mathscr{M}_p$ on $\mathcal{C}$ involving a Lagrange parameter and the variation of length due to the length constraint in $\mathcal{C}$. But the Lagrange parameter depends on γ_K, so that we cannot speak of infinitely many solutions of the *same* variational equation in that case.

Let us mention recent work of Jason Cantarella, Jennifer Ellis, Joseph H.G. Fu, and Matt Mastin on the principle of symmetric criticality for ropelength which can be used to construct ropelength-critical points different from known minima as long as one finds representatives with symmetries in the same knot class; see [31].

5.2.2. On the gradient flow for integral Menger curvature

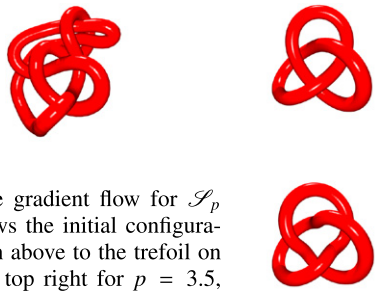
A finer tool to investigate the energy landscape of integral Menger curvature would be a gradient flow, i.e., a time-dependent partial differential equation of the type $\dot\gamma = V$ for a family of curves $\gamma = \gamma(s, \tau)$ depending on time τ, where the velocity field V is proportional to the gradient of $\mathscr{M}_p$. A solution γ of this equation would describe the flow in the direction of steepest descent of $\mathscr{M}_p$. Due to the complexity of the gradient of $\mathscr{M}_p$ (derived by Hermes in [11, Section 2.5]) nothing is known about the existence of such a solution, not even for short time.

Problem 5.7. Does the evolution equation describing the gradient flow for integral Menger curvature $\mathscr{M}_p$ has a solution — at least for a short time interval? Is there a chance to prove long-time existence, and what is the asymptotic behaviour of the solution as time tends to the boundary of the existence interval?

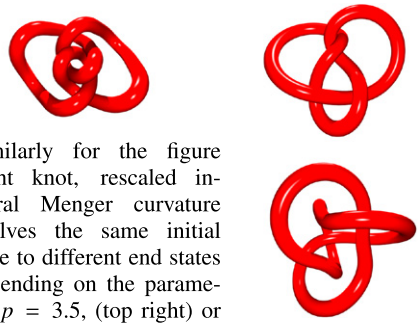
The only known existence results about gradient flows for knot energies are the contributions of Blatt [68,69] (based on earlier work of He [56]) on the Möbius energy $\mathscr{E}_{\mathrm{Möb}}$, and on related knot energies introduced by O'Hara. Blatt proved long-time existence and convergence to local minima as time $\tau \to \infty$; in case of the Möbius energy, however, under the additional assumption that the initial curve is sufficiently close to a (possibly different) local minimum of $\mathscr{E}_{\mathrm{Möb}}$.

Hermes fully discretized the very complicated evolution equation for the rescaled integral Menger curvature $\mathscr{S}_p$ in space and time and implemented a powerful and reliable numerical scheme to compute the gradient flow for $\mathscr{S}_p$; see [11, Chapter 3]. His impressive simulations demonstrate that integral Menger curvature can be used to flow initially highly complicated knotted configurations to "optimal" representatives, presumably local $\mathscr{S}_p$-minimizers, and to untangle complex structures, as long as $p > 3$. Hermes' examples also show that for p large the $\mathscr{S}_p$-flow tends to produce nicest embeddings close to the ideal shapes produced by RIDGERUNNER, the algorithm devised by Ted Ashton, Jason Cantarella, Michael Piatek, and Eric Rawdon [12] to produce ideal knots and links. For p close to 3, however, the smoothing effects seem to dominate the self-repulsion effects, so that the final configurations do look different, although the knot class is preserved during both flows; see

[16] By the aforementioned characterization of finite energy curves by Blatt [37,38] it suffices to have simple regular curves γ in a certain fractional Sobolev space, perturbed by h in the same Sobolev space; see [11, Theorem 2.33].



The gradient flow for $\mathscr{S}_p$ 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 $p = 50$.



Similarly for the figure eight knot, rescaled integral Menger curvature evolves the same initial state to different end states depending on the parameter $p = 3.5$, (top right) or $p = 50$ (bottom right).

Fig. 17. Different parameters $p > 3$ lead to different final configurations for the gradient flow of the rescaled integral Menger curvature $\mathscr{S}_p$, but the knot type is preserved.
Source: Images by courtesy of T. Hermes.

Fig. 17. In addition, the data produced by Hermes' flow strongly suggest that integral Menger curvature is basic as indicated in Fig. 18.

5.3. Energies of sheets, surfaces and submanifolds

It is natural to ask whether the integral Menger curvature $\mathscr{M}_p$ (or other related energies that were discussed in Section 3) can be extended to surfaces in $\mathbb{R}^3$, or, more generally, to m-dimensional sets in $\mathbb{R}^n$, with similar regularizing and self-avoidance effects as in the curve case. The answer turns out to be positive; surprisingly, one of the crucial difficulties is *the choice of the integrand*.

5.3.1. High-dimensional integral Menger curvatures

For the sake of simplicity, let us first describe such an extension of $\mathscr{M}_p$, and a few of its properties, for two-dimensional surfaces in $\mathbb{R}^3$.

The most natural generalization of $\mathscr{M}_p$ to two-dimensional closed surfaces $\Sigma \subset \mathbb{R}^3$ would be to replace the circumcircle radius $R(x, y, z)$ of three points x, y, z by the circumsphere radius $R(\xi, x, y, z)$ of the tetrahedron $T := (\xi, x, y, z)$ spanned by the four non-coplanar points ξ, x, y, z. This radius is given by

$$\frac{1}{2R(T)} = \frac{\left|\langle z_3, z_1 \times z_2\rangle\right|}{\left||z_1|^2 z_2 \times z_3 + |z_2|^2 z_3 \times z_1 + |z_3|^2 z_1 \times z_2\right|}, \tag{45}$$

where $z_1 = \xi - z, z_2 = x - z, z_3 = y - z$. This would lead to a possible variant of integral Menger curvature for surfaces,

$$\int_\Sigma \int_\Sigma \int_\Sigma \int_\Sigma \frac{d\mathscr{H}^2(\xi)\, d\mathscr{H}^2(x)\, d\mathscr{H}^2(y)\, d\mathscr{H}^2(z)}{R^p(\xi, x, y, z)}, \tag{46}$$

which, however, has several unpleasant disadvantages. Although the integrand is constant if Σ happens to be a round sphere — there are smooth surfaces with straight nodal lines (such as the graph of the function $f(x, y) = xy$) where the integrand is not pointwise bounded. The reason is that on a surface, close to *every* point, there are lots of small spheres intersecting the surface transversally, along a curve, see Fig. 19. This is a problem since we want to consider arbitrarily large p, and we envision a whole family of integral Menger curvatures that are finite on *any* closed smooth surface for *any* value of p. The naive generalization (46) fails to satisfy this requirement.

Let us go back to formula (10) for $1/R$,

$$\frac{1}{R(x, y, z)} = \frac{2\,\mathrm{dist}(z, L_{xy})}{|x - z|\,|y - z|} = \frac{4\,\mathrm{Area}\,\triangle(x, y, z)}{|x - z|\,|y - z|\,|x - y|},$$

where L_{xy} denotes the straight line through x and y. The right hand side expresses, in metric terms, what sort of behaviour is penalized by integral Menger curvature: if there are lots of small, nearly equilateral, triangles with vertices on γ, the energy has to be large. On small triangles, with all edges $\approx d$ (up to, say, a factor of 2), the integrand $1/R$ becomes bounded only if the area is $\approx d^3$, i.e. only if the (shortest) height of the triangle is much smaller than the longest edge. This way, one is tempted to consider various 4-*point-integrands* that measure the degree of flatness of tetrahedra T with all four vertices on the surface.

J.C. Léger [70, p. 833] proposed a general integrand of that type for m-dimensional sets; for $m = 2$ his choice is

$$K_{\text{Lég}}(\xi, x, y, z,) = \frac{\mathrm{dist}(\xi, \langle x, y, z\rangle)}{|\xi - x|\,|\xi - y|\,|\xi - z|} \tag{47}$$

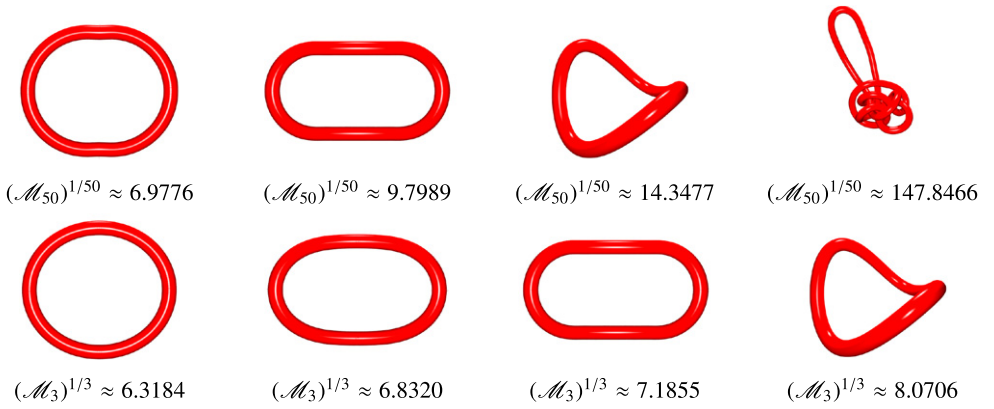


Fig. 18. A sample of T. Hermes' computations of integral Menger curvature of various unknots in $\mathcal{C}$ different from the circle for $p = 50$ and $p = 3$. Even the slightest perturbations of the circle which, for each $p > 0$, satisfies $(\mathscr{M}_p(\text{circle}))^{1/p} \approx 6.2832$ lead to a significant increase of integral Menger curvature for $p = 50$ (top row) strongly suggesting that $\mathscr{M}_p$ is basic for $p > 3$. Even for $p = 3$ where no analytic results are available yet, integral Menger curvature seems to have the circle as the unique minimizer (bottom row). Nota bene: the numerical gradient flow of the rescaled energy $\mathscr{S}_{50}$ does deform the complicated unknot on the top right into a perfectly round circle.
Source: Images by courtesy of T. Hermes.

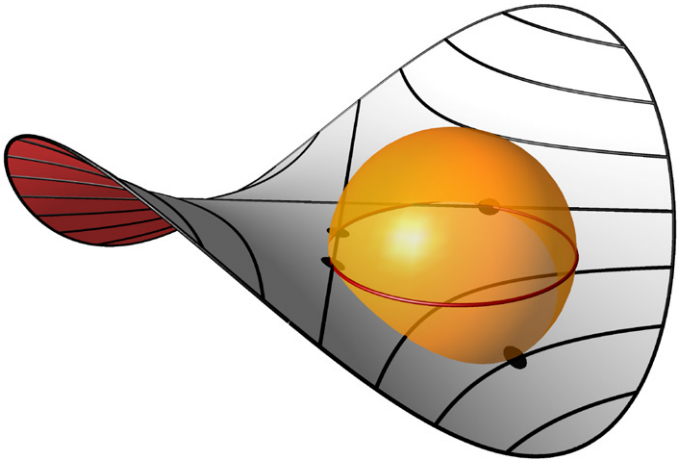


Fig. 19. The graph of $(x, y) \mapsto xy$ near zero and a sphere of radius R intersecting it at four non-coplanar points. Three of them, and the centre of the sphere, are in the (x, y)-plane. Similar spheres can be found in smaller scales, closer to $(0, 0, 0)$, making $1/R$ (essentially) unbounded. Perturbing the four points in little dark patches, we do not change R too much. An argument based on scaling shows that for this particular surface $1/R^p$ cannot be integrable over $\Sigma \times \Sigma \times \Sigma \times \Sigma$ for $p \geq 8$.

where $\langle x, y, z\rangle$ denotes the affine 2-plane through generic non-collinear points $x, y, z \in \mathbb{R}^3$. However, for this particular integrand, due to the lack of symmetry with respect to permutations of the 4 points, the situation is even worse than for inverse powers of the circumsphere radius: even the energy of a round sphere,

$$\int_{\mathbb{S}^2}\int_{\mathbb{S}^2}\int_{\mathbb{S}^2}\int_{\mathbb{S}^2} K^p_{\text{Lég}}\, d\mathscr{H}^2(\xi)\, d\mathscr{H}^2(x)\, d\mathscr{H}^2(y)\, d\mathscr{H}^2(z)$$

becomes infinite for *all* sufficiently large p; see [71, Appendix B]. This singular behaviour is caused by small tetrahedra for which the plane through (x, y, z) is almost perpendicular to the surface.

A whole series of high-dimensional geometric curvatures measuring the flatness of simplices have been introduced by Gilad Lerman and J. Tyler Whitehouse in their pioneering work [72,73] dealing with m-rectifiability and least square approximation of m-regular measures. Their discrete curvatures are based, roughly speaking, on the so-called polar sine function scaled by some power of the diameter of the simplex, and can be used to obtain powerful and very general characterizations of rectifiability of measures. (In [73, Sections 1.5 and 6] the authors also note that the integrand suggested by Léger does not fit into their setting.) However, for surfaces the discrete curvatures of Lerman and Whitehouse, e.g. c_{MT} in [73, p. 327], scale like the inverse of the *cube* of length. This enforces too much singularity for our purposes, see [71, Sections 1 and 5]. Namely, it turns out that for any integrand $\mathcal{K}_s(T)$ satisfying

$$\mathcal{K}_s(T) \approx \frac{h_{\min}(T)}{(\operatorname{diam} T)^{2+s}}, \quad s > 0,$$

i.e. scaling like the inverse of length to some power *bigger than one*, the corresponding surface energy

$$\mathscr{E}(\Sigma) := \int_{\Sigma^4} \mathcal{K}_s(T)^p \, d\mu,$$

where $d\mu = d\mathscr{H}^2 \otimes d\mathscr{H}^2 \otimes d\mathscr{H}^2 \otimes d\mathscr{H}^2$ is the natural measure on $\Sigma^4 = \Sigma \times \Sigma \times \Sigma \times \Sigma$, has the following property: for all $p > 24/s$ the only surface of finite energy is a flat plane!

Motivated by all this, we have been led to consider another 4-point symmetric integrand, with fewer cancellations in the denominator. For a tetrahedron $T = (\xi, x, y, z)$ with all vertices on Σ consider the function

$$\mathcal{K}(T) := \begin{cases} \dfrac{\text{Volume}\,(T)}{\text{Area}\,(T)\,(\text{diam}\,T)^2} & \text{if the vertices of } T \text{ are not coplanar,} \\ 0 & \text{otherwise,} \end{cases} \tag{48}$$

where the total area Area (T) of T, i.e., the sum of the areas of all four triangular faces of T, could also be replaced by another factor $(\text{diam}\,T)^2$ (cf. formula (52) in general dimensions). Thus, up to a constant factor $\mathcal{K}$ is the ratio of the minimal height of T to the square of its diameter. The main – and crucial – difference with the curvatures defined in [73] is that our $\mathcal{K}$ scales like the inverse of length. The corresponding *integral Menger curvature for two-dimensional surfaces* $\Sigma \subset \mathbb{R}^3$, defined as

$$\mathscr{M}_p(\Sigma) := \int_\Sigma \int_\Sigma \int_\Sigma \int_\Sigma \mathcal{K}^p(T) \, d\mathscr{H}^2 \otimes d\mathscr{H}^2 \otimes d\mathscr{H}^2 \otimes d\mathscr{H}^2(T), \tag{49}$$

is finite for all C^2-surfaces for all finite p, since $\mathcal{K}(T)$ is bounded on the set of all nondegenerate tetrahedra with vertices on such a surface; we refer to [71, Appendix A] for details.

This energy is well defined for a broad admissible class $\mathscr{A}$ of nonsmooth surfaces, including all closed Lipschitz surfaces (i.e., boundaries of domains that are locally a graph of a Lipschitz functions) and some other surfaces that are not even topological submanifolds of $\mathbb{R}^3$, e.g. a sphere with the north and south pole glued together (or, in other words, the horn torus depicted in Fig. 3(c)), or whole infinite stacks of concatenated spheres or boxes; see the examples in [71].

Our paper [71] contains several results which explain topological, measure theoretic and analytic consequences of the finiteness of $\mathscr{M}_p$. The scale invariant exponent here is $p = 8$; for $p > 8$ one can control the flatness and bending of the surface, excluding self-intersections, wrinkles, folds along lines, conical or cuspidal singularities etc. The picture is analogous to the one for integral Menger curvature for curves in Section 2.3. Surfaces Σ with $\mathscr{M}_p(\Sigma) < \infty$ for $p > 8$ turn out to be C^1-smooth and have well defined tangent planes at *every* point. Moreover, there is a length scale $R_0 \approx E^{-1/(p-8)}$ depending *only* on p and on the energy bound E such that below this scale every surface Σ with integral Menger curvature $\mathscr{M}_p(\Sigma) \le E$ is a nearly flat graph over a disc. Here is a more precise formulation.

Theorem 5.8. *If $\mathscr{M}_p(\Sigma) \le E$ for some $p > 8$, then for each point $x \in \Sigma$ and each radius $r < R_0$ the intersection $\Sigma \cap B(x, r)$ coincides with Graph$f \cap B(x, r)$, where $f : \mathbb{R}^2 \equiv T_x\Sigma \to \mathbb{R} \equiv (T_x\Sigma)^\perp$ is a function defined on the tangent plane to Σ at x; the function f is of class $C^{1,\alpha}$ for $\alpha = 1 - \frac{8}{p}$ and satisfies a uniform estimate*

$$\|f\|_{C^{1,\alpha}} \le C(p) E^{1/p}, \tag{50}$$

with the constant $C(p)$ depending only on p.

Thanks to [38], we know that the exponent $1 - \frac{8}{p}$ is best possible.

To prove this, one has to adapt the argument from Section 2.3 from curves to surfaces *and* overcome one crucial difficulty which is absent in the case of curves. Namely, we need to know first that if $r < R_0$ and $\mathscr{M}_p(\Sigma) \le E$, then the intersection $\Sigma \cap B(x, r)$ has surface measure comparable to the flat disc of radius r, so that, say

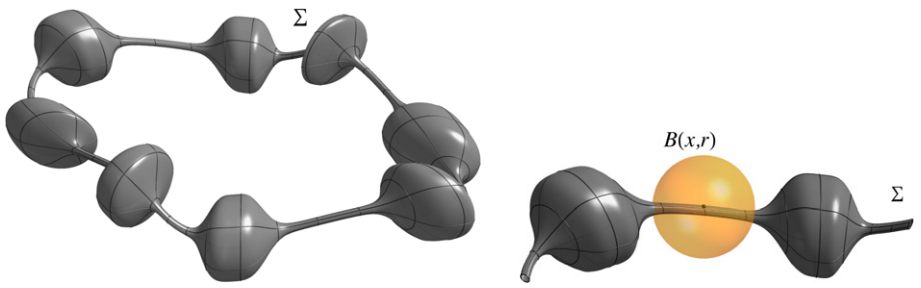
$$\frac{\mathscr{H}^2(\Sigma \cap B(x, r))}{\pi r^2} \ge \frac{1}{2} \quad \text{for all } r < R_0 = \delta(p) E^{-1/(p-8)}. \tag{51}$$

Note that this property – which we refer to as *uniform Ahlfors regularity* – cannot be guaranteed even by requiring that Σ be a priori smooth. Even for a very smooth surface with long thin tentacles or tubes the ratio $\mathscr{H}^2(\Sigma \cap B(x, r))/\pi r^2$ can be as small as one wishes, cf. Fig. 20. (For curves the situation is different: if $x \in \gamma$ and $\text{diam}\gamma > 2r$, then we certainly have $\mathscr{H}^1(\gamma \cap B(x, r)) \ge r$, since γ has to reach the exterior of $B(x, r)$ from its centre.)

Since the estimate (50) is *uniform*, it can be used to obtain compactness results and, as corollaries, the existence of area minimizers under the constraint of bounded energy and fixed genus, and of energy minimizers in a given isotopy class of surfaces with uniformly bounded area.

Theorem 5.9. *If $p > 8$, and $\{\Sigma_j\}$ is a sequence of closed, compact and connected Lipschitz surfaces, all of them containing $0 \in \mathbb{R}^3$ and satisfying the bounds*

$$\mathscr{M}_p(\Sigma_j) \le E \quad \text{and} \quad \mathscr{H}^2(\Sigma_j) \le A \quad \text{for all } j \in \mathbb{N}$$



(a) If $r \gg E^{-1/(p-8)}$, then the intersection of Σ and the ball $B(x, r)$ centred on Σ does not have to be a graph.

(b) The ratio $\mathscr{H}^2(\Sigma \cap B(x, r))/\pi r^2$ can be made arbitrarily small just by making the tube thinner; this would increase $\mathscr{M}_p(\Sigma)$.

Fig. 20. A torus with thin tubes and thick bumps has large integral Menger curvature.

then there is a compact $C^{1,1-8/p}$-manifold Σ without boundary embedded in $\mathbb{R}^3$, and a subsequence $\{\Sigma_{j'}\}$ converging to Σ in C^1 as $j' \to \infty$, and such that

$$\mathscr{M}_p(\Sigma) \le \liminf_{j'\to\infty} \mathscr{M}_p(\Sigma_{j'}).$$

(Instead of the uniform area bound one could use a uniform diameter bound in the above theorem.)

Theorem 5.10. *The class $\mathscr{C}_E(M_g)$ of all closed, compact and connected Lipschitz surfaces $\Sigma \subset \mathbb{R}^3$, ambiently isotopic to a fixed reference surface $M_g \subset \mathbb{R}^3$ of genus g and satisfying the constraint $\mathscr{M}_p(\Sigma) \le E$, contains – for each fixed genus g, each M_g and each energy bound E – a surface of least area.*

Similarly, the class $\mathscr{C}_A(M_g)$ of closed, compact and connected Lipschitz surfaces $\Sigma \subset \mathbb{R}^3$ ambiently isotopic to a fixed reference surface M_g of genus g, and satisfying a uniform area bound $\mathscr{H}^2(\Sigma) \le A$, contains a surface Σ minimizing integral Menger curvature $\mathscr{M}_p$ on $\mathscr{C}_A(M_g)$.

Following [71–73], and sharpening the higher-dimensional techniques developed in [74], Sławomir Kolasiński [75,76] has studied a general integral Menger curvature $\mathscr{M}_p$, defined for a wide class of nonsmooth m-dimensional surfaces in $\mathbb{R}^n$. His $\mathscr{M}_p$ is defined as the integral over $m+2$ copies of Σ of the multipoint integrand

$$\mathcal{K}_{SK}(x_0, x_1, \ldots, x_{m+1}) := \frac{\mathscr{H}^{m+1}\big(\operatorname{conv}(x_0, x_1, \ldots, x_{m+1})\big)}{\operatorname{diam}\big(x_0, x_1, \ldots, x_{m+1}\big)^{m+2}}. \tag{52}$$

Here, $x_0, x_1, \ldots, x_{m+1} \in \Sigma$ are vertices of an $(m+1)$-dimensional simplex $T = \operatorname{conv}(x_0, x_1, \ldots, x_{m+1})$. Again, $\mathcal{K}$ scales like the inverse of length. For $m = 2$ the integrand differs slightly from (48); nevertheless, all the results from [71] do also hold for (52) with $m = 2$. The integral Menger curvature of an m-dimensional set Σ is given by

$$\mathscr{M}_p(\Sigma) = \underbrace{\int_\Sigma \int_\Sigma \ldots \int_\Sigma}_{m+2 \text{ times}} \mathcal{K}_{SK}(x_0, x_1, \ldots, x_{m+1})^p \, d\mathscr{H}^m(x_0) \, d\mathscr{H}^m(x_1) \, \ldots \, d\mathscr{H}^m(x_{m+1}). \tag{53}$$

The large class $\mathscr{A}$ of admissible m-dimensional sets considered by Kolasiński in [75] is strongly related but not identical to the admissibility class considered in [74]. It contains all m-dimensional closed Lipschitz submanifolds of $\mathbb{R}^n$ and all sets $\Sigma = f(M)$ where M is an abstract C^1 manifold and $f: M \to \mathbb{R}^n$ an immersion. It also contains all finite unions of embedded closed C^1 manifolds. One of the main results of [75] is the following.

Theorem 5.11. *If $\Sigma \subset \mathbb{R}^n$ is an m-dimensional admissible surface with $\mathscr{M}_p(\Sigma) \le E$ for some $p > m(m+2)$, then Σ is an embedded manifold of class $C^{1,\alpha}$, $\alpha = 1 - m(m+2)/p$. Moreover, there is a length scale $R_0 \approx E^{-1/(p-m(m+2))}$, depending only on p and the energy bound E, such that for each $x \in \Sigma$ and each radius $r < R_0$ the intersection $\Sigma \cap B(x, r)$ is a (nearly flat) m-dimensional disc: it equals, up to an isometry, Graph $f \cap B(0, r)$ where $f: \mathbb{R}^m \to \mathbb{R}^{n-m}$ is a function of class $C^{1,\alpha}$, with*

$$\|f\|_{C^{1,\alpha}} \lesssim \mathscr{M}_p(\Sigma)^{1/p}.$$

Again, the exponent $\alpha = 1 - m(m+2)/p$ is best possible here.

This regularity result – again, due to uniform estimates on the patch size R_0 and on the norm of the graph representation f – can be used to deduce that, for each $p > m(m+2)$, each E and each A, there are only finitely many C^1-smooth manifolds M satisfying a uniform volume bound $\mathscr{H}^m(M) \le A$ and a uniform integral Menger curvature bound $\mathscr{M}_p(M) \le E$.[17]

[17] This is work in progress [77]. The result can be viewed as an analogue of Anderson–Cheeger finiteness theorems in (smooth) Riemannian geometry, but here in a setting which is 'below the C^2 category'; see [78] and the references therein.

5.3.2. Other high-dimensional energies

In our papers [79,80] we have introduced the concept of thickness $\triangle[X]$ for a large class of nonsmooth parametric surfaces $X\colon \Sigma \to \mathbb{R}^3$, where Σ is a smooth fixed reference surface. As for curves, a uniform lower bound $\triangle[X] \geq \theta > 0$ provides the surface with "thickness"; it is equivalent to the requirement that the so-called *reach*[18] of the set $X(\Sigma) \subset \mathbb{R}^3$ be at least θ, see [80, Lemma 6.2]. The normal vector to a surface with $\triangle[X] \geq \theta$ turns out to be (locally, at scales $\approx \theta$) Lipschitz continuous with constant $\approx 1/\theta$. This is the reason why families of 'thick surfaces' with uniform area bounds (or uniform diameter bounds) are compact in the C^1-topology. As a result, one can minimize area under thickness and genus constraints; more precisely, one can prove that each class of compact, closed surfaces of fixed genus, global curvature bounded from below by $\theta > 0$, and ambiently isotopic to a fixed reference surface, contains at least one surface of minimal area. For more details, also for surfaces with nonempty boundary, we refer to [79,80], and to related work of Alexander Nabutovsky [81] on thick knotted hyperspheres.

Let us also briefly mention that Kolasiński [75] obtains geometric regularity results for high-dimensional integral curvatures generalizing the $\mathscr{I}_p$-energy (23) for curves. Setting

$$\mathcal{K}^{(l)}(x_0, \ldots, x_{l-1}) = \sup_{x_l, x_{l+1}, \ldots, x_{m+1} \in \Sigma} \mathcal{K}_{SK}(x_0, \ldots, x_{l-1}, x_l, \ldots, x_{m+1}) \tag{54}$$

where $l = 1, \ldots, m+1$, he considers the integral

$$\mathscr{E}_p^{(l)} = \underbrace{\int_\Sigma \ldots \int_\Sigma}_{l \text{ times}} \mathcal{K}^{(l)}\big(x_0, \ldots, x_{l-1}\big)^p \, d\mathscr{H}^m(x_0) \ldots d\mathscr{H}^m(x_{l-1}). \tag{55}$$

Analogues of Theorem 5.11 discussed above hold true for each of those energies. For C^1 manifolds, Blatt and Kolasiński [38] give an equivalent condition (expressed in terms of the so-called fractional Sobolev spaces) for finiteness of these energies. We refer to [75,38] for more details.

In [74], we have studied a high-dimensional counterpart of the tangent-point energy $\mathscr{E}_p$ defined for curves by (28), and obtained self-avoidance and regularity results analogous to (v).

Finally, in the joint work with Kolasiński [82] we study the energy $\mathscr{E}_p^{(1)}$ defined by (55), and a related energy where the integrand is expressed in terms of the size of spheres tangent to Σ at one point and passing through another point of Σ. It turns out that for a each of these two energies is finite if and only if the set Σ (which a priori might be nonsmooth and have self-intersections, cusps, folds etc.) is a manifold of class $W^{2,p}$, i.e. locally a graph of a C^1 function which has second order distributional derivatives in L^p.

All the open problems on curves mentioned in the previous sections immediately extend to hard problems in higher dimension and codimension, such as finding nontrivial (if not sharp) lower bounds for $\mathscr{M}_p(\Sigma)$, optimal regularity of $\mathscr{M}_p$-minimizing surfaces, and – even more difficult to tackle – the question if there is a reasonable notion of a gradient flow for higher dimensional integral Menger curvatures.

Acknowledgements

The idea to write this survey was born while both authors participated in the *Knotted fields* program at the Kavli Institute for Theoretical Physics in the summer of 2012. They are grateful to KITP for providing excellent working conditions and for partially supporting this research under the NSF Grant No. PHY11-25915. Before 2012, various parts of this research have been partially supported by several institutions, including Deutsche Forschungsgemeinschaft, Polish Ministry of Science and Higher Education, and Centro di Ricerca Matematica "Ennio De Giorgi" at the Scuola Normale Superiore di Pisa.

The first author was partially supported by a grant from NCN. The second author was partially supported by the DFG grant Geometric curvature energies (Mo966/4-1).

References

[1] Louis H. Kauffman, The mathematics and physics of knots, Rep. Progr. Phys. 68 (12) (2005) 2829–2857.
[2] Uroš Tkalec, Miha Ravnik, Simon Čopar, Slobodan Žumer, Igor Muševič, Reconfigurable knots and links in chiral nematic colloids, Science 333 (6038) (2011) 62–65.
[3] Shinji Fukuhara, Energy of a knot, in: A Fête of Topology, Academic Press, Boston, MA, 1988, pp. 443–451.
[4] Jonathan Simon, Energy functions for knots: beginning to predict physical behavior, in: Mathematical Approaches to Biomolecular Structure and Dynamics (Minneapolis, MN, 1994), in: IMA Vol. Math. Appl., vol. 82, Springer, New York, 1996, pp. 39–58.
[5] Gregory Buck, Jonathan Simon, Energy and length of knots, in: Lectures at KNOTS '96 (Tokyo), in: Ser. Knots Everything, vol. 15, World Sci. Publ., River Edge, NJ, 1997, pp. 219–234.
[6] 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), in: Contemp. Math., vol. 304, Amer. Math. Soc., Providence, RI, 2002, pp. 181–186.

18 The concept of *reach* has been introduced by Federer [43].

[7] Jun O'Hara, Energy of Knots and Conformal Geometry, in: Series on Knots and Everything, vol. 33, World Scientific Publishing Co. Inc., River Edge, NJ, 2003.
[8] Louis H. Kauffman, Following knots down their energy gradients, Symmetry 4 (2) (2012) 276–284.
[9] Oscar Gonzalez, John H. Maddocks, Global curvature, thickness, and the ideal shapes of knots, Proc. Natl. Acad. Sci. USA 96 (9) (1999) 4769–4773. (electronic).
[10] Michael H. Freedman, Zheng-Xu He, Zhenghan Wang, Möbius energy of knots and unknots, Ann. of Math. (2) 139 (1) (1994) 1–50.
[11] Tobias Hermes, Analysis of the first variation and a numerical gradient flow for integral Menger curvature. Ph.D. Thesis, RWTH Aachen University, 2012. Available at http://darwin.bth.rwth-aachen.de/opus3/volltexte/2012/4186/.
[12] Ted Ashton, Jason Cantarella, Michael Piatek, Eric J. Rawdon, Knot tightening by constrained gradient descent, Exp. Math. 20 (1) (2011) 57–90.
[13] Jun O'Hara, Energy of a knot, Topology 30 (2) (1991) 241–247.
[14] Guy David, Analytic capacity, Calderón–Zygmund operators, and rectifiability, Publ. Mat. 43 (1) (1999) 3–25.
[15] Xavier Tolsa, Analytic capacity, rectifiability, and the Cauchy integral, in: International Congress of Mathematicians. Vol. II, Eur. Math. Soc., Zürich, 2006, pp. 1505–1527.
[16] Pertti Mattila, Search for geometric criteria for removable sets of bounded analytic functions, Cubo 6 (4) (2004) 113–132.
[17] Karl Menger, Untersuchungen über allgemeine Metrik, Math. Ann. 103 (1) (1930) 466–501.
[18] Leonard M. Blumenthal, Karl Menger, Studies in Geometry, W.H. Freeman and Co., San Francisco, Calif, 1970.
[19] Jason Cantarella, Robert B. Kusner, John M. Sullivan, On the minimum ropelength of knots and links, Invent. Math. 150 (2) (2002) 257–286.
[20] Oscar Gonzalez, John H. Maddocks, Friedemann Schuricht, Heiko von der Mosel, Global curvature and self-contact of nonlinearly elastic curves and rods, Calc. Var. Partial Differential Equations 14 (1) (2002) 29–68.
[21] O. Gonzalez, R. de la Llave, Existence of ideal knots, J. Knot Theory Ramifications 12 (1) (2003) 123–133.
[22] Justyna Baranska, Piotr Pieranski, Eric J. Rawdon, Ropelength of tight polygonal knots, in: Physical and Numerical Models in Knot Theory, in: Ser. Knots Everything, vol. 36, World Sci. Publ., Singapore, 2005, pp. 293–321.
[23] M. Carlen, B. Laurie, J.H. Maddocks, J. Smutny, Biarcs, global radius of curvature, and the computation of ideal knot shapes, in: Physical and Numerical Models in Knot Theory, in: Ser. Knots Everything, vol. 36, World Sci. Publ., Singapore, 2005, pp. 75–108.
[24] E.L. Starostin, A constructive approach to modelling the tight shapes of some linked structures, Forma 18 (4) (2003) 263–293.
[25] Jason Cantarella, Joseph H.G. Fu, Rob Kusner, John M. Sullivan, Nancy C. Wrinkle, Criticality for the Gehring link problem, Geom. Topol. 10 (2006) 2055–2116. (electronic).
[26] Henryk Gerlach, Heiko von der Mosel, What are the longest ropes on the unit sphere? Arch. Ration. Mech. Anal. 201 (1) (2011) 303–342.
[27] Henryk Gerlach, Heiko von der Mosel, On sphere-filling ropes, Amer. Math. Monthly 118 (10) (2011) 863–876.
[28] Henryk Gerlach, Construction of sphere-filling ropes, 2009. Available at http://www.littleimpact.de/permanent/math/sphere_filling/.
[29] Friedemann Schuricht, Heiko von der Mosel, Characterization of ideal knots, Calc. Var. Partial Differential Equations 19 (3) (2004) 281–305.
[30] Jason Cantarella, Joseph H.G. Fu, Robert Kusner, John M. Sullivan, Ropelength criticality, 2011. arXiv:1102.3234v2 [math.DG].
[31] Jason Cantarella, Jennifer Ellis, Joseph H.G. Fu, Matt Mastin, Symmetric criticality for tight knots, 2012. arXiv:1208.3879v3 [math.DG].
[32] Friedemann Schuricht, Heiko von der Mosel, Euler–Lagrange equations for nonlinearly elastic rods with self-contact, Arch. Ration. Mech. Anal. 168 (1) (2003) 35–82.
[33] Sebastian Scholtes, For which positive p is the integral Menger curvature $\mathcal{M}_p$ finite for all simple polygons? 2011. arXiv:1202.0504v1.
[34] Paweł Strzelecki, Marta Szumańska, Heiko von der Mosel, Regularizing and self-avoidance effects of integral Menger curvature, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) 9 (1) (2010) 145–187.
[35] Marta Szumańska, Integral versions of Menger curvature: smoothing potentials for rectifiable curves, Ph.D. Thesis, Institute of Mathematics, University of Warsaw, 2009 (in Polish).
[36] Sławomir Kolasiński, Marta Szumańska, Minimal Hölder regularity implying finiteness of integral Menger curvature, Manuscripta Math. 141 (1–2) (2013) 125–147.
[37] Simon Blatt, A note on integral Menger curvature, 2011. Preprint.
[38] Simon Blatt, Sławomir Kolasiński, Sharp boundedness and regularizing effects of the integral Menger curvature for submanifolds, Adv. Math. 230 (3) (2012) 839–852.
[39] Robert A. Adams, Sobolev Spaces, in: Pure and Applied Mathematics, vol. 65, Academic Press, New York, London, 1975, [A subsidiary of Harcourt Brace Jovanovich, Publishers].
[40] Peter W. Jones, Rectifiable sets and the traveling salesman problem, Invent. Math. 102 (1) (1990) 1–15.
[41] Guy David, Stephen Semmes, Analysis of and on Uniformly Rectifiable Sets, in: Mathematical Surveys and Monographs, vol. 38, American Mathematical Society, Providence, RI, 1993.
[42] Paweł Strzelecki, Heiko von der Mosel, Tangent-point self-avoidance energies for curves, J. Knot Theory Ramifications 21 (5) (2012) 28 pages.
[43] Herbert Federer, Curvature measures, Trans. Amer. Math. Soc. 93 (1959) 418–491.
[44] Jun O'Hara, Energy functionals of knots. II, Topology Appl. 56 (1) (1994) 45–61.
[45] Morris W. Hirsch, Differential Topology, in: Graduate Texts in Mathematics, vol. 33, Springer-Verlag, New York, 1976.
[46] Philipp Reiter, All curves in a C^1-neighbourhood of a given embedded curve are isotopic, Preprint Nr. 4, Institut für Mathematik, RWTH Aachen University, 2005.
[47] Simon Blatt, Note on continuously differentiable isotopies, Preprint Nr. 34, Institut für Mathematik, RWTH Aachen University, 2009.
[48] O. Gonzalez, J.H. Maddocks, J. Smutny, Curves, circles, and spheres, in: Physical Knots: Knotting, Linking, and Folding Geometric Objects in $\mathbb{R}^3$ (Las Vegas, NV, 2001), in: Contemp. Math., vol. 304, Amer. Math. Soc., Providence, RI, 2002, pp. 195–215.
[49] Paweł Strzelecki, Marta Szumańska, Heiko von der Mosel, A geometric curvature double integral of Menger type for space curves, Ann. Acad. Sci. Fenn. Math. 34 (1) (2009) 195–214.
[50] Malte Laurens Kampschulte, The symmetrized tangent-point energy. Master's thesis, RWTH Aachen University, 2012.
[51] Paweł Strzelecki, Heiko von der Mosel, On rectifiable curves with L^p-bounds on global curvature: self-avoidance, regularity, and minimizing knots, Math. Z. 257 (2007) 107–130.
[52] Paweł Strzelecki, Marta Szumańska, Heiko von der Mosel, On some knot energies involving Menger curvature, Topology Appl. (in press). arXiv:1209.1527v2 [math.CA].
[53] István Fáry, Sur la courbure totale d'une courbe gauche faisant un nœud, Bull. Soc. Math. France 77 (1949) 128–138.
[54] J.W. Milnor, On the total curvature of knots, Ann. of Math. (2) 52 (1950) 248–257.
[55] Jayanth R. Banavar, Oscar Gonzalez, John H. Maddocks, Amos Maritan, Self-interactions of strands and sheets, J. Stat. Phys. 110 (1–2) (2003) 35–50.
[56] Zheng-Xu He, The Euler–Lagrange equation and heat flow for the Möbius energy, Comm. Pure Appl. Math. 53 (4) (2000) 399–431.
[57] Simon Blatt, Philipp Reiter, Armin Schikorra, Hard analysis meets critical knots (Stationary points of the Moebius energy are smooth), 2012. arXiv:1202.5426v2.
[58] Gregory Buck, Jonathan Simon, Thickness and crossing number of knots, Topology Appl. 91 (3) (1999) 245–257.
[59] R.A. Litherland, J. Simon, O. Durumeric, E. Rawdon, Thickness of knots, Topology Appl. 91 (3) (1999) 233–244.
[60] Friedemann Schuricht, Heiko von der Mosel, Global curvature for rectifiable loops, Math. Z. 243 (1) (2003) 37–77.
[61] Herbert Federer, Geometric measure theory, in: Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.
[62] Jason Cantarella, Robert B. Kusner, John M. Sullivan, On the minimum ropelength of knots and links, Nature 392 (1998) 237.
[63] Colin C. Adams, An elementary introduction to the mathematical theory of knots, in: The Knot Book, American Mathematical Society, Providence, RI, 2004, Revised reprint of the 1994 original.
[64] Gerhard Burde, Heiner Zieschang, Knots, second ed., in: de Gruyter Studies in Mathematics, vol. 5, Walter de Gruyter & Co., Berlin, 2003.

[65] Jenelle Marie McAtee, Knots of constant curvature, 2004. arXiv:math/0403089v1 [math.GT].
[66] Philipp Reiter, Repulsive knot energies and pseudodifferential calculus. Ph.D. Thesis, RWTH Aachen University, 2009. Available at http://darwin.bth.rwth-aachen.de/opus3/volltexte/2009/2848/.
[67] Philipp Reiter, Regularity theory for the Möbius energy, Commun. Pure Appl. Anal. 9 (5) (2010) 1463–1471.
[68] Simon Blatt, The gradient flow of the Möbius energy near local minimizers, Calc. Var. Partial Differential Equations 43 (3–4) (2012) 403–439.
[69] Simon Blatt, The gradient flow of O'Hara's knot energies, 2012 (in preparation).
[70] Jean-Christophe Léger, Menger curvature and rectifiability, Ann. of Math. (2) 149 (3) (1999) 831–869.
[71] Paweł Strzelecki, Heiko von der Mosel, Integral Menger curvature for surfaces, Adv. Math. 226 (2011) 2233–2304.
[72] Gilad Lerman, J. Tyler Whitehouse, High-dimensional Menger-type curvatures. Part I: geometric multipoles and multiscale inequalities, Rev. Mat. Iberoam. 27 (2) (2011) 493–555.
[73] Gilad Lerman, J. Tyler Whitehouse, High-dimensional Menger-type curvatures. II. d-separation and a menagerie of curvatures, Constr. Approx. 30 (3) (2009) 325–360.
[74] Paweł Strzelecki, 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, 2011. arXiv:1102.3642 ; J. Geom. Anal., http://dx.doi.org/10.1007/s12220-011-9275-z (in press).
[75] Sławomir Kolasiński, Geometric Sobolev-like embedding using high-dimensional Menger-like curvature, Trans. Amer. Math. Soc. (2012) (in press) arXiv:1205.4112v1.
[76] Sławomir Kolasiński, Integral Menger curvature for sets of arbitrary dimension and codimension. Ph.D. Thesis, Institute of Mathematics, University of Warsaw, 2011. arXiv:1011.2008v4.
[77] Sławomir Kolasiński, Paweł Strzelecki, Heiko von der Mosel, Compactness and finiteness results for sets with equibounded integral Menger curvature, 2013 (in preparation).
[78] Michael T. Anderson, Jeff Cheeger, Diffeomorphism finiteness for manifolds with Ricci curvature and $L^{n/2}$-norm of curvature bounded, Geom. Funct. Anal. 1 (3) (1991) 231–252.
[79] Paweł Strzelecki, Heiko von der Mosel, On a mathematical model for thick surfaces, in: Physical and Numerical Models in Knot Theory, in: Ser. Knots Everything, vol. 36, World Sci. Publ., Singapore, 2005, pp. 547–564.
[80] Paweł Strzelecki, Heiko von der Mosel, Global curvature for surfaces and area minimization under a thickness constraint, Calc. Var. Partial Differential Equations 25 (4) (2006) 431–467.
[81] Alexander Nabutovsky, Non-recursive functions, knots "with thick ropes", and self-clenching "thick" hyperspheres, Comm. Pure Appl. Math. 48 (4) (1995) 381–428.
[82] Sławomir Kolasiński, Paweł Strzelecki, Heiko von der Mosel, Characterizing $W^{2,p}$-submanifolds by p-integrability of global curvatures, Geom. Funct. Anal. 23 (3) (2013) 937–984.