The Proceedings of the National Academy of Sciences, USA, **96**, (1999) 4769–4773.

Global Curvature, Thickness and the Ideal Shapes of Knots

Oscar Gonzalez & John H. Maddocks

Département de Mathématiques
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland

Classification: Physical Sciences (Applied Mathematics/Biophysics)

Manuscript Information: text pages (12), figures (2), tables (0)

Word/Character Counts: abstract (70 words), article (38887 chars)

Corresponding Author: JHM (Phone: +41 21 693 2762;
Fax: +41 21 693 5530; E-mail: maddocks@dma.epfl.ch)

Global Curvature, Thickness and the Ideal Shapes of Knots

Oscar Gonzalez & John H. Maddocks

Département de Mathématiques
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland

Abstract

The global radius of curvature of a space curve is introduced. This function is related to, but distinct from, the standard local radius of curvature, and is connected to various physically appealing properties of a curve. In particular, the global radius of curvature function provides a concise characterization of the thickness of a curve, and of certain ideal shapes of knots as have been investigated within the context of DNA [1, 2, 3].

1 Introduction

Any smooth, non-self-intersecting curve can be thickened into a smooth, non-self-intersecting tube of constant radius centered on the curve. If the curve is a straight line there is no upper bound on the tube radius, but for nonstraight curves there is a critical radius above which the tube either ceases to be smooth or exhibits self-contact. This critical radius is an intrinsic property of the curve called its *thickness* or *normal injectivity radius* [4, 5]. If one considers the class of smooth, non-self-intersecting closed curves of a prescribed knot type and unit length, one may ask which curve in this class is the thickest. The thickness of such a curve is an intrinsic property of the knot, and the curve itself provides a certain *ideal shape* or representation of the knot type. Approximations of ideal shapes in this sense have been found via a series of computer experiments [1, 2]. These shapes were seen to have intriguing physical features, and even a correspondence to time-averaged shapes of knotted DNA molecules in solution [1, 2, 3]. In this article we introduce the notion of the *global radius of curvature* function for a curve. We show that this function provides a simple characterization of curve thickness, and further use it to derive an elementary necessary condition that any ideal shape of a knotted curve must satisfy.

The presentation is structured as follows. In Section 2 we define the global radius of curvature function for a space curve, and in Section 3 we show that the thickness of a smooth curve is equal to the minimum value of its global radius of curvature function. In Section 4 we introduce a definition of the ideal shape of a smooth, knotted curve and then show that ideal shapes, by necessity, have the property that their global radius of curvature function is constant, except possibly on straight portions of the curve. In Section 5 we introduce a definition of global radius of curvature appropriate for discretized curves (as occur in some numerical computations for example), discuss discrete analogs of thickness and ideality, and present some numerical examples showing that slightly corrected versions of previously computed ideal shapes [1, 2] satisfy our necessary condition. Finally, in Section 6, we discuss connections between global radius of curvature, various measures of energy and the writhing number of a space curve [6].

2 Global Radius of Curvature

By a *curve* $\mathcal{C}$ we mean a continuous three-dimensional vector function $\boldsymbol{q}(s)$ of a real variable s with $0 \leq s \leq L$. The curve $\mathcal{C}$ is *smooth* if the function $\boldsymbol{q}(s)$ is continuously differentiable to any order and if the tangent vector $\boldsymbol{q}'(s)$ is nonzero for all s. In the smooth case we typically interpret s as the arclength parameter. A curve $\mathcal{C}$ is *closed* if $\boldsymbol{q}(L) = \boldsymbol{q}(0)$, in which case we interpret the parameter s modulo L. Moreover, if $\mathcal{C}$ is smoothly closed, the derivatives of all orders of $\boldsymbol{q}(s)$ agree at $s = 0$ and $s = L$. Finally, a curve $\mathcal{C}$ is *simple* if it has no self-intersections, that is, $\boldsymbol{q}(s_1) = \boldsymbol{q}(s_2)$ only when $s_1 = s_2$.

Our definition of global radius of curvature is based on the elementary facts [7] that any three non-collinear points $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ in three-dimensional space define a unique circle (the circumcircle), and the radius of this circle (the circumradius) can be written as

$$r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \frac{|\boldsymbol{x}-\boldsymbol{y}||\boldsymbol{x}-\boldsymbol{z}||\boldsymbol{y}-\boldsymbol{z}|}{4\mathcal{A}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})} \tag{1}$$

where $\mathcal{A}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ is the area of the triangle with vertices $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$, $|\boldsymbol{x}-\boldsymbol{y}|$ is the Euclidean distance between the points $\boldsymbol{x}$ and $\boldsymbol{y}$, and so on. When the points $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are distinct, but collinear, the circumcircle degenerates into a straight line and we assign a value of infinity to $r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Note that any three non-collinear points not only define a unique circumcircle, but also a unique circumsphere which contains the circumcircle as a great circle.

Given three non-collinear points $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ there are various formulae for computing the area $\mathcal{A}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. One example is

$$\mathcal{A}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \frac{1}{2}|\boldsymbol{x}-\boldsymbol{z}||\boldsymbol{y}-\boldsymbol{z}||\sin\theta_{\boldsymbol{xzy}}| \tag{2}$$

where $\theta_{\boldsymbol{xzy}}$ is the angle between the vectors $\boldsymbol{x}-\boldsymbol{z}$ and $\boldsymbol{y}-\boldsymbol{z}$. This particular expression shows the connection between the value of the circumradius function $r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ and the standard sine rule from elementary geometry.

When $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are points on a simple, smooth curve $\mathcal{C}$, the domain of the function $r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ can be extended by continuous limits to all triples of points on $\mathcal{C}$. For example, if $\boldsymbol{x} = \boldsymbol{q}(s)$, $\boldsymbol{y} = \boldsymbol{q}(\sigma)$ and $\boldsymbol{z} = \boldsymbol{q}(\tau)$ are three distinct points on $\mathcal{C}$, then it is straightforward to show that

$$\lim_{\tau\to\sigma} r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \frac{|\boldsymbol{x}-\boldsymbol{y}|}{2|\sin\theta_{\boldsymbol{xy}'}|} \tag{3}$$

where $\theta_{\boldsymbol{xy}'}$ is the angle between the vector $\boldsymbol{x}-\boldsymbol{y} \neq \boldsymbol{0}$ and the tangent vector to $\mathcal{C}$ at $\boldsymbol{y}$. We denote this limit by $r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y})$ and note that the limit circumcircle passes through $\boldsymbol{x}$ and is tangent to $\mathcal{C}$ at $\boldsymbol{y}$. Similar expressions and interpretations hold for the other single limits. For the double limit $\sigma, \tau \to s$ we obtain the familiar result

$$\lim_{\sigma,\tau \to s} r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \rho(\boldsymbol{x}) \tag{4}$$

where $\rho(\boldsymbol{x})$ is the standard local radius of curvature of $\mathcal{C}$ at $\boldsymbol{x}$. We denote this limit by $r(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x})$ and note that the limit circumcircle is actually the osculating circle to $\mathcal{C}$ at $\boldsymbol{x}$.

Given a curve $\mathcal{C}$ we define the *global radius of curvature* $\rho_G(\boldsymbol{x})$ at each point $\boldsymbol{x}$ of $\mathcal{C}$ by

$$\rho_G(\boldsymbol{x}) := \inf_{\substack{\boldsymbol{y},\boldsymbol{z} \in \mathcal{C} \\ \boldsymbol{x} \neq \boldsymbol{y} \neq \boldsymbol{z} \neq \boldsymbol{x}}} r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}). \tag{5}$$

When $\mathcal{C}$ is simple and smooth, the infimum in (5) can be replaced by a minimum over all $\boldsymbol{y}, \boldsymbol{z}$ on $\mathcal{C}$, namely

$$\rho_G(\boldsymbol{x}) = \min_{\boldsymbol{y},\boldsymbol{z} \in \mathcal{C}} r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), \tag{6}$$

and the function $\rho_G(\boldsymbol{x})$ is a continuous function of $\boldsymbol{x}$ on $\mathcal{C}$. To obtain (6) from (5), one exploits the continuity properties of $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ and the fact that $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ admits the limits (3) and (4) when $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are restricted to lie on $\mathcal{C}$.

The function ρ_G can be interpreted as a generalization, indeed a globalization, of the standard local radius of curvature function. From (6) and (4), we see that global radius of curvature is bounded by local radius of curvature in the sense that $0 \leq \rho_G(\boldsymbol{x}) \leq \rho(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{C}$. Also, just as with local radius of curvature, the global radius of curvature function is infinite when $\mathcal{C}$ is a straight line. Conversely, if on a curve $\mathcal{C}$ there is a point $\boldsymbol{x}$ such that $\rho_G(\boldsymbol{x})$ is infinite, then ρ_G is infinite at all points and $\mathcal{C}$ is a straight line.

The optimality conditions associated with the minimization in (6) imply a certain geometrical characterization of the global radius of curvature. In particular, depending on the point $\boldsymbol{x}$ on $\mathcal{C}$, the number $\rho_G(\boldsymbol{x})$ may be the local radius of curvature, or the strictly smaller radius of a circle containing $\boldsymbol{x}$ and another distinct point $\boldsymbol{y}$ of the curve at which the circle is tangent, as illustrated in Figure 1a,b. Thus, to determine $\rho_G(\boldsymbol{x})$, one need only consider the minimization in (6) with the restriction $\boldsymbol{y} = \boldsymbol{z}$.

The above conclusion may be reached by geometrical arguments as follows. Suppose $\rho_G(\boldsymbol{x})$ were achieved by a pair of distinct points $\boldsymbol{y}$ and $\boldsymbol{z}$, each distinct from $\boldsymbol{x}$. Unless the curve $\mathcal{C}$ is tangent to the circumsphere at either $\boldsymbol{y}$ or $\boldsymbol{z}$, or both, we obtain an immediate contradiction, for otherwise there are points $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{z}}$ on $\mathcal{C}$ close to $\boldsymbol{y}$ and $\boldsymbol{z}$ for which $r(\boldsymbol{x}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{z}}) < r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. If the tangency is at $\boldsymbol{y}$, the circle through $\boldsymbol{x}$ and tangent to $\mathcal{C}$ at $\boldsymbol{y}$ lies on the circumsphere and so has radius less than or equal to the great circle radius, implying $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}) \leq r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. This shows that the minimum in (6) is never exclusively achieved by three distinct points on $\mathcal{C}$, and so is always achieved by limits such as (3) or (4). Similar arguments show that not all the limit cases of the form (3) need be considered. In particular, one need not consider limits such as $r(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y})$, but only $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y})$ with the case $\boldsymbol{y} = \boldsymbol{x}$ of local radius of curvature being possible.

If on a simple, smooth curve $\mathcal{C}$ the function ρ_G is not constant, then certain points $\boldsymbol{y}$ and $\boldsymbol{z}$ are excluded from achieving the minimum in the definition (6) of $\rho_G(\boldsymbol{x})$. More precisely, let a and d be the minimum and maximum of ρ_G on $\mathcal{C}$, let c be any number in the open interval (a, d) and let E be the set of those $\boldsymbol{x}$ on $\mathcal{C}$ such that $\rho_G(\boldsymbol{x}) < c$. Then, for any $\boldsymbol{x}$ in the set E, the minimum in (6) can only be achieved by points $\boldsymbol{y}$ and $\boldsymbol{z}$ in the set E. That is, if $\rho_G(\boldsymbol{x}) = r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y})$, then $\rho_G(\boldsymbol{y}) \leq \rho_G(\boldsymbol{x})$. This property, which follows from the definition of ρ_G and the fact that $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is symmetric in its arguments at distinct points, will be referred to as the *lower interaction property* of the global radius of curvature.

To any curve $\mathcal{C}$ we associate a number $\Delta[\mathcal{C}]$ defined by

$$\Delta[\mathcal{C}] := \inf_{\boldsymbol{x} \in \mathcal{C}} \rho_G(\boldsymbol{x}). \tag{7}$$

When $\mathcal{C}$ is simple and smooth, the function $\rho_G(\boldsymbol{x})$ is a continuous function of $\boldsymbol{x}$ on $\mathcal{C}$, and the infimum in (7) can be replaced by a minimum, namely

$$\Delta[\mathcal{C}] = \min_{\boldsymbol{x} \in \mathcal{C}} \rho_G(\boldsymbol{x}). \tag{8}$$

From (8) and (6) we see that on a simple, smooth curve $\Delta[\mathcal{C}]$ is the minimum value of the circumradius function $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ over all triplets of points. This observation leads to the following physical interpretation. Any spherical shell of radius less than $\Delta[\mathcal{C}]$ cannot intersect $\mathcal{C}$ in three or more points (counting tangency points twice). In effect, a billiard ball of radius less than $\Delta[\mathcal{C}]$ cannot find a stable resting place in $\mathcal{C}$, for there is always enough room for it to pass through the curve, as illustrated in Figure 2b.

Arguments involving the circumsphere can be used to demonstrate that, for a given smooth curve $\mathcal{C}$, the number $\Delta[\mathcal{C}]$ is either the minimum local radius of curvature, or the strictly smaller radius of a sphere that contains no portion of the curve in its interior and is tangent to the curve at two diametrically opposite points $\boldsymbol{x}$ and $\boldsymbol{y}$. At such points we have the symmetry property $r(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y}) = r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y})$. Moreover, $\boldsymbol{x}$ and $\boldsymbol{y}$ must be points of closest approach of $\mathcal{C}$ in the sense that $\boldsymbol{x} \neq \boldsymbol{y}$ and the vector $\boldsymbol{x} - \boldsymbol{y}$ is orthogonal to the tangent vectors to $\mathcal{C}$ at both $\boldsymbol{x}$ and $\boldsymbol{y}$.

As a real-valued function on the vector space of twice continuously differentiable curves $\boldsymbol{q}(s)$, $\Delta[\mathcal{C}]$ has various continuity properties with respect to the norm

$$||\mathcal{C}|| = \max_{0 \leq s \leq L} \{|\boldsymbol{q}(s)|, |\boldsymbol{q}'(s)|, |\boldsymbol{q}''(s)|\}.$$

In particular, $\Delta[\mathcal{C}]$ is continuous at any simple curve $\mathcal{C}$ which is not a straight line, while for straight lines $\Delta[\mathcal{C}]$ is infinite. Moreover, $\Delta[\mathcal{C}_k]$ tends to zero for any sequence $\mathcal{C}_k$ of smooth, simple curves that tends to a self-intersecting curve.

3 Thickness of a Curve

The thickness of a simple, smooth space curve $\mathcal{C}$ may be defined as follows [4, 5]. Given a point $\boldsymbol{x}$ on $\mathcal{C}$ and a real number $\eta > 0$, let $\mathcal{D}(\boldsymbol{x}, \eta)$ denote the circular disk of radius η centered at $\boldsymbol{x}$ and contained in the normal plane to $\mathcal{C}$ at $\boldsymbol{x}$. For sufficiently small η, the disks $\mathcal{D}(\boldsymbol{x}, \eta)$ are pairwise disjoint and their union form a smooth solid tube $\mathcal{T}[\mathcal{C}, \eta]$ around $\mathcal{C}$. If $\mathcal{C}$ is a straight line there is no upper bound on the tube radius η, but if $\mathcal{C}$ is curved there is a critical radius $\eta_*[\mathcal{C}]$ above which the tube either ceases to be smooth or exhibits self-contact. This critical radius is called the *thickness* or *normal injectivity radius* of $\mathcal{C}$. Simple geometrical considerations show that

$$\eta_*[\mathcal{C}] = \min \left\{ \min_{\boldsymbol{x} \in \mathcal{C}} \rho(\boldsymbol{x}), d_*[\mathcal{C}]/2 \right\} \tag{9}$$

where $d_*[\mathcal{C}]$ is a minimum distance of closest approach defined as

$$d_*[\mathcal{C}] := \min_{\boldsymbol{x},\boldsymbol{y} \in \Omega} |\boldsymbol{x} - \boldsymbol{y}|. \tag{10}$$

Here Ω is the set of all pairs of points $\boldsymbol{x}, \boldsymbol{y}$ on $\mathcal{C}$ such that $\boldsymbol{x} \neq \boldsymbol{y}$, and such that the vector $\boldsymbol{x} - \boldsymbol{y}$ is orthogonal to the tangent vectors to $\mathcal{C}$ at both $\boldsymbol{x}$ and $\boldsymbol{y}$.

It can now be seen that the minimum global radius of curvature $\Delta[\mathcal{C}]$ is precisely the thickness $\eta_*[\mathcal{C}]$. That is to say, $\Delta[\mathcal{C}]$ is either the minimum local radius of curvature or half of the minimum distance of closest approach, whichever is smaller. However, in contrast to the characterization in (9), the quantity $\Delta[\mathcal{C}]$ given in (8) *simultaneously* captures the possibility that curve thickness may be controlled by local curvature effects, or by the distance of closest approach of non-adjacent points on $\mathcal{C}$. Thus, in addition to the interpretation of $\Delta[\mathcal{C}]$ in terms of circumspheres tangent to the curve as illustrated in Figure 2b, we see that $\Delta[\mathcal{C}]$ is also the radius of the thickest smooth tube that can be centered on $\mathcal{C}$, as illustrated in Figure 2a.

4 Ideal Shapes of Knots

Using the global radius of curvature function ρ_G and the thickness Δ we can formulate a concise mathematical definition of an ideal shape of a knotted curve. In particular, let $\mathcal{K}$ denote the set of all simple, smooth curves $\mathcal{C}$ of a specified knot type with fixed length $L > 0$, and consider the problem of finding those curves $\mathcal{C}^*$ in $\mathcal{K}$ satisfying

$$\Delta[\mathcal{C}^*] = \sup_{\mathcal{C} \in \mathcal{K}} \Delta[\mathcal{C}]. \tag{11}$$

We call any curve $\mathcal{C}^*$ in $\mathcal{K}$ an *ideal shape* if it achieves the supremum in (11). As illustrated in Figure 2a, this definition corresponds precisely to the intuitive notion of the thickest tube of fixed length that can be tied into a given knot [1, 2].

We are not aware of any result guaranteeing the existence of a smooth ideal shape for an arbitrary knot type (in particular, curves that maximize $\Delta[\mathcal{C}]$ may or may not be smooth). Nevertheless, we can derive a necessary condition, implied by (11), that any smooth ideal shape must satisfy. Given any curve $\mathcal{C}$ in $\mathcal{K}$ let $\mathcal{S}_\mathcal{C}$ denote the set of all points $\boldsymbol{x}$ on $\mathcal{C}$ for which $\rho(\boldsymbol{x})$ is infinite, that is, $\mathcal{S}_\mathcal{C}$ is the set of straight segments of $\mathcal{C}$. Then a curve $\mathcal{C}$ can be ideal only if there is a constant $a > 0$ such that $\rho_G(\boldsymbol{x}) = a$ for all points on $\mathcal{C}\backslash\mathcal{S}_\mathcal{C}$, and $\rho_G(\boldsymbol{x}) \geq a$ for all points on $\mathcal{S}_\mathcal{C}$. That is to say, a smooth knotted curve $\mathcal{C}$ can be ideal only if its global radius of curvature function is constant and minimal on every curved segment of $\mathcal{C}$.

The above conclusion may be reached by a contradiction argument as follows. Let $\mathcal{C}^*$ be an ideal shape in $\mathcal{K}$ with arclength parametrization $\boldsymbol{q}^*(s)$, $0 \leq s \leq L$, and assume ρ_G is not constant on a curved segment of $\mathcal{C}^*$. Let a and d be the minimum and maximum of ρ_G on $\mathcal{C}^*$, and note that, by continuity, there is a number $c > a$ for which the set

$$F^* = \{\boldsymbol{x} \in \mathcal{C}^* \backslash \mathcal{S}_{\mathcal{C}^*} \mid \rho_G(\boldsymbol{x}) \geq c\}$$

is non-empty. Next, consider any number b in the open interval (a, c) and let

$$E^* = \{\boldsymbol{x} \in \mathcal{C}^* \mid \rho_G(\boldsymbol{x}) < b\}.$$

For any $\boldsymbol{x}$ in the set E^* recall that the lower interaction property of ρ_G implies the minimum in (6) is only achieved by points $\boldsymbol{y}$ and $\boldsymbol{z}$ that are also in the set E^*.

Let F_q^* and E_q^* be those subsets of $[0, L]$ corresponding to F^* and E^* under the parametrization $\boldsymbol{q}^*(s)$, and consider a second curve $\mathcal{C}^{**}$ with parametrization $\boldsymbol{q}^{**}(s)$, $0 \leq s \leq L$. For $s \in [0, L] \backslash F_q^*$ let $\boldsymbol{q}^{**}(s) = \boldsymbol{q}^*(s)$, and for $s \in F_q^*$ let

$$\boldsymbol{q}^{**}(s) = \boldsymbol{q}^*(s) + \epsilon\phi(s)\boldsymbol{q}^{*\prime\prime}(s)$$

where $\epsilon > 0$ is a parameter and $\phi(s)$ is a smooth, bounded function with compact support in F_q^*. Then $\mathcal{C}^{**}$ differs from $\mathcal{C}^*$ only in the interior of F_q^* and, for all $\epsilon > 0$ sufficiently small, $\mathcal{C}^{**}$ is strictly shorter than $\mathcal{C}^*$. Moreover, since $c > b$, the lower interaction property of ρ_G together with the continuity properties of $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ imply that $E_q^{**} = E_q^*$ for all sufficiently small $\epsilon > 0$, in which case $\mathcal{C}^{**}$ and $\mathcal{C}^*$ have the same minimum value of ρ_G. Thus, we can construct a curve $\mathcal{C}^{**}$ of the same knot type as $\mathcal{C}^*$, the same minimum value of ρ_G, but of strictly smaller length. Rescaling the length of $\mathcal{C}^{**}$ to L yields a curve in $\mathcal{K}$ whose minimum value of ρ_G is greater than that of $\mathcal{C}^*$, contradicting the hypothesis that $\mathcal{C}^*$ was ideal. Moreover, the constant value of ρ_G on curved segments of $\mathcal{C}^*$ must be the minimum value of ρ_G; otherwise, the same argument can be made.

5 Discrete Curves

In many cases the object of study is not a smooth curve, but rather a discrete curve, as in some numerical computations for example. By a *discrete curve* $\mathcal{C}_n$ we mean a list of distinct points or nodes $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ in three-dimensional space. To any discrete curve $\mathcal{C}_n$ one can associate a continuous, piecewise linear curve $\mathcal{C}$ by connecting $\boldsymbol{q}_1$ to $\boldsymbol{q}_2$ with a straight line and so on. Here we extend the definitions of the global radius of curvature, thickness and the ideal shape of a knot to this discrete case.

Given a discrete curve we define the global radius of curvature ρ_G at each node $\boldsymbol{q}_i$ of $\mathcal{C}_n$ by

$$\rho_G(\boldsymbol{q}_i) = \min_{\substack{1 \leq j,k \leq n \\ i \neq j \neq k \neq i}} r(\boldsymbol{q}_i, \boldsymbol{q}_j, \boldsymbol{q}_k).$$

That is, $\rho_G(\boldsymbol{q}_i)$ is the radius of the smallest circle containing $\boldsymbol{q}_i$ and two other distinct nodes $\boldsymbol{q}_j$ and $\boldsymbol{q}_k$. In the discrete case, we interpret the local radius of curvature ρ at a node $\boldsymbol{q}_i$ of $\mathcal{C}_n$ as $\rho(\boldsymbol{q}_i) = r(\boldsymbol{q}_{i-1}, \boldsymbol{q}_i, \boldsymbol{q}_{i+1})$. Finally, we can associate a thickness $\Delta[\mathcal{C}_n]$ to $\mathcal{C}_n$ by the expression

$$\Delta[\mathcal{C}_n] = \min_{1\leq i\leq n} \rho_G(\boldsymbol{q}_i).$$

That is, $\Delta[\mathcal{C}_n]$ is the radius of the smallest circle containing three distinct nodes $\boldsymbol{q}_i$, $\boldsymbol{q}_j$ and $\boldsymbol{q}_k$ of $\mathcal{C}_n$.

A mathematical definition of an ideal shape for knots formed by discrete curves can be formulated using $\Delta[\mathcal{C}_n]$. Let $\mathcal{K}_n$ denote the set of all discrete curves $\mathcal{C}_n$ with the property that the associated closed, piecewise linear curve $\mathcal{C}$ is simple, of a specified knot type and of a prescribed length $L > 0$. For n sufficiently large the set $\mathcal{K}_n$ is non-empty (a sufficiently fine discretization of a smooth curve of the given knot type is in this set), and one may consider the problem of finding those discrete curves $\mathcal{C}_n^*$ in $\mathcal{K}_n$ satisfying

$$\Delta[\mathcal{C}_n^*] = \sup_{\mathcal{C}_n\in\mathcal{K}_n} \Delta[\mathcal{C}_n]. \tag{12}$$

We call any $\mathcal{C}_n^*$ in $\mathcal{K}_n$ an ideal shape if it achieves the supremum in (12).

Just as in the smooth case, we can derive a necessary condition, implied by (12), that any ideal curve in $\mathcal{K}_n$ must satisfy. Given any $\mathcal{C}_n$ in $\mathcal{K}$ let $\mathcal{S}_n$ denote the set of all nodes $\boldsymbol{q}_i$ in $\mathcal{C}_n$ for which $\boldsymbol{q}_i$ and its two adjacent-index neighbors are collinear. Then a discrete curve $\mathcal{C}_n$ can be ideal only if there is a constant $a > 0$ such that $\rho_G(\boldsymbol{q}_i) = a$ for all nodes in $\mathcal{C}_n\backslash\mathcal{S}_n$, and $\rho_G(\boldsymbol{q}_i) \geq a$ for all nodes in $\mathcal{S}_n$. That is to say, a knotted discrete curve $\mathcal{C}_n$ can be ideal only if its global radius of curvature function is constant and minimal on every curved sequence of nodes in $\mathcal{C}_n$. This conclusion may be reached using a curve-shortening argument analogous to that in the smooth case.

Using the discrete global radius of curvature we were able to verify that previously computed ideal shapes of knots [1, 2] (more precisely, slightly corrected versions of them) are in agreement with our necessary condition. In Figure 1 we compare global and local radius of curvature for both a generic and an ideal 3_1 torus knot. For a generic 3_1 torus knot shape, ρ_G varies in a general manner with respect to arclength as shown by curve 1 in plot **d**. However, for data corresponding to an ideal shape [1], ρ_G varies according to our necessary condition as shown by curve 2 in plot **d**. The data for the ideal shape contains 160 points nearly uniformly spaced in arclength and was provided by the authors of [1, 2], who computed it using a Metropolis Monte Carlo procedure.

In Figure 2 we present results for a composite $3_1\#3_1$ knot. The data for this knot shape contains 286 points nearly uniformly spaced in arclength and was provided, as in the case of the 3_1 knot, by the authors of [1, 2]. The two upward spikes in both curves 1 and 2 in plot **c** occur at nearly straight regions of the knot as indicated by the dark bands in panel **a**. This behavior is in concordance with our necessary condition for idealness. However, the downward spikes in curve 2 are due to small regions of high local curvature, are in disagreement with our analytical characterization, and imply that the original shape computed in [2] is slightly non-ideal. Working with the original knot shape data, we performed a local smoothing in these regions of high local curvature and were able to correct, or improve, the original shape. The corrected shape not only satisfies our global curvature condition, but also has a larger value of $\Delta[\mathcal{C}]$ as can be seen by comparing curve 1 (corrected shape) and curve 2 (original shape) in plot **c**.

The corrected shape is shown in panels **a** and **b** and is visually identical to the original one computed in [2].

The data in Figures 1 and 2 fully support our necessary condition on global radius of curvature. In particular, global radius of curvature is constant to within 0.1 percent for the entire arclength interval in the case of the ideal 3_1 knot, and to within 0.4 percent for the three obvious intervals in the case of the ideal $3_1\#3_1$ knot. Furthermore, the data also suggest that, in contrast, local radius of curvature is not particularly simple on ideal shapes. For example, the oscillations in local radius of curvature around the arclength values of 0.4 and 0.9 in plot **d** of Figure 2 are well-resolved by the discretization and thus appear not to be numerical artifacts. Each of the two regions of oscillation occur over more than 20 nodes and we were unable to smooth these oscillations without causing significant decreases in the value of $\Delta[\mathcal{C}]$. Nevertheless, it is ultimately unclear as to what the discrete data of [1, 2] implies about ideal shapes in the continuous limit, that is, as the number of discretization points becomes infinite, and we are unaware of any results in this direction.

6 Discussion

We have shown that the notion of global radius of curvature provides a concise characterization of the thickness of a curve, and of certain ideal shapes of knots. The circumradius and global radius of curvature functions can apparently also be used in other ways. Here we mention applications in the study of knot energies, and describe a connection with the writhing number of a space curve.

Various authors have considered the concept of a *knot energy* [4, 8], which, for our purposes, will mean a functional $E[\mathcal{C}]$ that is defined and finite for any smooth, simple closed curve $\mathcal{C}$, and which tends to infinity as $\mathcal{C}$ tends to a non-simple curve. In other words, $E[\mathcal{C}]$ is a knot energy if it separates knot types by infinite energy barriers. The thickness energy $E[\mathcal{C}] = 1/\Delta[\mathcal{C}]$ is one example, but to evaluate this functional one must solve a minimization problem. Here we discuss *integral* knot energies, in which the functional $E[\mathcal{C}]$ is defined as the integral of an appropriate density function over the curve $\mathcal{C}$.

The most intuitive approach to the construction of an integral knot energy is the following. Given an arbitrary simple, closed curve $\mathcal{C}$ with parametrization $\boldsymbol{q}(s)$, let $f(s,\sigma) = |\boldsymbol{q}(s) - \boldsymbol{q}(\sigma)|$ be the pairwise Euclidean distance function for $\mathcal{C}$. Then, for any number $m \geq 2$, a candidate knot energy functional $E[\mathcal{C}]$ would be the double integral of $f^{-m}(s,\sigma)$. The basic idea here is that, for $m \geq 2$, this integral would tend to infinity as $\boldsymbol{q}(s)$ tends to $\boldsymbol{q}(\sigma)$ with $s \neq \sigma$, thus providing the required infinite energy barrier. However, such an integral would be divergent due to nearest-neighbor effects since $f(s,\sigma) = 0$ when $s = \sigma$. Thus, one needs to either *regularize* the integrand $f^{-m}(s,\sigma)$ by subtracting something equally divergent as $s \to \sigma$, or *mollify* the integrand using a multiplicative factor that tends to zero at an appropriate rate as $s \to \sigma$ [4, 8].

The circumradius and global radius of curvature functions introduced in this article lead to families of integral knot energies that do not require explicit regularization or mollification. To consider some natural examples, first note that the functions $1/r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$, $1/r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y})$ and $1/\rho_G(\boldsymbol{x})$ are all bounded, continuous functions as $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ vary on a smooth, simple closed curve $\mathcal{C}$. For any $p \geq 2$, the L_p-norms of these functions lead to the candidate energy functionals

$$U_{p,3}[\mathcal{C}] = \left(\iiint \frac{1}{r^p(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})} \, \mathrm{d}\mathcal{C}_{\boldsymbol{x}} \mathrm{d}\mathcal{C}_{\boldsymbol{y}} \mathrm{d}\mathcal{C}_{\boldsymbol{z}} \right)^{1/p}$$

$$U_{p,2}[\mathcal{C}] = \left(\iint \frac{1}{d^p(\boldsymbol{x},\boldsymbol{y})} \, \mathrm{d}\mathcal{C}_{\boldsymbol{x}} \mathrm{d}\mathcal{C}_{\boldsymbol{y}} \right)^{1/p}$$

$$U_{p,1}[\mathcal{C}] = \left(\int \frac{1}{\rho_G^p(\boldsymbol{x})} \, \mathrm{d}\mathcal{C}_{\boldsymbol{x}} \right)^{1/p}$$

where the integrals are taken over the curve $\mathcal{C}$ and $d(\boldsymbol{x},\boldsymbol{y}) = r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y})$. Using the definition of the L_∞-norm and the fact that the integrands are bounded and continuous, we obtain the results

$$\lim_{p\to\infty} U_{p,i}[\mathcal{C}] = 1/\Delta[\mathcal{C}], \qquad (i = 1, 2, 3)$$

which connect these families of integral energies to the thickness energy. For $p = 2$ the functional $U_{p,2}[\mathcal{C}]$, and its symmetric counterpart

$$\widehat{U}_{p,2}[\mathcal{C}] = \left(\iint \frac{1}{[d(\boldsymbol{x},\boldsymbol{y})d(\boldsymbol{y},\boldsymbol{x})]^{p/2}} \, \mathrm{d}\mathcal{C}_{\boldsymbol{x}} \mathrm{d}\mathcal{C}_{\boldsymbol{y}} \right)^{1/p},$$

can be shown to be equivalent to knot energies recently studied by Buck and Simon [4].

There is also a connection between the non-symmetric doublet function $d(\boldsymbol{x},\boldsymbol{y})$ and the writhing number of a space curve [6]. In the Gauss integral form, the writhing number $\mathrm{Wr}[\mathcal{C}]$ of an oriented, smooth, simple space curve $\mathcal{C}$ is

$$\mathrm{Wr}[\mathcal{C}] = \frac{1}{4\pi} \iint \frac{\boldsymbol{e} \cdot (\boldsymbol{t}_{\boldsymbol{y}} \times \boldsymbol{t}_{\boldsymbol{x}})}{|\boldsymbol{y} - \boldsymbol{x}|^2} \, \mathrm{d}\mathcal{C}_{\boldsymbol{x}} \mathrm{d}\mathcal{C}_{\boldsymbol{y}}$$

where $\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{x})/|\boldsymbol{y} - \boldsymbol{x}|$, and $\boldsymbol{t}_{\boldsymbol{x}}$ and $\boldsymbol{t}_{\boldsymbol{y}}$ are the unit tangents to $\mathcal{C}$ at $\boldsymbol{x}$ and $\boldsymbol{y}$. Using the doublet function $d(\boldsymbol{x},\boldsymbol{y})$ one then has

$$\mathrm{Wr}[\mathcal{C}] = \frac{1}{4\pi} \iint \frac{\sin\psi(\boldsymbol{x},\boldsymbol{y})}{d(\boldsymbol{x},\boldsymbol{y})d(\boldsymbol{y},\boldsymbol{x})} \, \mathrm{d}\mathcal{C}_{\boldsymbol{x}} \mathrm{d}\mathcal{C}_{\boldsymbol{y}}$$

where $\psi(\boldsymbol{x},\boldsymbol{y})$ denotes the angle, with an appropriate sign convention, between the two planes containing the circumcircles associated with $d(\boldsymbol{x},\boldsymbol{y})$ and $d(\boldsymbol{y},\boldsymbol{x})$.

References

[1] Katritch, V., Bednar, J., Michoud, D., Scharein, R.G., Dubochet, J. & Stasiak, A. (1996) *Nature (London)* **384**, 142–145.

[2] Katritch, V., Olson, W.K., Pieranski, P., Dubochet, J. & Stasiak, A. (1997) *Nature (London)* **388**, 148–151.

[3] Stasiak, A., Katritch, V., Bednar, J., Michoud, D. & Dubochet, J. (1996) *Nature (London)* **384**, 122.

[4] Buck, G. & Simon, J. (1997) in *Lectures at Knots 96*, ed. Suzuki, S. (World Scientific, Singapore).

[5] Litherland, R.A., Simon, J., Durumeric, O. & Rawdon, E. (1996) *Topology and its Applications (Elsevier, Amsterdam)* **91**, 233–244.

[6] Fuller, F.B. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 815–819.

[7] Coxeter, H.S.M. (1969) *Introduction to Geometry* (Wiley, New York), 2nd Ed.

[8] Simon, J. (1996) in *Mathematical Approaches to Biomolecular Structure and Dynamics*, eds. Mesirov, J.P., Schulten, K. & Sumners, D.W. (Springer, New York), pp. 39–58.

Acknowledgments. It is a pleasure to thank A. Stasiak for posing questions leading to this work, V. Katritch and P. Pieranski for making their knot data available, R. Paffenroth for his assistance in use of the DataViewer graphics package and POV-Ray ray-tracing program and D. Singer for his insightful comments. OG was partially supported by the US National Science Foundation.

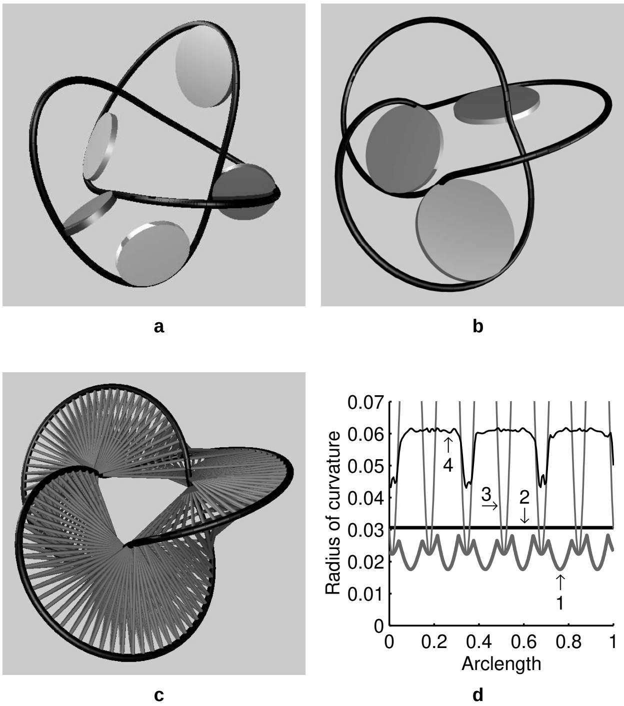

Figure 1: Interpretation of the global radius of curvature ρ_G at various points on a non-ideal and an ideal shape of a 3_1 torus knot. **a**, generic shape produced using a simple parametric representation. The light grey discs indicate points where ρ_G can be associated with the local radius of curvature, and the dark grey discs indicate points where ρ_G can be associated with a distance of closest approach. **b**, numerically computed ideal shape. Here ρ_G corresponds to a distance of closest approach at all points. **c**, same shape as in **b** but with a different visualization of the global radius of curvature. Each of the several spokes emanating from a point on the curve represents the diameter of a disc that realizes ρ_G at that point. **d**, global and local radius of curvature plots for the shapes in **a** and **b**. Curves 1 and 3 are global and local radius of curvature, respectively, for the non-ideal shape **a**. Curves 2 and 4 are global and local radius of curvature, respectively, for the ideal shape **b**. (Curve 3 is nearly periodic but its upper limits are not contained within the plot range.)

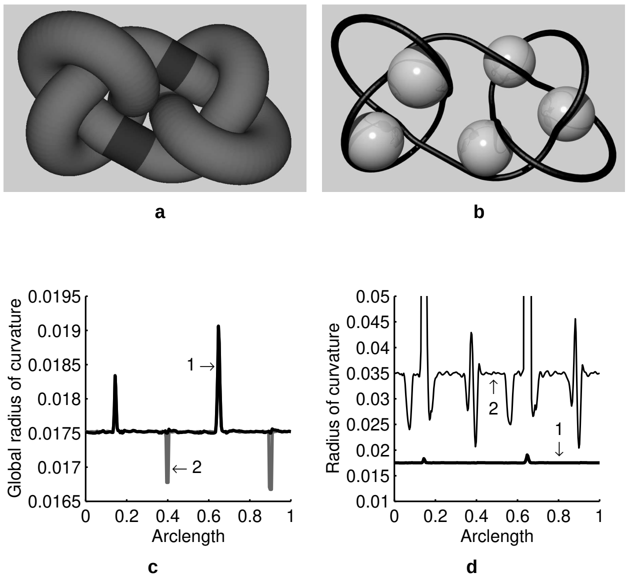

Figure 2: Interpretation of the minimum global radius of curvature for a numerically computed ideal $3_1\#3_1$ knot. **a**, the tube interpretation. The minimal value of ρ_G is the radius of the tube shown here. The dark bands on the tube indicate regions of near-zero curvature (straight portions). **b**, the sphere interpretation. Any spherical shell of radius less than the minimum value of ρ_G cannot intersect the curve at three or more points (counting tangencies twice). The spheres shown here have radius equal to the minimum value of ρ_G. **c**, global radius of curvature plot for shape in **a**. Curve 2 corresponds to raw data from Katritch *et al.* [2] and curve 1 corresponds to a corrected shape. Curves 1 and 2 are nearly identical except for the two downward spikes in curve 2. **d**, comparison of global radius of curvature (curve 1) and local radius of curvature (curve 2) for the corrected shape.