

# Generalization of distance to higher dimensional objects

Steven S. Plotkin<sup>†</sup>

Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC, Canada V6T 1Z1

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved June 26, 2007 (received for review September 7, 2006)

**The measurement of distance between two objects is generalized to the case where the objects are no longer points but are one-dimensional. Additional concepts such as nonextensibility, curvature constraints, and noncrossing become central to the notion of distance. Analytical and numerical results are given for some specific examples, and applications to biopolymers are discussed.**

The distance, as conventionally defined between two zero-dimensional objects (points)  $A$  and  $B$  at positions  $\mathbf{r}_A$  and  $\mathbf{r}_B$ , is the minimal arclength travelled in the transformation from  $A$  to  $B$ . A transformation  $\mathbf{r}(t)$  between  $A$  and  $B$  is a vector function that may be parametrized by a scalar variable  $t$ :  $0 \leq t \leq T$ ,  $\mathbf{r}(0) = \mathbf{r}_A$ ,  $\mathbf{r}(T) = \mathbf{r}_B$ , and the distance travelled is a functional of  $\mathbf{r}(t)$ . The (minimal) transformation  $\mathbf{r}^*(t)$  is an object of dimension one higher than  $A$  or  $B$ ; i.e., it yields a distance that is one-dimensional. The distance  $\mathcal{D}^*$  is found through the variation of the functional (1):

$$\mathcal{D}^* = \mathcal{D}[\mathbf{r}^*(t)] \text{ where } \mathbf{r}^*(t) \text{ satisfies} \quad [1a]$$

$$\delta \int_0^T dt (g_{\mu\nu} \dot{x}^\mu(t) \dot{x}^\nu(t))^{1/2} = \delta \int_0^T dt \sqrt{\dot{\mathbf{r}}^2} = 0. \quad [1b]$$

Here,  $\dot{x} = dx/dt$ ,  $\dot{\mathbf{r}} = d\mathbf{r}/dt$ , and we have let  $g_{\mu\nu} = \delta_{\mu\nu}$  (Euclidean metric). The boundary conditions mentioned above are present at the end points of the integral. The Einstein summation convention will be used where convenient (e.g., eq. 1b); however, all the analysis here deals with spatial coordinates,  $\nu = 1, 2, 3$ , on a Euclidean metric. Generalizations to dimension higher than 3, as well as non-Euclidean metrics, are straightforward to incorporate into the formalism.

On a Euclidean metric, the minimal distance becomes the diagonal of a hyper-cube. However, formulated as above, the solutions minimizing  $\mathcal{D}$  are infinitely degenerate, because particles moving at various speeds but tracing the same trajectory over the total time  $T$  all give the same distance. To circumvent this problem, what is typically done is to let one of the space variables (e.g.,  $x$ ) become the independent variable. However, for higher dimensional objects or zero dimensional objects on a manifold with nontrivial topology, there is no guarantee that the dependent variables ( $y, z$ ) constitute single valued functions of  $x$ . Alternatively, one can study the “time” trajectory of the parametric curve defined above, but under a gauge that fixes the speed to a constant,  $v_o$ , for example. One can either fix the gauge from the outset with Lagrange multipliers, or choose a gauge that may simplify the problem after finding the extremum equations. The latter is often simpler in practice.

To be specific, the effective Lagrangian  $\mathcal{L}$  appearing in the above problem is  $\sqrt{\dot{\mathbf{r}}^2}$ , and the Euler–Lagrange (EL) equations are

$$\frac{d}{dt} \left( \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}} \right) = 0 \quad \text{or} \quad \frac{d}{dt} \left( \frac{\dot{\mathbf{r}}}{|\dot{\mathbf{r}}|} \right) = \dot{\hat{\mathbf{v}}} = 0, \quad [2]$$

with  $\hat{\mathbf{v}}$  the unit vector in the direction of the velocity. The boundary conditions are

$$\mathbf{r}^*(0) = \mathbf{r}_A \quad \text{and} \quad \mathbf{r}^*(T) = \mathbf{r}_B. \quad [3]$$

Because the derivative of a unit vector is always orthogonal to that vector, Eq. 2 says that the direction of the velocity cannot change, and therefore straight line motion results. Applying the boundary conditions gives  $\hat{\mathbf{v}} = (\mathbf{r}_B - \mathbf{r}_A)/|\mathbf{r}_B - \mathbf{r}_A|$ . However, any function  $\mathbf{v}(t) = |v_o(t)|\hat{\mathbf{v}}$  satisfying the boundary conditions is a solution, so long as  $\int_0^T dt |v_o(t)| = |\mathbf{r}_B - \mathbf{r}_A|$ . This is the infinite degeneracy of solutions mentioned above. Then  $\mathbf{r}^*(t) = \mathbf{r}_A + (\mathbf{r}_B - \mathbf{r}_A)/|\mathbf{r}_B - \mathbf{r}_A| \int_0^t dt |v_o(t)|$ , and  $\mathcal{D}^* = \int_0^T dt \sqrt{\dot{\mathbf{r}}^{*2}} = \int_0^T dt |v_o(t)| = |\mathbf{r}_B - \mathbf{r}_A|$ . At this point we could fix the parameterization by choosing  $|v_o(t)| = |\mathbf{r}_B - \mathbf{r}_A|/T$  (constant speed), for example.

The extremum is a minimum, as can be shown by analyzing the eigenvalues of the matrix  $\partial^2 \mathcal{D}/\partial x_\nu(t) \partial x_\mu(t') = -\delta_{\mu\nu} \delta''(t - t')$ . Diagonalizing by Fourier transform gives positive elements  $+\omega_n^2 \delta_{\mu\nu} \delta(\omega_n - \omega_n)$  for the stability matrix and thus positive eigenvalues.

In what follows, we generalize the notion of distance to higher dimensional objects, specifically space curves. We will see many of the above themes reiterated, as well as some fundamentally new features that emerge when one treats the space curves as nonextensible, having some persistence length or curvature constraint, and noncrossing or unable to pass through themselves. We provide analytical and numerical results for some prototypical examples for nonextensible chains, and we lay the foundations for treating curvature and noncrossing constraints.

## Distance Metric for One-Dimensional Objects

The distance  $\mathcal{D}^*$  between two one-dimensional objects (which we refer to as space curves or strings)  $A$  and  $B$  having configurations  $\mathbf{r}_A(s)$  and  $\mathbf{r}_B(s)$ ,  $0 \leq s \leq L$ , is obtained from the transformation from  $A$  to  $B$  that minimizes the integrated distance travelled. By integrated distance we mean the cumulative arclength all elements of the string had to move in the transformation from  $A$  to  $B$ . For the transformation to exist, strings  $A$  and  $B$  must have the same length (although this condition may be relaxed by allowing specific extensions or contractions). For the distance to be finite, open space curves must be finite in length. For closed non-crossing space curves,  $A$  and  $B$  must be in the same topological class for the transformation to exist. Describing the transformation  $\mathbf{r}(s, t)$  requires two scalar parameters, one for arc length  $s$  along the string and another measuring progress as in the zero-dimensional case, say  $t$ :  $0 \leq t \leq T$ , so that  $\mathbf{r}(s, 0) = \mathbf{r}_A(s)$  and  $\mathbf{r}(s, T) = \mathbf{r}_B(s)$ . The distance travelled is a functional of the vector function  $\mathbf{r}(s, t)$ . The minimal transformation  $\mathbf{r}^*(t, s)$  is an object of dimension one higher than  $A$  or  $B$ , i.e., it yields a distance that is two-dimensional. The problem does not map to a simple soap film, since there are many configuration pairs that have zero area between them but nonzero

Author contributions: S.S.P. performed research and wrote the paper.

The author declares no conflict of interest.

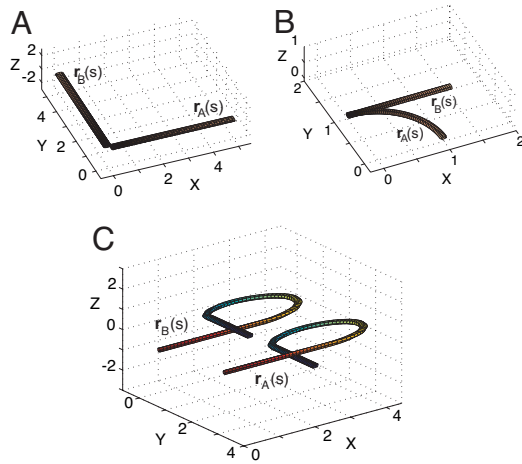
This article is a PNAS Direct Submission.

Abbreviation: EL, Euler–Lagrange; MRSD, mean root square distance.

<sup>†</sup>E-mail: steve@physics.ubc.ca.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0607833104/DC1](http://www.pnas.org/cgi/content/full/0607833104/DC1).

© 2007 by The National Academy of Sciences of the USA



**Fig. 1.** Three representative pairs of curves. (A) Straight line curve rotated by  $\pi/2$ . (B) One string has a finite radius of curvature, the other is straight. (C) A canonical example where noncrossing is important; the curves are displaced for easy visualization but should be imagined to be superimposed.

distance travelled, e.g., a straight line displaced along its own axis, or that in Fig. 1C. The analogue to a higher-dimensional surface of minimal area when the “time”  $t$  is included is closer but inexact (see footnote below).

We can construct the effective Lagrangian along the same lines as the zero-dimensional case. Using the shorthand  $\mathbf{r} \equiv \mathbf{r}(s, t)$ ,  $\dot{\mathbf{r}} \equiv \partial\mathbf{r}/\partial t$ ,  $\mathbf{r}' \equiv \partial\mathbf{r}/\partial s$ , the distance travelled is<sup>‡</sup>

$$\mathcal{D} = \int_0^L ds \int_0^T dt \sqrt{\dot{x}^v \dot{x}_v} = \int_0^L ds \int_0^T dt \sqrt{\dot{\mathbf{r}}^2}. \quad [4]$$

However, to meaningfully represent the distance a string must move to reconfigure itself from conformation  $A$  to  $B$ , the transformation must be subject to several auxiliary conditions.

The first of these is nonextensibility. Points along the space curve cannot move independently of one another but are constrained to integrate to fixed length, so the curve cannot stretch or contract. Thus there is a Lagrange multiplier  $\lambda(s, t)$  weighting the (nonholonomic) constraint:

$$\sqrt{\mathbf{r}'^2} = 1. \quad [5]$$

This constraint ensures a parameterization of the string with unit tangent vector  $\hat{\mathbf{t}} = \mathbf{r}'$ , so that the total length of the string is  $L = \int_0^L ds \sqrt{\mathbf{r}'^2} = \int_0^L ds$ . In the language of differential geometry, the space curve is a unit-speed curve.

<sup>‡</sup>The distance-metric action in Eq. 4 bears a strong resemblance to the Nambu–Goto action for a classical relativistic string (2):  $S_{\text{NG}}[\mathbf{r}(s, t)] = \int d\sigma d\tau \sqrt{(\dot{\mathbf{r}}\mathbf{r}')^2 - (\dot{\mathbf{r}}^2(\mathbf{r}')^2)}$ , where  $\mathbf{r}$  in  $S_{\text{NG}}$  is now a four-vector and the dot product is the relativistic dot product. This action is physically interpreted as the (Lorentz invariant) world-sheet area of the string. If Eq. 4 could be mapped by suitable choice of gauge to the minimization of the Nambu–Goto action, one could exploit here the same reparameterization invariance that results in wave equation solutions to the equations of motion for the classical relativistic string, by choosing a parameterization such that  $\dot{\mathbf{r}}\mathbf{r}' = 0$  (for the purely geometrical problem, the discriminant under the square root in the action has opposite sign). Unfortunately, however, because the velocity in the distance-metric action is a 3-velocity rather than a 4-velocity, our action only accumulates area when parts of the string move in 3-space, in contrast to the Nambu–Goto action that accumulates area even for a static string. The distance-metric action Eq. 4 has a lower symmetry than that for the classical relativistic string.  $\mathcal{D}^*$  cannot depend on the time the transformation took, whereas the world sheet area does. Conversely, if we take, for example, configuration  $A$  at  $t = 0$  to be a straight line of length  $L$ , and configurations  $B$  at  $t = T$  to be the same straight line but displaced along its own axis by varying amounts  $d$ , the geometrical area for all transformations would be  $LT$ , whereas the distances  $\mathcal{D}_{AB}^*$  for each transformation would be  $Ld$ .

If the constraint (Eq. 5) were not present in Eq. 4, each point along the space curve could follow a straight line path from  $A$  to  $B$  and the problem of minimizing the distance would be trivial. Equivalently, setting  $\lambda = 0$  should reduce the problem to a sum of straight lines analogously to the zero-dimensional case above.

As in the case of distance between points, one can fix the  $t$ -parameterization from the outset by introducing a Lagrange multiplier  $\alpha(t)$  that fixes the total distance covered per time  $\int_0^L ds \sqrt{\dot{\mathbf{r}}^2}$  to a known function  $f(t)$ . Although this approach removes the infinite degeneracy mentioned above, as a global isoperimetric condition it reduces the symmetry of the problem. For example, there would then be no conservation law that could be written to capture the invariance of the effective Lagrangian with respect to the independent variable  $t$ . For these reasons, we choose to leave the answer as unparameterized with respect to  $t$ , analogous to the point-distance case above.

**Ideal Chains.** There are many examples of transformations between two strings  $A$  and  $B$  where chain noncrossing is unimportant (e.g., Fig. 1A and B). Here we derive the EL equations for this case.

From Eqs. 4 and 5, the extrema of the distance  $\mathcal{D}$  are found from

$$\begin{aligned} \delta\mathcal{D} &= \delta \int_0^L \int_0^T ds dt \mathcal{L}(\dot{\mathbf{r}}, \mathbf{r}') = 0 \\ \text{or } \delta \int_0^L \int_0^T ds dt \left( \sqrt{\dot{\mathbf{r}}^2} - \lambda \sqrt{\mathbf{r}'^2} \right) &= 0. \end{aligned} \quad [6]$$

Performing the variation gives

$$\begin{aligned} \delta\mathcal{D} &= \int_0^L ds [\mathbf{p}_t \cdot \delta\mathbf{r}]_0^T + \int_0^T dt [\mathbf{p}_s \cdot \delta\mathbf{r}]_0^L \\ &\quad - \int_0^L \int_0^T ds dt \delta\mathbf{r} \cdot \left[ \frac{d\mathbf{p}_s}{ds} + \frac{d\mathbf{p}_t}{dt} \right] = 0, \end{aligned} \quad [7]$$

where the generalized momenta  $\mathbf{p}_t$  and  $\mathbf{p}_s$  are given by:

$$\mathbf{p}_t = \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{r}}} = \hat{\mathbf{v}} \quad \text{and} \quad \mathbf{p}_s = \frac{\partial\mathcal{L}}{\partial\mathbf{r}'} = -\lambda\hat{\mathbf{t}}, \quad [8]$$

where  $\hat{\mathbf{v}}$  is again the unit velocity vector, and  $\hat{\mathbf{t}}$  is the unit tangent to the curve.

The EL equation follows from the last term in Eq. 7, and yields a partial differential equation for the minimal transformation  $\mathbf{r}^*(s, t)$ :

$$(\dot{\mathbf{r}}^2)\ddot{\mathbf{r}} - (\dot{\mathbf{r}}\ddot{\mathbf{r}})\dot{\mathbf{r}} = |\dot{\mathbf{r}}|^3(\lambda\mathbf{r}'' + \lambda'\mathbf{r}'), \quad [9]$$

where we have used the facts that  $|\mathbf{r}'| = 1$  and  $\mathbf{r}' \cdot \mathbf{r}'' \equiv \hat{\mathbf{t}} \cdot \boldsymbol{\kappa} = 0$ , since the tangent is always orthogonal to the curvature at any given point along a space curve.

Eq. 9 can be written in terms easier to understand intuitively by using the unit velocity vector  $\hat{\mathbf{v}}$ , tangent  $\hat{\mathbf{t}}$ , and curvature  $\boldsymbol{\kappa}$ :<sup>§</sup>

$$\dot{\hat{\mathbf{v}}} = \lambda\boldsymbol{\kappa} + \lambda'\hat{\mathbf{t}}. \quad [10]$$

Comparing Eqs. 10 and 2, we confirm that setting the Lagrange multiplier  $\lambda$  corresponding to the nonextensibility condition to zero

<sup>§</sup>The invariance of the Lagrangian to  $(s, t)$  leads to conservation laws by Noether’s theorem (1), which here take the form of divergence conditions. However, these generally contain no new information beyond the EL equations and can be obtained by dotting Eq. 10 with either  $\mathbf{r}'$  to give  $\lambda' = \dot{\hat{\mathbf{v}}} \cdot \hat{\mathbf{t}}$  or  $\dot{\mathbf{r}}$  to give  $\mathbf{v} \cdot (\lambda\hat{\mathbf{t}}) = 0$ .

results in straight line solutions for all points along the space curve. Conversely, the condition that the space curve form a contiguous object results generally in nonzero deviation from straight line motion. So in comparing various extremal solutions to Eq. 10, the minimal solution tends to minimize  $|\lambda|$  everywhere.

The boundary conditions are obtained from the first two terms in Eq. 7. Because the initial and final configurations are specified, the variation  $\delta\mathbf{r}$  vanishes at  $t = 0, T$ , and the corresponding boundary conditions, or initial and final conditions, are:

$$\mathbf{r}^*(s, 0) = \mathbf{r}_A(s) \text{ and } \mathbf{r}^*(s, T) = \mathbf{r}_B(s). \quad [11]$$

Because the end points of the string are free during the transformation,  $\delta\mathbf{r} \neq 0$  at  $s = 0, L$ , and so the conjugate momenta must vanish:  $\mathbf{p}_s(0, t) = \mathbf{p}_s(L, t) = 0$ . This means that  $\lambda\hat{\mathbf{t}} = 0$  at the end points. However, because  $\hat{\mathbf{t}}$  cannot be zero, the only way this can occur is for  $\lambda(0, t) = \lambda(L, t) = 0$ . The Lagrange multiplier, which represents the conjugate force or tension to ensure an inextensible chain, must vanish at the end points of the string. If  $\lambda = 0$ , the EL Eq. 10 gives  $\dot{\hat{\mathbf{v}}} = \lambda'\hat{\mathbf{t}}$  at the end points. However, because  $\hat{\mathbf{v}}$  is a unit vector,  $\dot{\hat{\mathbf{v}}}$  is orthogonal to  $\hat{\mathbf{v}}$  (or  $\mathbf{v}$ ), and we have finally the boundary conditions at the end points of the string:

$$\lambda'\mathbf{v} \cdot \hat{\mathbf{t}} = 0 \quad (\text{at the end points}). \quad [12]$$

Eq. 12 has three possible solutions. One is that  $\mathbf{v} \cdot \hat{\mathbf{t}} = 0$  or equivalently  $\hat{\mathbf{r}} \cdot \mathbf{r}' = 0$ , which corresponds to pure rotation of the end points. It is worth mentioning that the end points of the classical relativistic string also move transversely to the string. Moreover because of the Minkowski metric the end points must also move at the speed of light. Here, however, because Lorentz invariance is not at issue, additional solutions are possible. The end points of our string can be at rest,  $\mathbf{v} = 0$ , and satisfy Eq. 12. The last solution of Eq. 12 is for  $\lambda' = 0$ . Because  $\lambda$  also vanishes at the end points, Eq. 10 gives  $\dot{\hat{\mathbf{v}}} = 0$ , or straight line motion. In summary, the three possible boundary conditions for the string end points are:

$$\mathbf{v} \cdot \hat{\mathbf{t}} = 0 \quad (\text{pure rotation}) \quad [13a]$$

$$\mathbf{v} = 0 \quad (\text{at rest}) \quad [13b]$$

$$\dot{\hat{\mathbf{v}}} = 0 \quad (\text{straight line motion}). \quad [13c]$$

Whether an extremal transformation is a minimum can be determined by examining the second variation of the functional (Eq. 6):

$$\delta^2\mathcal{D} = \frac{1}{2} \int_0^L \int_0^T ds dt [\delta\hat{\mathbf{r}} \cdot \mathbf{I} \cdot \delta\hat{\mathbf{r}} + \delta\mathbf{r}' \cdot \mathbf{\Lambda} \cdot \delta\mathbf{r}'], \quad [14]$$

where  $\mathbf{I}_{ij} = (\hat{\mathbf{r}}^2\delta_{ij} - \hat{r}_i\hat{r}_j)/|\hat{\mathbf{r}}|^3$  and  $\mathbf{\Lambda}_{ij} = -\lambda(s, t)\delta_{ij}$ , and  $\delta\mathbf{r}'$  and  $\delta\hat{\mathbf{r}}$  are the  $s$  and  $t$  derivatives of the variation  $\delta\mathbf{r}$  from the extremal path.

We now apply these concepts to some specific examples.

**Examples. Translations.** If two space curves differ by a translation,  $\mathbf{r}_B(s) = \mathbf{r}_A(s) + \mathbf{d}$  with  $\mathbf{d}$  a constant vector. The appropriate boundary condition for the end points is Eq. 13c. The points along the string can all satisfy Eq. 10 with  $\dot{\hat{\mathbf{v}}} = 0$  and  $\lambda = 0$  everywhere (because  $\hat{\mathbf{t}}, \kappa \neq 0$ ), and straight line motion results:  $\mathbf{r}^*(s, t) = \mathbf{r}_A(s) + (\mathbf{r}_B(s) - \mathbf{r}_A(s))t/T$ . The distance  $\mathcal{D}^* = L|\mathbf{d}|$ . This is the one-dimensional analogue to Eqs. 2 and 3.

**Piece-wise linear space curves.** Suppose initially the curvature of some section of the string is zero. Then, taking the dot product of  $\mathbf{v}$  with Eq. 10, we see that Eq. 12 holds for all points along the string. So the string either rotates or translates (or remains at rest if that

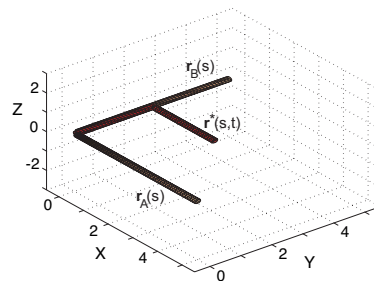
segment has completed the transformation). If both  $\mathbf{r}_A$  and  $\mathbf{r}_B$  are straight lines as in Fig. 1A, Eq. 12 holds for both. It is then reasonable to seek solutions  $\mathbf{r}^*$  of the EL equation such that Eq. 12 holds for all  $(s, t)$ .

Consider the two space curves shown in Fig. 1A with  $\mathbf{r}_A(s) = s\hat{\mathbf{x}}$  and  $\mathbf{r}_B(s) = s\hat{\mathbf{y}}$ , both with curvature  $\kappa = 0$ . We first investigate rotation from  $A$  to  $B$ . This transformation satisfies the EL equation so appears to be extremal:  $\mathbf{r} = s\hat{\mathbf{r}} = s(\cos\omega t\hat{\mathbf{x}} + \sin\omega t\hat{\mathbf{y}})$ . The velocity  $\dot{\mathbf{r}} = s\omega\hat{\theta}$ , so the distance  $D[\mathbf{r}_{\text{ROT}}(s, t)] = \pi L^2/4$ . Taking the dot product of  $\hat{\mathbf{t}}$  with Eq. 10 gives  $\lambda' = \hat{\mathbf{t}} \cdot \dot{\hat{\mathbf{v}}} = -\omega$  or  $\lambda(s, t) = \lambda_0 - \omega s$ . For the transformation to be extremal, the conjugate momenta must also vanish at the string end points, or  $\lambda(0, t) = \lambda(L, t) = 0$ . This is impossible to achieve with this functional form, so the transformation is not extremal, unless we include the subsidiary condition here that  $\mathbf{r}_A(0, t) = \mathbf{r}_B(0, t)$ . Then the end point of the string at  $s = 0$  is determined, and the variations  $\delta\mathbf{r}(0, t)$  must vanish. Now only  $\lambda(L, t) = 0$ , and so  $\lambda(s, t) = \omega(L - s)$ . The transformation is extremal.

Whether it is a minimum can be determined by examining the second variation (Eq. 14). For the transformation  $\mathbf{r}_{\text{ROT}}(s, t)$ , the matrix  $\mathbf{I}$  in Eq. 14 is nonnegative definite, a necessary condition for a local minimum (1); however,  $\mathbf{\Lambda}$  is negative definite, so the character of the extremum is determined by the interplay of the two terms in Eq. 14. Variations  $\delta\mathbf{r}$  that preserve  $\mathbf{r}'^2 = 1$  or  $2\hat{\mathbf{t}} \cdot \delta\mathbf{r}' = 0$  are satisfied in this example by  $\delta\mathbf{r} = f(s, t)\hat{\theta}$ , where  $f(s, t)$  must satisfy the boundary conditions  $\delta\mathbf{r}(0, t) = \delta\mathbf{r}(s, 0) = \delta\mathbf{r}(s, T) = 0$ . We thus let the variations have the functional form:  $\delta\mathbf{r} = \varepsilon \sin(ks) \sin(n\pi t/T)\hat{\theta}$ , where  $\hat{\theta} = -\sin\omega t\hat{\mathbf{x}} + \cos\omega t\hat{\mathbf{y}}$ ,  $n$  is a positive integer, and  $k$  is unrestricted. Inserting this functional form for the variations into Eq. 14 gives  $\delta^2\mathcal{D} = (\varepsilon^2\pi/8)F(kL)$ , where  $F(x)$  is a nonpositive, monotonically decreasing function, with a maximum of zero at  $kL = 0$ . In fact to lowest order  $F(kL) \approx -(\pi\varepsilon^2/2160)(kL)^6$ . The extremum corresponding to pure rotation of curve  $\mathbf{r}_A$  into  $\mathbf{r}_B$  is a maximum!

The only other solution to Eqs. 10 and 12 for all  $(s, t)$  is for each point  $s$  on  $\mathbf{r}_A(s)$  to be connected to a corresponding point on  $\mathbf{r}_B(s)$  by a straight line, corresponding to Eq. 13c. Eq. 12 holds everywhere because  $\lambda'(s, t) = 0$ . Because  $\lambda$  is zero at the boundaries, it is thus zero everywhere.

An intermediate configuration then has the shape of a piecewise linear curve with a right angle “kink” at  $s^*(t)$  (see Fig 2). As  $t$  progresses, the kink propagates along curve  $\mathbf{r}_B$ , and the “free” part of the chain follows straight line diagonal motion, shrinking as its left end is overlaid onto curve  $\mathbf{r}_B$ . The solution for the velocity at all  $(s, t)$  is given by  $\mathbf{v}(s, t) = v_o(t)\Theta(s - s^*(t))\hat{\mathbf{e}}_v$  where  $s^*(t)$  is the position of the tangent discontinuity in Fig. 2, which goes from  $s^*(0) = 0$  to  $s^*(T) = L$  as  $t$  goes from 0 to  $T$ .  $\hat{\mathbf{e}}_v$  is a unit vector along the direction of the velocity,  $\hat{\mathbf{e}}_v = (-\hat{\mathbf{x}} + \hat{\mathbf{y}})/\sqrt{2}$ , and  $v_o(t)$  is a speed which can be taken constant. By simple geometry,  $v_o = \sqrt{2}s^*$ .



**Fig. 2.** The minimal transformation from  $A$  to  $B$  in Fig. 1A involves the propagation of a kink along curve  $B$ . The end point of the curve at intermediate states satisfies  $x + y = L$ , the equation for a straight line. A similar linear equation holds for any point on the curve; thus, no solution with shorter distance can exist. An intermediate configuration is shown in red. Alternative transformations are possible with kinks along  $A$  as well as multiple kinks (see text).

Because  $s^*(T) = L$ ,  $v_o = \sqrt{2}L/T$  and  $s^*(t) = Lt/T$ . The total distance travelled from Eq. 4 is then  $\mathcal{D}^* = L^2/\sqrt{2}$ .

Because the transformation involves straight line motion, it is minimal. This can be seen from the second variation Eq. 14. The shape of the curve at all times is given by

$$\mathbf{r}^*(s, t) = s\Theta(Lt/T - s)\hat{\mathbf{y}} + (Lt/T)\Theta(s - Lt/T)\hat{\mathbf{y}} + (s - Lt/T)\Theta(s - Lt/T)\hat{\mathbf{x}}. \quad [15]$$

Taking variations from the extremal path as before, let  $\delta\mathbf{r} = \varepsilon \sin k(s - Lt/T) \sin(n\pi t/T)\Theta(s - Lt/T)\hat{\mathbf{y}}$ . These variations only act on the “free” part of the string and preserve a unit tangent to first order. The matrix  $\Lambda$  in Eq. 14 is zero for straight line transformations where  $\lambda = 0$ . The quadratic form  $\delta\mathbf{r} \cdot \mathbf{I} \cdot \delta\mathbf{r}$  is nonnegative, and results in a second variation  $\delta^2\mathcal{D} = \varepsilon^2(32\sqrt{2})^{-1}[(kL)^2 + (n\pi)^2(1 - \text{sinc}^2(kL))]$ , which is nonnegative, monotonically increasing in  $kL$ , and quadratic to lowest order, with a minimum of zero at  $kL = 0$ . The transformation is indeed minimal.

Likewise, the minimal distance to fold a string of length  $L$  upon itself starting from a straight line (to form a hairpin) is  $\mathcal{D}^* = L^2/4$ . **Solution degeneracy.** In the above example, one can piece together various rotations and translations for parts or all of the chain while still satisfying the EL equations. This infinity of extrema renders the solution of Eq. 9 by direct numerical integration very difficult. For these reasons we apply a method based on analytic geometry to obtain numerical solutions, described in more detail below.

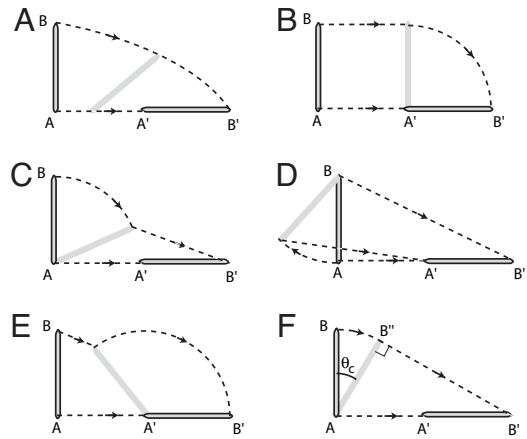
There is also an infinite degeneracy of solutions having the minimal distance in the above example. To see a second minimal transformation, imagine running the above solution backward in time, so the kink propagates from  $s = L$  to  $s = 0$  along  $\mathbf{r}_B$ . However, this solution should hold forward in time for the original problem if we permute  $\mathbf{r}_B$  and  $\mathbf{r}_A$ . Now, intermediate states  $\mathbf{r}^*$  first run along  $\hat{\mathbf{x}}$ , then  $\hat{\mathbf{y}}$ . But then we can introduce multiple right-angle kinks in various places, without causing the trajectories in the transformation to deviate from straight lines, so that intermediate states look like staircases. Because there are an infinite number of possible staircases in the continuum limit, there is an infinite degeneracy. This can lead to a tangent vector  $\mathbf{r}'$  whose magnitude is length-scale dependent, and less than unity until  $s \rightarrow 0$ . For example, an intermediate configuration can be drawn in Fig. 2 that appears as a straight diagonal line from  $\mathbf{r}^*(0, t)$  to  $\mathbf{r}^*(L, t)$ , until  $s \rightarrow 0$  when an infinite number of step discontinuities are revealed. This problem is resolved in practice through finite-size effects involving different critical angles of rotation described below. In the continuum limit, it is resolved by introducing curvature constraints.

**Curvature constraints.** In applications to polymer physics, chains have a stiffness characterized by bending potential in the analysis that is proportional to the square of the local curvature. Here, we may choose to characterize stiffness by introducing a constraint on the configurations of the space curve, so that the curvature simply cannot exceed a given number:

$$V_\kappa(\mathbf{r}'') = A\Theta(|\mathbf{r}''| < \kappa_C) \quad (A > 1). \quad [16]$$

This term lifts the infinite degeneracy mentioned above, as each near-kink (with curvature  $\kappa$  now  $= \kappa_C$ ) would result in slight deviations from linear motion in the above example, and thus an additional cost in the effective action. Other functional forms for  $V_\kappa$  are also possible. For some applications, a more conventional stiffness potential of the form  $V_\kappa(\mathbf{r}'') = 1/2A_\kappa \mathbf{r}''^2$  may be more appropriate. However, the action would no longer consist of a true distance functional, and its minimization would involve the detailed interplay of the parameter  $A_\kappa$  favoring globally minimal curvature with other factors affecting distance in the problem.

**Discrete chains.** Strings with a finite number of elements (chains) provide a more accurate representation of real-world systems such as biopolymers. Discretization is also essential for numerical solutions in these more realistic cases. Monomers on a discretized chain



**Fig. 3.** Transformations between two rigid rods. *A* undergoes simultaneous translation and rotation and so is not extremal. *B* is extremal and minimal. The rod cannot rotate any less given that it translates first. However, this transformation is a weak or local minimum. *C–E* are extremal in bulk but not minimal because they violate corner conditions (A. Mohazab and S.S.P., unpublished data). *F* is the global minimum. It rotates the minimal amount, and both *A* and *B* move monotonically toward *A'*, *B'*. A purely straight-line transformation exists but involves moving point *A* away from *A'* before moving toward it (similar to *D*), thus covering a larger distance than the minimal transformation.

travel along a curved metric (3), and Lagrange multipliers explicitly account for this fact here.

We start by discretizing the string into a chain of  $N$  links each with length  $ds = L/N$ , so that Eq. 4 becomes  $(ds) \int dt \sum_{i=1}^{N+1} \sqrt{\dot{\mathbf{r}}_i^2}$ , with each  $\mathbf{r}_i(t)$  a function of  $t$  only. The total distance is the accumulated distance of all the points joining the links, plus that of the end points, all times  $ds$ . This approach is essentially the method of lines for solving Eq. 10: the PDE becomes a set of  $N + 1$  coupled ODEs.

Eq. 5 becomes  $N$  constraint equations added to the Lagrangian:  $\sum_{i=1}^N \lambda_{i,i+1} \sqrt{(\mathbf{r}_{i+1} - \mathbf{r}_i)^2}$ . We rewrite this strictly for convenience as  $\sum (1/2) \lambda_{i,i+1} \mathbf{r}_{i+1/i}^2$ , where  $\mathbf{r}_{i+1/i} \equiv \mathbf{r}_{i+1} - \mathbf{r}_i$ , and  $|\mathbf{r}_{i+1/i}| = L/N$ .

The PDE in Eq. 10 then becomes  $N + 1$  coupled (vector) ODEs, each of the form

$$\dot{\hat{\mathbf{v}}}_i + \lambda_{i-1,i} \mathbf{r}_{i/i-1} - \lambda_{i,i+1} \mathbf{r}_{i+1/i} = 0, \quad [17]$$

with  $\lambda_{0,1} = \lambda_{N+1,N+2} = 0$ . Eq. 17 is consistent with Eq. 10 after suitable definitions, for example the curvature at point  $i$  after discretization is given by  $(\mathbf{r}_{i+1/i} - \mathbf{r}_{i/i-1})/ds^2$ .

**One link.** We turn to the simplest problem of one link with end points  $A$  and  $B$  (see Fig. 3), for which the action reads  $L \int_0^T dt (\sqrt{\dot{\mathbf{r}}_A^2} + \sqrt{\dot{\mathbf{r}}_B^2} - (\lambda(t)/2) \mathbf{r}_{B/A}^2)$ . Points  $A$  and  $B$  have boundary conditions  $\mathbf{r}_A(0) = \mathbf{A}$ ,  $\mathbf{r}_B(0) = \mathbf{B}$ ,  $\mathbf{r}_A(T) = \mathbf{A}'$ ,  $\mathbf{r}_B(T) = \mathbf{B}'$ . The link in our problem is taken to have a direction, so point  $A$  cannot transform to point  $B'$ . The EL equations become:

$$\begin{aligned} \dot{\hat{\mathbf{v}}}_A - \lambda \mathbf{r}_{B/A} &= 0 & \text{or} & & \lambda \mathbf{v}_A \cdot \mathbf{r}_{B/A} &= 0 \\ \dot{\hat{\mathbf{v}}}_B + \lambda \mathbf{r}_{B/A} &= 0 & & & \lambda \mathbf{v}_B \cdot \mathbf{r}_{B/A} &= 0, \end{aligned} \quad [18]$$

where the orthogonality of  $\mathbf{v}$  and  $\dot{\hat{\mathbf{v}}}$  has been used.

Reminiscent of Eq. 12, Eqs. 18 each have three solutions. For point  $A$ , these are: (i)  $\mathbf{v}_A \cdot \mathbf{r}_{B/A} = 0$ , or pure rotation of  $A$  about  $B$ , (ii)  $\mathbf{v}_A = 0$  or point  $A$  is stationary, or (iii)  $\lambda = 0$  and thus  $\dot{\hat{\mathbf{v}}}_A = 0$  from the EL equations, indicating straight-line motion. Moreover,  $i$  implies  $\mathbf{v}_B = 0$ , or both points rotate about a common center,  $ii$  implies  $\mathbf{v}_B \cdot \mathbf{r}_{B/A} = 0$  or  $B$  rotates, and  $iii$  implies  $\dot{\hat{\mathbf{v}}}_B = 0$  as well, so that both points move in straight lines. An extremal transformation thus involves either straight line motion, or rotations of one point about the other at rest (or common center), see Fig. 3 *B–F*.

The Lagrange multiplier may be found from the dot product of the EL equation for  $B$  with  $\mathbf{r}_{B/A}$ , which gives  $-ds^2\lambda = \mathbf{r}_{B/A} \cdot \hat{\mathbf{v}}_B$ . Thus, when  $B$  moves in a straight line  $\lambda = 0$ . When  $B$  rotates about  $A$ , its acceleration  $\mathbf{a}_B$  follows from rigid body kinematics as  $\mathbf{a}_A + \boldsymbol{\alpha} \times \mathbf{r}_{B/A} - \omega^2 \mathbf{r}_{B/A}$ , where  $\omega$  and  $\boldsymbol{\alpha}$  are the angular velocity and acceleration, respectively, and  $\mathbf{a}_A = 0$ . Thus  $\lambda = 1/L$ .

The minimal solution is the one that involves the minimal amount of rotation (and monotonic approach to  $A'B'$ ). This may be obtained from analytic geometry: for the example configurations in Fig. 3*F*, point  $B$  rotates about point  $A$  until  $B''$ , where the straight line  $B''B'$  is tangent to the circle of radius  $ds = L$  about  $A$ . The distance (over  $ds$ ) is  $AA' + L\theta_c + B''B'$ , where  $\sin \theta_c = L/(L+AA')$  and  $B''B' = \sqrt{(AA')^2 + 2L(AA')}$ , so, for example, if  $AA' = 2L$ ,  $\mathcal{D} \approx 5.168 L^2$ .

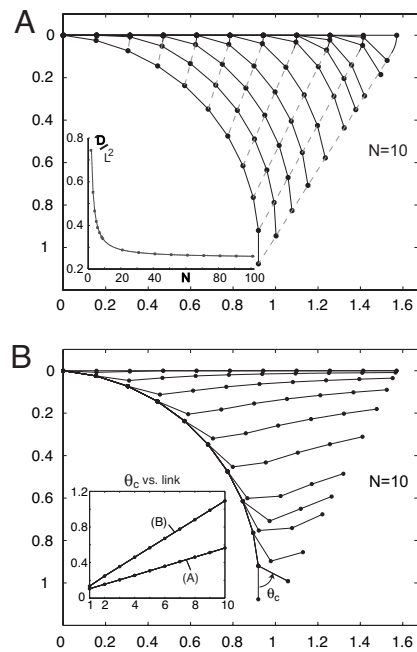
**Chains with curvature.** We can now investigate the transformation shown in Fig. 1*B* with the above methods. This is the canonical example when at least one of the space curves has nonzero curvature  $\boldsymbol{\kappa}$ . Let  $\mathbf{r}_A = R \sin(\pi s/2L)\hat{\mathbf{x}} + R \cos(\pi s/2L)\hat{\mathbf{y}}$  and  $\mathbf{r}_B = s\hat{\mathbf{x}} + R\hat{\mathbf{y}}$ , with  $0 \leq s \leq L$  and  $R = 2L/\pi$ . We then discretize the chain into  $N$  segments. According to Eq. 17, the end point velocities  $\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_{i+1}$  obey EL equations of the same form as Eqs. 18, and thus either rotate or translate. The situation for these links is analogous to Fig. 3*B* and *F*, in that the angle the link must rotate depends on the order of translation and rotation. The geometry in Fig. 1*B* is analogous to transformations  $A'B' \rightarrow AB$  in Fig. 3*B* and *F*, in that the critical angle  $\theta_c$  the link must rotate is smaller if translation occurs first.

Fig. 4 shows the two minimal solutions thus obtained. The transformation in Fig. 4*A* undergoes translation away from curve  $\mathbf{r}_A$ , and rotation at  $\mathbf{r}_B$ . It is the global minimum. The transformation in Fig. 4*B* rotates from  $\mathbf{r}_A$  through a larger critical angle (see Fig. 4*B* *Inset*), and then translates to  $\mathbf{r}_B$ . Both solutions have a soliton-like kink that propagates across either space-curve  $\mathbf{r}_B$  or  $\mathbf{r}_A$ .

The minimal transformation follows these steps: (i) Link  $r_{2/1}$  rotates about  $\mathbf{r}_1, \mathbf{v}_1 = 0, \mathbf{v}_2 \cdot \mathbf{r}_{2/1} = 0$ , and the Lagrange multiplier representing the conjugate “force”  $\lambda_{12} \neq 0$ . During this rotation, nodes 3, 4, ... move in straight lines formed by their initial values  $\mathbf{r}_{A3}, \mathbf{r}_{A4}, \dots$  and the tangent points to circles of radius  $ds$  centered at  $\mathbf{r}_{B2}, \mathbf{r}_{B3}, \dots$ . The corresponding Lagrange constraint forces  $\lambda_{23}, \lambda_{34}, \dots$  are all zero. Links  $r_{3/2}, r_{4/3}, \dots$  all adjust their orientation to ensure straight-line motion of their end points (dashed lines in Fig. 4*A*), except for  $r_2$ , which follows a curved path. (ii) When link  $r_{2/1}$  completes its rotation, it coincides with curve  $\mathbf{r}_B$ , and the process starts again with link  $r_{3/2}$ , which begins its rotation about  $\mathbf{r}_2$ , whereas nodes 4, 5, ... move in straight lines. This process continues until the final link  $r_{N+1/N}$  rotates into place on  $\mathbf{r}_B$ . The transformation in Fig. 4*B* is essentially the time-reverse of the above, but starting at curve  $\mathbf{r}_B$  and ending on  $\mathbf{r}_A$ .

For ideal chains without curvature constraints, the distances obtained from the two transformations in Fig. 4*A* and *B* differ nonextensively as the number of links  $N \rightarrow \infty$ . Moreover, the distance for each transformation itself differs nonextensively from the mean root square distance  $MRSD = N^{-1} \sum_{i=1}^N \sqrt{(\mathbf{r}_{Ai} - \mathbf{r}_{Bi})^2}$  as  $N \rightarrow \infty$ .<sup>11</sup> Specifically, the distance travelled by straight line motion scales as  $ds NL \sim L^2$ , whereas the distance travelled by rotational motion scales as  $ds (N\theta_c ds) \sim L^2/N$ .

On the other hand, curvature constraints as in Eq. 16 become more severe on consecutive links as  $N \rightarrow \infty$ , and can yield extensive corrections to the distance. Specifically, the increase in distance  $\Delta\mathcal{D}$  due to curvature constraints scales like the radius of curvature  $R$  times  $N$ , because every node is affected by the rounded kink as it propagates. So  $\Delta\mathcal{D} \sim ds NR \sim LR$ . The importance of this effect then depends on how  $R$  compares to  $L$  (the ratio of the persistence



**Fig. 4.** Two minimal transformations between the curves shown in Fig. 1*B*, for  $N = 10$  links. (A) The global minimal transformation  $r^*(s, t)$ , with  $\mathcal{D}^* \approx 0.330 L^2$ . (B) A local minimum with  $\mathcal{D} \approx 0.335 L^2$ . In A, links with one end touching curve  $\mathbf{r}_B$  rotate; the others translate first from  $\mathbf{r}_A$ , rotating only when one end of a link has touched  $\mathbf{r}_B$ . In B, they rotate first from  $\mathbf{r}_A$  and then translate into  $\mathbf{r}_B$ . Dashed lines in A show the paths travelled for each bead. (A *Inset*) The total distance travelled as a function of the number of links  $N$ , with various  $N$  plotted as filled circles to indicate the rapid decrease and asymptotic limit to  $\mathcal{D}_\infty \approx 0.251 L^2$ . (B *Inset*) The minimal angle each link must rotate during the transformation; it is less for the transformation in A. Animations of these transformations are provided as [supporting information \(SI\) Movies 1–4](#).

length to the total length). It does not vanish as  $N \rightarrow \infty$ . Non-crossing constraints described below also yield extensive corrections to the distance travelled.

**Noncrossing Space Curves.** The minimal transformation may be qualitatively different when chain crossing is explicitly disallowed. Fig. 1*C* illustrates a pair of curves that differ only by the order of chain crossing. They are displaced in Fig. 1*C* for easier visualization but should be imagined to overlap so the quantity  $\int_0^L |\mathbf{r}_A - \mathbf{r}_B| \approx 0$ , i.e., if they were ghost chains their distance would be nearly zero, and most existing metrics give zero distance between these curve pairs (see Table 1).

Analogous to the construction of Alexander polynomials for knots, if we form the orthogonal projection of these space curves onto a plane, there will be double points indicating one part of the curve crossing over or under another. If we trace the curve in an arbitrary but fixed direction, each double point occurs twice, once as underpass and once as an overpass. We may call the part of the curve between two consecutive passes a bridge. If the bridge ends in an overpass we assign it +1, if the bridge ends in an underpass we assign it -1, so traversing from the left in Fig. 1*C*, curve  $\mathbf{r}_B$  has (+1) sense, and curve  $\mathbf{r}_A$  (-1). For transformations obeying noncrossing, a bridge can undergo change in sense  $\pm 2$  to zero by moving from under or over the chain, whereas bridges in ghost chains undergo changes of sense by crossing from  $\pm 1$  to  $\mp 1$  directly.

The non-crossing condition means that the Lagrangian for the minimal transformation now depends on the position  $\mathbf{r}(s, t)$  of the space curve, which may be accounted for using an Edwards potential:  $V_{NC}([\mathbf{r}(s, t)]) = \int_0^L ds_1 \int_0^L ds_2 \delta(\mathbf{r}(s_1, t) - \mathbf{r}(s_2, t))$ . In practice, a Gaussian may be used to approximate the delta function, with

<sup>11</sup>The MRSD is always less than or equal to the rmsd between structures, as can be shown by applying Hölder’s inequality.

**Table 1. Values of the distance for various examples considered here, compared to other metrics**

Curve pair	$\mathcal{D}^*$ ( $L^2$ )	rmsd <sup>†</sup> ( $L$ )	(1-Q) <sup>‡</sup>	$\chi^{\S}$
Trivial translation	$ d /L$	$ d /L$	0	0
"L-curves", Fig. 1A	$1/\sqrt{2}$	$\sqrt{2}/3$	— <sup>¶</sup>	0
Straight line to hairpin	1/4	$1/\sqrt{6}$	1	1/2
"C-curve" - straight line, Fig. 4A	0.330	0.371	— <sup>¶</sup>	0.417
"C-curve" - straight line, Fig. 1A <sup>  </sup>	0.251	0.334	— <sup>¶</sup>	1
"Over/under" curves, Fig. 1C	$(\ell/L)^2$	$\approx 0$	0 <sup>††</sup>	0
Single link, Fig. 3F <sup>†††</sup>	5.168	$\sqrt{7}^{\S\S}$	— <sup>¶¶</sup>	— <sup>¶¶¶</sup>

<sup>†</sup>rmsd  $\equiv \sqrt{N^{-1} \sum_i (\mathbf{r}_{Ai} - \mathbf{r}_{Bi})^2}$ .

<sup>‡</sup>Fraction of shared contacts  $A$  has with  $B$ ; see refs. 7 and 8 for definitions.

<sup>§</sup>Structural overlap function equal to 1 minus the fraction of residue pairs with similar distances in structures  $A$  and  $B$ . The formula in ref. 9 is used.

<sup>¶</sup>0/0 or undefined.

<sup>||</sup>In the continuum limit.

<sup>††</sup>Assuming a contact is made at the junction.

<sup>†††</sup>For  $AA' = 2 \times$  link length.

<sup>§§</sup> $\mathcal{D} >$  rmsd here because rmsd contains a factor of 2, whereas  $\mathcal{D}$  did not. An "effective distance" for the rod could divide by 2.

<sup>¶¶</sup>Undefined for a single link.

a variance that may be adjusted to account for the thickness or volume of the chain.

The EL equation now becomes

$$(V_{NC})_{\mathbf{r}} = (\mathcal{L}_{\mathbf{r}})_{\mathbf{s}} + (\mathcal{L}_{\dot{\mathbf{r}}})_{\mathbf{t}} - [(V_{\kappa})_{\mathbf{r}}]_{\mathbf{ss}}, \quad [19]$$

where the curvature potential in Eq. 16 has been included, and the notation  $(\mathcal{L}_{\mathbf{r}})_{\mathbf{s}} \equiv (d/ds)(\partial\mathcal{L}/\partial\mathbf{r}')$  has been used. Eq. 10 is now modified to

$$\hat{\mathbf{v}}_{\mathbf{t}} = (\lambda\hat{\mathbf{t}})_{\mathbf{s}} + \nabla V_{NC} + [(V_{\kappa})_{\mathbf{r}}]_{\mathbf{ss}}. \quad [20]$$

To access various conformations, the minimal transformation must now abide by the nontrivial geometrical constraints that are induced by non-crossing. In general, this renders the problem difficult; however, the example in Fig. 1C is simple enough to propose a mechanism for the minimal transformation consistent with the developments above, without explicitly solving the EL equations in this case. In analogy with the hairpin transformation described below Eq. 15, the transformation here involves essentially forming and then unforming a hairpin.  $\mathbf{r}_A(N)$  (the blue end of curve  $\mathbf{r}_A$  in Fig. 1C) propagates back along its own length until it reaches the junction, where it then rotates over it to become the overpass (this "rotation" takes essentially zero distance in the continuum limit). The curve then doubles back, following its path in reverse to its starting point. This transformation is fully consistent with the allowed extremal rotations and translations of the discretized chain. The distance in the continuum limit is  $D = \int_0^1 ds (2s) = \ell^2$ , where  $\ell$  is the length of the shorter arm extending from the junction in Fig. 1C.

## Discussion

The distance between finite objects of any dimension  $d$  is a variational problem, and may be calculated by minimizing a vector functional of  $d + 1$  independent variables. Here we formulated the

problem for space curves, where the function  $\mathbf{r}^*(s, t)$  defining the transformation from curve  $\mathbf{r}_A$  to curve  $\mathbf{r}_B$  gives the minimal distance  $\mathcal{D}$ .

We provided a general recipe for the solution to the problem through the calculus of variations. For simple cases, the solution is analytically tractable. Direct numerical methods are difficult due to multiple extrema. We employed a method that interpreted the discretized EL equations geometrically to obtain minimal solutions. The various solutions obtained here are summarized in Table 1 and compared with other similarity measures currently used.

The distance metric may be generalized to higher dimensional manifolds; for example, a two-dimensional surface needs three independent parameters to describe the transformation. The distance becomes  $\mathcal{D} = \int du \int dv \int dt |\dot{\mathbf{r}}|$  and the constant unit area condition becomes  $|\partial\mathbf{r}/\partial u \times \partial\mathbf{r}/\partial v| = 1$ .

The question of a distance metric between configurations of a biopolymer has occupied the minds of many in the protein folding community for some time (c.f., for example, refs. 4–8). Such a metric is of interest for comparison between folded structures, as well as to quantify how close an unfolded or partly folded structure is to the native. Chan and Dill (5) investigated the minimum number of moves necessary to transform one lattice structure to another, in particular while breaking the smallest number of hydrogen bonds. Leopold *et al.* (4) investigated the minimum number of monomers that had to be moved to transform one compact conformation to another. Falicov and Cohen (6) investigated structural comparison by rotation and translation until the minimal area surface by triangulation was obtained between two protein structures.

The present theoretical framework allows computation of a minimal distance between proteins of the same length by rotating and translating until  $\mathcal{D}$  is minimized, as done in the calculation of rmsd. Comparison between different length proteins would involve the further optimization with respect to insertions or deletions.

It is interesting to ask which folded structures have the largest, or smallest, average distance  $\langle\mathcal{D}\rangle$  from an ensemble of random coil structures, and also whether the accessibility of these structures in terms of  $\mathcal{D}$  translates to their folding rates. It can also be determined whether the distance to a structure correlates with kinetic proximity in terms of its probability  $p_F$  to fold before unfolding (7), by calculating  $\langle\mathcal{D}p_F\rangle$ . The question of the most accessible or least accessible structure may be formulated variationally as a free-boundary or variable end-point problem.

It is an important future question to address whether the entropy of paths to a particular structure is as important as the minimal distance. In this sense, it may be the finite "temperature" ( $\beta < \infty$ ) partition function  $Z(\beta) = \int d[\mathbf{r}(s, t)] \exp(-\beta\mathcal{D}[\mathbf{r}(s, t)])$ , i.e., the sum over paths weighted by their "actions," which is the most important quantity in determining the accessibility between structures. This has an analogue to the quantum string: we investigated only  $Z(\infty)$  here. We hope that this work proves useful in laying the foundations for unambiguously defining distance between biomolecular structures in particular and high-dimensional objects in general.

We are grateful to Ali Mohazab, Moshe Schecter, Matt Choptuik, and Bill Unruh for insightful discussions. This work was supported by the Natural Sciences and Engineering Research Council and the A. P. Sloan Foundation.

- Gelfand IM, Fomin SV (2000) *Calculus of Variations* (Dover, New York).
- Zwiebach B (2004) *A First Course in String Theory* (Cambridge Univ Press, New York).
- Grosberg AY (2004) in *Computational Soft Matter: From Synthetic Polymers to Proteins*, eds Attig N, Binder K, Grubmüller H, Kremer K (John von Neumann Institut für Computing, Bonn), NIC series, Vol 23, pp 375–399.
- Leopold PE, Montal M, Onuchic JN (1992) *Proc Natl Acad Sci USA* 89:8721–8725.

- Chan HS, Dill KA (1994) *J Chem Phys* 100:9238–9257.
- Falicov A, Cohen FE (1996) *J Mol Biol* 258:871–892.
- Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich ES (1998) *J Chem Phys* 108:334–350.
- Cho SS, Levy Y, Wolynes PG (2006) *Proc Natl Acad Sci USA* 103:586–591.
- Veitshans T, Klimov D, Thirumalai D (1996) *Folding Des* 2:1–22.