Geometry of thermodynamic control

Patrick R. Zulkowski, David A. Sivak, Gavin E. Crooks, and Michael R. DeWeese

The geometry of thermodynamic control

Patrick R. Zulkowski*

Department of Physics, University of California, Berkeley, CA 94720 and
Redwood Center for Theoretical Neuroscience, University of California, Berkeley CA 94720

David A. Sivak[†] and Gavin E. Crooks[‡]

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Michael R. DeWeese[§]

Department of Physics, University of California, Berkeley, CA 94720
Redwood Center for Theoretical Neuroscience, University of California, Berkeley CA 94720 and
Helen Wills Neuroscience Institute, University of California, Berkeley CA 94720

A deeper understanding of nonequilibrium phenomena is needed to reveal the principles governing natural and synthetic molecular machines. Recent work has shown that when a thermodynamic system is driven from equilibrium then, in the linear response regime, the space of controllable parameters has a Riemannian geometry induced by a generalized friction tensor. We exploit this geometric insight to construct closed-form expressions for minimal-dissipation protocols for a particle diffusing in a one dimensional harmonic potential, where the spring constant, inverse temperature, and trap location are adjusted simultaneously. These optimal protocols are geodesics on the Riemannian manifold, and reveal that this simple model has a surprisingly rich geometry. We test these optimal protocols via a numerical implementation of the Fokker-Planck equation and demonstrate that the friction tensor arises naturally from a first order expansion in temporal derivatives of the control parameters, without appealing directly to linear response theory.

I. INTRODUCTION

There has been considerable progress in the study of nonequilibrium processes in recent years. For example, fluctuation theorems relating the probability of an increase to that of a comparable decrease in entropy during a finite time period have been derived [1–5] and experimentally verified [6–9] in a variety of contexts. Moreover, other new fundamental relationships between thermodynamic quantities that remain valid even for systems driven far from equilibrium, such as the Jarzynski equality [10–13], have also been established. Interestingly, some of these ideas were independently developed in parallel within the machine learning community [14], as ideas from nonequilibrium statistical mechanics are increasingly finding applications to learning and inference problems [15, 16].

For macroscopic systems, the properties of optimal driving processes have been investigated using thermodynamic length, a natural measure of the distance between pairs of equilibrium thermodynamics states [17–22], with extensions to microscopic systems involving a metric of Fisher information [23, 24]. Recently, a linear-response framework has been proposed for protocols that minimize the dissipation during nonequilibrium perturbations of microscopic systems. In the resulting geometric formulation, a generalized friction tensor induces a Riemannian manifold structure on the space of parameters, and optimal protocols trace out geodesics of this friction tensor [25].

In this article, we make use of Riemannian geometry theorems to simplify the problem of optimizing protocols. To illustrate the power of these geometric ideas, we consider a simple, but previously unsolved, stochastic system and calculate closed-form expressions for optimal protocols. We test the accuracy of our approximation by numerically comparing our optimal protocols against naive protocols using the Fokker-Planck equation. We conclude by demonstrating that our inverse diffusion tensor framework arises naturally from a first order expansion in temporal derivatives of the control parameters, without appealing directly to linear response theory.

II. DERIVATION OF THE EXCESS POWER FOR VARIABLE TEMPERATURE

For a physical system at equilibrium in contact with a thermal bath, the probability distribution over microstates x is given by the canonical ensemble

$$\pi(x|\lambda) \equiv \exp \beta \left[F(\boldsymbol{\lambda}) - E(x, \boldsymbol{\lambda}) \right] , \qquad (1)$$

where $\beta = (k_{\mathrm{B}} T)^{-1}$ is the inverse temperature in natural units, $F(\boldsymbol{\lambda})$ is the free energy, and $E(x, \boldsymbol{\lambda})$ is the system energy as a function of the microstate x and a collection of experimentally controllable parameters $\boldsymbol{\lambda}$.

* pzulkowski@berkeley.edu
† dasivak@lbl.gov; Present address: Center for Systems and Synthetic Biology, University of California, San Francisco, CA 94158; david.sivak@ucsf.edu
‡ gcrooks@lbl.gov
§ deweese@berkeley.edu

In equilibrium, the thermodynamic state of the system (the probability distribution over microstates) is completely specified by values of the control parameters, but out of equilibrium the system's probability distribution over microstates fundamentally depends on the history of the control parameters $\boldsymbol{\lambda}$, which we denote by the control parameter protocol $\boldsymbol{\Lambda}$. We assume the protocol to be sufficiently smooth to be twice-differentiable.

The usual expressions for heat and work [26–29] assume that the temperature of the heat bath is held constant over the course of the nonequilibrium protocol. Following the development of methods to handle time-varying temperature described in section 1.5 of [30], and preceding Eq. (4) of [31], we argue that the unitless energy $\beta E(x, \boldsymbol{\lambda})$ (normalized by the natural scale of equilibrium thermal fluctuations, $k_{\mathrm{B}}T = \beta^{-1}$, set by equipartition) is the fundamental thermodynamic quantity. Thus when generalizing to a variable heat bath temperature, we arrive at the following definition for the average instantaneous rate of (unitless) energy flow into the system:

$$\left\langle \frac{d}{dt}\left(\beta E(x, \boldsymbol{\lambda})\right) \right\rangle_{\boldsymbol{\Lambda}} , \tag{2}$$

where angled brackets with subscript indicate a nonequilibrium average dependent on the protocol $\boldsymbol{\Lambda}$. For constant β, this reduces to the standard thermodynamic definition [25]. With this definition, we can prove that for systems obeying Fokker-Planck dynamics, excess work is guaranteed to be non-negative for any path, which is not true of the naive definition (see § III). Nonetheless, a deeper understanding of the subtleties involved in our modified energy flow definition (Eq. (2)) calls out for further study.

Eq. (2) may be written as

$$\left\langle \frac{dx^T}{dt} \cdot \frac{\partial \left(\beta E\right)}{\partial x}(x, \boldsymbol{\lambda}) \right\rangle_{\boldsymbol{\Lambda}} + \left\langle \frac{d\boldsymbol{\lambda}^T}{dt} \cdot \frac{\partial \left(\beta E\right)}{\partial \boldsymbol{\lambda}}(x, \boldsymbol{\lambda}) \right\rangle_{\boldsymbol{\Lambda}} . \tag{3}$$

The first term represents energy flux due to fluctuations of the system at constant parameter values and naturally defines heat flux for nonequilibrium systems. The second term, associated with an energy flux due to changes of the external parameters, defines nonequilibrium average power in the general setting of time-variable bath temperature.

The average excess power exerted by the external agent on the system, over and above the average power on a system at equilibrium, is

$$\beta(t_0)\mathcal{P}_{\mathrm{ex}}(t_0) \equiv -\left[\frac{d\boldsymbol{\lambda}^T}{dt}\right]_{t_0} \cdot \langle \triangle \boldsymbol{X} \rangle_{\boldsymbol{\Lambda}} . \tag{4}$$

Here $\boldsymbol{X} \equiv -\frac{\partial(\beta E)}{\partial \boldsymbol{\lambda}}$ are the forces conjugate to the control parameters $\boldsymbol{\lambda}$, and $\triangle \boldsymbol{X}(t_0) \equiv \boldsymbol{X}(t_0) - \langle \boldsymbol{X} \rangle_{\boldsymbol{\lambda}(t_0)}$ is the deviation of $\mathbf{X}(t_0)$ from its current equilibrium value.

Applying linear response theory [32],

$$\langle \triangle \boldsymbol{X}(t_0) \rangle_{\boldsymbol{\Lambda}} \approx \int_{-\infty}^{t_0} \boldsymbol{\chi}(t_0 - t') \cdot [\boldsymbol{\lambda}(t_0) - \boldsymbol{\lambda}(t')] \, dt' \tag{5}$$

where $\chi_{ij}(t) \equiv -d\Sigma_{ij}^{\boldsymbol{\lambda}(t_0)}(t)/dt$ represents the response of conjugate force X_i at time t to a perturbation in control parameter λ^j at time zero, and

$$\Sigma_{ij}^{\boldsymbol{\lambda}(t_0)}(t) \equiv \langle \delta X_j(0)\delta X_i(t) \rangle_{\boldsymbol{\lambda}(t_0)} . \tag{6}$$

For protocols that vary sufficiently slowly [25], the resulting mean excess power is

$$\beta(t_0)\mathcal{P}_{\mathrm{ex}}(t_0) \approx \left[\frac{d\boldsymbol{\lambda}^T}{dt}\right]_{t_0} \cdot g(\boldsymbol{\lambda}(t_0)) \cdot \left[\frac{d\boldsymbol{\lambda}}{dt}\right]_{t_0} , \tag{7}$$

for inverse diffusion tensor

$$g_{ij} \equiv \beta(t_0)\zeta_{ij}(\boldsymbol{\lambda}(t_0)) = \int_0^\infty dt' \langle \delta X_j(0)\delta X_i(t') \rangle_{\boldsymbol{\lambda}(t_0)} , \tag{8}$$

where ζ_{ij} is the friction tensor in control parameter space from [25]. We will construct geodesics using this inverse diffusion tensor g_{ij}.

III. THE MODEL SYSTEM AND ITS INVERSE DIFFUSION TENSOR

We consider a particle (initially at equilibrium) in a one-dimensional harmonic potential diffusing under inertial Langevin dynamics, with equation of motion

$$m\ddot{y} + k(t)\left(y - y_0(t)\right) + \zeta^c \dot{y} = F(t) , \tag{9}$$

for Gaussian white noise $F(t)$ satisfying

$$\langle F(t) \rangle = 0 , \quad \langle F(t)F(t') \rangle = \frac{2\zeta^c}{\beta(t)}\delta(t - t') . \tag{10}$$

Here ζ^c is the Cartesian friction coefficient. We take as our three control parameters: the inverse temperature of the bath β, the location of the harmonic potential minimum y_0, and the stiffness of the trap k [see Fig. 1(a)]. The conjugate forces are

$$\boldsymbol{X} = \left(\beta k\left(y - y_0\right), -\frac{p^2}{2m} - \frac{k}{2}\left(y - y_0\right)^2, -\frac{\beta}{2}\left(y - y_0\right)^2\right). \tag{11}$$

This model can be experimentally realized as, for instance, a driven torsion pendulum [33, 34].

The excess work

$$\langle (\beta W)_{\mathrm{ex}} \rangle \equiv \int_{t_a}^{t_b} dt \, \beta(t)\mathcal{P}_{\mathrm{ex}}(t) \tag{12}$$

is non-negative. Assuming the system begins in equilibrium, the relative entropy $D\big[f \parallel \pi(x|\lambda(t))\big]$ corresponds to the available energy in the system due to being out of
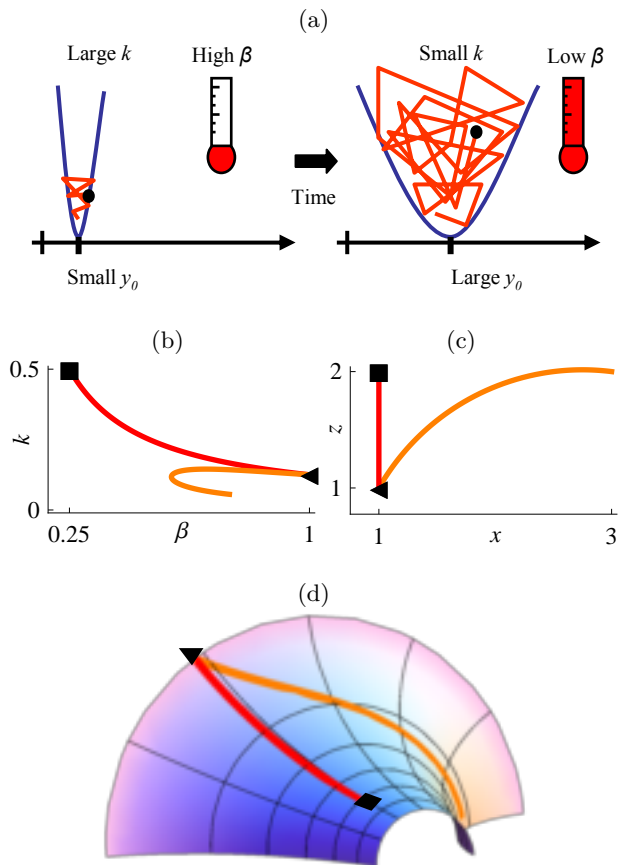
FIG. 1. **(a)** Our model system. A particle (black dot) diffusing in a harmonic potential with adjustable spring constant k, position y_0, and inverse temperature $\beta = \frac{1}{k_B T}$ (indicated by thermometers). **(b)** Representative optimal protocols (orange and red curves) plotted for two of the three control parameters, k and β. An optimal protocol (*e.g.*, red curve) results in the minimum dissipation for any path taking the system from one particular state (black square) to another (black triangle) in a fixed amount of time. **(c)** A change of variables $\{\beta, k\} \rightarrow \{z, x\}$ (Eq. (34)) reveals that our model system has an underlying structure described by hyperbolic geometry, represented here as the Poincaré half-plane, in which geodesics form half circles (orange curve) or vertical lines (red line). **(d)** A piece of the (z, x) manifold may be isometrically embedded as a saddle in $\mathbb{R}^3$. The distortions in each of these two optimal paths as shown in panels (b) and (c) reflect the curvature of this manifold.

equilibrium [35], and bounds the excess work from below. Here, π is the equilibrium distribution (Eq. (1)) defined by parameters $\boldsymbol{\lambda}(t)$, and $f \equiv f(y, p, t)$ is the nonequilibrium probability distribution. The time derivative of the relative entropy may be written as

$$\frac{d}{dt} D\big[f \,||\, \pi(x|\lambda(t))\big] = \int \frac{\partial f}{\partial t} \log\left(f/\pi(x|\lambda(t))\right) + \beta(t)\mathcal{P}_{ex}(t), \tag{13}$$

which follows from the identity

$$\int f\left[\frac{\partial}{\partial t} \log\left(\pi(x|\lambda(t))\right)\right] = \frac{d\boldsymbol{\lambda}^T}{dt} \cdot \langle\triangle \mathbf{X}(t)\rangle_{\boldsymbol{\Lambda}}. \tag{14}$$

The first term of Eq. (13) simplifies to

$$-\frac{\zeta^c}{\beta} \int e^{-\frac{\beta p^2}{m}} \left[\frac{1}{f}\left(\frac{\partial}{\partial p}\left(e^{\frac{\beta p^2}{2m}} f\right)\right)^2\right] \leq 0. \tag{15}$$

Integrating Eq. (13) from 0 to t_0 proves the relative entropy bounds the excess work from below. Since this quantity is always non-negative, so is the excess work; in fact, for any finite-duration path visiting more than one point in parameter space, it is strictly positive, yielding a well-behaved metric in our geometrical formalism. See [36] for related results in the special case of constant temperature. Note that, unlike our modified definition for work, the naive definition $\int_{t_a}^{t_b} dt\, \mathcal{P}_{\text{ex}}(t)$ may be negative for particular protocols that vary β.

Calculation of the time correlation functions in Eq. (8) requires knowledge of the dynamics for fixed control parameters. We may write any solution to the equation of motion as a sum $y_h + y_p$ of a homogeneous part y_h, which depends on the initial conditions and is independent of $F(t)$, and a particular part y_p, which has vanishing initial conditions but depends linearly on $F(t)$ (see, for instance, Theorem 3.7.1 in [37]). Explicitly, we may write

$$y_p(t) = \int_0^t \left(\frac{y_h^{(1)}(s)y_h^{(2)}(t) - y_h^{(1)}(t)y_h^{(2)}(s)}{y_h^{(1)}(s)\frac{d}{ds}y_h^{(2)}(s) - y_h^{(2)}(s)\frac{d}{ds}y_h^{(1)}(s)}\right)\frac{F(s)}{m}ds \tag{16}$$

where $y_h^{(i)}(t)$ for $i = 1, 2$ are independent solutions of the homogeneous equation. It follows immediately that

$$y_h(t) = C_1 y_h^{(1)}(t) + C_2 y_h^{(2)}(t) \tag{17}$$

where the constants C_1, C_2 are determined by initial conditions.

For Gaussian white noise $F(t)$, it is easy to show that the particular piece y_p does not contribute to the equilibrium time correlation function $\langle \delta X_j(0)\delta X_i(t)\rangle$. For simplicity and without loss of generality, consider the correlation function $\langle \delta y(t)^2 \delta y(0)^2\rangle$. Expanding this expression,

$$\langle \delta y(t)^2 \delta y(0)^2\rangle = \langle y(t)^2 y(0)^2\rangle - \langle y(t)^2\rangle\langle y(0)^2\rangle, \tag{18}$$

and substituting $y(t) = y_h(t) + y_p(t)$, we find

$$\langle \delta y(t)^2 \delta y(0)^2\rangle = \langle y_h(t)^2 y(0)^2\rangle + \langle y_p(t)^2 y(0)^2\rangle$$
$$- \langle y_h(t)^2\rangle\langle y(0)^2\rangle - \langle y_p(t)^2\rangle\langle y(0)^2\rangle \tag{19}$$
$$+ 2\Big(\langle y_h(t)y_p(t)y(0)^2\rangle - \langle y_h(t)y_p(t)\rangle\langle y(0)^2\rangle\Big).$$

Angled brackets denote an average over noise and initial conditions.

According to Eq. (16), the particular solution y_p does not depend on the initial conditions. It follows immediately that

$$\langle y_p(t)^2 y(0)^2 \rangle - \langle y_p(t)^2 \rangle \langle y(0)^2 \rangle = 0 \ . \quad (20)$$

Furthermore, since y_h depends only on the initial conditions and is independent of the noise,

$$\langle y_h(t) y_p(t) y(0)^2 \rangle - \langle y_h(t) y_p(t) \rangle \langle y(0)^2 \rangle = 0 \ , \quad (21)$$

follows from the assumption that $\langle F(t) \rangle = 0$. To summarize,

$$\langle \delta y(t)^2 \delta y(0)^2 \rangle = \langle \delta y_h(t)^2 \delta y(0)^2 \rangle \ . \quad (22)$$

For each of the time correlation functions needed to compute the inverse diffusion tensor, it is generally true that $y_h(t)$ may be substituted in the average for $y(t)$.

Without loss of generality, let us assume for the moment that $(\zeta^c)^2 - 4km > 0$. If we define

$$r_\pm = \frac{\zeta^c}{2m} \pm \frac{1}{2} \sqrt{\left(\frac{\zeta^c}{m} \right)^2 - \frac{4k}{m}} \ , \quad (23)$$

then the homogeneous solution with initial conditions $\{y(0), p(0) = m\dot{y}(0)\}$ is given by

$$
\begin{aligned}
y_h(t) = y_0 &+ \frac{p(0) + mr_- \left(y(0) - y_0 \right)}{m(r_- - r_+)} e^{-r_+ t} \\
&+ \frac{p(0) + mr_+ \left(y(0) - y_0 \right)}{m(r_+ - r_-)} e^{-r_- t} \ ,
\end{aligned}
\quad (24)
$$

where y_0 is the fixed trap position. For convenience, let us define $Y \equiv y - y_0$. Assuming that the initial conditions $\{y(0), p(0)\}$ are distributed according to the equilibrium Boltzmann distribution $\pi[y(0), p(0)] \propto e^{-\beta E[y(0), p(0)]}$ for $E[y, p] = \frac{p^2}{2m} + \frac{1}{2} kY^2$, we obtain the following identities:

$$\langle \delta Y_h^2(t) \delta Y^2(0) \rangle = \frac{2}{(k\beta)^2 (r_+ - r_-)^2} \left(r_- e^{-r_+ t} - r_+ e^{-r_- t} \right)^2 \quad (25a)$$

$$\langle \delta \dot{Y}_h^2(t) \delta Y^2(0) \rangle = \frac{2 r_+^2 r_-^2}{k^2 \beta^2 (r_+ - r_-)^2} \left(e^{-r_+ t} - e^{-r_- t} \right)^2 \quad (25b)$$

$$\langle \delta Y_h^2(t) \delta p^2(0) \rangle = \frac{2}{\beta^2 (r_+ - r_-)^2} \left(e^{-r_+ t} - e^{-r_- t} \right)^2 \quad (25c)$$

$$\langle \delta \dot{Y}_h^2(t) \delta p^2(0) \rangle = \frac{2}{\beta^2 (r_+ - r_-)^2} \left(r_- e^{-r_- t} - r_+ e^{-r_+ t} \right)^2 \ . \quad (25d)$$

Integrating these expressions, we obtain

$$\int_0^\infty dt \langle \delta Y_h^2(t) \delta Y^2(0) \rangle = \frac{m}{k^2 \beta^2 \zeta^c} \left(1 + \frac{(\zeta^c)^2}{km} \right) \quad (26a)$$

$$\int_0^\infty dt \langle \delta \dot{Y}_h^2(t) \delta Y^2(0) \rangle = \frac{1}{k\beta^2 \zeta^c} \quad (26b)$$

$$\int_0^\infty dt \langle \delta Y_h^2(t) \delta p^2(0) \rangle = \frac{m^2}{k\beta^2 \zeta^c} \quad (26c)$$

$$\int_0^\infty dt \langle \delta \dot{Y}_h^2(t) \delta p^2(0) \rangle = \frac{m}{\zeta^c \beta^2} \ . \quad (26d)$$

Thus the inverse diffusion tensor is

$$
g_{ij} = \frac{m}{4\zeta^c}
\begin{pmatrix}
\frac{4(\zeta^c)^2}{m} \beta & 0 & 0 \\
0 & \frac{1}{\beta^2} \left(4 + \frac{(\zeta^c)^2}{km} \right) & \frac{1}{\beta k} \left(2 + \frac{(\zeta^c)^2}{km} \right) \\
0 & \frac{1}{\beta k} \left(2 + \frac{(\zeta^c)^2}{km} \right) & \frac{1}{k^2} \left(1 + \frac{(\zeta^c)^2}{km} \right)
\end{pmatrix} ,
\quad (27)
$$

which endows the space $-\infty < y_0 < \infty, 0 < \beta < \infty, 0 < k < \infty$ with a Riemannian structure.

IV. BRIEF REVIEW OF RIEMANNIAN GEOMETRY.

We recall some definitions from Riemannian geometry and establish notation (see [38–40] for details). For a smooth Riemannian manifold M endowed with metric tensor g, the Christoffel symbols are defined as

$$\Gamma^j_{ik} \equiv \frac{1}{2} g^{jl} \left(\partial_i g_{kl} + \partial_k g_{il} - \partial_l g_{ik} \right) \quad (28)$$

where g^{ij} denotes the matrix inverse of the metric. We employ the Einstein summation convention here (and assume it throughout). The Riemann tensor, constructed from the Christoffel symbols, measures the curvature of the manifold M and is given in local coordinates by

$$R^i_{\ jkl} \equiv \partial_k \Gamma^i_{jl} - \partial_l \Gamma^i_{jk} + \Gamma^m_{jl} \Gamma^i_{mk} - \Gamma^m_{jk} \Gamma^i_{ml}. \quad (29)$$

Contracting indices gives the Ricci tensor R_{ij} and the Ricci scalar R,

$$R_{ij} = R^l_{\ ilj} \ , \ R = g^{ij} R_{ij} \ , \quad (30)$$

which are useful for determining the curvature content of the manifold M. Geodesics are defined in local coordinates by

$$\frac{d^2 \lambda^i}{d\tau^2} + \Gamma^i_{jk} \frac{d\lambda^j}{d\tau} \frac{d\lambda^k}{d\tau} = 0 \ . \quad (31)$$

V. OPTIMAL PROTOCOLS

Though one can write down the geodesic equations for the metric Eq. (27) in the (y_0, β, k) coordinate system, more insight is gained by finding a suitable change of

coordinates. Consider the lower right 2×2 block of the metric Eq. (27) which is the metric tensor for the two-dimensional (β, k) submanifold. A direct calculation of this submanifold's Ricci scalar yields $R = -2\zeta^c/m$ which is constant and always strictly negative.

Theorems from Riemannian geometry [38] imply that this constant negative-curvature submanifold is isometrically related to the hyperbolic plane. In our construction, we choose the Poincaré half-plane representation of the hyperbolic plane, which is described by $\{(z,x) \in \mathbb{R}^2, x > 0\}$ with metric tensor given by the line element $ds^2 = g_{ij}dx^i dx^j = \frac{dx^2 + dz^2}{x^2}$. The geodesics of the hyperbolic plane (see Fig. 1) are half-circles with centers on the z-axis and lines perpendicular to the z-axis. Fig. 1(c) shows two geodesics in (z,x) coordinates. The portion of the hyperbolic plane $\{(z,x) \in \mathbb{R}^2, x > 1, z \in [0,\pi]\}$ may be isometrically embedded in $\mathbb{R}^3$ using the map

$$\left(\frac{1}{x}\cos z, \log\left(\sqrt{x^2-1} + x \right) - \frac{\sqrt{x^2-1}}{x}, \frac{1}{x}\sin z \right).$$
(32)

The geodesics of Fig. 1 (c) and the part of the hyperbolic plane containing them are embedded in $\mathbb{R}^3$ in Fig. 1(d).

The line element associated with the submanifold metric tensor,

$$ds^2 = \frac{m}{4\zeta^c} \left[\frac{1}{\beta^2}\left(4 + \frac{(\zeta^c)^2}{km}\right)d\beta^2 + \right.$$
(33)
$$\left. \frac{2}{\beta k}\left(2 + \frac{(\zeta^c)^2}{km}\right)d\beta\,dk + \frac{1}{k^2}\left(1 + \frac{(\zeta^c)^2}{km}\right)dk^2 \right],$$

is coordinate-invariant since it measures geometric distances. Thus we may construct an explicit coordinate transformation,

$$x \equiv \frac{1}{2\beta\zeta^c}\sqrt{\frac{m}{k}} \;,\; z \equiv \frac{1}{4\beta k}\;,$$
(34)

to demonstrate the equivalence of the submanifold with a portion of the Poincaré plane. Note that x is proportional to the classical partition function of the system in equilibrium, and z is proportional to the equilibrium variance of $y - y_0$. Inverting Eq. (34), and substituting into Eq. (33) gives the metric tensor in (z,x)-coordinates,

$$ds^2 = \frac{m}{\zeta^c}\frac{dx^2 + dz^2}{x^2}\;.$$
(35)

The line element corresponding to the metric of the full three-dimensional manifold in Eq. (27) is

$$ds^2 = \frac{m}{\zeta^c}\frac{z\,dy_0^2 + dx^2 + dz^2}{x^2}$$
(36)

in (y_0, z, x) coordinates. To fully exploit the machinery of Riemannian geometry to find closed-form geodesics, we look for Killing fields of Eq. (36). In general [38–40], isometries of a metric are generated by the Killing vector fields K which are themselves characterized by the Killing equation

$$\nabla_i K_j + \nabla_j K_i = \partial_i K_j + \partial_j K_i - 2\Gamma_{ij}^k K_k = 0\;.$$
(37)

While directly solving this system of equations may be difficult, certain characterizations of Killing vectors help circumvent this difficulty. For instance, if in a given coordinate system the metric tensor components are independent of a coordinate x^i, then the coordinate vector ∂_{x^i} is a Killing field [40]. Hence, ∂_{y_0} is clearly a Killing vector field. Examining the full set of Killing equations shows that

$$K = y_0\partial_{y_0} + 2x\partial_x + 2z\partial_z$$
(38)

is also a Killing vector field. There may be more solutions to the Killing equation yet to be discovered.

In general [40], for Killing vector K_i the quantity $K_i\frac{d\lambda^i}{d\tau}$ is conserved along the geodesic described by $\boldsymbol{\lambda}$. This follows from

$$\frac{d}{d\tau}\left(K_i\frac{d\lambda^i}{d\tau} \right) = \nabla_i K_j\frac{d\lambda^i}{d\tau}\frac{d\lambda^j}{d\tau} + K_i\frac{D}{d\tau}\frac{d\lambda^i}{d\tau} = 0.$$
(39)

The first term of the equation vanishes by the definition of the Killing field and the second term vanishes by the geodesic equation Eq. (31). For the three-dimensional inverse diffusion tensor, we have the following two conserved quantities associated with Killing fields:

$$\frac{z(\tau)}{x^2(\tau)}\frac{dy_0}{d\tau}\;,\; \frac{2}{x(\tau)}\frac{dx}{d\tau} + \frac{z(\tau)}{x^2(\tau)}y_0(\tau)\frac{dy_0}{d\tau} + \frac{2z(\tau)}{x^2(\tau)}\frac{dz}{d\tau}.$$
(40)

To solve the geodesic equations, note that the velocity of the geodesic (*i.e.*, its tangent vector) must have constant norm [38–40]. For convenience, we choose the norm so that

$$1 = \frac{1}{x^2(\tau)}\left(\left(\frac{dz}{d\tau}\right)^2 + \left(\frac{dx}{d\tau}\right)^2 \right) + \frac{c_1^2}{r^2}\frac{x^2(\tau)}{z(\tau)}$$
(41)

where we have used the first conserved quantity of Eq. (40). We combine this with the full geodesic equation for $x(\tau)$, to decouple $x(\tau)$ from $y_0(\tau)$ and $z(\tau)$:

$$\frac{d^2x}{d\tau^2} - \frac{2}{x(\tau)}\left(\frac{dx}{d\tau}\right)^2 + x(\tau) = 0\;,$$
(42)

which has solution

$$x(\tau) = r\,\mathrm{sech}(\tau)\;.$$
(43)

When $z(\tau)$ is constant, the geodesic equation for z implies that y_0 is also constant, giving a geodesic straight line in the constant-z submanifold.

When $z(\tau)$ is not constant, Eqs. (41) and (43) imply

$$\frac{x^4(\tau)}{r^2} = \left(\frac{dz}{d\tau}\right)^2 + \frac{c_1^2}{r^2}\frac{x^4(\tau)}{z(\tau)}\;,$$
(44)

which integrates to

$$z(\tau) = h^{-1}(c_2 - r\tanh(\tau)),\qquad(45)$$

where

$$h(\xi) \equiv \xi\sqrt{1-\tfrac{c_1^2}{\xi}}+\tfrac{1}{2}c_1^2\log\left(2\xi\bigl(1+\sqrt{1-\tfrac{c_1^2}{\xi}}\bigr)-c_1^2\right).\ (46)$$

The Killing conserved quantities of Eq. (40), together with $x(\tau)$ and $z(\tau)$, yield

$$y_0(\tau) = E - c_1\log\Bigg[-c_1^2 + 2h^{-1}(c_2-r\tanh(\tau))\times$$
$$\left(1+\sqrt{1-\frac{c_1^2}{h^{-1}(c_2-r\tanh(\tau))}}\right)\Bigg].\quad(47)$$

Let $(y_{0,i},x_i,z_i)$ and $(y_{0,f},x_f,z_f)$ denote the endpoints of the geodesic. Define $\triangle\lambda \equiv \lambda_f - \lambda_i$ and $\bar\lambda \equiv \frac{\lambda_i+\lambda_f}{2}$ for $\lambda \in \{y_0,x,z\}$. Defining $\bar h \equiv \frac{h(z_f)+h(z_i)}{2}$ and $\triangle h \equiv h(z_f)-h(z_i)$, the constant c_2 may be written as

$$c_2 = \bar h + \bar x\frac{\triangle x}{\triangle h}\qquad(48)$$

and r is given by

$$r^2 = x_i^2 + \frac{1}{4}\left(\triangle h + 2\frac{\triangle x}{\triangle h}\bar x\right)^2.\qquad(49)$$

The constant E is given by

$$E = y_{0,i} + c_1\log\left(-c_1^2 + 2z_i\left(1+\sqrt{1-\frac{c_1^2}{z_i}}\right)\right)\quad(50)$$

and c_1 is determined by the equation

$$\triangle y_0 = -c_1\Bigg[\log\left(-c_1^2+2z_f\left(1+\sqrt{1-\frac{c_1^2}{z_f}}\right)\right)-$$
$$\log\left(-c_1^2+2z_i\left(1+\sqrt{1-\frac{c_1^2}{z_i}}\right)\right)\Bigg].\quad(51)$$

The parameter τ ranges between the values

$$\tau_i = \mathrm{sgn}\left(\triangle h + 2\frac{\triangle x}{\triangle h}\bar x\right)\mathrm{sech}^{-1}\left(\frac{x_i}{r}\right)\quad(52)$$

and

$$\tau_f = -\mathrm{sgn}\left(\triangle h - 2\frac{\triangle x}{\triangle h}\bar x\right)\mathrm{sech}^{-1}\left(\frac{x_f}{r}\right).\quad(53)$$

When y_0 is held fixed, the geodesics are precisely those of the hyperbolic plane as expected. Furthermore, these geodesics are necessarily minimizing by virtue of the constant, negative Ricci scalar [38, 39]. Several example geodesics are displayed in Fig. 2.

VI. COMPUTING DISSIPATION NUMERICALLY

We validate the optimality of these geodesics by calculating excess work directly from the Fokker-Planck equation. In full generality, the mean excess work as a functional of the protocol $\boldsymbol\lambda(t) = (y_0(t),\beta(t),k(t))$ is

$$\langle(\beta W)_{\mathrm{ex}}\rangle \equiv \int_0^{t_f} dt\ \beta\mathcal{P}_{\mathrm{ex}}\qquad(54a)$$
$$= \int_0^{t_f} dt\left(\dot\beta\frac{\langle p^2\rangle}{2m} + \frac{1}{2}\langle(y-y_0)^2\rangle\left(k\dot\beta + \dot k\beta\right)\right.$$
$$\left.+ k\beta\dot y_0\langle y_0-y\rangle - \frac{\dot\beta}{\beta} - \frac{\dot k}{2k}\right).\qquad(54b)$$

Here angled brackets denote averages over the nonequilibrium probability density $f(y,p,t)$.

Standard arguments [32] yield the Fokker-Planck equation for the time evolution of $f(y,p,t)$,

$$\frac{\partial f}{\partial t}+\frac{p}{m}\frac{\partial f}{\partial y}-k(t)[y-y_0(t)]\frac{\partial f}{\partial p}-\frac{\zeta^c}{m}\frac{\partial[pf]}{\partial p}-\frac{\zeta^c}{\beta(t)}\frac{\partial^2 f}{\partial p^2}=0.\ (55)$$

By integrating Eq. (55) against y,p, etc., we find a system
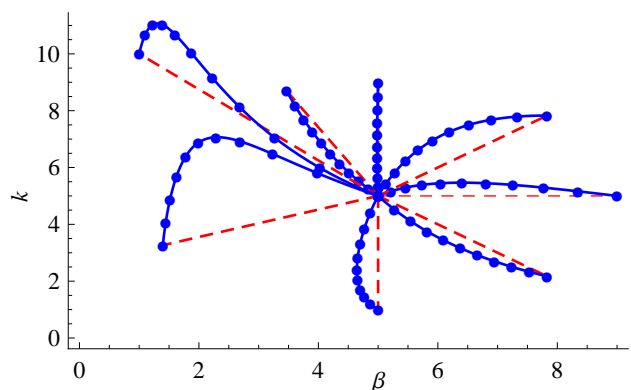

FIG. 2. Optimal protocols differ substantially from linear interpolation (red dashed lines). Blue solid curves represent geodesics of the inverse diffusion tensor, and are thus optimal protocols for transitioning the system from one state to another in a fixed amount of time. Blue dots indicate points separated by equal times along each of the eight optimal paths shown.

of equations for relevant nonequilibrium averages:

$$\frac{d}{dt}\langle y\rangle = \frac{\langle p\rangle}{m} \tag{56a}$$

$$\frac{d}{dt}\langle p\rangle = -\frac{\zeta^c}{m}\langle p\rangle - k\langle y - y_0\rangle \tag{56b}$$

$$\frac{d}{dt}\langle py\rangle = \frac{\langle p^2\rangle}{m} - k\langle y^2\rangle - \frac{\zeta^c}{m}\langle py\rangle + ky_0\langle y\rangle \tag{56c}$$

$$\frac{d}{dt}\langle y^2\rangle = \frac{2}{m}\langle py\rangle \tag{56d}$$

$$\frac{d}{dt}\langle p^2\rangle = -2k\langle p\,(y - y_0)\rangle - 2\frac{\zeta^c}{m}\langle p^2\rangle + \frac{2\zeta^c}{\beta}. \tag{56e}$$

Following the derivation of the friction tensor in [25] would require us to use linear response theory and to supplement the system Eq. (56) by initial conditions

$$\langle y\rangle\,(0) = y_0(0) \tag{57a}$$

$$\langle p\rangle\,(0) = 0 \tag{57b}$$

$$\langle y^2\rangle\,(0) = y_0(0)^2 + \frac{1}{k(0)\beta(0)} \tag{57c}$$

$$\langle py\rangle\,(0) = 0 \tag{57d}$$

$$\langle p^2\rangle\,(0) = \frac{m}{\beta(0)}\,. \tag{57e}$$

We solve these equations numerically and compare a geodesic protocol with naive protocols in Fig. 3.

This system has three natural dimensionless quantities

$$A \equiv \frac{m}{\zeta^c\triangle t}\,,\ B \equiv \frac{\zeta^c}{\tilde{k}\triangle t}\,,\ M \equiv \frac{\zeta^c\,(\triangle t)^3}{\tilde{l}^2 m^2\tilde{\beta}}\,, \tag{58}$$

dependent upon characteristic scales for (inverse) temperature $\tilde{\beta}$, length $\tilde{l}$, spring constant $\tilde{k}$ and the protocol duration $\triangle t$. These suggest at least two plausible measures of distance from equilibrium [25]. A corresponds to the ratio of two timescales, the timescale $\frac{m}{\zeta^c}$ for frictional damping and the timescale of the perturbation protocol $\triangle t$. Likewise, B is the ratio of two powers during changes of y_0, the dissipative power $\zeta^c(\Delta y_0/\triangle t)^2$ and the elastic power $\tilde{k}(\Delta y_0)^2/\triangle t$. As A decreases and as B decreases, the system will remain closer to equilibrium during the course of the nonequilibrium perturbation, and hence our near-equilibrium approximation will be more accurate.

This intuition is confirmed in our numerical calculations: with $A \ll 1$ and $B \ll 1$, the dissipation of geodesic protocols obtained numerically via Fokker-Planck agrees with the inverse diffusion tensor approximation to better than .1% (see Fig. 3). Note that, while the inverse diffusion tensor approximation is excellent for optimal protocols and small deviations thereof, it can deviate substantially from the exact result for large deviations from the geodesic.

VII. THE INVERSE DIFFUSION TENSOR ARISES NATURALLY FROM THE FOKKER-PLANCK EQUATION

If we neglect terms involving derivatives of protocols of degree two and higher, we may find an approximate solution to the Fokker-Planck system:

$$\langle y\rangle \approx y_0 - \frac{\zeta^c}{k}\dot{y}_0 \tag{59a}$$

$$\langle p\rangle \approx m\dot{y}_0 \tag{59b}$$

$$\langle py\rangle \approx my_0\dot{y}_0 - \frac{m}{2}\left(\frac{\dot{k}}{\beta k^2} + \frac{\dot{\beta}}{\beta^2 k}\right) \tag{59c}$$

$$\langle p^2\rangle \approx \frac{m}{\beta} + \frac{m^2}{\zeta^c}\left(\frac{\dot{k}}{2\beta k} + \frac{\dot{\beta}}{\beta^2}\right) \tag{59d}$$

$$\langle y^2\rangle \approx y_0^2 + \frac{1}{\beta k} - \frac{2\zeta^c}{k}y_0\dot{y}_0 \tag{59e}$$

$$+ \dot{k}\left(\frac{m}{\zeta^c}\frac{1}{2\beta k^2} + \frac{\zeta^c}{2\beta k^3}\right) + \dot{\beta}\left(\frac{m}{\zeta^c}\frac{1}{\beta^2 k} + \frac{\zeta^c}{2\beta^2 k^2}\right).$$

Substituting this into the expression for mean excess power Eq. (54b), we recover Eq. (7). The argument above suggests that the emergence of the inverse diffusion tensor from the Fokker-Planck equation may follow from a perturbation expansion in small parameters.

VIII. DISCUSSION

We have employed geometric techniques to find optimal protocols for a simple, but previously unsolved, stochastic system. Calculation of the Ricci scalar for a submanifold pointed to a change of coordinates that identified the submanifold with the hyperbolic plane and greatly simplified the metric for the full three-dimensional manifold. This simplification, combined with the identification of a Killing field, permitted calculation of an exact closed-form expression for geodesics. Exact calculations using the Fokker-Planck equation confirmed that geodesics in the (β, k)-submanifold do indeed produce less dissipation than any comparison protocol we tested.

In addition to being useful for identifying optimal protocols, we expect that the Ricci scalar will turn out to have an important physical interpretation. Riemannian geometry has been useful for the study of thermodynamic length of macroscopic systems [41, 42], and there has been some speculation about the role of the Ricci scalar in that setting [42], but the interpretation of R arising from the inverse diffusion tensor remains ambiguous. We hope that further study of these geometrical ideas extended to nonequilibrium systems will help clarify its role.

It would also be interesting to establish a physical interpretation for the conserved quantities arising from Killing fields in this context. We found two conserved quantities (see Eq. (40)), which may be the only ones, but this model could have as many as six, given that there might be as many as six unique globally smooth Killing fields for this three-dimensional model system. (In general, there are at most $\frac{1}{2}n\,(n+1)$ independent
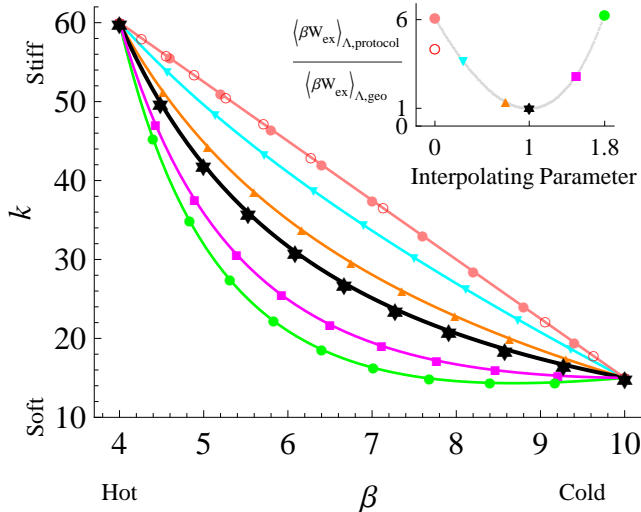
FIG. 3. Geodesics describe protocols that outperform naive straight line paths in parameter space. A geodesic between two fixed points in the (β, k)-plane (black) and several comparison protocols are pictured above. The comparison protocols were generated via a linear interpolation between the constant speed straight line (pink) and the geodesic. The tick marks represent points separated by equal times. The solid pink dots correspond to the constant speed parametrization of the line whereas the open red circles correspond to the optimal parametrization along this straight path. The ratio of excess work to that of the geodesic protocol is: 6.12 (pink circle), 4.37 (red open circle), 3.67 (cyan downward-pointing triangle), 1.38 (orange upward-pointing triangle), 1.00 (black star (geodesic)), 2.86 (magenta square), and 6.29 (green circle). These ratios are plotted in the inset figure along with a graph of the ratio as a function of the interpolating parameter (light gray curve). All excess work values were calculated using the Fokker-Planck system Eq. (56). Here, $A = 10^{-2}, B = 10^{-3}, M = 1$, placing the system within the near-equilibrium regime and ensuring accuracy of the inverse diffusion tensor approximation.

globally smooth Killing fields where n is the dimension of the manifold [40].)

In the course of developing our framework, we encountered four distinct measures of the departure from equilibrium. The first two were dimensionless parameters, A and B, which have relatively straightforward physical interpretations — the timescale for frictional dissipation relative to the protocol duration and the ratio of the dissipative power to elastic power, respectively (see discussion following Eq. (58)).

The third was the disagreement between dissipation computed assuming linear response theory and the true dissipation. Empirically, we found that our linear response approximation was consistently accurate for all parameter regimes we tested in which both dimensionless

parameters A and B were small, at least for protocols not too far from geodesics. Conversely, the linear response approximation appeared to break down for many cases we tested with at least one of these parameters of order unity or greater. However, the full extent of validity of the linear response approximation is not clear to us, suggesting an important direction for future research.

Finally, we found that truncating to first order in temporal derivatives of the control parameters in our model was sufficient to yield the same inverse diffusion tensor formalism we originally derived using linear response theory. While it is plausible that these two types of linear approximations are directly related, further exploration is needed to uncover the relationship between linear response theory and truncating the model equations to first order in temporal derivatives.

Our results are novel in three distinct ways. First, we included β as a control parameter, which is a natural extension of thermodynamic length (e.g. [41, 43]) that is amenable to direct experimental confirmation. Our work generalizes the construction of [25] and opens up new experimental avenues for testing the validity of the framework.

Secondly, our geodesic protocols optimize dissipation for simultaneous variation of all three adjustable parameters; to our knowledge, no previous study has reported optimal protocols for any model system with three control parameters. In [44, 45], Seifert and coworkers elegantly derived the exact optimal protocols for perturbing the position y_0 and spring constant k separately, for both over-damped and under-damped Langevin dynamics. In [46], Aurell and coworkers discuss the simultaneous variation of the stiffness and the location of the trap. We note that our method misses the protocol jumps found in their analysis due to our smoothness assumptions on the protocols. When this restriction on the differentiability of the curve is imposed, we found that any component of the optimal protocol $\big(y_0(t), \beta(t), k(t)\big)$ generically depends on all components of both endpoints due to the non-trivial geometry of the parameter space.

Finally, we successfully brought the machinery of Riemannian geometry to bear on a small-scale, nonequilibrium thermodynamic problem, revealing a surprisingly rich geometric structure. Concepts such as Killing vector fields, coordinate invariance and the Ricci scalar proved indispensable in the construction of optimal protocols. These results are encouraging and this approach may prove useful for understanding the constraints on the non-equilibrium thermodynamic efficiency of biological and synthetic molecular machines.

[1] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).
[2] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).
[3] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).
[4] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[5] T. Hatano and S. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).
[6] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).
[7] D. M. Carberry, J. C. Reid, G. M. Wang, E. M. Sevick, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **92**, 140601 (2004).
[8] N. Garnier and S. Ciliberto, Phys. Rev. E **71**, 060101 (2005).
[9] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat. Phys. **6**, 988 (2010).
[10] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[11] J. T. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr, and C. Bustamante, Science **296**, 1832 (2002).
[12] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[13] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010).
[14] R. M. Neal, Stat. Comput. **11**, 125 (2001).
[15] J. P. Nilmeier, G. E. Crooks, D. D. L. Minh, and J. D. Chodera, P. Natl. Acad. Sci. USA **108**, E1009 (2011).
[16] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, Phys. Rev. Lett. **109**, 120604 (2012).
[17] F. Weinhold, J. Chem. Phys. **63**, 2479 (1975).
[18] G. Ruppeiner, Phys. Rev. A **20**, 1608 (1979).
[19] F. Schlögl, Z. Phys. B **59**, 449 (1985).
[20] P. Salamon, J. Nulton, and E. Ihrig, J. Chem. Phys. **80**, 436 (1984).
[21] P. Salamon and R. S. Berry, Phys. Rev. Lett. **51**, 1127 (1983).
[22] D. Brody and N. Rivier, Phys. Rev. E **51**, 1006 (1995).
[23] G. E. Crooks, Phys. Rev. Lett. **99**, 100602 (4) (2007).
[24] J. Burbea and C. R. Rao, J. Multivariate Anal. **12**, 575 (1982).
[25] D. A. Sivak and G. E. Crooks, Phys. Rev. Lett. **108**, 190602 (2012).
[26] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).
[27] K. Sekimoto, Prog. Theor. Phys. Suppl. **130**, 17 (1998).
[28] C. Jarzynski, Acta. Phys. Pol. B **29**, 1609 (1998).
[29] A. Imparato, L. Peliti, G. Pesce, G. Rusciano, and A. Sasso, Phys. Rev. E **76**, 050101 (2007).
[30] G. E. Crooks, *Excursions in Statistical Dynamics*, Ph.D. thesis, University of California at Berkeley (1999).
[31] C. Jarzynski, Journal of Statistical Physics **96**, 415 (1999).
[32] R. Zwanzig, *Nonequilibrium statistical mechanics* (Oxford University Press, New York, 2001).
[33] F. Douarche, S. Ciliberto, A. Petrosyan, and I. Rabbiosi, Europhys. Lett. **70**, 593 (2005).
[34] S. Ciliberto, S. Joubaud, and A. Petrosyan, J. Stat. Mech.: Theor. Exp. , P12003 (2010).
[35] D. A. Sivak and G. E. Crooks, Phys. Rev. Lett. **108**, 150601 (2012).
[36] S. Vaikuntanathan and C. Jarzynski, Europhys. Lett. **87**, 60005 (6 pp.) (2009).
[37] W. Boyce and R. DiPrima, *Elementary differential equations and boundary value problems*, seventh ed. (Wiley, New York, 2000).
[38] M. P. do Carmo, *Riemannian geometry* (Birkhäuser, Boston, 1992).
[39] J. Jost, *Riemannian geometry and geometric analysis*, 4th ed. (Springer, Berlin, 2005).
[40] S. M. Carroll, *Spacetime and geometry: An introduction to general relativity*, 1st ed. (Addison Wesley, San Francisco, 2004).
[41] D. C. Brody and D. W. Hook, J. Phys. A **42**, 023001 (33) (2009).
[42] G. Ruppeiner, Am. J. Phys. **78**, 1170 (2010).
[43] D. K. Shenfeld, H. Xu, M. P. Eastwood, R. O. Dror, and D. E. Shaw, Phys. Rev. E **80**, 046705 (4) (2009).
[44] A. Gomez-Marin, T. Schmiedl, and U. Seifert, J. Chem. Phys. **129**, 024114 (8) (2008).
[45] T. Schmiedl and U. Seifert, Phys. Rev. Lett. **98**, 108301 (2007).
[46] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, Phys. Rev. Lett. **106**, 250601 (4) (2011).