

Beyond Linear Response: Equivalence between Thermodynamic Geometry and Optimal Transport

Adrienne Zhong^{1,2,*} and Michael R. DeWeese^{1,2,3}

¹*Department of Physics, University of California, Berkeley, Berkeley, CA, 94720*

²*Redwood Center For Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA, 94720*

³*Department of Neuroscience, University of California, Berkeley, Berkeley, CA, 94720*

(Dated: April 29, 2024)

A fundamental result of thermodynamic geometry is that the optimal, minimal-work protocol that drives a nonequilibrium system between two thermodynamic states in the slow-driving limit is given by a geodesic of the friction tensor, a Riemannian metric defined on control space. For overdamped dynamics in arbitrary dimensions, we demonstrate that thermodynamic geometry is equivalent to L^2 optimal transport geometry defined on the space of equilibrium distributions corresponding to the control parameters. We show that obtaining optimal protocols past the slow-driving or linear response regime is computationally tractable as the sum of a friction tensor geodesic and a counterdiabatic term related to the Fisher information metric. These geodesic-counterdiabatic optimal protocols are exact for parameteric harmonic potentials, reproduce the surprising non-monotonic behavior recently discovered in linearly-biased double well optimal protocols, and explain the ubiquitous discontinuous jumps observed at the beginning and end times.

Introduction.—A consequence of the Second Law of Thermodynamics is that finite-time processes require work to be irretrievably lost as dissipation. Recent studies in stochastic thermodynamics have aimed to characterize minimal-work protocols, which have applications for nanoscopic engineering [1–12] and for understanding biophysical systems [13–18]. In this Letter we unify disparate geometric approaches and arrive at a novel framework for obtaining and better understanding thermodynamically optimal protocols.

The problem statement is: Given a configuration space $x \in \mathbb{R}^d$, inverse temperature β , and potential energy function $U_\lambda(x)$ parameterized by $\lambda \in \mathcal{M}$, what is the optimal protocol $\lambda^*(t)$ connecting the parameter values λ_i and λ_f in a finite time τ that minimizes the work

$$W[\lambda(t)] = \int_0^\tau \frac{d\lambda^\mu}{dt} \left\langle \frac{\partial U_\lambda}{\partial \lambda^\mu} \right\rangle dt? \quad (1)$$

Here $\mathcal{M} \subseteq \mathbb{R}^n$ is an orientable m -dimensional manifold, locally resembling \mathbb{R}^m everywhere with $m \leq n$. We use Greek indices to denote local coordinates of $\lambda \in \mathcal{M}$, and the Einstein summation convention (i.e., repeated Greek indices within a term are implicitly summed). The ensemble average $\langle \cdot \rangle$ is over trajectories $X(t)|_{t \in [0, \tau]}$ that start in equilibrium with λ_i and evolve via some specified Hamiltonian or Langevin dynamics under $U_{\lambda(t)}|_{t \in [0, \tau]}$.

Schmiedl and Seifert [19] showed that optimal protocols minimizing Eq. (1) have intriguing *discontinuous jumps* at the beginning and end times, which have proven to be ubiquitous [5, 19–29]. Furthermore, optimal protocols can even be *non-monotonic* in time [5, 26].

Sivak and Crooks demonstrated through linear response [30] that in the slow-driving limit ($\tau \gg \tau_R$, an appropriate relaxation time-scale), optimal protocols are *geodesics* of a symmetric positive-definite [31] friction tensor defined in terms of equilibrium time-correlation

functions. Treating the friction tensor as a Riemannian metric induces a geometric structure on the space of control parameters, known as “Thermodynamic geometry.” This approach is computationally tractable, as the friction tensor can be obtained through measurement, and geodesics on \mathcal{M} can be determined by solving an ordinary differential equation. Geodesic protocols have been studied for a variety of systems including the Ising model [32–35], barrier crossing [15–17], bit-erasure [2, 36], and nanoscopic heat engines (after allowing temperature to be controlled) [9–12, 37], but unfortunately their performance can degrade past the slow-driving regime [26].

Alternatively, when the ensemble of trajectories is additionally constrained to end in equilibrium with λ_f , finding the work-minimizing protocol for overdamped dynamics is equivalent to the Benamou-Brenier formulation of the L^2 optimal transport problem [38, 39]—finding the dynamical mapping between the two distributions that has minimal integrated squared distance [40]—which itself yields a Riemannian-geometric structure [41–43]. The Benamou-Brenier solution is a time-dependent distribution and time-dependent velocity field that solves a continuity equation, which in this Letter we explicitly identify as a desired probability density evolution and the additional counterdiabatic forcing needed to effectuate its faster-than-quasistatic time evolution (as studied in so-called Engineered Swift Equilibration [44–46], Counterdiabatic Driving [47], and Shortcuts to Adiabaticity [48–56]). Remarkably, optimal protocols obtained in this manner are exact for arbitrary protocol durations τ [38]. Unfortunately, this approach involves solving coupled PDEs in configuration space (\mathbb{R}^d), which is typically infeasible for dimension $d \gtrsim 5$. Furthermore, the control space \mathcal{M} must be sufficiently expressive in order to implement the optimal transport solution, which is often overly-restrictive [57].

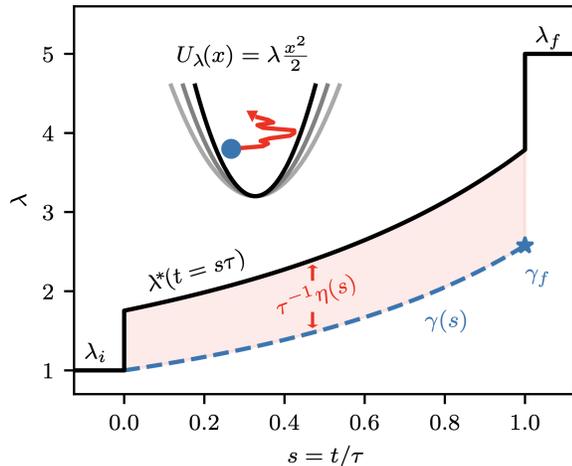


FIG. 1. Alg. 1. reproduces the exact optimal protocol (black) for the variable-stiffness harmonic oscillator (depicted here, $\lambda_i = 1$, $\lambda_f = 5$, and $\tau = 0.5$) found in [19], as the sum of geodesic (dashed blue) and counterdiabatic (shaded red) components. The geodesic $\gamma(s)$ connects $\gamma(0) = \lambda_i$ to $\gamma(1) = \lambda_f$ (blue star) solving Eq. (21).

For overdamped dynamics in arbitrary dimension, [37] showed that the friction tensor may be obtained via a perturbative expansion of the Benamou-Brenier objective function. Here we derive an even stronger result, that thermodynamic geometry is in fact *equivalent* to optimal transport geometry, in the sense that the friction tensor and the Benamou-Brenier problem restricted to equilibrium distributions parameterized by λ have identical geodesics and geodesic distances. Surprisingly, we find that a counterdiabatic component may be calculated using the Fisher information metric from information geometry [58–60]. We demonstrate that protocols obtained by adding this counterdiabatic term to thermodynamic geometry geodesics are analytically exact for parametric harmonic oscillators, reproduce recently discovered non-monotonic behavior in certain optimal protocols [26], and satisfactorily explain the origin of jumps at beginning and end times.

Preliminaries.—For each $\lambda \in \mathcal{M}$ there is a corresponding equilibrium distribution

$$\rho_\lambda^{\text{eq}}(x) = \exp\{-\beta[U_\lambda(x) - F(\lambda)]\}, \quad (2)$$

where $F(\lambda) = -\beta^{-1} \ln \int \exp[-\beta U_\lambda(x')] dx'$ is the equilibrium free energy of the potential energy $U_\lambda(\cdot)$. For ease of notation we will denote $\rho_i^{\text{eq}} = \rho_{\lambda_i}^{\text{eq}}$, $\rho_f^{\text{eq}} = \rho_{\lambda_f}^{\text{eq}}$, and $\Delta F = F(\lambda_f) - F(\lambda_i)$.

We consider overdamped Langevin equations, such that trajectories $X(t) \in \mathbb{R}^d$ follow the stochastic ODE

$$dX(t) = -\nabla U_{\lambda(t)}(X(t)) dt + \sqrt{2\beta^{-1}} dB(t), \quad (3)$$

where $B(t) \in \mathbb{R}^d$ is an instantiation of standard Brownian motion [61]. Here we will consider only isothermal

protocols, so WLOG we set $\beta = 1$.

The probability density $\rho(x, t)$ corresponding to Eq. (3) undergoes a time-evolution expressible either as a Fokker-Planck equation or a continuity equation of a gradient field

$$\frac{\partial \rho}{\partial t} = \mathcal{L}_{\lambda(t)} \rho \quad \text{or} \quad \frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \nabla \phi), \quad (4)$$

where \mathcal{L}_λ is the Fokker-Planck operator [62]

$$\mathcal{L}_\lambda \rho = \nabla^2 \rho + \nabla \cdot (\rho \nabla U_\lambda), \quad (5)$$

while ϕ is a scalar field that depends on both ρ and λ

$$\phi(x, t) = \ln \rho(x, t) + U_{\lambda(t)}(x, t). \quad (6)$$

The adjoint operator $\mathcal{L}_\lambda^\dagger$ acts on a scalar field $\psi(x)$ via [63]

$$[\mathcal{L}_\lambda^\dagger \psi](x) = \rho_\lambda^{\text{eq}}(x)^{-1} \nabla \cdot [\rho_\lambda^{\text{eq}}(x) \nabla \psi(x)]. \quad (7)$$

Finally, $f_\mu(x) := -\partial U_\lambda(x) / \partial \lambda^\mu$ is the conjugate force to λ^μ . The excess conjugate force is then

$$\delta f_\mu(x) = -\left[\frac{\partial U_\lambda(x)}{\partial \lambda^\mu} - \left\langle \frac{\partial U_\lambda}{\partial \lambda^\mu} \right\rangle_\lambda^{\text{eq}} \right] = \frac{\partial \ln \rho_\lambda^{\text{eq}}(x)}{\partial \lambda^\mu}. \quad (8)$$

Thermodynamic geometry.—In the slow-driving limit, the excess work, defined as the work (Eq. (1)) minus the equilibrium free energy difference $W_{\text{ex}} = W - \Delta F$, is [64]

$$W_{\text{ex}}[\lambda(t)] \approx \int_0^\tau \frac{d\lambda^\mu}{dt} \frac{d\lambda^\nu}{dt} g_{\mu\nu}(\lambda(t)) dt, \quad (9)$$

where

$$g_{\mu\nu}(\lambda) = \int_0^\infty \langle \delta f_\mu(X(t')) \delta f_\nu(X(0)) \rangle_\lambda^{\text{eq}} dt' \quad (10)$$

is the symmetric positive-definite [31] friction tensor. Here, $\langle \cdot \rangle_\lambda^{\text{eq}}$ denotes an equilibrium average (i.e., $X(0) \sim \rho_\lambda^{\text{eq}}$, and trajectories undergo Langevin dynamics (Eq. (3)) with constant λ).

Remarkably, the friction tensor induces a Riemannian geometry on control space (\mathcal{M}, g) known as “Thermodynamic geometry,” with squared thermodynamic length between $\lambda_A, \lambda_B \in \mathcal{M}$ given by minimizing the path action

$$\mathcal{T}^2(\lambda_A, \lambda_B) = \min_{\lambda(s)|_{s \in [0,1]}} \left\{ \int_0^1 \frac{d\lambda^\mu}{ds} \frac{d\lambda^\nu}{ds} g_{\mu\nu}(\lambda(s)) ds \right\} \quad \text{satisfying } \lambda(0) = \lambda_A, \lambda(1) = \lambda_B. \quad (11)$$

In the slow-driving limit, optimal protocols $\lambda^*(t)$ connecting λ_i and λ_f in time τ are time-rescaled versions of geodesics of Eq. (11), and the optimal excess work scales inversely with protocol time $W_{\text{ex}}^* \approx \mathcal{T}^2(\lambda_i, \lambda_f) / \tau$ [64, 65].

While this geometric framework is both mathematically elegant and computationally tractable, geodesic protocols are fundamentally approximate; their performance often degrades for sufficiently small protocol times, in some cases performing even worse than a linear interpolation protocol [26].

Optimal transport geometry.—Optimal transport is

$$\mathcal{W}_2^2[\rho_A, \rho_B] = \min_{\rho_s, \phi_s|_{s \in [0,1]}} \left\{ \int_0^1 \int \rho_s(x) |\nabla \phi_s(x)|^2 dx ds \mid \text{satisfying } \frac{\partial \rho_s}{\partial s} = \nabla \cdot (\rho_s \nabla \phi_s), \rho_0 = \rho_A, \rho_1 = \rho_B \right\}. \quad (12)$$

Here $\rho_s(\cdot)|_{s \in [0,1]}$ is a trajectory of configuration space probability densities $\mathcal{P}(\mathbb{R}^d)$ [66], and $\phi_s(\cdot)|_{s \in [0,1]}$ is a trajectory of scalar fields that yield gradient velocity fields $v_s = -\nabla \phi_s$ satisfying the continuity equation $\partial_s \rho_s = -\nabla \cdot (\rho_s v_s)$ [67]. This so-called Benamou-Brenier formulation of optimal transport (Eq. (12)) reveals a Riemannian structure on the space of probability distributions known as Otto calculus [41–43]: on this manifold of probability distributions $M := \mathcal{P}(\mathbb{R}^d)$, a “point” is a probability distribution $\rho \in M$, a “tangent space vector” is a gradient velocity field identifiable (up to a constant offset) by a scalar field $\phi \in T_\rho(M)$, and geodesics are the argmin of Eq. (12).

For overdamped dynamics, the work-minimizing protocol satisfying boundary conditions $\rho(\cdot, 0) = \rho_i^{\text{eq}}$ and $\rho(\cdot, \tau) = \rho_f^{\text{eq}}$ is a time-scaled solution of Eq. (12) for $\rho_A = \rho_i^{\text{eq}}, \rho_B = \rho_f^{\text{eq}}$, assuming sufficiently expressive control [38]. (See the SM for a concise derivation.) From the continuity equation form of $\partial \rho(\cdot, t)/\partial t$ (Eqs. (4) and (6)), the optimal protocol $\lambda^*(t)$ for finite time τ can be expressed in terms of ρ_s^* and ϕ_s^* that solve Eq. (12), as satisfying (up to a constant offset)

$$U_{\lambda^*(t)}(x) = -\ln \rho_{t/\tau}^*(x) + \tau^{-1} \phi_{t/\tau}^*(x). \quad (13)$$

The first term corresponds to the Benamou-Brenier geodesic ρ_s^* , and the second one with ϕ_s^* is a counterdiabatic term that drives the probability distribution solving Eq. (4) to match the geodesic $\rho(\cdot, t) = \rho_{t/\tau}^*$ (see [68]).

Remarkably, this solution is *exact* for any finite τ , and it provides a geometric interpretation for these work-minimizing protocols as optimal transport geodesics connecting ρ_i^{eq} to ρ_f^{eq} . Through the time-scaling $t = \tau s$, it follows that $W_{\text{ex}}^* = \mathcal{W}_2^2[\rho_i^{\text{eq}}, \rho_f^{\text{eq}}]/\tau$ is a *tight* lower bound for excess dissipation in this additionally-constrained setting [69, 70]. However, there are two important caveats to this approach: first, solving Eq. (12) involves PDEs on configuration space, which generally for dimension $d \gtrsim 5$ is computationally intractable (although, see [71–74] for sophisticated modern machine learning methods, as well as [57]). Second, the control parameters must be sufficiently expressive in the sense that for all $t \in (0, \tau)$ there

traditionally formulated as finding the transport map sending a distribution ρ_A to another ρ_B that minimizes an integrated (L^2) squared distance. This minimal integrated squared distance defines the squared L^2 -Wasserstein metric distance between probability distributions, which was shown in [40] to also be the minimum of a path action

has to be a $\lambda \in \mathcal{M}$ that satisfies Eq. (13). Worse yet, there might not be *any* admissible protocols that can satisfy the terminal constraint $\rho(\cdot, \tau) = \rho_f^{\text{eq}}$ [57].

Without the terminal condition, this problem is no longer over-constrained. The optimal excess work can be expressed as a minimum over $\rho_f = \rho(\cdot, \tau)$ (see the SM),

$$W_{\text{ex}}^* = \min_{\rho_f} \mathcal{W}_2^2[\rho_i^{\text{eq}}, \rho_f]/\tau + D_{\text{KL}}(\rho_f | \rho_f^{\text{eq}}), \quad (14)$$

where the additional KL-divergence cost

$$D_{\text{KL}}(\rho_A | \rho_B) := \int \rho_A(x) \ln \frac{\rho_A(x)}{\rho_B(x)} dx, \quad (15)$$

is the dissipation from the equilibration $\rho_f \rightarrow \rho_f^{\text{eq}}$ that occurs for $t > \tau$ (see [75]). Optimal protocols $\lambda^*(t)$ are also obtained via Eqs. (12) and (13), but now with $\rho_A = \rho_i^{\text{eq}}$ and $\rho_B = \rho_f^*$ that minimizes Eq. (14). Without the restrictive terminal constraint, protocols that approximate Eq. (13) are allowed (in the case of limited expressivity), and may be near-optimal in performance [28, 33].

Demonstrating equivalence of geometries.—We start by expressing Eq. (10) with the time-propagator (e.g., see Ch 4.2 of [62]):

$$\begin{aligned} g_{\mu\nu}(\lambda) &= \int_0^\infty \int \rho_\lambda^{\text{eq}}(x) \delta f_\mu(x) e^{\mathcal{L}_\lambda^\dagger t'} \delta f_\nu(x) dx dt' \\ &= - \int \rho_\lambda^{\text{eq}}(x) \delta f_\mu(x) \{\mathcal{L}_\lambda^\dagger\}^{-1} [\delta f_\nu](x) dx. \end{aligned} \quad (16)$$

The second line comes from taking the time-integral, where the inverse operator $\{\mathcal{L}_\lambda^\dagger\}^{-1}$ is defined in terms of a properly constructed Green’s function (Eq. (40) in [76]). This expression is the lowest order tensor found in a perturbative expansion of the Fokker-Planck equation [76].

By formally defining $\phi_\mu = \{\mathcal{L}_\lambda^\dagger\}^{-1} \delta f_\mu$ as (up to a constant offset) the scalar field solving $\mathcal{L}_\lambda^\dagger \phi_\mu = \delta f_\mu$, it is straightforward to show with Eqs. (7) and (8) that, for

any protocol $\lambda(s)|_{s \in [0,1]}$,

$$\frac{\partial \rho_{\lambda(s)}^{\text{eq}}}{\partial s} = \nabla \cdot (\rho_{\lambda(s)}^{\text{eq}} \nabla \phi_s), \quad \text{where } \phi_s(x) = \frac{d\lambda^\mu}{ds} \phi_\mu(x). \quad (17)$$

$$\mathcal{T}^2(\lambda_A, \lambda_B) = \min_{\lambda(s), \phi_s|_{s \in [0,1]}} \left\{ \int_0^1 \int \rho_{\lambda(s)}^{\text{eq}}(x) |\nabla \phi_s(x)|^2 dx ds \mid \text{satisfying } \frac{\partial \rho_{\lambda(s)}^{\text{eq}}}{\partial s} = \nabla \cdot (\rho_{\lambda(s)}^{\text{eq}} \nabla \phi_s), \lambda(0) = \lambda_A, \lambda(1) = \lambda_B \right\}. \quad (18)$$

This is our first major result: this expression is equivalent to the squared L^2 -Wasserstein distance (Eq. (12)) with the constraint that $\rho_s|_{s \in [0,1]}$ is a trajectory of equilibrium distributions $\rho_{\lambda(s)}^{\text{eq}}|_{s \in [0,1]}$. In other words, thermodynamic geometry induced by the friction tensor (Eq. (10)) on \mathcal{M} is *equivalent* to optimal transport geometry restricted to the equilibrium distributions $\mathcal{P}_{\mathcal{M}}^{\text{eq}}(\mathbb{R}^d)$ corresponding to \mathcal{M} [77], and thus share the same geodesics and geodesic distances.

Up till now, thermodynamic geometry has prescribed optimal protocols as friction tensor geodesics joining λ_i and λ_f , which are approximate for finite τ . Optimal transport solutions require solving PDEs, but yield exact optimal protocols containing both geodesic and counterdiabatic components (Eq. (13)). Our unification of geometries suggests that thermodynamic geometry protocols may be made exact by including a counterdiabatic term.

Geodesic-counterdiabatic optimal protocols.—From here we consider the control-affine parameterization

$$U_\lambda(x) = U_{\text{fixed}}(x) + U_{\text{offset}}(\lambda) + \lambda^\mu U_\mu(x), \quad (19)$$

and control space $\lambda \in \mathcal{M} = \mathbb{R}^m$. It follows from the equivalence of thermodynamic and optimal transport geometries that the optimal protocol should have the form

$$\lambda^*(t) = \gamma(t/\tau) + \tau^{-1} \eta(t/\tau), \quad (20)$$

namely the sum of a geodesic term and a counterdiabatic term that correspond to the two terms in Eq. (13), where $\rho_s^*|_{s \in [0,1]}$ and $\phi_s^*|_{s \in [0,1]}$ solve Eq. (12) with $\rho_A = \rho_i^{\text{eq}}$ and $\rho_B = \rho_f^*$ from Eq. (14). Here, $\gamma(s)$ will be a geodesic of $g(\lambda)$ joining $\gamma(0) = \lambda_i$ to $\gamma(1) = \lambda_f$, where

$$\lambda_f = \arg \min_{\lambda} \mathcal{T}^2(\lambda_i, \lambda) / \tau + D_{\text{KL}}(\rho_{\lambda}^{\text{eq}} | \rho_f^{\text{eq}}). \quad (21)$$

We show in Appendix A that the counterdiabatic term is

$$\eta(s) = h^{-1}(\gamma(s)) g(\gamma(s)) \left[\frac{d\gamma(s)}{ds} \right], \quad (22)$$

where, intriguingly, h is the Fisher information metric [78]

$$h_{\mu\nu}(\lambda) = \int \rho_{\lambda}^{\text{eq}}(x) \delta f_\mu(x) \delta f_\nu(x) dx, \quad (23)$$

Applying $\delta f_\mu = \mathcal{L}_\lambda^\dagger \phi_\mu$ and Eq. (7) to Eq. (16) shows that the thermodynamic distance (Eq. (11)) may be expressed as

which also induces a Riemannian geometry on the space of parametric equilibrium probability distributions (\mathcal{M}, h) known as “Information geometry” [58–60]. Eq. (22) is exact in cases of sufficient expressivity (i.e., when Eq. (13) can be satisfied); otherwise, $\eta(s)$ is the full solution projected onto \mathcal{M} .

This is our second major result: the equivalence between thermodynamic and optimal transport geometries implies that optimal protocols beyond linear response require counterdiabatic forcing, and can be obtained via:

Algorithm 1 Geodesic-counterdiabatic opt. protocols.

Input: λ_i, λ_f , protocol time τ , metrics $g_{\mu\nu}(\lambda)$, $h_{\mu\nu}(\lambda)$ (Eqs. (10), (23)), KL divergence $D_{\text{KL}}(\cdot | \rho_f^{\text{eq}})$ (Eq. (15)).

- 1: Solve geodesic $\gamma(s)|_{s \in [0,1]}$ connecting $\gamma(0) = \lambda_i$ and $\gamma(1) = \lambda_f$ (obtained from Eq. (21)) under $g_{\mu\nu}$.
- 2: Calculate counterdiabatic term $\eta(s) = h^{-1} g [d\gamma/ds]$
- 3: Return optimal protocol $\lambda^*(t) = \gamma(t/\tau) + \tau^{-1} \eta(t/\tau)$

We emphasize that this procedure does not require solving any configuration-space PDEs. Moreover, in the limit $\tau \rightarrow \infty$, the counterdiabatic component in Eq. (20) vanishes and Eq. (21) is solved by $\gamma_f = \lambda_f$, and thus geodesic protocols from thermodynamic geometry are reproduced.

Examples.—We show in Appendix B that Alg. 1. reproduces exact optimal protocols solved in [19] for controlling a parametric harmonic potential. Fig. 1. illustrates an optimal protocol for $U_\lambda(x) = \lambda x^2/2$. Notice that at $t = 0$ the counterdiabatic term is suddenly turned on, while at $t = \tau$ the geodesic ends at $\gamma_f \neq \lambda_f$ and the counterdiabatic term is suddenly turned off. Seen in this light, the discontinuous jumps in optimal protocols $\lambda^*(t)$ arise from the sudden turning on and off of counterdiabatic forcing, and the discontinuity of the geodesic at $t = \tau$. We note that starting in equilibrium at $t = 0$ and suddenly ending control at $t = \tau$ are both unnatural in biological settings; these generic discontinuous jumps can be seen as artifacts of the imposed boundary conditions.

Surprisingly, *non-monotonic* optimal protocols have been found for the linearly-biased double well [26]

$$U_\lambda(x) = E_0[(x^2 - 1)^2/4 - \lambda x], \quad (24)$$

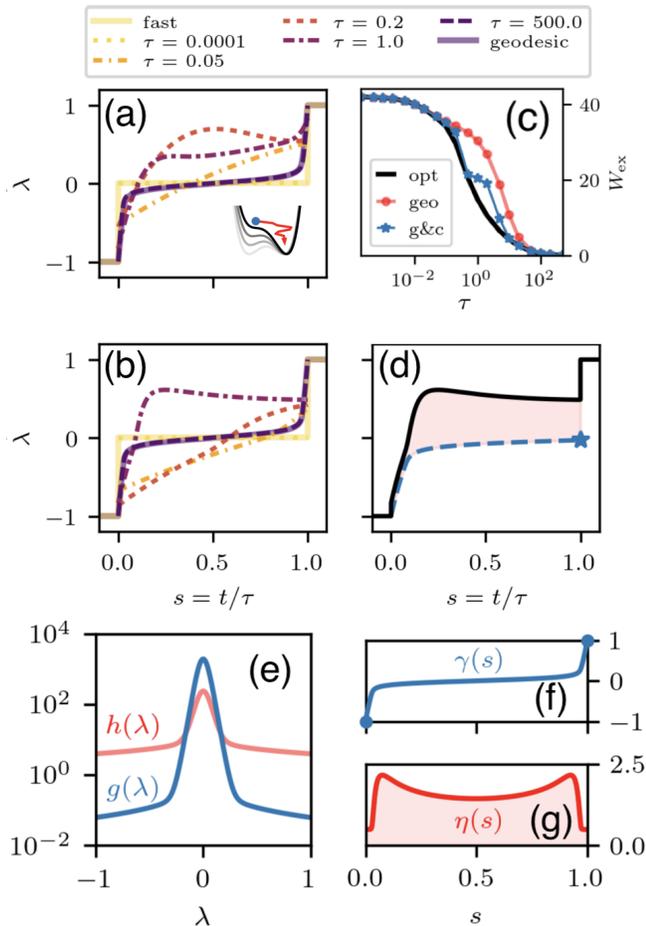


FIG. 2. (a) Exact optimal protocols obtained in [26] from solving PDEs, for the linearly-biased double well (Eq. (24); $E_0 = 16$) for different protocol durations τ including the fast protocol $\tau \rightarrow 0$ [25] (solid yellow) and the friction tensor geodesic protocol (dark purple). (b) Geodesic-counterdiabatic protocols numerically obtained from Alg. 1. (c) Geodesic-counterdiabatic protocols (blue stars) outperform the geodesic protocol (red circles) for all τ ; cf. performance of exact optimal protocols (black). We examine the reduction in performance at $\tau \sim 2$ in the SM. (d) The $\tau = 1$ protocol numerically obtained via Alg. 1 (here $\gamma_f = 0.0291$), same coloring as Fig. 1. (e) The friction and Fisher information tensors yield (f) the geodesic $\gamma(s)$ (here $\lambda_A = -1$, $\lambda_B = 1$) and (g) the non-monotonic counterdiabatic forcing $\eta(s)$.

for certain values of E_0 and τ (e.g., $\tau = 0.2$ in Fig. 2(a) for $E_0 = 16$). Fig. 2(b) illustrates protocols numerically obtained from Alg. 1 (details given in the SM). Due to the limited expressivity of the controls, these protocols are not identical to the exact optimal protocols obtained by solving PDEs [26] (Fig. 2(a)). However, they reproduce significant properties (e.g., discontinuous jumps and non-monotonicity, becoming exact in $\tau \rightarrow 0$ and $\tau \rightarrow \infty$), and lead to improved performance over geodesic protocols (Fig. 2(c)). Fig. 2(d) illustrates the non-monotonic $\tau = 1$ protocol as a sum of geodesic and counterdiabatic terms.

The tensors g and h (Fig. 2(e)) yield necessarily monotonic geodesics (Fig. 2(f)), and non-monotonic counterdiabatic forcing (Fig. 2(g)) that leads to non-monotonic optimal protocols.

Discussion.—We have demonstrated the equivalence between overdamped thermodynamic geometry on \mathcal{M} —previously seen as an approximate framework—and L^2 optimal transport geometry on equilibrium distributions $\mathcal{P}_{\mathcal{M}}^{\text{eq}}(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$. The resulting geodesic-counterdiabatic optimal protocols from Alg. 1 are exact for parametric harmonic traps, and explain both the ubiquitous discontinuous jumps and the non-monotonic behavior observed in optimal protocols.

We note that [56] presents a geodesic-counterdiabatic PDEs approach for underdamped dynamics. Additionally, underdamped optimal control has recently been related to a modified optimal transport problem [79, 80]. We expect that the metric tensor in [56], the friction tensor (Eq. (10)) for underdamped dynamics [65], and the optimal transport specified in [79, 80] may also be geometrically unified through methods similar to ours.

An interesting future direction will be to apply our findings to heat engines [9–12, 37] and active matter systems [18, 81], which have been studied with approximate geodesic protocols. We hope that the insight that minimal-work protocols require both geodesic and counterdiabatic components will prove to be useful in understanding the cyclic and fundamentally non-equilibrium processes that occur in biological systems.

Acknowledgements.—This work greatly benefited from conversations with Adam Frim. A.Z. was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program. MRD thanks Steve Strong and Fenrir LLC for supporting this project. This work was supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Contract No. W911NF-20-1-0151.

Appendix A: Counterdiabatic driving expression.—In this Appendix we derive our expression for the counterdiabatic term (Eq. (22)). Per definition, the counterdiabatic term $\phi_s(x) = \eta^\mu(s)U_\mu(x)$ is constructed to solve the continuity equation

$$\frac{\partial \rho_{\gamma(s)}^{\text{eq}}}{\partial s}(x) = \nabla \cdot (\rho_{\gamma(s)}^{\text{eq}}(x) \nabla \phi_s(x)). \quad (25)$$

We can divide by $\rho_\lambda^{\text{eq}}(x)$ and plug in $\partial_s \ln \rho_{\gamma(s)}^{\text{eq}} = [d\gamma(s)/ds] \delta f_\mu(x)$ to obtain

$$\frac{d\gamma^\mu}{ds} \delta f_\mu(x) = -\eta^\nu(s) [\mathcal{L}_{\gamma(s)}^\dagger \delta f_\nu](x), \quad (26)$$

where we have used the fact that $U_\nu(x) = -\delta f_\nu + \text{const}$, and that the adjoint operator (Eq. (7)) satisfies $\mathcal{L}_\lambda^\dagger[\psi + c] = \mathcal{L}_\lambda^\dagger[\psi]$ for any scalar field $\psi(x)$ and constant $c \in \mathbb{R}$.

Due to the limited expressivity of available controls in a given problem, it might not be possible to satisfy

$$\frac{d\gamma^\mu}{ds} \int \left[-\rho_{\gamma(s)}^{\text{eq}}(x) \delta f_\alpha(x) \{\mathcal{L}_{\gamma(s)}^\dagger\}^{-1} \right] \delta f_\mu(x) dx = -\eta^\nu(s) \int \left[-\rho_{\gamma(s)}^{\text{eq}}(x) \delta f_\alpha(x) \{\mathcal{L}_{\gamma(s)}^\dagger\}^{-1} \right] [\mathcal{L}_{\gamma(s)}^\dagger \delta f_\nu](x) dx, \quad (27)$$

which yields

$$g_{\alpha\mu}(\gamma(s)) \left[\frac{d\gamma^\mu}{ds} \right] = h_{\alpha\nu}(\gamma(s)) \eta^\nu(s), \quad (28)$$

using the friction tensor and Fisher information metric expressions Eq. (16) and Eq. (23). Optimal transport geometry measures “horizontal” displacement while information geometry measures “vertical” displacement [82], so g and h can be seen to give the conversion between the left and right hand sides of Eq. (25) [83].

Because the Fisher information metric is symmetric and positive-definite (assuming [31]), we can apply its inverse to Eq. (28), thus reproducing Eq. (22).

Appendix B: Analytic reproduction of harmonic oscillator optimal protocols.—In this Appendix we show that Alg. 1 exactly reproduces the optimal protocols for harmonic potentials first found in [19].

We first consider a variable-center harmonic potential in one dimension $U_\lambda(x) = (x - \lambda)^2/2$, both the friction and Fisher information tensors are spatially constant $g(\lambda) = f(\lambda) = 1$ due to translation symmetry. For any λ_f , the KL-divergence given by $D_{\text{KL}}(\rho_\lambda^{\text{eq}}|\rho_f^{\text{eq}}) = \int \rho_\lambda^{\text{eq}}(x) \ln[\rho_\lambda^{\text{eq}}(x)/\rho_f^{\text{eq}}(x)] dx = (\lambda - \lambda_f)^2/2$, while the squared thermodynamic length is given by $\mathcal{T}^2(\lambda_i, \lambda) = (\lambda - \lambda_i)^2$. Without loss of generality we fix $\lambda_i = 0$. Following Alg. 1, we obtain $\gamma_f = (1 + 2/\tau)^{-1} \lambda_f$ with geodesic $\gamma(s) = s \gamma_f$, and $\eta(s) = \gamma_f$. Ultimately, this yields the optimal protocol

$$\lambda^*(t) = \underbrace{[\lambda_f/(2 + \tau)] t}_{\text{geodesic}} + \underbrace{1/(2 + \tau)}_{\text{counterdiabatic}}, \quad \text{for } t \in (0, \tau), \quad (29)$$

which yields the original analytic solution reported by Schmiedl and Seifert $\lambda^*(t) = \lambda_f(1 + t)/(2 + \tau)$, Eq. (9) in [19] (note that they use t to denote protocol duration and τ to denote time, which is swapped with respect to our notation). However, we now have a refined interpretation of this optimal protocol: as consisting of a geodesic component that connects λ_i to $\gamma_f \neq \lambda_f$, and a counterdiabatic component necessary to achieve the geodesic trajectory for finite protocol durations τ .

We now solve for the optimal protocol for a variable-stiffness harmonic trap $U_\lambda(x) = \lambda x^2/2$. We will defer solving for γ_f until after obtaining the analytic form for $\lambda^*(t)$. The friction tensor for this potential has previously been shown to be $g(\lambda) = 1/4\lambda^3$ [64], and

Eq. (26). However, this potential insolubility is resolved by applying a projection operator to both sides

the Fisher information metric can be calculated to be $h(\lambda) = \langle x^4/4 \rangle_\lambda^{\text{eq}} - (\langle x^2/2 \rangle_\lambda^{\text{eq}})^2 = 1/2\lambda^2$. As shown in [64], by switching to standard deviation coordinates $\sigma = \lambda^{-1/2}$, the friction tensor is constant $\tilde{g}(\sigma) = 1$, and thus geodesics are $\sigma(s) = (1 - s)\sigma_A + s\sigma_B$ with thermodynamic length $\tilde{\mathcal{T}}^2(\sigma_A, \sigma_B) = (\sigma_B - \sigma_A)^2$, where $\sigma_A = \lambda_i^{-1/2}$ and $\sigma_B = \gamma_f^{-1/2}$. This yields the geodesic

$$\gamma(s) = [(1 - s)\sigma_A + s\sigma_B]^{-2}, \quad (30)$$

with the corresponding counterdiabatic term given by

$$\eta(s) = (\sigma_A - \sigma_B)/[(1 - s)\sigma_A + s\sigma_B]. \quad (31)$$

Plugging these into $\lambda^*(t) = \gamma(t/\tau) + \tau^{-1}\eta(t/\tau)$ leads to the interpretable optimal protocol expression

$$\lambda^*(t) = \underbrace{[1/\sigma^*(t)]^2}_{\text{geodesic}} + \underbrace{\tau^{-1}[\Delta\sigma/\sigma^*(t)]}_{\text{counterdiabatic}}, \quad (32)$$

where $\sigma^*(t) = \sigma_i + (t/\tau)\Delta\sigma$ is the linear interpolation between the endpoints $\sigma_i = \lambda_i^{-1/2}$ and $\sigma_i + \Delta\sigma = \gamma_f^{-1/2}$.

Finally we solve for γ_f via

$$\gamma_f = \arg \min_\lambda \mathcal{T}^2(\lambda_i, \lambda)/\tau + D_{\text{KL}}(\rho_\lambda^{\text{eq}}|\rho_f^{\text{eq}}), \quad (33)$$

using $\mathcal{T}^2(\lambda_i, \lambda) = (\lambda_i^{-1/2} - \lambda^{-1/2})^2$ (recall that $\lambda^{-1/2} = \sigma$ is the standard deviation) and $D_{\text{KL}}(\rho_\lambda^{\text{eq}}|\rho_f^{\text{eq}}) = (1/2)[(\lambda_f/\lambda - 1) + \ln(\lambda/\lambda_f)]$ [84]. This yields

$$\gamma_f = (\sqrt{1 + 2\lambda_i\tau + \lambda_i\lambda_f\tau^2} - 1)^2/\lambda_i\tau^2, \quad (34)$$

or when expressed in σ coordinates

$$\Delta\sigma = \sigma_i \left(\frac{1 + \lambda_f\tau - \sqrt{1 + \lambda_k\tau + \lambda_i\lambda_f\tau^2}}{2 + \lambda_f\tau} \right). \quad (35)$$

Substituting this expression into Eq. (32) reproduces the exact optimal protocol, Eqs. (18) and (19) in [19]. (Note that they use t to denote protocol duration and τ to denote time, which is swapped with respect to our notation.)

Fig. 1 illustrates the obtained exact optimal protocol for this problem (Eqs. (32) and (35)) as a sum of geodesic and counterdiabatic components. This clarifies the origin of the jumps in optimal protocols: at $t = 0$, the counterdiabatic component is suddenly turned on, and at $t = \tau$ it is abruptly turned off.

-
- * adrzhong@berkeley.edu
- [1] G. Diana, G. B. Bagci, and M. Esposito, *Physical Review E* **87**, 012111 (2013).
 - [2] P. R. Zulkowski and M. R. DeWeese, *Physical Review E* **89**, 052140 (2014).
 - [3] K. Proesmans, J. Ehrich, and J. Bechhoefer, *Physical Review Letters* **125**, 100602 (2020).
 - [4] K. Proesmans, J. Ehrich, and J. Bechhoefer, *Physical Review E* **102**, 032105 (2020).
 - [5] S. Whitelam, *Physical Review E* **108**, 044138 (2023).
 - [6] T. Schmiedl and U. Seifert, *Europhysics letters* **81**, 20003 (2007).
 - [7] I. A. Martínez, É. Roldán, L. Dinis, and R. A. Rica, *Soft matter* **13**, 22 (2017).
 - [8] D. Martin, C. Nardini, M. E. Cates, and É. Fodor, *Europhysics Letters* **121**, 60005 (2018).
 - [9] P. Abiuso and M. Perarnau-Llobet, *Physical review letters* **124**, 110606 (2020).
 - [10] K. Brandner and K. Saito, *Physical review letters* **124**, 040602 (2020).
 - [11] A. G. Frim and M. R. DeWeese, *Physical Review E* **105**, L052103 (2022).
 - [12] A. G. Frim and M. R. DeWeese, *Physical Review Letters* **128**, 230601 (2022).
 - [13] P. Geiger and C. Dellago, *Physical Review E* **81**, 021127 (2010).
 - [14] C. Dellago and G. Hummer, *Entropy* **16**, 41 (2013).
 - [15] D. A. Sivak and G. E. Crooks, *Physical Review E* **94**, 052106 (2016).
 - [16] J. N. Lucero, A. Mehdizadeh, and D. A. Sivak, *Physical Review E* **99**, 012119 (2019).
 - [17] S. Blaber and D. A. Sivak, *Europhysics Letters* **139**, 17001 (2022).
 - [18] L. K. Davis, K. Proesmans, and É. Fodor, *Physical Review X* **14**, 011012 (2024).
 - [19] T. Schmiedl and U. Seifert, *Physical review letters* **98**, 108301 (2007).
 - [20] The discontinuities in the protocol are also taken into account in the integral Eq. (1).
 - [21] Gomez-Marin et al found that optimal protocols for under-damped dynamics have even worse regularity [85]; at starting and end times, they have Dirac impulses!
 - [22] H. Then and A. Engel, *Physical Review E* **77**, 041105 (2008).
 - [23] M. V. Bonança and S. Deffner, *Physical Review E* **98**, 042103 (2018).
 - [24] P. Nazé, S. Deffner, and M. V. Bonança, *Journal of Physics Communications* **6**, 083001 (2022).
 - [25] S. Blaber, M. D. Louwerse, and D. A. Sivak, *Physical Review E* **104**, L022101 (2021).
 - [26] A. Zhong and M. R. DeWeese, *Physical Review E* **106**, 044135 (2022).
 - [27] A. Rolandi, M. Perarnau-Llobet, and H. J. Miller, *New Journal of Physics* **25**, 073005 (2023).
 - [28] M. C. Engel, J. A. Smith, and M. P. Brenner, *Physical Review X* **13**, 041032 (2023).
 - [29] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, *Europhysics letters* **89**, 20003 (2010).
 - [30] R. Zwanzig, *Nonequilibrium statistical mechanics* (Oxford university press, 2001).
 - [31] We assume the parameterization of U_λ is such that ρ_λ^{eq} are different for different λ ; otherwise, the tensor could be semi-positive definite.
 - [32] G. M. Rotskoff and G. E. Crooks, *Physical Review E* **92**, 060102 (2015).
 - [33] T. R. Gingrich, G. M. Rotskoff, G. E. Crooks, and P. L. Geissler, *Proceedings of the National Academy of Sciences* **113**, 10263 (2016).
 - [34] G. M. Rotskoff, G. E. Crooks, and E. Vanden-Eijnden, *Physical Review E* **95**, 012148 (2017).
 - [35] M. D. Louwerse and D. A. Sivak, *The Journal of Chemical Physics* **156** (2022).
 - [36] M. Scandi, D. Barker, S. Lehmann, K. A. Dick, V. F. Maisi, and M. Perarnau-Llobet, *Physical Review Letters* **129**, 270601 (2022).
 - [37] S. Chennakesavalu and G. M. Rotskoff, *Physical Review Letters* **130**, 107101 (2023).
 - [38] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, *Physical review letters* **106**, 250601 (2011).
 - [39] S. Ito, *Information Geometry* **7**, 441 (2024).
 - [40] J.-D. Benamou and Y. Brenier, *Numerische Mathematik* **84**, 375 (2000).
 - [41] F. Otto, (2001).
 - [42] P. Kniefacz, *Otto calculus or the weak Riemannian structure of $(\mathcal{P}_2(M), \mathcal{W}_2)$* , Ph.D. thesis, Technische Universität Wien (2017).
 - [43] C. Villani, *Topics in optimal transportation*, Vol. 58 (American Mathematical Soc., 2021) see Ch. 8.
 - [44] I. A. Martínez, A. Petrosyan, D. Guéry-Odelin, E. Trizac, and S. Ciliberto, *Nature physics* **12**, 843 (2016).
 - [45] A. Baldassarri, A. Puglisi, and L. Sesta, *Physical Review E* **102**, 030105 (2020).
 - [46] A. G. Frim, A. Zhong, S.-F. Chen, D. Mandal, and M. R. DeWeese, *Physical Review E* **103**, L030102 (2021).
 - [47] S. Iram, E. Dolson, J. Chiel, J. Pelesko, N. Krishnan, Ö. Güngör, B. Kuznets-Speck, S. Deffner, E. Ilker, J. G. Scott, *et al.*, *Nature Physics* **17**, 135 (2021).
 - [48] E. Torrontegui, S. Ibáñez, S. Martínez-Garaot, M. Modugno, A. del Campo, D. Guéry-Odelin, A. Ruschhaupt, X. Chen, and J. G. Muga, in *Advances in atomic, molecular, and optical physics*, Vol. 62 (Elsevier, 2013) pp. 117–169.
 - [49] C. Jarzynski, *Physical Review A* **88**, 040101 (2013).
 - [50] S. Campbell and S. Deffner, *Physical review letters* **118**, 100601 (2017).
 - [51] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, *Reviews of Modern Physics* **91**, 045001 (2019).
 - [52] C. A. Plata, A. Prados, E. Trizac, and D. Guéry-Odelin, *Physical Review Letters* **127**, 190605 (2021).
 - [53] D. Guéry-Odelin, C. Jarzynski, C. A. Plata, A. Prados, and E. Trizac, *Reports on Progress in Physics* **86**, 035902 (2023).
 - [54] A. Patra and C. Jarzynski, *Physical Review Research* **3**, 013087 (2021).
 - [55] G. Li, H. Quan, and Z. Tu, *Physical Review E* **96**, 012144 (2017).
 - [56] G. Li, J.-F. Chen, C. Sun, and H. Dong, *Physical Review Letters* **128**, 230603 (2022).
 - [57] J. Klinger and G. M. Rotskoff, arXiv preprint arXiv:2402.17931 (2024).
 - [58] R. A. Fisher, *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathe-*

- mathematical or physical character **222**, 309 (1922).
- [59] C. R. Rao, in *Breakthroughs in Statistics: Foundations and basic theory* (Springer, 1992) pp. 235–247.
- [60] S.-i. Amari, *Information geometry and its applications*, Vol. 194 (Springer, 2016).
- [61] B. Oksendal, *Stochastic differential equations: an introduction with applications* (Springer Science & Business Media, 2013).
- [62] H. Risken and H. Risken, *Fokker-planck equation* (Springer, 1996).
- [63] The operator $\mathcal{L}_\lambda^\dagger$ is known as the Backward Kolmogorov operator or the “generator” of Eq. (3). By substituting in Eq. (2) into Eq. (7), one readily recovers the perhaps more familiar expression $\mathcal{L}_\lambda^\dagger \psi = \nabla^2 \psi - \nabla U_\lambda \cdot \nabla \psi$ [62].
- [64] D. A. Sivak and G. E. Crooks, Physical review letters **108**, 190602 (2012).
- [65] P. R. Zulkowski, D. A. Sivak, G. E. Crooks, and M. R. DeWeese, Physical Review E **86**, 041148 (2012).
- [66] For technical reasons, each probability distribution must have finite second moment [43].
- [67] In the original formulation the optimization is over trajectories of effective vector fields $v_s|_{s \in [0,1]}$ as opposed to scalar fields $\phi_s|_{s \in [0,1]}$ [40]; however, it is straightforward to prove that under optimality, each v_s must be a gradient field [39, 40].
- [68] Note that the two terms $-\ln \rho_s^*$ and ϕ_s^* corresponding exactly to time-reversal-symmetric $U_0(\cdot, t)$ and time-reversal-asymmetric $U_1(\cdot, t)$ found in Eqs. (6)-(8) in [55] and Eq. (8) in [86]; also, see Eq. (6) in [4] and Eq. (2) in the SI of [44] for dimension $d = 1$.
- [69] A. Dechant and Y. Sakurai, arXiv preprint arXiv:1912.08405 (2019).
- [70] M. Nakazato and S. Ito, Physical Review Research **3**, 043093 (2021).
- [71] C. Beck, S. Becker, P. Grohs, N. Jaafari, and A. Jentzen, Journal of Scientific Computing **88**, 1 (2021).
- [72] J. Han, A. Jentzen, and W. E, Proceedings of the National Academy of Sciences **115**, 8505 (2018).
- [73] Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart, Journal of Computational Physics **447**, 110668 (2021).
- [74] M. S. Albergo, N. M. Boffi, and E. Vanden-Eijnden, arXiv preprint arXiv:2303.08797 (2023).
- [75] Interestingly, this cost function is equivalent to the celebrated JKO scheme used to study convergence properties of the Fokker-Planck equation [87], as observed in [88], with effective time-step $h = \tau/2$. We study the time-scaling of these unconstrained-end-state optimal protocols in a separate upcoming work [89].
- [76] N. S. Wadia, R. V. Zarcone, and M. R. DeWeese, Physical Review E **105**, 034130 (2022).
- [77] This is entirely analogous to how both the unit ball $B^2 = \{(x, y, z) | x^2 + y^2 + z^2 \leq 1\}$ and unit sphere $S^2 = \{(x, y, z) | x^2 + y^2 + z^2 = 1\}$ can be seen as submanifolds of Euclidean space \mathbb{R}^3 ; while geodesics connecting two points in the unit ball are equivalent to straight-line Euclidean geodesics, on the unit sphere they differ.
- [78] Traditionally the Fisher information metric is denoted as $h_{\mu\nu}(\lambda) = \int \rho_\lambda^{\text{eq}} (\partial \ln \rho_\lambda^{\text{eq}} / \partial \lambda^\mu) (\partial \ln \rho_\lambda^{\text{eq}} / \partial \lambda^\nu) dx$ (which can be seen as equivalent via Eq. (8)).
- [79] R. Sabbagh, O. M. Miangolarra, and T. T. Georgiou, arXiv preprint arXiv:2312.07788 (2023).
- [80] J. Sanders, M. Baldwin, and P. Muratore-Ginanneschi, arXiv preprint arXiv:2403.00679 (2024).
- [81] P. R. Zulkowski, D. A. Sivak, and M. R. DeWeese, PloS one **8**, e82754 (2013).
- [82] G. Khan and J. Zhang, Information Geometry **5**, 47 (2022).
- [83] One may choose other projection operators to act on Eq. (26), but then non-conventional tensors would have to be defined. Furthermore, the particular interpretation for Eq. (28), that the geometries associated with g and h convert between horizontal and vertical displacements, will be lost.
- [84] bayerj (<https://stats.stackexchange.com/users/2860/bayerj>), “Kl divergence between two univariate gaussians,” Cross Validated, uRL:<https://stats.stackexchange.com/q/7440> (version: 2022-12-27), <https://stats.stackexchange.com/q/7440>.
- [85] A. Gomez-Marin, T. Schmiedl, and U. Seifert, The Journal of chemical physics **129** (2008).
- [86] A. Zhong, B. Kuznets-Speck, and M. R. DeWeese, arXiv preprint arXiv:2304.12287 (2023).
- [87] R. Jordan, D. Kinderlehrer, and F. Otto, SIAM journal on mathematical analysis **29**, 1 (1998).
- [88] Y. Chen, T. T. Georgiou, and A. Tannenbaum, IEEE transactions on automatic control **65**, 2979 (2019).
- [89] A. Zhong and M. R. DeWeese, (unpublished).

Supplementary Information for “Beyond Linear Response: Equivalence between Thermodynamic Geometry and Optimal Transport”

Adrienne Zhong^{1,2,*} and Michael R. DeWeese^{1,2,3}

¹*Department of Physics, University of California, Berkeley, Berkeley, CA, 94720*

²*Redwood Center For Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA, 94720*

³*Department of Neuroscience, University of California, Berkeley, Berkeley, CA, 94720*

(Dated: April 29, 2024)

I. OPTIMAL TRANSPORT FORMULATION OF OPTIMAL, WORK-MINIMIZING PROTOCOLS

We present a concise derivation the optimal transport formulation of minimum-work protocols. For simplicity here we use the notation $\rho_t = \rho(\cdot, t)$ and $U_t = U_{\lambda(t)}(\cdot)$. For now, we consider the initial conditions $\rho_0 = \rho_i^{\text{eq}}$ and $U_0 = U_i$, and the terminal condition $U_\tau = U_f$, *without* imposing $\rho_\tau = \rho_f^{\text{eq}}$.

Recall that the Fokker-Planck equation giving the time-evolution for ρ_t may be written as the continuity equation

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla \phi_t), \quad \text{where} \quad \phi_t := U_t + \ln \rho_t. \quad (\text{S1})$$

The ensemble work rate is defined as

$$\frac{dW}{dt} = \int \rho_t(x) \frac{dU_t(x)}{dt} dx. \quad (\text{S2})$$

By applying the identity $\frac{d}{dt} \int \rho_t U_t dx = \int (\partial_t \rho_t) U_t dx + \int \rho_t (\partial_t U_t) dx$ [S1], algebraic manipulation yields

$$\begin{aligned} \frac{dW}{dt} &= - \int \frac{\partial \rho_t}{\partial t} U_t dx + \frac{d}{dt} \int \rho_t U_t dx \\ &= - \int \frac{\partial \rho_t}{\partial t} (U_t + \ln \rho_t + 1) dx + \frac{d}{dt} \int \rho_t (U_t + \ln \rho_t) dx \\ &= \int \rho_t |\nabla (U_t + \ln \rho_t)|^2 dx + \frac{d}{dt} \int \rho_t (U_t + \ln \rho_t) dx. \end{aligned} \quad (\text{S3})$$

Here in the second line we have added and subtracted $\frac{d}{dt} \int \rho_t \ln \rho_t dx = \int (\partial_t \rho_t) (\ln \rho_t + 1) dx$, and in the third line we have plugged in the Fokker-Planck equation Eq. (S1) and integrated by parts in x .

Finally, the work is defined as the time-integral of the work rate

$$W = \int_0^\tau \frac{dW}{dt} dt = \int_0^\tau \int \rho_t |\nabla \phi_t|^2 dx dt + \int \rho_t (U_t + \ln \rho_t) dx \Big|_{t=0}^\tau, \quad (\text{S4})$$

where we have used the definition of ϕ_t in Eq. (S1).

After noting that the KL divergence from a distribution ρ to another an equilibrium distribution ρ_λ^{eq} can be equivalently written as

$$D_{\text{KL}}(\rho | \rho_\lambda^{\text{eq}}) := \int \rho(x) \ln \left[\frac{\rho(x)}{\rho_\lambda^{\text{eq}}(x)} \right] dx = \int \rho (\ln \rho + U_\lambda) dx - F(\lambda) \quad (\text{S5})$$

(recall $F(\lambda) := - \ln \int \exp(-U_\lambda(x)) dx$ is the equilibrium free energy), the excess work $W_{\text{ex}} = W - \Delta F$ can be written as

$$W_{\text{ex}} = \int_0^\tau \int \rho_t |\nabla \phi_t|^2 dx dt + D_{\text{KL}}(\rho_\tau | \rho_f^{\text{eq}}), \quad (\text{S6})$$

*Electronic address: adrzhong@berkeley.edu

where we have applied the boundary conditions $\rho_0 = \rho_i^{\text{eq}}, U_0 = U_i$, and $U_\tau = U_f$.

The first term in Eq. (S6) represents the dissipation occurring within the protocol from $t = 0$ and $t = \tau$ [S2] and is optimized by a Benamou-Brenier solution between ρ_0 and ρ_τ , while the second term represents the dissipation occurring for $t > \tau$ (i.e., after the protocol) as the distribution relaxes from ρ_τ to $\rho_\infty = \rho_f^{\text{eq}}$. The change of variables $s = t/\tau$ thus yields $W_{\text{ex}}^* = \min_{\rho_f} \mathcal{W}_2^2[\rho_i^{\text{eq}}, \rho_f]/\tau + D_{\text{KL}}(\rho_f|\rho_f^{\text{eq}})$, which was noted in [S3] to be equivalent to the JKO scheme used to study convergence properties of the Fokker-Planck equation [S4] with effective time-step $h = \Delta t/2$.

If the additional terminal condition $\rho_\tau = \rho_f^{\text{eq}}$ is additionally imposed, then the second KL divergence term goes away, reproducing the Benamou-Brenier cost function multiplied by $1/\tau$.

II. NUMERICAL STUDY OF LINEARLY-BIASED DOUBLE WELL OPTIMAL PROTOCOLS

In this section, we provide details of our numerical study of the linearly-biased double well

$$U_\lambda(x) = E_0[(x^2 - 1)^2/4 - \lambda x], \quad (\text{S7})$$

for $E_0 = 16, \lambda_i = -1, \lambda_f = 1$, and analyze the reduction in performance around $\tau \sim 2$ seen in Fig S1(a) (a higher resolution version of Fig. 2(e)). We propose an explanation for the geodesic-counterdiabatic protocols overshooting the optimal protocols (Fig S2(b)) that correspond to a reduction in performance for $\tau \sim 2$.

A. Lattice discretization implementation

In order to calculate $g(\lambda), h(\lambda), \mathcal{T}^2(\lambda_i, \lambda), D_{\text{KL}}(\rho_\lambda^{\text{eq}}|\rho_f^{\text{eq}})$, as well as to measure the performance $W_{\text{ex}}[\lambda(t)]$, we base our numerical implementation on the lattice-discretized Fokker-Planck equation method introduced in [S5]. This is the same discretization scheme used by [S6], which allows a direct comparison of numerical results.

We discretize the one-dimensional configuration space as an N -state lattice with spacing Δx and reflecting boundaries at $x_b = \pm(N-1)\Delta x/2$. Following [S6], we use $\Delta x = 0.025$ and $x_b = \pm 3$. The probability density may be represented as a vector $\boldsymbol{\rho}(t) = (\rho^1(t), \rho^2(t), \dots, \rho^N(t))$ via $\rho(x, t) = (\Delta x)^{-1}[\boldsymbol{\rho}(t)]^{l(x)}$, where $l(x) = \lfloor x/\Delta x + N/2 \rfloor$. Likewise, the potential energy may be represented as a covector $\mathbf{U}_\lambda = (U_1(\lambda), U_2(\lambda), \dots, U_N(\lambda))^T$ with $U_\lambda(x) = [\mathbf{U}_\lambda]_{l(x)}$, yielding the equilibrium probability vector $\boldsymbol{\rho}_\lambda^{\text{eq}} = Z(\lambda)^{-1}(\exp(-U_1(\lambda)), \exp(-U_2(\lambda)), \dots, \exp(-U_N(\lambda)))$, where $Z(\lambda) = \sum_i \exp(-U_i(\lambda))$ is the normalization constant. The excess conjugate force is given by the covector

$$\delta \mathbf{f}(\lambda) = -\frac{\partial \mathbf{U}_\lambda}{\partial \lambda} + \frac{\partial \mathbf{U}_\lambda}{\partial \lambda} \cdot \boldsymbol{\rho}_\lambda^{\text{eq}}, \quad (\text{S8})$$

where the second term is a dot product with the equilibrium probability vector.

The Fokker-Planck equation is represented as the master equation

$$\dot{\boldsymbol{\rho}} = \mathcal{L}_\lambda \boldsymbol{\rho}, \quad (\text{S9})$$

where \mathcal{L}_λ is a transition rate matrix on which we impose the following form

$$[\mathcal{L}_\lambda]_{ij}^i = \begin{cases} (\Delta x)^{-2} e^{(U_j(\lambda) - U_i(\lambda))/2} & |i - j| = 1 \\ -(\Delta x)^{-2} \sum_{k \neq j} e^{(U_j(\lambda) - U_k(\lambda))/2} & i = j \\ 0 & \text{else.} \end{cases} \quad (\text{S10})$$

Taking the continuum limit $N \rightarrow \infty, \Delta x \rightarrow 0$ with constant x_b yields the Fokker-Planck equation $\partial_t \rho = \mathcal{L}_\lambda \rho$ [S6].

For this lattice discretization, The friction tensor and Fisher information metric are given by

$$g(\lambda) = -\sum_{i=1}^N [\boldsymbol{\rho}_\lambda^{\text{eq}}]^i [\delta \mathbf{f}]_i [(\mathcal{L}_\lambda^T)^{-1} \delta \mathbf{f}]_i \quad \text{and} \quad h(\lambda) = \sum_{i=1}^N [\boldsymbol{\rho}_\lambda^{\text{eq}}]^i [\delta \mathbf{f}]_i [\delta \mathbf{f}]_i, \quad (\text{S11})$$

where $(\mathcal{L}_\lambda^T)^{-1}$ is the inverse of the adjoint Fokker-Planck matrix with its zero-mode removed [S7]. The KL-divergence is

$$D_{\text{KL}}(\rho_\lambda^{\text{eq}}|\rho_f^{\text{eq}}) = \sum_{i=1}^N [\boldsymbol{\rho}_\lambda^{\text{eq}}]^i \ln \left[\frac{[\boldsymbol{\rho}_\lambda^{\text{eq}}]^i}{[\boldsymbol{\rho}_f^{\text{eq}}]^i} \right], \quad (\text{S12})$$

and the squared thermodynamic distance (cf., Eq. (5) in [S8]) takes the form

$$\mathcal{T}^2(\lambda_i, \lambda) = \left[\int_{\lambda_i}^{\lambda} \sqrt{g(\lambda)} d\lambda \right]^2. \quad (\text{S13})$$

We numerically compute the integral as a trapezoid sum on interval $[\lambda_i, \lambda]$ split into 1000 even subintervals.

Geodesic-counterdiabatic protocols of duration τ are obtained by first solving for γ_f through minimizing

$$\gamma_f = \arg \min_{\lambda} \mathcal{T}^2(\lambda_i, \lambda)/\tau + D_{\text{KL}}(\rho_{\lambda}^{\text{eq}}|\rho_f^{\text{eq}}) \quad (\text{S14})$$

using Eq. (S12) and (S13) above, and then computing the time-discretized geodesic for $k = 0, 1, 2, \dots, K$

$$\gamma(\cdot) = \{(s_0 = 0, \gamma_0 = \lambda_i), (s_1, \gamma_1), (s_2, \gamma_2), \dots, (s_{K-1}, \gamma_{K-1}), (s_K = 1, \gamma_K = \gamma_f)\} \quad (\text{S15})$$

using equally spaced $\gamma_k = [1 - (k/K)]\lambda_i + (k/K)\gamma_f = \lambda_i + k\Delta\gamma$, and variable timesteps $s_{k+1} - s_k = \alpha\Delta\gamma\sqrt{g(\gamma_k) + g(\gamma_{k+1})}/2$ with scaling factor α chosen so that $s_K = 1$. Finally, the time-discretized $\eta(\cdot)$ is obtained via

$$\eta(s) = h^{-1}(\gamma(s))g(\gamma(s))\frac{d\gamma(s)}{ds} \quad (\text{S16})$$

where $d\gamma(s)/ds$ is computed using finite differences.

Following [S6], we use $K = 1000$. This way of computing geodesic-counterdiabatic protocols is consistent with how geodesic protocols were obtained in [S6], allowing for a direct comparison of performance.

B. Phase transition for γ_f

An interesting property we found for this problem is that the geodesic endpoint γ_f obtained as the argmin of the objective function Eq. (S14) has a discontinuity at $\tau \approx 3.3$ (Fig. S2(a)). This is due to the fact that the objective function, computed using Eqs. (S11)-(S13), is not convex, and has multiple local minima for $\tau \sim 3$ (Fig. S2(b)).

C. Measuring performance

For a time-discretized protocol $\lambda(t) = \{(t_0 = 0, \lambda_0 = \lambda_i), (t_1, \lambda_1), \dots, (t_{K-1}, \lambda_{K-1}), (t_K = \tau, \lambda_K)\}$, the trajectory for $\rho(t)$ is obtained by integrating Eq. (S9)

$$\rho_{k+1} = \exp[\mathcal{L}_{k+1/2}(t_{k+1} - t_k)] \rho_k \quad \text{with} \quad \rho_0 = \rho_i^{\text{eq}}, \quad (\text{S17})$$

where $\mathcal{L}_{k+1/2}$ is the transition rate matrix Eq. (S10) for $\lambda_{k+1/2} = (\lambda_k + \lambda_{k+1}/2)$.

Finally, the work for a particular protocol is calculated through the first law of thermodynamics

$$W[\lambda(t)] = \Delta E - Q[\lambda(t)] = [\mathbf{U}_f \cdot \rho_K - \mathbf{U}_i \cdot \rho_i^{\text{eq}}] - \sum_{k=0}^{K-1} \mathbf{U}_{k+1/2} \cdot [\rho_{k+1} - \rho_k], \quad (\text{S18})$$

and the excess work is $W_{\text{ex}} = W[\lambda(t)] - \Delta F$. Because we are considering $\lambda_i = -1$ and $\lambda_f = 1$ for Eq. (S7), we have $\Delta F = 0$.

D. Reduction in performance around $\tau \sim 2$

Though they still outperform geodesic protocols, geodesic-counterdiabatic protocols obtained via Alg. 1 exhibit a noticeable decrease in performance compared with the exact optimal limited-control solutions computed in [S6] using optimal control theory on PDEs. It is important to note the assumptions made for Alg. 1:

1. The complete L^2 optimal transport solution, a Wasserstein geodesic $\rho_s^*|_{s \in [0,1]} \in \mathcal{P}(\mathbb{R}^d)$, is closely approximated by a trajectory of equilibrium distributions $\rho_{\lambda(s)}^{\text{eq}}|_{s \in [0,1]} \in \mathcal{P}_{\mathcal{M}}^{\text{eq}}(\mathbb{R}^d)$, i.e., a thermodynamic geometry geodesic.

2. In the case of limited expressivity of available controls, the continuity equation $\partial_s \rho_{\gamma(s)}^{\text{eq}} = \eta^\mu(s) \nabla \cdot (\rho_{\gamma(s)}^{\text{eq}} \nabla U_\mu)$ is approximately satisfied by the counterdiabatic driving term obtained via Eq. (S16), so that the time-dependent probability distribution $\rho(\cdot, t) \in \mathcal{P}(\mathbb{R}^d)$ solving the Fokker-Planck equation (Eq. (S1)) under the geodesic-counterdiabatic protocol $\lambda^*(t) = \gamma(t/\tau) + \tau^{-1} \eta(t/\tau)$ closely approximates the geodesic path of equilibrium distributions $\rho_{\gamma(t/\tau)}^{\text{eq}} \in \mathcal{P}_{\mathcal{M}}^{\text{eq}}(\mathbb{R}^d)$.

Both of these assumptions hold for the parametric harmonic oscillator, but do not for the linearly-biased double well (Eq. (S7), $E_0 = 16$) for $\tau \sim 2$. In the case that one or both of these assumptions are broken, Alg. 1 may produce a protocol that poorly approximates the true optimal protocol $\lambda^*(t)$. In the case that the first assumption is broken, the global optimal limited-control protocol $\lambda^*(t)$ (i.e., obtained using optimal control on PDEs [S6]) may yield a Fokker-Planck equation solution $\rho^*(\cdot, t)|_{t \in [0, \tau]}$ that better approximates the optimal transport solution $\rho_s^*|_{s \in [0, 1]}$ by not being on the equilibrium manifold, i.e., $\rho^*(\cdot, t)|_{t \in [0, \tau]} \notin \mathcal{P}(\mathbb{R}^d)$.

We believe that this existence of multiple local minima for γ_f (Fig. S2(b)) indicates that near values of this “critical protocol duration” $\tau \approx 3.3$, the first assumption is indeed broken. The objective function yielding the terminal time $\rho_1^* = \rho_f^*$ for the optimal transport solution

$$\rho_f^* = \arg \min_{\rho_f} \mathcal{W}_2^2[\rho_i, \rho_f]/\tau + D_{\text{KL}}(\rho_f | \rho_f^{\text{eq}}) \quad (\text{S19})$$

has only a single minimum, as constructed and proved in [S4], and therefore is unlikely to represent an equilibrium distribution for any choice of control parameter values $\rho_f^* \notin \mathcal{P}_{\mathcal{M}}^{\text{eq}}(\mathbb{R}^d)$. Thus, for $\tau \sim 2$, the optimal transport solution $\rho_s^*|_{s \in [0, 1]}$ is not well approximated by an equilibrium distribution trajectory, and so Alg. 1 does not produce a protocol that approximates the optimal protocol $\lambda^*(t)$ obtained in [S6] (Fig. S1(b)), leading to a decrease in performance (Fig. S1(a)). Nevertheless optimal protocols obtained via Alg. 1 still noticeably outperform the geodesic protocol that connect λ_i to λ_f (Fig. S1(a)).

[S1] Note, this identity is equivalently the First Law of Thermodynamics $dE/dt = dQ/dt + dW/dt$.

[S2] S. Ito, *Information Geometry* **7**, 441 (2024).

[S3] Y. Chen, T. T. Georgiou, and A. Tannenbaum, *IEEE transactions on automatic control* **65**, 2979 (2019).

[S4] R. Jordan, D. Kinderlehrer, and F. Otto, *SIAM journal on mathematical analysis* **29**, 1 (1998).

[S5] V. Holubec, K. Kroy, and S. Steffenoni, *Physical Review E* **99**, 032117 (2019).

[S6] A. Zhong and M. R. DeWeese, *Physical Review E* **106**, 044135 (2022).

[S7] N. S. Wadia, R. V. Zarcone, and M. R. DeWeese, *Physical Review E* **105**, 034130 (2022).

[S8] G. E. Crooks, *Physical Review Letters* **99**, 100602 (2007).

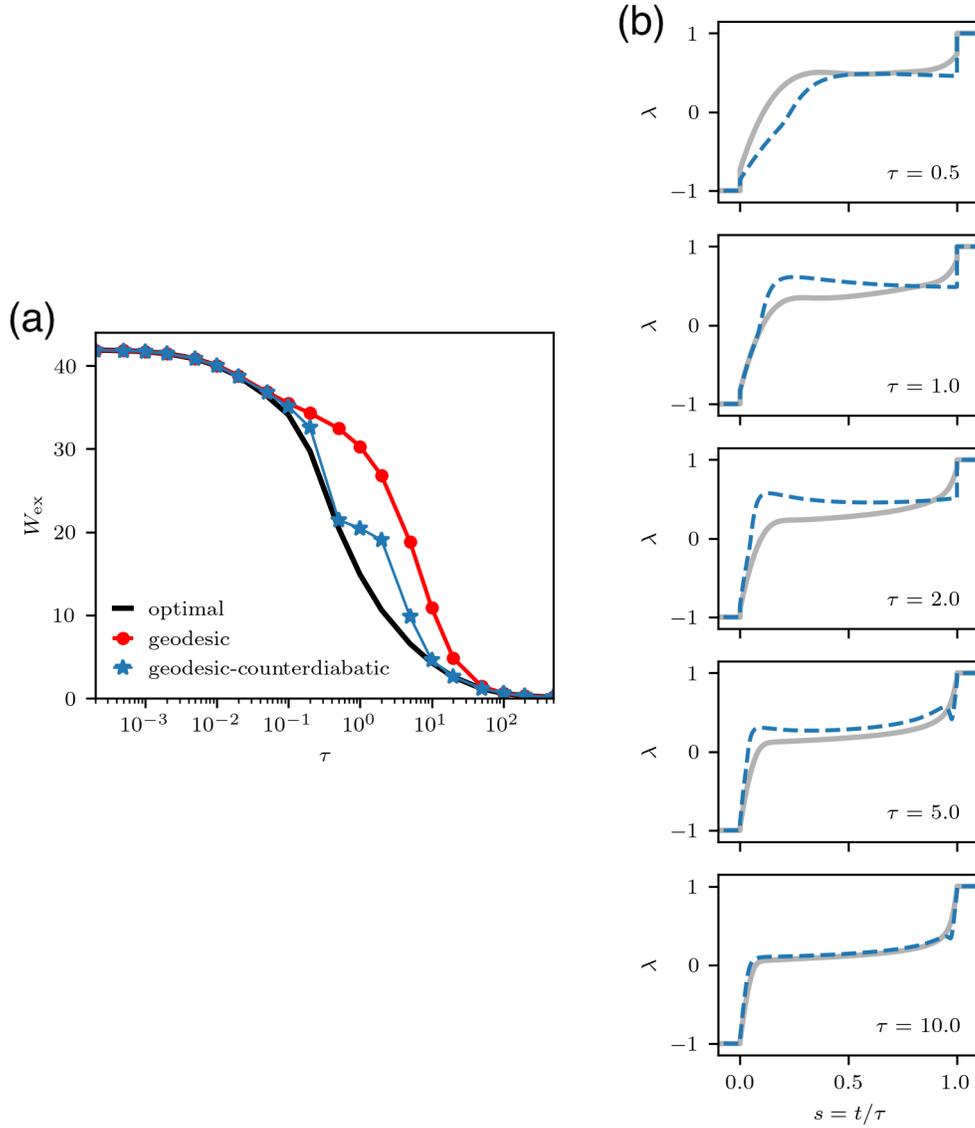


FIG. S1: (a) Performance of optimal protocols from [S6] (black), geodesic (red dots), and geodesic-counterdiabatic protocols (blue stars), higher resolution reproduction of Fig.2(c) from main text. (b) Optimal protocols from [S6] (filled grey line) and geodesic-counterdiabatic protocols from Alg. 1 (broken blue line) for τ between 0.5 to 10.0, corresponding to the decrease in performance “peak” in (a). We see that for $\tau \sim 2$, the geodesic-counterdiabatic protocols overshoot the optimal protocols, corresponding to a reduction in performance.

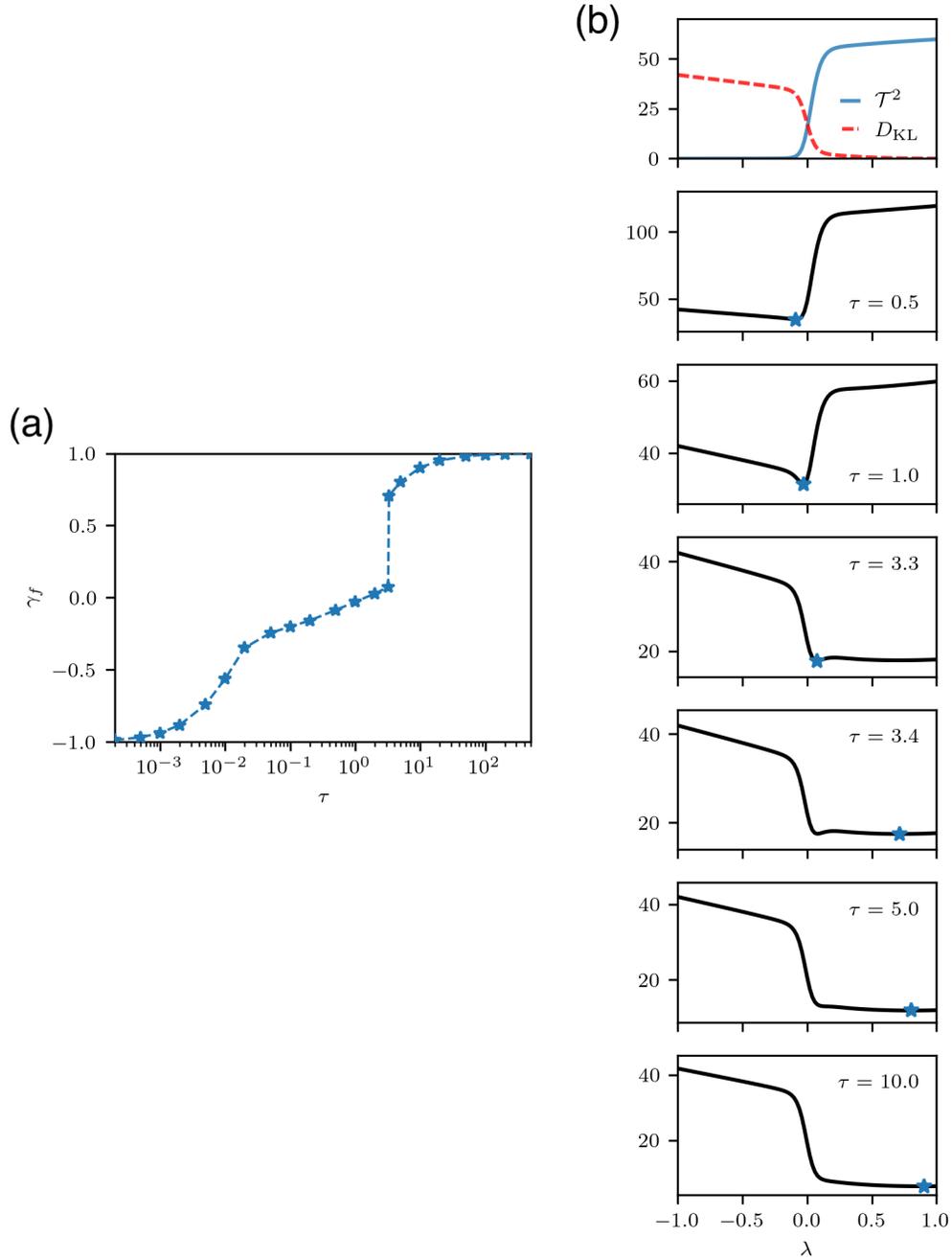


FIG. S2: (a) The geodesic endpoint for geodesic-counterdiabatic protocols $\gamma_f = \arg \min_{\lambda} \mathcal{T}^2(\lambda_i, \lambda)/\tau + D_{\text{KL}}(\rho_{\lambda}^{\text{eq}}|\rho_f^{\text{eq}})$, as a function of protocol duration τ . At $\tau \approx 3.3$ there is a discontinuity from $\gamma_f = 0.073$ to $\gamma_f = 0.706$. (b) The functions $\mathcal{T}^2(\lambda_i, \lambda)$ (solid blue) and $D_{\text{KL}}(\rho_{\lambda}^{\text{eq}}|\rho_f^{\text{eq}})$ (broken red), and their sum $\mathcal{T}^2(\lambda_i, \lambda)/\tau + D_{\text{KL}}(\rho_{\lambda}^{\text{eq}}|\rho_f^{\text{eq}})$ for various values of τ . The blue star indicates the minimum, yielding γ_f , which suddenly jumps from $\gamma_f = 0.073$ at $\tau = 3.3$ to $\gamma_f = 0.706$ at $\tau = 3.4$, due to the existence of multiple local minima in the objective function.