

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

The Geometry of Thermodynamic Control

### Permalink

<https://escholarship.org/uc/item/7tr08578>

### Author

Zulkowski, Patrick

### Publication Date

2014

Peer reviewed|Thesis/dissertation

The Geometry of Thermodynamic Control

By

Patrick Russell Zulkowski

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael DeWeese, Chair

Professor Holger Müller

Professor David Chandler

Spring 2014

The Geometry of Thermodynamic Control

Copyright © 2014

by

Patrick Russell Zulkowski

## Abstract

The Geometry of Thermodynamic Control

by

Patrick Russell Zulkowski

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Michael R. DeWeese, chair

Living systems are distinguished by their self-organization. Given the entropic driving force embodied in the second law of thermodynamics, creating and maintaining such organization requires staying far from equilibrium. Furthermore, since selective advantage may be incurred by energetically-efficient operation, evolution may have sculpted biological components to interact so as to reduce the energy wasted during transitions. Therefore, a deeper understanding of the principles governing biological molecular machines and their synthetic counterparts may be achieved by cultivating a set of tools to explore the optimization of finite-time nonequilibrium transitions of mesoscopic systems.

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. Optimal protocols are equivalent to geodesics in the geometric sense.

We exploit this geometric insight to construct closed-form expressions for minimal-dissipation protocols for a colloidal particle diffusing in a one dimensional harmonic potential. These protocols are verified numerically. We also calculate and numerically verify protocols optimizing the Hatano-Sasa  $Y$ -value (a quantity relevant for transitions between nonequilibrium steady states and similar to dissipated work) for a colloidal particle dragged through solution by a translating optical trap with two controllable parameters. Finally, in an application that has particular relevance to small-scale information processing systems, we calculate maximally efficient erasure cycles for deletion of a single classical bit of information. The system storing this bit consists of an overdamped Brownian colloidal particle diffusing in a one-dimensional double-well potential separated by a potential barrier stabilizing the memory.

## Acknowledgments

This thesis would not have been possible without the guidance, patience and support of Professor Michael DeWeese, my research advisor at Berkeley. Mike has a contagious enthusiasm for science and an unparalleled dedication to his students. I am truly grateful for the opportunity to work with him on the projects forming the basis of this dissertation. I will always consider him one of my great mentors and will look back at the time I had as his student with warmth and gratitude.

I would also like to thank Professors Holger Müller and David Chandler who graciously agreed to take part in the committee overseeing my dissertation. Their input was invaluable in shaping this thesis. My thanks go to Professor Ori Ganor who participated in my qualifying examination along with Professors Müller and Chandler. I also want to express my gratitude towards Professor Holger Müller for the mentorship he provided during the semesters we taught together and for his support beyond the classroom.

Finally, but not least in my mind, I wish to thank my beloved wife Alleen. Without her love, strength and unending optimism I would not have seen the end of this journey. This work is dedicated to her.

*The best is yet to be ...*

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Inverse Diffusion Tensor</b>	<b>5</b>
2.1	Positivity of the average Y-value . . . . .	6
2.2	The geometric approximation . . . . .	8
<b>3</b>	<b>Geometry of Thermodynamic Control</b>	<b>12</b>
3.1	Derivation of the excess power for variable temperature . . . . .	12
3.2	The model system and its inverse diffusion tensor . . . . .	14
3.3	Optimal protocols . . . . .	18
3.4	Computing dissipation numerically . . . . .	22
3.5	The inverse diffusion tensor arises naturally from the Fokker-Planck equation . . .	24
3.6	Discussion . . . . .	24
<b>4</b>	<b>Optimal Control of Nonequilibrium Steady State Transitions</b>	<b>28</b>
4.1	The model system and its inverse diffusion tensor . . . . .	29
4.2	Optimal protocols . . . . .	31
4.3	Optimal straight-line protocols . . . . .	31
4.4	Computing the Y-value numerically . . . . .	33
4.5	The inverse diffusion tensor arises naturally from the Fokker-Planck equation . . .	35
4.6	Discussion . . . . .	36
<b>5</b>	<b>Optimal Finite-Time Erasure Cycles</b>	<b>39</b>
5.1	Introduction . . . . .	39

5.2	Model of classical information erasure . . . . .	40
5.3	Exact optimizer . . . . .	45
5.4	Inverse diffusion tensor-based approximation . . . . .	49
5.5	Beyond erasure . . . . .	53
5.6	Discussion . . . . .	53
<b>6</b>	<b>Inverse Diffusion Tensor for General Potentials</b>	<b>55</b>
6.1	General construction . . . . .	56
6.1.1	Example: Erasure Model . . . . .	60
6.1.2	Example: Ratchet model with optical trap coupling . . . . .	63
6.2	Derivative truncation . . . . .	66
<b>A</b>	<b>Riemannian Geometry</b>	<b>76</b>
A.1	Metric Tensor . . . . .	76
A.2	Curvature Tensor . . . . .	78
A.3	Geodesics . . . . .	81
A.4	Killing Vectors . . . . .	81
<b>B</b>	<b>Fokker-Planck Equations</b>	<b>83</b>
B.1	Derivation of a Fokker-Planck Equation . . . . .	83

# List of Figures

3.1	Model system . . . . .	15
3.2	Optimal protocols . . . . .	22
3.3	Numerical evaluation of geodesic performance . . . . .	25
4.1	Optimal protocols outperform constant- $k$ protocols tested in Ref. [1]. . . . .	34
4.2	Geodesics outperform naive (constant-speed) straight-line paths . . . . .	36
5.1	Erasure model system . . . . .	41
5.2	Optimally efficient erasure cycles . . . . .	52
5.3	Beyond erasure . . . . .	53
A.1	Parallel transport on the sphere . . . . .	79

# Chapter 1

## 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 [2, 3, 4, 5, 6] and experimentally verified [7, 8, 9, 10] 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 [11, 12, 13, 14], have also been established. Interestingly, some of these ideas were independently developed in parallel within the machine learning community [15], as ideas from nonequilibrium statistical mechanics are increasingly finding applications to learning and inference problems [16, 17].

A unifying picture of nonequilibrium statistical mechanics remains elusive due to the broad range of nonequilibrium phenomena. In this thesis, we narrow our focus considerably to a specific optimization scheme for nonequilibrium transitions of small-scale systems with an eye towards applications in two very important (and possibly related) realms of nonequilibrium physics: living systems and small-scale information processing systems.

Living systems are distinguished by their self-organization. Given the entropic driving force embodied in the second law of thermodynamics, creating and maintaining such organization requires staying far from equilibrium [18], typically by coupling to nonequilibrium gradients. For example, ATP-driven molecular motors (*e.g.*, kinesin) are forced away from equilibrium by cellular maintenance of a chemical potential difference between ATP and ADP [19], and the rotary  $F_0$ - $F_1$  ATP synthase operates out of equilibrium due to cellular maintenance of an electrochemical gradient across the inner mitochondrial membrane [20]. That these molecular-scale machines typically operate out of thermodynamic equilibrium results in a major obstacle to quantitative understanding of their thermodynamics.

Since selective advantage may be incurred by energetically-efficient operation, evolution may

have sculpted biological components to interact so as to reduce the energy wasted during transitions. Therefore, a deeper understanding of the principles governing biological molecular machines and their synthetic counterparts may be achieved by cultivating a set of tools to explore the optimization of finite-time nonequilibrium transitions of mesoscopic systems.

Optimization schemes for thermodynamic processes occurring in finite time will be needed for applications in which energetic or entropic costs are undesirable [21, 22] and are not limited to purely small-scale biological problems. Another important class of such processes consists of mesoscopic information processing systems operating out of equilibrium. Optimization will aid technological development in the decades to come as computational demands begin to reach limits imposed by physical law [23].

Moreover, understanding these systems will provide insight into the foundations of nonequilibrium statistical mechanics. Investigations into the interplay between information and thermodynamics seem to have originated with Maxwell's hypothetical demon and its implications for the second law of thermodynamics [24]. Much ground-breaking work followed from the Maxwell demon paradox including Szilard's engine revealing a quantitative link between thermodynamic work and information [25], Landauer's observation of the physical nature of information [26] and Bennett's interpretation of the paradox in terms of the relation between logical and thermodynamic reversibility [27].

In recent times, research into nonequilibrium statistical mechanics of small-scale systems has shed more light on the thermodynamic role of information. Most notable is experimental verification [28] of the theoretical prediction of microscopic violations of Landauer's principle with the preservation of the principle on average [29], analogous to experimental and theoretical work on fluctuation theorems demonstrating that entropy absorbing processes can occur microscopically whereas the second law holds on average [30]. Research into feedback and measurement of mesoscopic nonequilibrium systems has improved our understanding of the role information plays in the second law [31, 32]. Other work has focused on developing techniques to optimize thermodynamic quantities arising in small-scale systems designed to store and erase classical information [33, 34, 35], including the derivation of a refined second law [35]. Recent work has also focused on the general problem of predicting optimal protocols to drive systems between stationary states with minimal dissipation [36, 37, 38, 39, 40, 41, 42, 43].

This discussion highlights the need for the development of optimization techniques applicable to small-scale nonequilibrium systems. 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 [44, 45, 46, 47, 48, 49], with extensions to microscopic systems involving a metric of Fisher information [50, 51]. 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 inverse

diffusion tensor induces a Riemannian manifold structure on the space of parameters, and optimal protocols trace out geodesics of this inverse diffusion tensor [52]. Ch. 2 of this thesis reproduces an argument for the general construction of this generalized inverse diffusion tensor, making available the mathematical details to the interested reader.

In Ch. 3, we make use of Riemannian geometry theorems and the groundwork laid down in Ch. 2 to calculate optimal protocols for a specific physical system consisting of a colloidal particle diffusing in a one dimensional harmonic potential, where the spring constant, inverse temperature, and trap location are adjusted simultaneously. That we can calculate closed-form expressions for optimal protocols for this model system illustrates the power of these geometric ideas. 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.

For constant ATP and ADP concentrations, or constant membrane potential, the dynamics of an ensemble of molecular motors will approximate a nonequilibrium steady state (NESS). Thus, biological systems are often better characterized as nonequilibrium steady states rather than equilibrium systems. Such NESS may change in response to changing environmental conditions and so, if we are to eventually see applications of this geometric framework to faithful models of biological processes, we must seek extensions to systems relaxing to stationary state distributions. Indeed, the study of NESS promises greater mathematical tractability than arbitrary nonequilibrium phenomena, while significantly relaxing the restrictive assumptions of equilibrium physics.

We take a first step towards this goal by optimizing the Hatano-Sasa Y-value, a quantity similar to dissipated work, for the paradigmatic model system tested in [1] and analyzed in [53] with an eye towards experimental tests. In Ch. 4, we calculate closed-form expressions for both the geodesic optimal protocol and the optimal straight-line protocol and test these protocols numerically via a system of equations derived from the Fokker-Planck equation. Finally, we propose a regime of validity of our approximation based on this numerical work. By measuring the average work required to drive this system along either optimal or naive paths through control parameter space, our results can be tested experimentally in a straightforward way using existing experimental techniques.

In Ch. 5, we optimize the efficiency of erasure cycles for a simple model system storing a classical bit of information. The system storing this bit consists of an overdamped Brownian colloidal particle diffusing in a one-dimensional double-well potential separated by a potential barrier stabilizing the memory. Information erasure inevitably leads to heat dissipation according to the Landauer principle. Minimizing this dissipation will be crucial for the development of small-scale information processing systems as alluded to above. We take as control parameters the height of the potential barrier and the potential difference of the two wells. We demonstrate close agreement between the exact optimal cycle and the protocol found using a linear-response framework.

Ch. 6 illustrates how the inverse diffusion tensor may be calculated and applied to systems influenced by non-linear deterministic forces. The paradigmatic models used in Ch. 3 and Ch. 4 involve a colloidal particle diffusing in a thermal bath while coupled to an optical trap represented by a harmonic potential. For the harmonic potential (and only for this potential), the inverse diffusion tensor may be calculated straightforwardly from the Langevin dynamics of the model system. This technique does not carry over to more general potentials. In Ch. 6, we show how the method of Laplace transform may be used to obtain an explicit expression for the inverse diffusion tensor directly from the Fokker-Planck equation. With this expression we compute optimal protocols for some idealized models to further illustrate the power of this approximation in determining optimal protocols.

For completeness, we include a chapter (Ch. A) on the essential notions of Riemannian geometry relevant to this thesis. This chapter is not intended to be a complete treatise of Riemannian geometry which is a firmly established branch of pure mathematics with a long and colorful history. Readers acquainted with General Relativity and/or String Theory should have enough experience with geometry to make the arguments presented in this thesis comprehensible. A firm grounding in nonequilibrium statistical mechanics may be achieved by reading [54]. However, we present a short discussion of the Fokker-Planck equation in Ch. B.

# Chapter 2

## The Inverse Diffusion Tensor

For the model systems considered in this thesis, we wish to minimize the nonequilibrium average of a quantity sometimes referred to as the Hatano-Sasa Y-value [6, 1] while driving a system from either an equilibrium state or a nonequilibrium steady state. During the driving process, 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  $\Lambda$ . We assume the protocol to be sufficiently smooth to be twice-differentiable. The Y-value for a given stochastic trajectory  $x(t)$  is defined as

$$Y \equiv \int_0^\tau dt \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(x(t), \boldsymbol{\lambda}(t)) \quad (2.1)$$

where  $\tau$  is the duration of the protocol,  $\phi \equiv -\ln \rho_{ss}$  and  $\rho_{ss}$  is the time-independent stationary distribution describing the system in the absence of driving after a sufficiently long relaxation time has elapsed. We refer to  $\phi$  as the nonequilibrium potential.

For systems relaxing to an equilibrium state described by Boltzmann's distribution, the Y-value equals the work dissipated during the driving process (scaled by  $\beta$ ). Suppose the ambient temperature is constant during the driving process; the more general case of time-dependent  $\beta$  is considered in Ch. 3. Suppose further that  $E(x, \boldsymbol{\lambda})$  represents the energy of a microstate of the system indexed by the continuous symbol  $x$  (e.g. one spatial dimension). The Boltzmann distribution is given by

$$\rho_{eq}(x, \boldsymbol{\lambda}) \equiv \frac{e^{-\beta E(x, \boldsymbol{\lambda})}}{Z(\boldsymbol{\lambda})} \quad (2.2)$$

where  $Z(\boldsymbol{\lambda}) = \int dx e^{-\beta E(x, \boldsymbol{\lambda})}$  and  $\beta \equiv (k_B T)^{-1}$  with  $k_B$  Boltzmann's constant. By definition of the nonequilibrium potential,

$$\phi(x, \boldsymbol{\lambda}) = -\ln \rho_{eq}(x, \boldsymbol{\lambda}) = \beta E(x, \boldsymbol{\lambda}) + \ln Z(\boldsymbol{\lambda}). \quad (2.3)$$

From

$$\frac{\partial}{\partial \boldsymbol{\lambda}} \ln Z(\boldsymbol{\lambda}) = \frac{1}{Z(\boldsymbol{\lambda})} \int dx \frac{\partial}{\partial \boldsymbol{\lambda}} (-\beta E(x, \boldsymbol{\lambda})) e^{-\beta E(x, \boldsymbol{\lambda})} \equiv \left\langle \frac{\partial}{\partial \boldsymbol{\lambda}} (-\beta E) \right\rangle_{\boldsymbol{\lambda}} \quad (2.4)$$

follows

$$Y = - \int_0^\tau dt \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \delta \mathbf{X}(t) \quad (2.5)$$

where we define

$$\mathbf{X}(t) = - \frac{\partial}{\partial \boldsymbol{\lambda}} (\beta E(x(t), \boldsymbol{\lambda}(t))) \quad (2.6)$$

and  $\delta \mathbf{X}(t) \equiv X(t) - \langle X \rangle_{\boldsymbol{\lambda}(t)}$ . The average  $\langle X \rangle_{\boldsymbol{\lambda}(t)}$  is taken with respect to the Boltzmann distribution characterized by the control parameters  $\boldsymbol{\lambda}$  at time  $t$ . Taking the nonequilibrium average over stochastic trajectories yields the average amount of work done on the system during the driving times  $\beta$  minus  $\beta F(\boldsymbol{\lambda}(\tau)) - \beta F(\boldsymbol{\lambda}(0))$ ; *i.e.* the average work dissipated scaled by  $\beta$ .

The remainder of this chapter establishes the general optimization scheme. First, we provide a heuristic argument for the positivity of the average Y-value. This should convince the reader that the optimization problem is well-posed for the applications we have in mind. Second, we describe in detail a linear response approximation originally presented in [52] and further elaborated in [37] that allows us to reinterpret the problem of optimizing the average Y-value over admissible protocols as the problem of computing geodesics on a Riemannian manifold.

## 2.1 Positivity of the average Y-value

General arguments based on a fluctuation theorem exist to establish the positivity of the average Y-value for nonequilibrium driving processes [6, 55, 30, 1]. In general,  $\langle e^{-Y} \rangle_{\Lambda} = 1$  and so, by Jensen's inequality,  $\langle Y \rangle_{\Lambda} \geq 0$ . However, a detour exploring the details of this theorem will take us too far afield. To keep the exposition as clear and as self-contained as possible, we provide a proof in the simple case of an overdamped colloidal particle diffusing in one dimension under the influence of a time-dependent potential  $U(x, \boldsymbol{\lambda}(t))$ . We assume the stationary state is given by the Boltzmann distribution. The proof is by no means limited to this special case as we will see in Ch. 3. However, it captures the essential physics without delving too deeply into complicated mathematics.

The nonequilibrium ensemble average of the Y-value is defined as

$$\langle Y \rangle_{\Lambda} \equiv \int_0^\tau dt \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \left\langle \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} \quad (2.7)$$

where, for the system considered here,

$$\left\langle \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} \equiv \int_{-\infty}^{\infty} dx \rho(x, t) \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(x, \boldsymbol{\lambda}(t)) \quad (2.8)$$

and the nonequilibrium distribution  $\rho(x, t)$  is governed by the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{\beta \gamma} \frac{\partial}{\partial x} \left[ \beta \partial_x U(x, \boldsymbol{\lambda}(t)) \rho + \frac{\partial \rho}{\partial x} \right]. \quad (2.9)$$

The constant  $\gamma$  is the Cartesian friction coefficient. We assume the bath temperature to be constant; this is not a limitation of the technique, merely a mathematical convenience.

Consider now the relative entropy  $D[\rho \parallel \rho_{eq}(x, \boldsymbol{\lambda}(t))]$  which corresponds to the available energy in the system due to being out of equilibrium [56] and is defined by

$$D[\rho \parallel \rho_{eq}(x, \boldsymbol{\lambda}(t))] \equiv \int_{-\infty}^{\infty} dx \rho(x, t) \ln \left( \frac{\rho(x, t)}{\rho_{eq}(x, \boldsymbol{\lambda}(t))} \right) \quad (2.10)$$

with

$$\rho_{eq}(x, \boldsymbol{\lambda}) = \frac{e^{-\beta U(x, \boldsymbol{\lambda})}}{Z(\boldsymbol{\lambda})}. \quad (2.11)$$

The time derivative of the relative entropy is

$$\begin{aligned} \frac{d}{dt} D[\rho \parallel \rho_{eq}(x, \boldsymbol{\lambda}(t))] &= \int_{-\infty}^{\infty} dx \frac{\partial \rho}{\partial t} \ln \left( \frac{\rho(x, t)}{\rho_{eq}(x, \boldsymbol{\lambda}(t))} \right) + \int_{-\infty}^{\infty} dx \rho(x, t) \left[ \frac{1}{\rho(x, t)} \frac{\partial \rho}{\partial t} + \frac{\partial \phi}{\partial t} \right] \\ &= \int_{-\infty}^{\infty} dx \frac{\partial \rho}{\partial t} \ln \left( \frac{\rho(x, t)}{\rho_{eq}(x, \boldsymbol{\lambda}(t))} \right) + \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \left\langle \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda}. \end{aligned} \quad (2.12)$$

The first term may be simplified by integrating by parts and using the Fokker-Planck equation:

$$\begin{aligned} \int_{-\infty}^{\infty} dx \frac{\partial \rho}{\partial t} \ln \left( \frac{\rho(x, t)}{\rho_{eq}(x, \boldsymbol{\lambda}(t))} \right) &= \int_{-\infty}^{\infty} dx \left\{ \frac{1}{\beta \gamma} \frac{\partial}{\partial x} \left[ \beta \partial_x U(x, \boldsymbol{\lambda}(t)) \rho + \frac{\partial \rho}{\partial x} \right] \right\} \ln \left( \frac{\rho(x, t)}{\rho_{eq}(x, \boldsymbol{\lambda}(t))} \right) \\ &= -\frac{1}{\beta \gamma} \int_{-\infty}^{\infty} dx \left[ \beta \partial_x U(x, \boldsymbol{\lambda}(t)) \rho + \frac{\partial \rho}{\partial x} \right] \left[ \frac{1}{\rho} \frac{\partial \rho}{\partial x} - \frac{1}{\rho_{eq}} \frac{\partial \rho_{eq}}{\partial x} \right] \\ &= -\frac{1}{\beta \gamma} \int_{-\infty}^{\infty} dx \left[ \beta \partial_x U(x, \boldsymbol{\lambda}(t)) \rho + \frac{\partial \rho}{\partial x} \right]^2 \frac{1}{\rho} \\ &\leq 0. \end{aligned} \quad (2.13)$$

Therefore, it follows immediately that

$$\int_0^\tau dt \frac{d}{dt} D[\rho \parallel \rho_{eq}(x, \boldsymbol{\lambda}(t))] \leq \langle Y \rangle_{\Lambda}. \quad (2.14)$$

Since we assume that our system begins in equilibrium,  $D[\rho \parallel \rho_{eq}(x, \boldsymbol{\lambda}(0))] = 0$ . Furthermore, since the relative entropy is always nonnegative, we conclude that the average  $Y$ -value is positive for the nonequilibrium driving process.

## 2.2 The geometric approximation

The material in this section is largely based on the arguments of [52]. However, we will fill in the mathematical details for the reader.

By definition,

$$\begin{aligned} \left\langle \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} &\equiv \int dx \rho(x, t) \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \\ &= \int dx_0 \rho_{ss}(x_0, \boldsymbol{\lambda}(0)) \int dx \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \rho(x, t | x_0, 0). \end{aligned} \quad (2.15)$$

Abusing notation, we define

$$\frac{\partial \phi}{\partial \lambda^i}(t) \equiv \int dx \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \rho(x, t | x_0, 0). \quad (2.16)$$

In this notation we suppress the dependence on the initial condition  $x_0$ . We may use dynamic linear response theory to first order in the magnitude of the externally imposed protocol to show

$$\left\langle \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t_0)) \right\rangle_{\Lambda} \approx \int_{-\infty}^{t_0} dt' \chi_{ij}(t_0 - t') [\lambda^j(t') - \lambda^j(t_0)] \quad (2.17)$$

where  $\chi$  is referred to as the linear response kernel [52, 54].

We may obtain an expression for the linear response kernel  $\chi_{ij}$  as follows: consider a system in a stationary state from time  $-\infty$  to time 0 characterized by control parameters  $\boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}$ . At time  $t = 0$ , assume the perturbation vanishes so that the control parameters have values  $\boldsymbol{\lambda}$ . For times  $t > 0$ , we see from Eq. (2.15) that

$$\left\langle \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} = \int dx_0 \frac{\partial \phi}{\partial \lambda^i}(t) \rho_{ss}(x_0, \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}). \quad (2.18)$$

Approximating  $\rho_{ss}(x_0, \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}) \approx \rho_{ss}(x_0, \boldsymbol{\lambda}) [1 - [\Delta\boldsymbol{\lambda}^T] \cdot \frac{\partial\phi}{\partial\boldsymbol{\lambda}}(x_0, \boldsymbol{\lambda})]$ , we see that

$$\left\langle \frac{\partial\phi}{\partial\lambda^i}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} \approx \int dx_0 \rho_{ss}(x_0, \boldsymbol{\lambda}) \left[ 1 - [\Delta\boldsymbol{\lambda}^T] \cdot \frac{\partial\phi}{\partial\boldsymbol{\lambda}}(x_0, \boldsymbol{\lambda}) \right] \frac{\partial\phi}{\partial\lambda^i}(t). \quad (2.19)$$

The first term vanishes since  $\rho_{ss}(x_0, \boldsymbol{\lambda})$  must be normalized.

Therefore, we find that

$$\left\langle \frac{\partial\phi}{\partial\lambda^i}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} \approx -\Delta\lambda^j \cdot \left\langle \frac{\partial\phi}{\partial\lambda^i}(t) \frac{\partial\phi}{\partial\lambda^j}(0) \right\rangle_{ss, \boldsymbol{\lambda}} \quad (2.20)$$

for this simple example.

If the linear response approximation holds in general, then it must reproduce this simple case. Consider the right-hand side of Eq. (2.17) and assume  $t_0 > 0$ . The integral may be broken up into the integral over  $t'$  from  $-\infty$  to 0 and the integral from 0 to  $t_0$ . In the latter integral,  $\lambda^j(t') - \lambda^j(t_0)$  vanishes and so this integral does not contribute. In the former integral,  $\lambda^j(t') - \lambda^j(t_0) = \Delta\lambda^j$  and may be brought outside the time integral. Making the change of variables  $s = t_0 - t'$ , we see that

$$\left\langle \frac{\partial\phi}{\partial\lambda^i}(x, \boldsymbol{\lambda}(t_0)) \right\rangle_{\Lambda} \approx \Delta\lambda^j \int_{t_0}^{\infty} ds \chi_{ij}(s). \quad (2.21)$$

Therefore, it follows that

$$\int_{t_0}^{\infty} ds \chi_{ij}(s) = - \left\langle \frac{\partial\phi}{\partial\lambda^i}(t_0) \frac{\partial\phi}{\partial\lambda^j}(0) \right\rangle_{ss, \boldsymbol{\lambda}}. \quad (2.22)$$

If the time correlation functions vanish for long times, we see that

$$\chi_{ij}(s) \equiv \frac{d}{ds} \left\langle \frac{\partial\phi}{\partial\lambda^i}(s) \frac{\partial\phi}{\partial\lambda^j}(0) \right\rangle_{ss, \boldsymbol{\lambda}(t_0)}. \quad (2.23)$$

Substituting this expression for the linear response kernel into Eq. (2.17), we have

$$\begin{aligned}
\left\langle \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t_0)) \right\rangle_{\Lambda} &\approx \int_{-\infty}^{t_0} dt' \chi_{ij}(t_0 - t') [\lambda^j(t') - \lambda^j(t_0)] \\
&= \int_0^{\infty} ds \chi_{ij}(s) [\lambda^j(t_0 - s) - \lambda^j(t_0)] \\
&= \int_0^{\infty} ds \frac{d}{ds} \left\langle \frac{\partial \phi}{\partial \lambda^i}(s) \frac{\partial \phi}{\partial \lambda^j}(0) \right\rangle_{ss, \boldsymbol{\lambda}(t_0)} [\lambda^j(t_0 - s) - \lambda^j(t_0)] \\
&= - \int_0^{\infty} ds \left\langle \frac{\partial \phi}{\partial \lambda^i}(s) \frac{\partial \phi}{\partial \lambda^j}(0) \right\rangle_{ss, \boldsymbol{\lambda}(t_0)} \frac{d}{ds} (\lambda^j(t_0 - s)) \\
&= \int_0^{\infty} ds \left\langle \frac{\partial \phi}{\partial \lambda^i}(s) \frac{\partial \phi}{\partial \lambda^j}(0) \right\rangle_{ss, \boldsymbol{\lambda}(t_0)} \frac{d\lambda^j}{dt}(t_0 - s)
\end{aligned} \tag{2.24}$$

When the control parameter velocities change on timescales slower than the relaxation time of the system's force fluctuations [52], we can make the approximation

$$\frac{d\lambda^j}{dt}(t_0 - s) \approx \frac{d\lambda^j}{dt}(t_0). \tag{2.25}$$

In conclusion,

$$\langle Y \rangle_{\Lambda} \approx \int_0^{\tau} dt \zeta_{ij}(\boldsymbol{\lambda}(t)) \frac{d\lambda^i}{dt}(t) \frac{d\lambda^j}{dt}(t) \tag{2.26}$$

with the inverse diffusion tensor defined by

$$\zeta_{ij}(\boldsymbol{\lambda}) \equiv \int_0^{\infty} dt' \left\langle \frac{\partial \phi}{\partial \lambda^i}(t') \frac{\partial \phi}{\partial \lambda^j}(0) \right\rangle_{ss, \boldsymbol{\lambda}}. \tag{2.27}$$

When the detailed balance condition is satisfied, the matrix  $\langle \partial_{\lambda^i} \phi(t') \partial_{\lambda^j} \phi(0) \rangle_{\boldsymbol{\lambda}}$  is a covariance matrix. Any covariance matrix is symmetric and positive-semidefinite [52]. It follows that the inverse diffusion tensor  $\zeta(\boldsymbol{\lambda})$  is symmetric, positive-semidefinite, and smoothly varying except at macroscopic phase transitions. Therefore the inverse diffusion tensor induces a Riemannian manifold structure on the space of thermodynamic states when the detailed balance condition holds. Furthermore, positive-semidefiniteness of the inverse diffusion tensor guarantees that the average  $Y$ -value is non-negative, consistent with the observations made in Ch. 2.1.

In general, the detailed balance condition is violated in NESS and so the matrix  $\langle \partial_{\lambda^i} \phi(t') \partial_{\lambda^j} \phi(0) \rangle_{\boldsymbol{\lambda}}$  may be asymmetric. Eq. (2.26) shows the use of Eq. (2.27) (specifically its symmetric part) as a metric tensor is not precluded. However,  $\langle \partial_{\lambda^i} \phi(t') \partial_{\lambda^j} \phi(0) \rangle_{\boldsymbol{\lambda}}$  is not necessarily a covariance matrix and so a general proof of positive-definiteness is lacking [52]. These considerations do not affect the models considered in this thesis but future work is needed to address this issue for the general case.

We conclude this chapter with a few observations first presented in [52]. First, by interpreting the inverse diffusion tensor  $\zeta(\lambda)$  as a Riemannian metric on the space of parameters, we see that protocols optimizing the average Y-value are equivalent to geodesics, or length minimizing curves, at least in the linear response regime. As we will see in subsequent chapters, this proves to be a very useful observation as the geodesic equation is invariant under coordinate transformations by virtue of the invariance of Eq. (2.26). By choosing an appropriate coordinate system, it is possible to compute exact, closed-form expressions for optimal protocols for the model systems considered in subsequent chapters. Second, the control parameter path of an optimal protocol is independent of the protocol duration. Increasing or decreasing the duration does not change the optimal path in the linear response regime. However, subsequent chapters will demonstrate that the validity of the approximation itself will be sensitive to the duration of the protocol compared to some relaxation time characteristic of the system. Finally, since the length of a tangent vector to a geodesic is constant along the path of the geodesic, the average Y-value along an optimal path accumulates at a constant rate. (See Ch. A for details.)

# Chapter 3

## Geometry of Thermodynamic Control

This chapter expands upon results originally published in [37]. Ch. 2 demonstrated 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 inverse diffusion 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.

### 3.1 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

$$\rho_{eq}(x, \boldsymbol{\lambda}) \equiv \exp \beta [F(\boldsymbol{\lambda}) - E(x, \boldsymbol{\lambda})] , \quad (3.1)$$

where  $\beta = (k_B T)^{-1}$  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  $\boldsymbol{\lambda}$ .

The usual expressions for heat and work [57, 58, 59, 60] 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 [61], and preceding Eq. (4) of [62], we argue that the unitless energy  $\beta E(x, \boldsymbol{\lambda})$  (normalized by the natural scale of equilibrium

thermal fluctuations,  $k_B T = \beta^{-1}$ , 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:

$$\left\langle \frac{d}{dt} \left( \beta E(x, \boldsymbol{\lambda}) \right) \right\rangle_{\boldsymbol{\Lambda}}, \quad (3.2)$$

where angled brackets with subscript indicate a nonequilibrium average dependent on the protocol  $\boldsymbol{\Lambda}$ . For constant  $\beta$ , this reduces to the standard thermodynamic definition [52]. 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. Nonetheless, a deeper understanding of the subtleties involved in our modified energy flow definition (Eq. (3.2)) calls out for further study.

Eq. (3.2) may be written as

$$\left\langle \frac{dx^T}{dt} \cdot \frac{\partial(\beta E)}{\partial x}(x, \boldsymbol{\lambda}) \right\rangle_{\boldsymbol{\Lambda}} + \left\langle \frac{d\boldsymbol{\lambda}^T}{dt} \cdot \frac{\partial(\beta E)}{\partial \boldsymbol{\lambda}}(x, \boldsymbol{\lambda}) \right\rangle_{\boldsymbol{\Lambda}}. \quad (3.3)$$

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

$$\beta(t_0) \mathcal{P}_{\text{ex}}(t_0) \equiv - \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right]_{t_0} \cdot \langle \delta \mathbf{X} \rangle_{\boldsymbol{\Lambda}}. \quad (3.4)$$

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

As observed in the introduction of Ch. 2, the time integral of this average excess power is precisely equal to the average of the Hatano-Sasa Y-value. We employ the machinery developed in Ch. 2 to approximate the resulting mean excess power as

$$\beta(t_0) \mathcal{P}_{\text{ex}}(t_0) \approx \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right]_{t_0} \cdot \boldsymbol{\zeta}(\boldsymbol{\lambda}(t_0)) \cdot \left[ \frac{d\boldsymbol{\lambda}}{dt} \right]_{t_0}, \quad (3.5)$$

for inverse diffusion tensor

$$\zeta_{ij}(\boldsymbol{\lambda}(t_0)) = \int_0^\infty dt' \langle \delta X_j(0) \delta X_i(t') \rangle_{\boldsymbol{\lambda}(t_0)}. \quad (3.6)$$

We will construct geodesics using this inverse diffusion tensor.

## 3.2 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

$$m\ddot{y} + k(t)(y - y_0(t)) + \gamma\dot{y} = F(t), \quad (3.7)$$

for Gaussian white noise  $F(t)$  satisfying

$$\langle F(t) \rangle = 0, \quad \langle F(t)F(t') \rangle = \frac{2\gamma}{\beta(t)}\delta(t - t'). \quad (3.8)$$

Here  $\gamma$  is the Cartesian friction coefficient. We take as our three control parameters: the inverse temperature of the bath  $\beta$ , the location of the harmonic potential minimum  $y_0$ , and the stiffness of the trap  $k$  [see Fig. 3.1(a)]. The conjugate forces are

$$\mathbf{X} = \left( \beta k (y - y_0), -\frac{p^2}{2m} - \frac{k}{2} (y - y_0)^2, -\frac{\beta}{2} (y - y_0)^2 \right). \quad (3.9)$$

This model can be experimentally realized as, for instance, a driven torsion pendulum [63, 64].

The excess work

$$\langle (\beta W)_{\text{ex}} \rangle \equiv \int_{t_a}^{t_b} dt \beta(t) \mathcal{P}_{\text{ex}}(t) \quad (3.10)$$

is non-negative. Assuming the system begins in equilibrium, the relative entropy  $D[f \parallel \rho_{\text{eq}}(x, \boldsymbol{\lambda}(t))]$  corresponds to the available energy in the system due to being out of equilibrium [56], and bounds the excess work from below. Here,  $\rho_{\text{eq}}$  is the equilibrium distribution 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

$$\frac{d}{dt} D[f \parallel \rho_{\text{eq}}(x, \boldsymbol{\lambda}(t))] = \int \frac{\partial f}{\partial t} \log(f/\rho_{\text{eq}}(x, \boldsymbol{\lambda}(t))) + \beta(t) \mathcal{P}_{\text{ex}}(t), \quad (3.11)$$

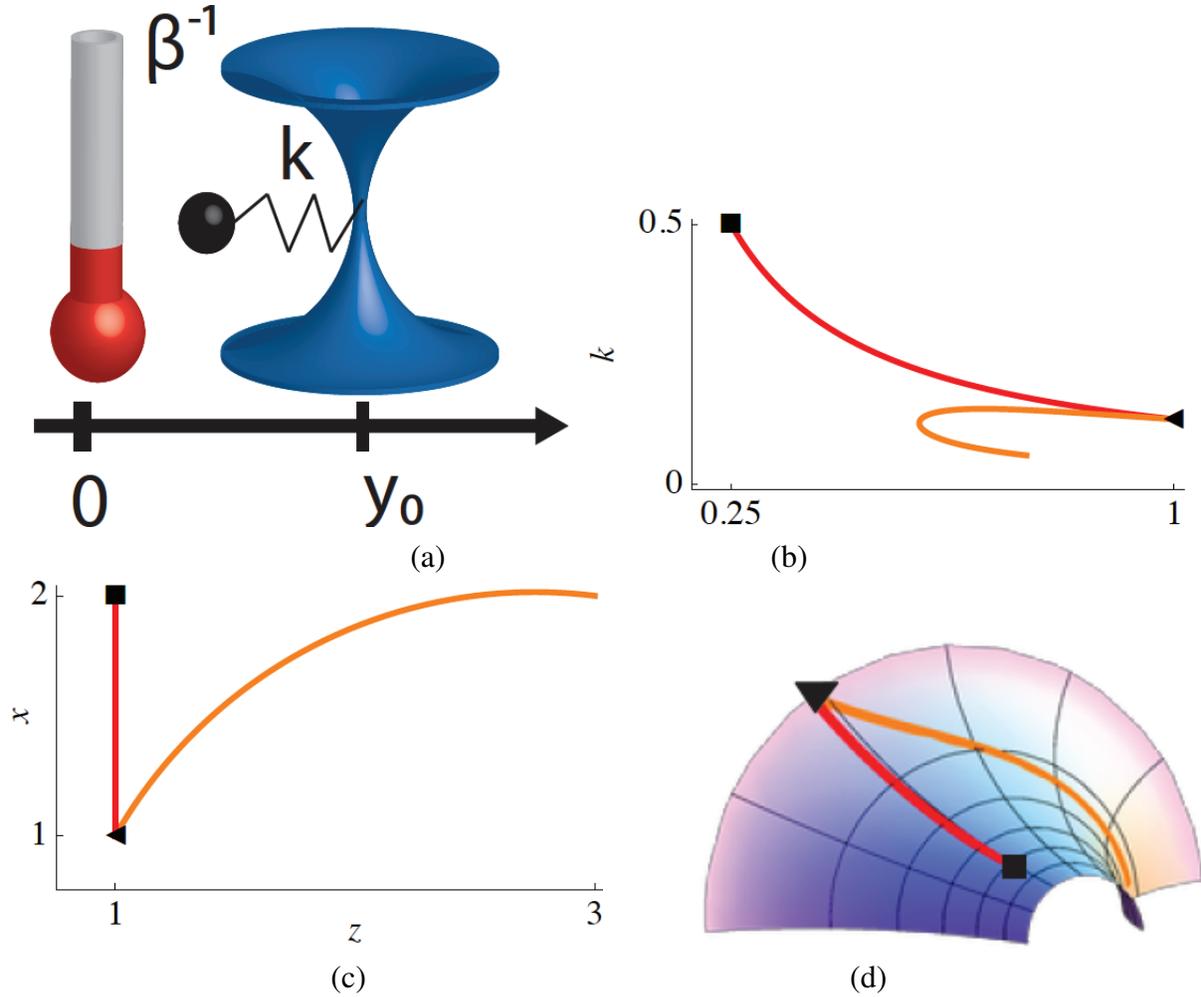


Figure 3.1: **(a)** Our model system. A particle (black dot) diffusing in an optical trap (blue) modeled by a harmonic potential with adjustable spring constant  $k$ , position  $y_0$ , and inverse bath temperature  $\beta = \frac{1}{k_B T}$  (indicated by the thermometer). **(b)** Representative optimal protocols (orange and red curves) plotted for two of the three control parameters,  $k$  and  $\beta$ . 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. (3.28)) 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.

which follows from the identity

$$\int f \left[ \frac{\partial}{\partial t} \log (\rho_{eq}(x, \boldsymbol{\lambda}(t))) \right] = \frac{d\boldsymbol{\lambda}^T}{dt} \cdot \langle \delta \mathbf{X}(t) \rangle_{\boldsymbol{\Lambda}}. \quad (3.12)$$

The first term of Eq. (3.11) simplifies to

$$-\frac{\gamma}{\beta} \int e^{-\frac{\beta p^2}{m}} \left[ \frac{1}{f} \left( \frac{\partial}{\partial p} \left( e^{\frac{\beta p^2}{2m}} f \right) \right)^2 \right] \leq 0. \quad (3.13)$$

Integrating Eq. (3.11) 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 [65] for related results in the special case of constant temperature. Note that, unlike our modified definition for work, the naive definition  $\int_{t_a}^{t_b} dt \mathcal{P}_{ex}(t)$  may be negative for particular protocols that vary  $\beta$ .

Calculation of the time correlation functions in Eq. (3.6) 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 [66]). Explicitly, we may write

$$y_p(t) = \int_0^t \left( \frac{y_h^{(1)}(s)y_h^{(2)}(t) - y_h^{(1)}(t)y_h^{(2)}(s)}{y_h^{(1)}(s)\frac{d}{ds}y_h^{(2)}(s) - y_h^{(2)}(s)\frac{d}{ds}y_h^{(1)}(s)} \right) \frac{F(s)}{m} ds \quad (3.14)$$

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

$$y_h(t) = C_1 y_h^{(1)}(t) + C_2 y_h^{(2)}(t) \quad (3.15)$$

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,

$$\langle \delta y(t)^2 \delta y(0)^2 \rangle = \langle y(t)^2 y(0)^2 \rangle - \langle y(t)^2 \rangle \langle y(0)^2 \rangle, \quad (3.16)$$

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

$$\begin{aligned} \langle \delta y(t)^2 \delta y(0)^2 \rangle &= \langle y_h(t)^2 y(0)^2 \rangle + \langle y_p(t)^2 y(0)^2 \rangle \\ &\quad - \langle y_h(t)^2 \rangle \langle y(0)^2 \rangle - \langle y_p(t)^2 \rangle \langle y(0)^2 \rangle \\ &\quad + 2 \left( \langle y_h(t) y_p(t) y(0)^2 \rangle - \langle y_h(t) y_p(t) \rangle \langle y(0)^2 \rangle \right). \end{aligned} \quad (3.17)$$

Angled brackets denote an average over noise and initial conditions.

According to Eq. (3.14), the particular solution  $y_p$  does not depend on the initial conditions. It follows immediately that

$$\langle y_p(t)^2 y(0)^2 \rangle - \langle y_p(t)^2 \rangle \langle y(0)^2 \rangle = 0. \quad (3.18)$$

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

$$\langle y_h(t) y_p(t) y(0)^2 \rangle - \langle y_h(t) y_p(t) \rangle \langle y(0)^2 \rangle = 0, \quad (3.19)$$

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

$$\langle \delta y(t)^2 \delta y(0)^2 \rangle = \langle \delta y_h(t)^2 \delta y(0)^2 \rangle. \quad (3.20)$$

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  $(\gamma)^2 - 4km > 0$ . The components of the inverse diffusion tensor calculated below are identical to those calculated with the assumption  $(\gamma)^2 - 4km \leq 0$ . If we define

$$r_{\pm} = \frac{\gamma}{2m} \pm \frac{1}{2} \sqrt{\left(\frac{\gamma}{m}\right)^2 - \frac{4k}{m}}, \quad (3.21)$$

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

$$y_h(t) = y_0 + \frac{p(0) + mr_-(y(0) - y_0)}{m(r_- - r_+)} e^{-r_+ t} + \frac{p(0) + mr_+(y(0) - y_0)}{m(r_+ - r_-)} e^{-r_- t}, \quad (3.22)$$

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 distribu-

tion  $\rho_{eq}[y(0), p(0)] \propto e^{-\beta E[y(0), p(0)]}$  for  $E[y, p] = \frac{p^2}{2m} + \frac{1}{2}kY^2$ , we obtain the following identities:

$$\langle \delta Y_h^2(t) \delta Y^2(0) \rangle = \frac{2}{(k\beta)^2 (r_+ - r_-)^2} (r_- e^{-r_+ t} - r_+ e^{-r_- t})^2 \quad (3.23a)$$

$$\langle \delta \dot{Y}_h^2(t) \delta Y^2(0) \rangle = \frac{2r_+^2 r_-^2}{k^2 \beta^2 (r_+ - r_-)^2} (e^{-r_+ t} - e^{-r_- t})^2 \quad (3.23b)$$

$$\langle \delta Y_h^2(t) \delta p^2(0) \rangle = \frac{2}{\beta^2 (r_+ - r_-)^2} (e^{-r_+ t} - e^{-r_- t})^2 \quad (3.23c)$$

$$\langle \delta \dot{Y}_h^2(t) \delta p^2(0) \rangle = \frac{2}{\beta^2 (r_+ - r_-)^2} (r_- e^{-r_- t} - r_+ e^{-r_+ t})^2. \quad (3.23d)$$

Integrating these expressions, we obtain

$$\int_0^\infty dt \langle \delta Y_h^2(t) \delta Y^2(0) \rangle = \frac{m}{k^2 \beta^2 \gamma} \left( 1 + \frac{(\gamma)^2}{km} \right) \quad (3.24a)$$

$$\int_0^\infty dt \langle \delta \dot{Y}_h^2(t) \delta Y^2(0) \rangle = \frac{1}{k \beta^2 \gamma} \quad (3.24b)$$

$$\int_0^\infty dt \langle \delta Y_h^2(t) \delta p^2(0) \rangle = \frac{m^2}{k \beta^2 \gamma} \quad (3.24c)$$

$$\int_0^\infty dt \langle \delta \dot{Y}_h^2(t) \delta p^2(0) \rangle = \frac{m}{\gamma \beta^2}. \quad (3.24d)$$

Thus the inverse diffusion tensor is

$$\zeta_{ij} = \frac{m}{4\gamma} \begin{pmatrix} \frac{4(\gamma)^2}{m} \beta & 0 & 0 \\ 0 & \frac{1}{\beta^2} \left( 4 + \frac{(\gamma)^2}{km} \right) & \frac{1}{\beta k} \left( 2 + \frac{(\gamma)^2}{km} \right) \\ 0 & \frac{1}{\beta k} \left( 2 + \frac{(\gamma)^2}{km} \right) & \frac{1}{k^2} \left( 1 + \frac{(\gamma)^2}{km} \right) \end{pmatrix}, \quad (3.25)$$

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

### 3.3 Optimal protocols

Though one can write down the geodesic equations for the metric Eq. (3.25) in the  $(y_0, \beta, k)$  coordinate system, more insight is gained by finding a suitable change of coordinates. Consider the lower right  $2 \times 2$  block of the metric Eq. (3.25) which is the metric tensor for the two-dimensional  $(\beta, k)$  submanifold. A direct calculation of this submanifold's Ricci scalar yields  $R = -2\gamma/m$  which is constant and always strictly negative.

Theorems from Riemannian geometry [67] 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  $ds^2 = \zeta_{ij} dx^i dx^j = \frac{dx^2 + dz^2}{x^2}$ . The geodesics of the hyperbolic plane (see Fig. 3.1) are half-circles with centers on the  $z$ -axis and lines perpendicular to the  $z$ -axis. Fig. 3.1(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

$$\left( \frac{1}{x} \cos z, \log \left( \sqrt{x^2 - 1} + x \right) - \frac{\sqrt{x^2 - 1}}{x}, \frac{1}{x} \sin z \right). \quad (3.26)$$

The geodesics of Fig. 3.1 (c) and the part of the hyperbolic plane containing them are embedded in  $\mathbb{R}^3$  in Fig. 3.1(d).

The line element associated with the submanifold metric tensor,

$$ds^2 = \frac{m}{4\gamma} \left[ \frac{1}{\beta^2} \left( 4 + \frac{(\gamma)^2}{km} \right) d\beta^2 + \frac{2}{\beta k} \left( 2 + \frac{(\gamma)^2}{km} \right) d\beta dk + \frac{1}{k^2} \left( 1 + \frac{(\gamma)^2}{km} \right) dk^2 \right], \quad (3.27)$$

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

$$x \equiv \frac{1}{2\beta\gamma} \sqrt{\frac{m}{k}}, \quad z \equiv \frac{1}{4\beta k}, \quad (3.28)$$

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. (3.28), and substituting into Eq. (3.27) gives the metric tensor in  $(z, x)$ -coordinates,

$$ds^2 = \frac{m}{\gamma} \frac{dx^2 + dz^2}{x^2}. \quad (3.29)$$

The line element corresponding to the metric of the full three-dimensional manifold in Eq. (3.25) is

$$ds^2 = \frac{m}{\gamma} \frac{z dy_0^2 + dx^2 + dz^2}{x^2} \quad (3.30)$$

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. (3.30). In general [68, 67, 69], isometries of a metric are generated by the Killing vector fields  $K$  which are themselves characterized by the

Killing equation

$$\partial_i K_j + \partial_j K_i - 2\Gamma_{ij}^k K_k = 0. \quad (3.31)$$

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  $\partial_{x^i}$  is a Killing field [69]. Hence,  $\partial_{y_0}$  is clearly a Killing vector field. Examining the full set of Killing equations shows that

$$K = y_0 \partial_{y_0} + 2x \partial_x + 2z \partial_z \quad (3.32)$$

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

In general [69], for Killing vector  $K_i$  the quantity  $K_i \frac{d\lambda^i}{dt}$  is conserved along the geodesic described by  $\lambda$ . For the three-dimensional inverse diffusion tensor, we have the following two conserved quantities associated with Killing fields:

$$\frac{z(t)}{x^2(t)} \frac{dy_0}{dt}, \quad \frac{2}{x(t)} \frac{dx}{dt} + \frac{z(t)}{x^2(t)} y_0(t) \frac{dy_0}{dt} + \frac{2z(t)}{x^2(t)} \frac{dz}{dt}. \quad (3.33)$$

To solve the geodesic equations, note that the velocity of the geodesic (*i.e.*, its tangent vector) must have constant norm [68, 67, 69]. (See Ch. A for details.) For convenience, we choose the norm so that

$$1 = \frac{1}{x^2(t)} \left( \left( \frac{dz}{dt} \right)^2 + \left( \frac{dx}{dt} \right)^2 \right) + \frac{c_1^2 x^2(t)}{r^2 z(t)} \quad (3.34)$$

where we have used the first conserved quantity of Eq. (3.33). We combine this with the full geodesic equation for  $x(t)$ , to decouple  $x(t)$  from  $y_0(t)$  and  $z(t)$ :

$$\frac{d^2 x}{dt^2} - \frac{2}{x(t)} \left( \frac{dx}{dt} \right)^2 + x(t) = 0, \quad (3.35)$$

which has solution

$$x(t) = r \operatorname{sech}(t). \quad (3.36)$$

When  $z(t)$  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  $z(t)$  is not constant, Eqs. (3.34) and (3.36) imply

$$\frac{x^4(t)}{r^2} = \left( \frac{dz}{dt} \right)^2 + \frac{c_1^2 x^4(t)}{r^2 z(t)}, \quad (3.37)$$

which integrates to

$$z(t) = h^{-1}(c_2 - r \tanh(t)), \quad (3.38)$$

where

$$h(\xi) \equiv \xi \sqrt{1 - \frac{c_1^2}{\xi^2}} + \frac{1}{2} c_1^2 \log \left( 2\xi \left( 1 + \sqrt{1 - \frac{c_1^2}{\xi^2}} \right) - c_1^2 \right). \quad (3.39)$$

The Killing conserved quantities of Eq. (3.33), together with  $x(t)$  and  $z(t)$ , yield

$$y_0(t) = E - c_1 \log \left[ -c_1^2 + 2h^{-1}(c_2 - r \tanh(t)) \times \left( 1 + \sqrt{1 - \frac{c_1^2}{h^{-1}(c_2 - r \tanh(t))}} \right) \right]. \quad (3.40)$$

Let  $(y_{0,i}, x_i, z_i)$  and  $(y_{0,f}, x_f, z_f)$  denote the endpoints of the geodesic. Define  $\Delta\lambda \equiv \lambda_f - \lambda_i$  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  $\Delta h \equiv h(z_f) - h(z_i)$ , the constant  $c_2$  may be written as

$$c_2 = \bar{h} + \bar{x} \frac{\Delta x}{\Delta h} \quad (3.41)$$

and  $r$  is given by

$$r^2 = x_i^2 + \frac{1}{4} \left( \Delta h + 2 \frac{\Delta x}{\Delta h} \bar{x} \right)^2. \quad (3.42)$$

The constant  $E$  is given by

$$E = y_{0,i} + c_1 \log \left( -c_1^2 + 2z_i \left( 1 + \sqrt{1 - \frac{c_1^2}{z_i}} \right) \right) \quad (3.43)$$

and  $c_1$  is determined by the equation

$$\Delta y_0 = -c_1 \left[ \log \left( -c_1^2 + 2z_f \left( 1 + \sqrt{1 - \frac{c_1^2}{z_f}} \right) \right) - \log \left( -c_1^2 + 2z_i \left( 1 + \sqrt{1 - \frac{c_1^2}{z_i}} \right) \right) \right]. \quad (3.44)$$

The parameter  $t$  ranges between the values

$$t_i = \text{sgn} \left( \Delta h + 2 \frac{\Delta x}{\Delta h} \bar{x} \right) \text{sech}^{-1} \left( \frac{x_i}{r} \right) \quad (3.45)$$

and

$$t_f = -\text{sgn} \left( \Delta h - 2 \frac{\Delta x}{\Delta h} \bar{x} \right) \text{sech}^{-1} \left( \frac{x_f}{r} \right). \quad (3.46)$$

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 [68, 67]. Several example geodesics are displayed in Fig. 3.2.

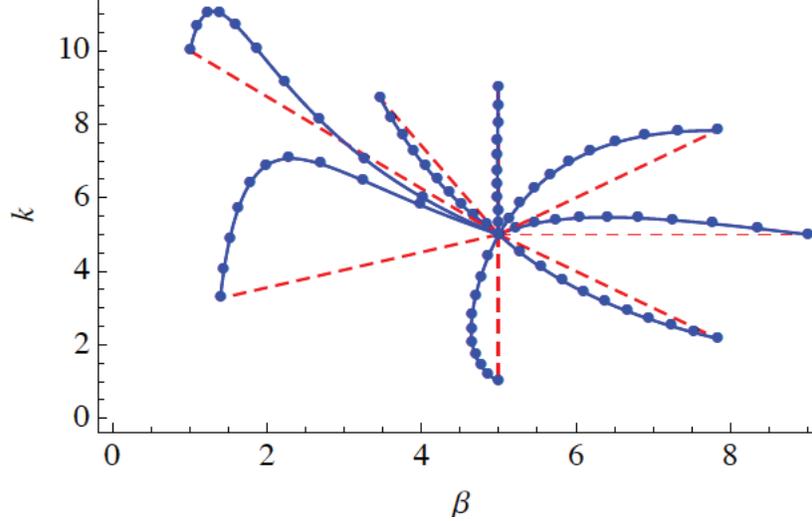


Figure 3.2: Optimal protocols differ substantially from linear interpolation (red dashed lines). Blue solid curves represent geodesics of the inverse diffusion tensor, and are thus optimal protocols for transitioning the system from one state to another in a fixed amount of time. Blue dots indicate points separated by equal times along each of the eight optimal paths shown.

### 3.4 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  $\lambda(t) = (y_0(t), \beta(t), k(t))$  is

$$\langle (\beta W)_{\text{ex}} \rangle \equiv \int_0^{t_f} dt \beta \mathcal{P}_{\text{ex}} = \int_0^{t_f} dt \left( \dot{\beta} \frac{\langle p^2 \rangle}{2m} + \frac{1}{2} \langle (y - y_0)^2 \rangle (k \dot{\beta} + \dot{k} \beta) + k \beta \dot{y}_0 \langle y_0 - y \rangle - \frac{\dot{\beta}}{\beta} - \frac{\dot{k}}{2k} \right). \quad (3.47)$$

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

Standard arguments [54] yield the Fokker-Planck equation for the time evolution of  $f(y, p, t)$ ,

$$\frac{\partial f}{\partial t} + \frac{p}{m} \frac{\partial f}{\partial y} - k(t) [y - y_0(t)] \frac{\partial f}{\partial p} - \frac{\gamma}{m} \frac{\partial [pf]}{\partial p} - \frac{\gamma}{\beta(t)} \frac{\partial^2 f}{\partial p^2} = 0. \quad (3.48)$$

By integrating Eq. (3.48) against  $y, p$ , etc., we find a system of equations for relevant nonequilibrium averages:

$$\frac{d}{dt}\langle y \rangle = \frac{\langle p \rangle}{m} \quad (3.49a)$$

$$\frac{d}{dt}\langle p \rangle = -\frac{\gamma}{m}\langle p \rangle - k\langle y - y_0 \rangle \quad (3.49b)$$

$$\frac{d}{dt}\langle py \rangle = \frac{\langle p^2 \rangle}{m} - k\langle y^2 \rangle - \frac{\gamma}{m}\langle py \rangle + ky_0\langle y \rangle \quad (3.49c)$$

$$\frac{d}{dt}\langle y^2 \rangle = \frac{2}{m}\langle py \rangle \quad (3.49d)$$

$$\frac{d}{dt}\langle p^2 \rangle = -2k\langle p(y - y_0) \rangle - 2\frac{\gamma}{m}\langle p^2 \rangle + \frac{2\gamma}{\beta}. \quad (3.49e)$$

Following the derivation of the friction tensor in [52] would require us to use linear response theory and to supplement the system Eq. (3.49) by initial conditions

$$\langle y \rangle(0) = y_0(0) \quad (3.50a)$$

$$\langle p \rangle(0) = 0 \quad (3.50b)$$

$$\langle y^2 \rangle(0) = y_0(0)^2 + \frac{1}{k(0)\beta(0)} \quad (3.50c)$$

$$\langle py \rangle(0) = 0 \quad (3.50d)$$

$$\langle p^2 \rangle(0) = \frac{m}{\beta(0)}. \quad (3.50e)$$

We solve these equations numerically and compare a geodesic protocol with naive protocols in Fig. 3.3.

This system has three natural dimensionless quantities

$$A \equiv \frac{m}{\gamma\Delta t}, \quad B \equiv \frac{\gamma}{\tilde{k}\Delta t}, \quad M \equiv \frac{\gamma(\Delta t)^3}{\tilde{l}^2 m^2 \tilde{\beta}}, \quad (3.51)$$

dependent upon characteristic scales for (inverse) temperature  $\tilde{\beta}$ , length  $\tilde{l}$ , spring constant  $\tilde{k}$  and the protocol duration  $\Delta t$ . These suggest at least two plausible measures of distance from equilibrium [52].  $A$  corresponds to the ratio of two timescales, the timescale  $\frac{m}{\gamma}$  for frictional damping and the timescale of the perturbation protocol  $\Delta t$ . Likewise,  $B$  is the ratio of two powers during changes of  $y_0$ , the dissipative power  $\gamma(\Delta y_0/\Delta t)^2$  and the elastic power  $\tilde{k}(\Delta y_0)^2/\Delta t$ . 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 dis-

sipation of geodesic protocols obtained numerically via Fokker-Planck agrees with the inverse diffusion tensor approximation to better than .1% (see Fig. 3.3). 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.

### 3.5 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:

$$\langle y \rangle \approx y_0 - \frac{\gamma}{k} \dot{y}_0 \quad (3.52a)$$

$$\langle p \rangle \approx m \dot{y}_0 \quad (3.52b)$$

$$\langle py \rangle \approx m y_0 \dot{y}_0 - \frac{m}{2} \left( \frac{\dot{k}}{\beta k^2} + \frac{\dot{\beta}}{\beta^2 k} \right) \quad (3.52c)$$

$$\langle p^2 \rangle \approx \frac{m}{\beta} + \frac{m^2}{\gamma} \left( \frac{\dot{k}}{2\beta k} + \frac{\dot{\beta}}{\beta^2} \right) \quad (3.52d)$$

$$\begin{aligned} \langle y^2 \rangle \approx & y_0^2 + \frac{1}{\beta k} - \frac{2\gamma}{k} y_0 \dot{y}_0 \quad (3.52e) \\ & + \dot{k} \left( \frac{m}{\gamma} \frac{1}{2\beta k^2} + \frac{\gamma}{2\beta k^3} \right) + \dot{\beta} \left( \frac{m}{\gamma} \frac{1}{\beta^2 k} + \frac{\gamma}{2\beta^2 k^2} \right). \end{aligned}$$

Substituting this into the expression for mean excess power Eq. (3.47), we recover Eq. (3.5). 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.

### 3.6 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  $(\beta, k)$  sub-

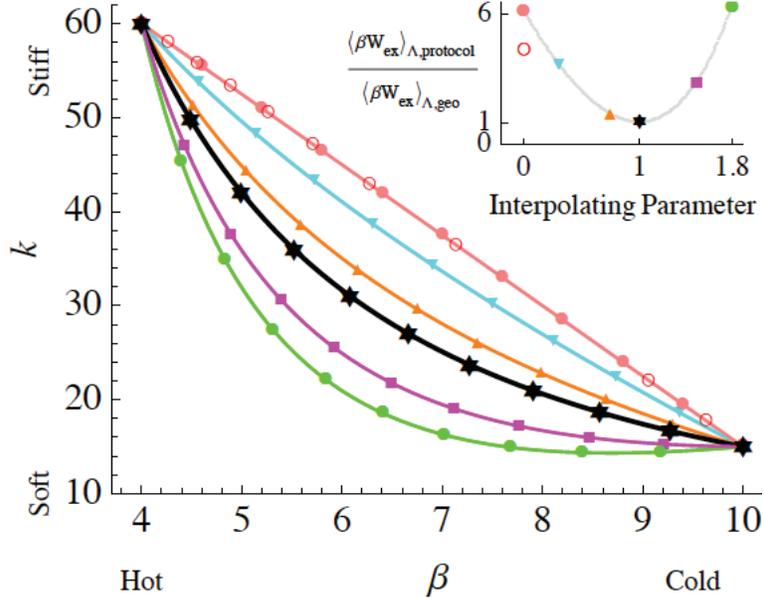


Figure 3.3: Geodesics describe protocols that outperform naive straight line paths in parameter space. A geodesic between two fixed points in the  $(\beta, k)$ -plane (black) and several comparison protocols are pictured above. The comparison protocols were generated via a linear interpolation between 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. (3.49). 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.

manifold 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 [40, 70], and there has been some speculation about the role of the Ricci scalar in that setting [70], 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. (3.33)), 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 globally smooth Killing fields where  $n$  is the dimension of the manifold [69].)

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. (3.51)).

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  $\beta$  as a control parameter, which is a natural extension of thermodynamic length (*e.g.* [71, 40]) that is amenable to direct experimental confirmation. Our work generalizes the construction of [52] 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 [41, 42], 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 [72], 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  $(y_0(t), \beta(t), k(t))$  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 indispens-

able 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.

## Chapter 4

# Optimal Control of Nonequilibrium Steady State Transitions

Living systems are distinguished by their self-organization. Biological systems fundamentally exist out of equilibrium in order to preserve organized structures and processes. Given the entropic driving force embodied in the second law of thermodynamics, creating and maintaining such organization requires staying far from equilibrium [18], typically by coupling to nonequilibrium gradients.

For example, ATP-driven molecular motors (*e.g.*, kinesin) are forced away from equilibrium by cellular maintenance of a chemical potential difference between ATP and ADP [19], and the rotary  $F_0$ - $F_1$  ATP synthase operates out of equilibrium due to cellular maintenance of an electrochemical gradient across the inner mitochondrial membrane [20]. For constant ATP and ADP concentrations, or constant membrane potential, the dynamics of an ensemble of such molecular motors will approximate a nonequilibrium steady state (NESS). Thus, biological systems are often better characterized as nonequilibrium steady states rather than equilibrium systems.

Such NESS may change in response to changing environmental conditions. Given that selective advantage may be incurred by energetically-efficient operation, evolution may have sculpted biological components to interact so as to reduce the energy wasted during transitions between NESS. Accordingly, optimizing such transitions may offer insights into the design principles of biological systems and guide the creation of synthetic molecular-scale machines.

In this chapter, we take a first step towards this goal by optimizing the Hatano-Sasa  $Y$ -value, a quantity similar to dissipated work, for the paradigmatic model system tested in [1] and analyzed in [53] with an eye towards experimental tests. We calculate and numerically verify optimal protocols for a colloidal particle dragged through solution by a translating optical trap with two controllable parameters. We propose a regime of validity of our approximation based on this nu-

merical work. By measuring the average work required to drive this system along either optimal or naive paths through control parameter space, our results can be tested experimentally in a straightforward way using existing experimental techniques.

## 4.1 The model system and its inverse diffusion tensor

We consider a particle with spatial coordinate  $x$  diffusing under Langevin dynamics subject to a one-dimensional harmonic potential, with equation of motion

$$\dot{x} = -\frac{k(t)}{\gamma}x + \eta(t) - v(t) , \quad (4.1)$$

for Gaussian white noise  $\eta(t)$  satisfying

$$\langle \eta(t) \rangle = 0 , \quad \langle \eta(t)\eta(t') \rangle = \frac{2}{\beta\gamma}\delta(t-t') . \quad (4.2)$$

Here  $\gamma$  is the Cartesian friction coefficient,  $k$  is the trap stiffness,  $v$  is the trap center velocity in the lab frame and  $x$  is the coordinate of the colloidal particle in the frame co-moving with the trap. The particle is initially in NESS due to constant trap velocity  $v$ .

As defined in Ch. 2, the Hatano-Sasa Y-value [6]

$$Y \equiv \int_0^\tau dt \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(x(t), \boldsymbol{\lambda}(t)) \quad (4.3)$$

arises in NESS transitions when the control parameters  $\boldsymbol{\lambda}$  are changed rapidly compared to the system's relaxation timescale. Here  $\phi(x, \boldsymbol{\lambda}) \equiv -\ln \rho_{ss}(x, \boldsymbol{\lambda})$  where  $\rho_{ss}(x, \boldsymbol{\lambda})$  is the steady state probability distribution and  $\tau$  is the protocol duration.

In some simple cases this corresponds to the system 'lagging' behind the changing control parameters. For transitions between equilibrium states this measure reduces to the standard dissipation governed by the Clausius inequality [73]. This measure of irreversibility (4.3) obeys a significant NESS fluctuation theorem that has been experimentally observed in our particular model system [1].

The ensemble average of the Y-value is

$$\langle Y \rangle_{\Lambda} \equiv \int_0^\tau dt \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \left\langle \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(x, \boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} . \quad (4.4)$$

Applying linear response theory [54, 52, 37] and assuming that the protocol varies sufficiently slowly [52], we arrive at an expression for the average Y-value

$$\langle Y \rangle_{\Lambda} \approx \int_0^{\tau} dt \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \boldsymbol{\zeta}(\boldsymbol{\lambda}(t)) \cdot \left[ \frac{d\boldsymbol{\lambda}}{dt} \right], \quad (4.5)$$

in terms of the control parameter velocities  $d\boldsymbol{\lambda}/dt$  and the inverse diffusion matrix  $\boldsymbol{\zeta}(\boldsymbol{\lambda})$  with entries

$$\zeta_{ij}(\boldsymbol{\lambda}) \equiv \int_0^{\infty} dt' \left\langle \frac{\partial \phi}{\partial \lambda^i}(t') \frac{\partial \phi}{\partial \lambda^j}(0) \right\rangle_{\boldsymbol{\lambda}}. \quad (4.6)$$

The angle brackets  $\langle \dots \rangle_{\boldsymbol{\lambda}}$  represent an average over noise followed by a stationary state average over initial conditions using the distribution  $\rho_{ss}(x, \boldsymbol{\lambda})$ . Note that, if  $\rho_{ss}(x, \boldsymbol{\lambda})$  is the equilibrium distribution, the inverse diffusion tensor of [37] is recovered. For details, the interested reader is referred to Ch. 2 of this thesis.

The steady-state distribution is given by  $\rho_{ss}(x, \boldsymbol{\lambda}) \equiv \sqrt{\frac{\beta k}{2\pi}} \exp\{-\frac{\beta}{2k} (kx + \gamma v)^2\}$  [1, 53]. The parameter space derivative of  $\phi$  is given by

$$\frac{\partial \phi}{\partial \boldsymbol{\lambda}} \equiv \left( \frac{\partial \phi}{\partial k}, \frac{\partial \phi}{\partial v} \right) \quad (4.7a)$$

$$= \left( -\frac{1}{2k} + \frac{\beta}{2} x^2 - \frac{\beta}{2} \left[ \frac{\gamma v}{k} \right]^2, \beta \gamma \left[ x + \frac{\gamma v}{k} \right] \right). \quad (4.7b)$$

In order to calculate the time correlation functions in Eq. (4.6), we solve Eq. (4.1) for constant  $k$  and  $v$ , giving

$$x(t) = x_0 e^{-\frac{k}{\gamma} t} + \int_0^t ds e^{-\frac{k}{\gamma}(t-s)} \eta(s) - \frac{\gamma v}{k} \left( 1 - e^{-\frac{k}{\gamma} t} \right). \quad (4.8)$$

Recalling that  $\eta(t)$  is Gaussian noise, Eq. (4.8) implies

$$\langle \partial_k \phi(t) \partial_k \phi(0) \rangle_{\boldsymbol{\lambda}} = \frac{\beta (\gamma v)^2}{k^3} e^{-\frac{k}{\gamma} t} + \frac{1}{2k^2} e^{-\frac{2k}{\gamma} t}, \quad (4.9a)$$

$$\langle \partial_k \phi(t) \partial_v \phi(0) \rangle_{\boldsymbol{\lambda}} = -\beta v \left( \frac{\gamma}{k} \right)^2 e^{-\frac{k}{\gamma} t}, \quad (4.9b)$$

$$\langle \partial_v \phi(t) \partial_k \phi(0) \rangle_{\boldsymbol{\lambda}} = -\beta v \left( \frac{\gamma}{k} \right)^2 e^{-\frac{k}{\gamma} t}, \quad (4.9c)$$

$$\langle \partial_v \phi(t) \partial_v \phi(0) \rangle_{\boldsymbol{\lambda}} = \frac{\beta \gamma^2}{k} e^{-\frac{k}{\gamma} t}. \quad (4.9d)$$

Integrating over time yields the inverse diffusion tensor:

$$\boldsymbol{\zeta}(k, v) = \begin{pmatrix} \frac{\gamma}{4k^4} [k + 4\beta (\gamma v)^2] & -\beta v \left[ \frac{\gamma}{k} \right]^3 \\ -\beta v \left[ \frac{\gamma}{k} \right]^3 & \beta \frac{\gamma^2}{k^2} \end{pmatrix}. \quad (4.10)$$

## 4.2 Optimal protocols

Though one can write down the geodesic equations for the metric [Eq. (4.10)] in the  $(k, v)$  coordinate system, more insight is gained by finding a suitable change of coordinates. A direct calculation of this metric's Ricci scalar yields  $R = 0$ , demonstrating that the underlying geometry is Euclidean [67].

The line element corresponding to the metric in Eq. (4.10) is

$$ds^2 = \frac{\gamma}{4k^4} [k + 4\beta(\gamma v)^2] dk^2 - 2\beta v \left(\frac{\gamma}{k}\right)^3 dk dv + \beta \frac{\gamma^3}{k^2} dv^2.$$

To find the explicit coordinate transformation making the Euclidean geometry manifest, we write the line element as

$$ds^2 = \beta\gamma^3 \left\{ \left[ d\left(\frac{v}{k}\right) \right]^2 + \left( \frac{dk}{2\sqrt{\beta}\gamma k^{\frac{3}{2}}} \right)^2 \right\}. \quad (4.11)$$

This suggests the coordinate transformation  $\xi = \frac{v}{k}$ ,  $\chi = \frac{1}{\gamma\sqrt{\beta k}}$ , so that

$$ds^2 = \beta\gamma^3 (d\xi^2 + d\chi^2). \quad (4.12)$$

In this coordinate system, geodesics are straight lines of constant speed. To find optimal protocols in  $(k, v)$  space, one simply transforms the coordinates of the endpoints into  $(\xi, \chi)$  space, connects these points by a straight line, and uses the inverse transformation to map the line onto a curve in  $(k, v)$  space. This follows from the invariance of the geodesic equation [67]. Explicitly, the optimal protocol joining  $(k_i, v_i)$  and  $(k_f, v_f)$  is

$$k(t) = \left[ \frac{1}{\sqrt{k_i}}(1 - T) + \frac{1}{\sqrt{k_f}}T \right]^{-2}, \quad (4.13a)$$

$$v(t) = k(t) \left[ \frac{v_i}{k_i}(1 - T) + T \frac{v_f}{k_f} \right], \quad (4.13b)$$

where  $T = \frac{t}{\tau}$ . Sample optimal protocols are pictured in Fig. 4.2.

## 4.3 Optimal straight-line protocols

In the absence of any particular information about the system's dynamical properties, a naive control strategy would change the control parameters at a constant rate, producing a straight line in control parameter space. The inverse diffusion tensor approximation [Eq. (4.5)] provides a recipe for choosing both a potentially nonlinear path through control parameter space, as well as a time-course along that path. The inverse diffusion tensor formalism can alternatively be used to optimize

the time-course along a straight-line control parameter path. Such a protocol provides a benchmark against which an optimal protocol [Eq. (4.13)] can be compared. For the model considered here, we will find that an optimal straight-line protocol can be substantially better than the most naive (constant-speed) straight-line protocol. Furthermore, straight-line protocols are relatively straightforward to test experimentally.

When  $k(t)$  is held fixed, a straightforward application of variational calculus demonstrates that a straight-line protocol in  $v(t)$  is *exactly* optimal and agrees with the predictions of the linear response approximation [Eq. (4.5)]. To see this, consider the explicit functional form the mean Y-value as recorded in Eq. (4.25). When  $\dot{k} = 0$ , this expression simplifies to

$$\beta\gamma \int_0^\tau dt \dot{v} \left[ \langle x \rangle_\Lambda + \gamma \frac{v}{k} \right]. \quad (4.14)$$

From Eq. (4.26) we may write

$$\dot{v} = - \left[ \frac{d^2 \langle x \rangle_\Lambda}{dt^2} + \frac{k}{\gamma} \frac{d \langle x \rangle_\Lambda}{dt} \right] \quad (4.15)$$

which allows us to write the mean Y-value as a functional of the ensemble moment  $\langle x \rangle_\Lambda$ :

$$\langle Y \rangle_\Lambda = \frac{\beta\gamma^2}{k} \int_0^\tau dt \left[ \frac{d^2 \langle x \rangle_\Lambda}{dt^2} + \frac{k}{\gamma} \frac{d \langle x \rangle_\Lambda}{dt} \right] \frac{d \langle x \rangle_\Lambda}{dt}. \quad (4.16)$$

Extrema of a functional of the form  $\int_0^\tau dt \mathcal{L}[f(t), \dot{f}(t), \ddot{f}(t)]$  satisfy the Euler-Lagrange equation

$$\partial_f \mathcal{L} - \frac{d}{dt} \partial_{\dot{f}} \mathcal{L} + \frac{d^2}{dt^2} \partial_{\ddot{f}} \mathcal{L} = 0. \quad (4.17)$$

Therefore, the optimal time course for the ensemble moment  $\langle x \rangle_\Lambda$  satisfies

$$-\frac{d}{dt} \left( \frac{d^2 \langle x \rangle_\Lambda}{dt^2} + \frac{2k}{\gamma} \frac{d \langle x \rangle_\Lambda}{dt} \right) + \frac{d^2}{dt^2} \left( \frac{d \langle x \rangle_\Lambda}{dt} \right) = 0 \quad (4.18)$$

or

$$\frac{d \langle x \rangle_\Lambda}{dt} = \text{const.} \quad (4.19)$$

Integrating this equation demonstrates that the optimal  $\langle x \rangle_\Lambda$  is linear in time. From Eq. (4.15), the optimal time course for  $v(t)$  must also be linear with coefficients determined by the endpoints of the protocol and the protocol duration.

In Ref. [1], the average Y-value was measured for three distinct experimental trials involving protocols with constant  $k$ . As summarized in Fig. 4.1, the optimal protocol, namely the naive straight line in the case of constant  $k$ , shows significantly reduced Y-value compared with the pro-

protocols used in each experimental trial. However, in terms of testing the performance of the optimal protocols [Eq. (4.13)],  $k_f \neq k_i$  is the more general case.

As in the case of finding globally optimal protocols, the problem of finding optimal straight line protocols simplifies dramatically in  $(\xi, \chi)$  coordinates. Using Eq. (4.12), we find

$$\langle Y \rangle_{\Lambda} \approx \beta \gamma^3 \int_0^{\tau} dt \left[ 1 + b^2 \chi^2(t) \right] \left( \frac{d\chi}{dt} \right)^2, \quad (4.20)$$

for

$$b \equiv 2\beta\gamma^2 \frac{k_f v_i - k_i v_f}{k_f - k_i}. \quad (4.21)$$

The Euler-Lagrange equation implies

$$\frac{d\chi}{dt} = \frac{\frac{1}{\tau} \int_{\chi_0}^{\chi_f} dz \sqrt{1 + b^2 z^2}}{\sqrt{1 + b^2 \chi^2(t)}}, \quad (4.22)$$

which determines an implicit expression for  $\chi(t)$ :

$$2b \left( \frac{t}{\tau} \right) \int_{\chi_0}^{\chi_f} dz \sqrt{1 + b^2 z^2} = b \left( \chi(t) \sqrt{1 + b^2 \chi^2(t)} - \chi_0 \sqrt{1 + b^2 \chi_0^2} \right) + \sinh^{-1} [b\chi(t)] - \sinh^{-1} [b\chi_0]. \quad (4.23)$$

The relation  $\chi = \frac{1}{\gamma\sqrt{\beta k}}$  determines an implicit expression for  $k(t)$ , and hence for  $v(t)$ .

## 4.4 Computing the Y-value numerically

We validate the optimality of the geodesics [Eq. (4.13)] and compare with optimal straight-line protocols by calculating the average Y-value directly by integrating in time the Fokker-Planck equation describing the dynamical evolution of the particle probability distribution [53],

$$\frac{\partial \rho}{\partial t} = \frac{k(t)}{\gamma} \frac{\partial}{\partial x} (x\rho) + v(t) \frac{\partial \rho}{\partial x} + \frac{1}{\beta\gamma} \frac{\partial^2 \rho}{\partial x^2}. \quad (4.24)$$

In full generality, the mean Y-value as a functional of the protocol  $\lambda(t) = (k(t), v(t))$  is

$$\int_0^{\tau} dt \left[ -\frac{\dot{k}}{2k} - \frac{\beta}{2} \left( \frac{\gamma v}{k} \right)^2 \dot{k} + \frac{\beta}{2} \dot{k} \langle x^2 \rangle_{\Lambda} + \beta \gamma \dot{v} \langle x \rangle_{\Lambda} + \beta \gamma^2 \frac{v}{k} \dot{v} \right]. \quad (4.25)$$

Here angled brackets denote averages over the nonequilibrium probability density  $\rho(x, t)$ .

By integrating Eq. (4.24) against  $x$  and  $x^2$ , we find a system of equations for relevant nonequi-

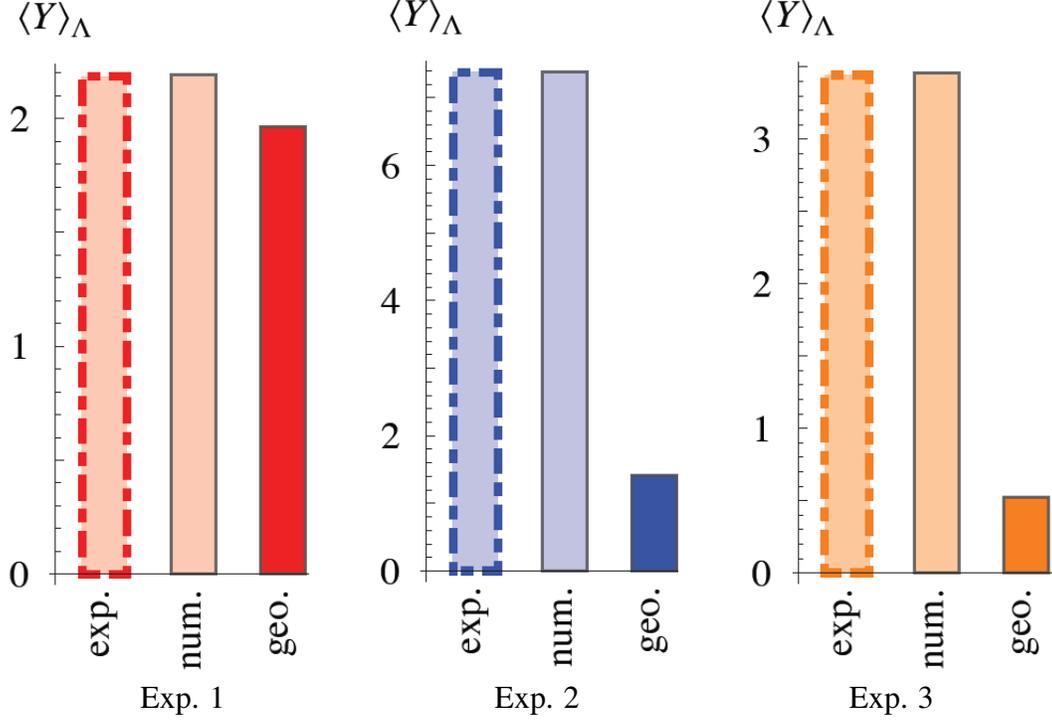


Figure 4.1: Experiment 1 (left, red) used a quarter-sine wave protocol to vary the trap speed; Experiments 2 (middle, blue) and 3 (right, orange) used an inverted three-quarters sine wave. Specifically,  $v(t) = 8.12 + 4.03 \sin(\pi t/2\tau)$ ,  $k = 4.25$ ,  $\tau = 0.06$ ,  $q = 0.20$  for Experiment 1,  $v(t) = 9.93 - 3.63 \sin(3\pi t/2\tau)$ ,  $k = 4.51$ ,  $\tau = 0.06$ ,  $q = 0.21$  for Experiment 2, and  $v(t) = 7.53 - 2.67 \sin(3\pi t/2\tau)$ ,  $k = 4.9$ ,  $\tau = 0.08$ ,  $q = 0.23$  for Experiment 3. Here, velocity is measured in  $\mu\text{m}/\text{s}$ ,  $\tau$  is the protocol duration measured in  $\text{s}$ ,  $k$  is the trap stiffness measured in  $\text{pN}/\mu\text{m}$ , and  $q \equiv k/\beta\gamma$  is measured in  $\text{pN} \mu\text{m}/\text{s}$ . The  $Y$ -value for these protocols (light color bar) and for the optimal protocols (solid color bar) were obtained numerically assuming  $\beta^{-1} = 4.6 \text{ pN nm}$  (red),  $\beta^{-1} = 4.45 \text{ pN nm}$  (blue),  $\beta^{-1} = 4.35 \text{ pN nm}$  (orange) respectively. These effective temperatures were chosen to give the best match between experiment and numerical calculation, and may differ from room temperature ( $\beta^{-1} = 4.14 \text{ pN nm}$ ) because of local heating by the optical trap [74]. We predict a significant reduction in  $Y$ -value for optimal protocol driving under the conditions of the three experiments described in Ref. [1].

librium averages:

$$\frac{d}{dt} \langle x \rangle_\Lambda = -\frac{k(t)}{\gamma} \langle x \rangle_\Lambda - v(t), \quad (4.26a)$$

$$\frac{d}{dt} \langle x^2 \rangle_\Lambda = -\frac{2k(t)}{\gamma} \langle x^2 \rangle_\Lambda - 2v(t) \langle x \rangle_\Lambda + \frac{2}{\beta\gamma}, \quad (4.26b)$$

supplemented by initial conditions

$$\langle x \rangle_{\Lambda}(0) = -\frac{\gamma v_i}{k_i}, \quad (4.27a)$$

$$\langle x^2 \rangle_{\Lambda}(0) = \frac{1}{\beta k_i} + \left[ \frac{\gamma v_i}{k_i} \right]^2. \quad (4.27b)$$

Note that for a more complex system the first and second moments  $\langle x \rangle$  and  $\langle x^2 \rangle$  are not sufficient to characterize the probability distribution, but time-dependent solutions are still accessible through standard (but more computationally intensive) numerical integration of the full Fokker-Planck equation (4.24) [75].

We solve these equations numerically and compare the performance of optimal straight lines against geodesics [Eq. (4.13)] and naive (constant-speed) straight-line protocols in Fig. 4.2. We selected endpoints and physical constants based on those used in the experiments of Ref. [1]; in this regime the inverse diffusion tensor approximation produces small relative error in Y-value. Though there is only a marginal difference in performance between the optimal straight-line protocol and the geodesic for both sets of endpoints, there is a substantial benefit in using either over the naive straight line protocol.

## 4.5 The inverse diffusion tensor arises naturally from the Fokker-Planck equation

If we neglect terms involving second- and higher-order temporal derivatives (an alternative near-equilibrium approximation), we obtain an approximate solution to the Fokker-Planck system:

$$\langle x \rangle_{\Lambda} \approx -\gamma \frac{v}{k} + \left( \frac{\gamma}{k} \right)^2 \dot{v} - \frac{v}{k} \left( \frac{\gamma}{k} \right)^2 \dot{k}, \quad (4.28a)$$

$$\langle x^2 \rangle_{\Lambda} \approx \gamma^2 \frac{v^2}{k^2} + \frac{1}{\beta k} + \frac{\gamma \dot{k}}{2\beta k^3} + \frac{2\gamma^3 v^2 \dot{k}}{k^4} - \frac{2\gamma^3 v \dot{v}}{k^3}. \quad (4.28b)$$

Substituting this into the expression for the mean Y-value [Eq. (4.25)], we recover Eq. (4.11). 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 [37].

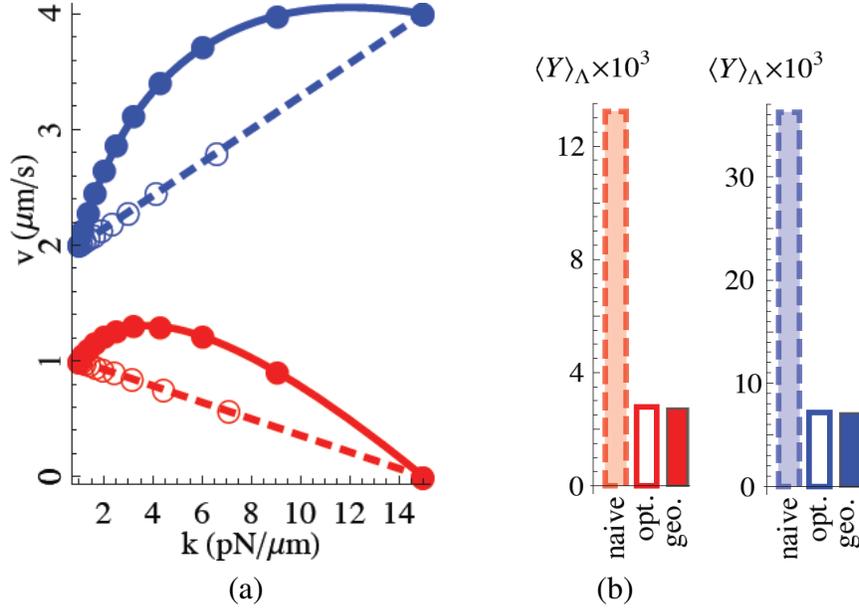


Figure 4.2: Geodesics describe protocols that outperform naive (constant-speed) straight-line paths in parameter space. Geodesics between fixed pairs of points in the  $(k, v)$ -plane and accompanying straight-line protocols are pictured in (a). The filled circles represent points separated by equal times. The open circles correspond to the optimal parametrization along the respective straight path. All mean  $Y$ -values were calculated using the Fokker-Planck system, Eq. (4.26). Here,  $\gamma = 0.1 \text{ pN s}/\mu\text{m}$  and  $\beta^{-1} = 4.6 \text{ pN nm}$  to approximate the experiments of Ref. [1]. The protocol duration is chosen to be  $\tau = 100 \text{ s}$  to ensure that the relative error  $|1 - \langle Y \rangle_{\Lambda}^{\text{approx}} / \langle Y \rangle_{\Lambda}|$  is less than 1.4% for all protocols. Protocol endpoints were selected for experimental accessibility [76]. The relative performance of naive straight-line, optimal straight-line, and geodesic protocols are summarized in (b).

## 4.6 Discussion

We have taken the first step towards extending the geometric framework for calculating optimal protocols presented in [52, 37] to systems relaxing to NESS. As energy-transducing biological systems are more faithfully described by NESS than by equilibrium statistics, this brings recent theoretical developments closer to the behavior of *in vivo* biological systems.

Using a linear response approximation, we found the optimal mean  $Y$ -value for a model system of a colloidal particle (initially in NESS) dragged through solution and subject to a time-dependent harmonic potential. We took as our control parameters the velocity and spring constant of the harmonic potential. Tools from Riemannian geometry revealed a useful coordinate transformation which greatly simplified the construction of optimal straight-line protocols as well as geodesic protocols. These optimal protocols were tested numerically and the small relative error in the  $Y$ -value approximation for experimentally relevant choices of parameters is encouraging.

Our predictions may be tested experimentally with existing hardware and methods. In Ref. [1],

the authors report on experiments performed with micron-sized polystyrene beads in solution. The harmonic potential is created by superposing the foci of two counterpropagating laser beams. The location of this trap was translated using a steerable mirror. The velocity  $v$  of the trap location was altered by changing the mirror’s angular rate of rotation, and the trap stiffness  $k$  can be manipulated by dynamically changing the intensity of the laser beam [77] or by passing the laser beam through a polarization filter and dynamically changing the polarization of the laser beam. Force is inferred from the rate of change of the momentum of light measured by position-sensitive photodetectors. Comparison of the average work incurred during different protocols would provide an experimental test of the optimal protocols predicted in this manuscript.

Using the inverse diffusion tensor approximation in general allows us access to the full power of Riemannian geometry in calculating optimal protocols. However, such experimental tests can assess the range of validity of the approximation. Our alternate derivation of the inverse diffusion tensor via a “derivative-truncation” expansion [37] suggests a greater robustness of the approximation.

In this chapter we provide concrete theoretical predictions for experiments — specifically, we find that geodesics, optimal straight-lines, and naive straight-line protocols all are substantially more efficient than the protocols tested in Trepagnier, *et al.* Moreover, we demonstrate that for simultaneous adjustment of  $k$  and  $v$ , optimal straight-line protocols can perform substantially better than naive (constant-speed) straight-line protocols. The necessary methodology and experimental apparatus are well-established [1] to not only verify our predictions but to push beyond the near-steady-state regime.

Given the greater generality embodied by the extension to NESS, and the accuracy of this approximation for a standard model system, optimal driving protocols derived in this framework promise greater applicability to models of biomolecular machines. Nevertheless, important hurdles remain: our model system experiences forces linear in position and has a steady-state distribution differing from the equilibrium one only in its average displacement. Molecular machines feature nonlinear force profiles, potentially nontrivial steady-state distributions, and often operate far from equilibrium. Thus our comparatively simple theoretical framework may need further elaboration to address the dynamics and efficiency of molecular machines with reasonable fidelity.

Furthermore, the relatively simple model system we treat in this chapter represents a new frontier for the analytical solution of optimal protocols under the inverse diffusion tensor approximation. For significantly more complicated models of greater biological interest, a simple general approach (in lieu of a search for an analytical solution) would be a fully numerical method, involving the calculation of the inverse diffusion tensor at a grid of points in control parameter space, analogous to the approach in [39].

Finally, there remains the important open question of what quantity or quantities are to be optimized in faithful models of biological processes. In this thesis chapter, we made the choice of optimizing the Y-value which has been experimentally studied in this particular model system [1] and may be optimized by the same geometric framework as in [52]. These qualities were advantageous to begin a clear and mathematically tractable first step towards optimization of steady state transitions.

However, it is possible and perhaps likely that a properly defined average dissipated heat will be the biologically relevant quantity to optimize rather than the Y-value. We anticipate that a geometric approach to optimization will be applicable to these more general systems and notions of heat production in a relevant regime of parameter values and protocol durations. However, a more general construction will have to take into account the so-called housekeeping heat [6, 55] which is generated in maintaining the steady state at given control parameter values. Future work is needed to address these issues properly.

# Chapter 5

## Optimal Finite-Time Erasure Cycles

### 5.1 Introduction

Optimization schemes for thermodynamic processes occurring in finite time will be needed for applications in which energetic or entropic costs are undesirable [21, 22]. An important class of such processes consists of mesoscopic information processing systems operating out of equilibrium. Optimization will aid technological development in the decades to come as computational demands approach limits imposed by physical law [23, 78], and it has implications for biomolecular processes [79].

Moreover, understanding these systems will provide insight into the foundations of nonequilibrium statistical mechanics. Investigations into the interplay between information and thermodynamics seem to have originated with Maxwell's hypothetical demon and its implications for the second law of thermodynamics [24]. Much ground-breaking work followed from the Maxwell demon paradox including Szilard's engine revealing a quantitative link between thermodynamic work and information [25], Landauer's observation of the physical nature of information [26] and Bennett's interpretation of the paradox in terms of the relation between logical and thermodynamic reversibility [27].

In recent times, research into nonequilibrium statistical mechanics of small-scale systems has shed more light on the thermodynamic role of information [31]. Most notable is experimental verification [28] of the theoretical prediction of microscopic violations of Landauer's principle with the preservation of the principle on average [29], analogous to experimental and theoretical work on fluctuations theorems demonstrating that entropy-reducing processes can occur microscopically whereas the second law holds on average [30]. Research into feedback and measurement of mesoscopic nonequilibrium systems has improved our understanding of the role information plays in the second law [31, 32]. Other work has focused on developing techniques to optimize thermodynamic quantities arising in small-scale systems designed to store and erase classical in-

formation [33, 34, 35], including the derivation of a refined second law [35]. Recent work has also focused on the general problem of predicting optimal protocols to drive systems between stationary states with minimal dissipation [36, 37, 80, 39, 40, 41, 42, 43]. Note that we use “dissipation” here to refer to all heat generation, including heat that could be recovered as useful work (*c.f.* [81]).

We have obtained closed-form expressions for the dissipation of maximally efficient cyclical protocols for a simple system designed to store and delete a classical bit of information. The system storing this bit consists of an overdamped Brownian colloidal particle diffusing in a one-dimensional double-square-well potential separated by a potential barrier stabilizing the memory. We take as control parameters the height of the potential barrier and the difference in minima of the two wells.

When our two simultaneously adjustable parameters are optimally controlled, we find that the dissipation falls off as the inverse of the cycle duration, asymptoting to the Landauer bound in the long duration limit. This is consistent with pioneering work on erasure for a similar model system [35] as well as a numerical [33] and theoretical [34] work on a single-level quantum dot, though it is unclear if predictions for high temperature behavior in that system [34] can be reconciled with our findings. However, unlike all previous studies, we have derived an explicit formula for the minimal dissipation, valid for arbitrary temperature, providing specific testable predictions for existing experimental setups [82, 28]. Our solutions are non-trivial in that both control parameters are continuously varying in time, but they are easily described, which complements previous results for a nonparametric model [35].

For durations that are long compared to a characteristic timescale, we obtain a simple expression for the dissipation that depends on the difference between the initial and final spatial distributions of the particle. Interestingly, the extra dissipation beyond the Landauer bound for the optimal finite-time protocol is proportional to the square of the Hellinger distance, which is always greater than zero for any nonzero change in the probabilities of finding the particle in the left or right potential well, unlike the Landauer bound itself, which can be zero or even negative depending on the change in entropy of the particle’s spatial distribution.

Finally, we demonstrate that the geometrical framework developed in previous chapters for finding optimal protocols based on the inverse diffusion tensor predicts nearly identical solutions to our exact optimal protocols in this parameter regime, which is an encouraging sign for finding optimal protocols in other model systems.

## 5.2 Model of classical information erasure

We consider the following model to represent a single classical bit of information: an overdamped Brownian colloidal particle diffusing in a one-dimensional double-well potential in con-

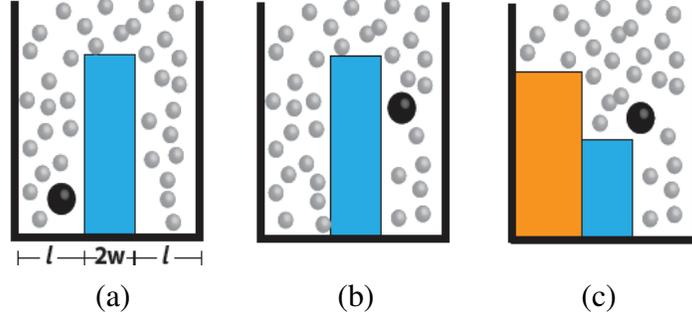


Figure 5.1: Double-well potential for storage of a single classical bit. (a,b) The system begins in thermal equilibrium with equipotential wells and a potential barrier of height much larger than the thermal fluctuation scale. Observing the particle (black dot) to the left (right) of the potential barrier corresponds to memory value 1 (0). The width  $2w$  of the central barrier and the width  $l$  of each well satisfy  $2w/l \ll 1$ . (c) Optimally-efficient erasure protocols are sought in which the “tilt”  $V_l$  (orange) and the barrier height  $V_b$  (blue) are control parameters. After the erasure step, the particle is much more likely to be in the right well, regardless of where it originated.

tact with a thermal bath of temperature  $T$  [29, 28] (Fig. 5.1). The wells are initially separated by a potential barrier whose height is much larger than the energy scale  $\beta^{-1} \equiv k_B T$  set by thermal fluctuations, ensuring stability of memory. The system is prepared so that the particle has equal probability of being found in either well. This may be achieved, for example, by selecting the initial position of the particle to be at the midpoint of the potential barrier and waiting a sufficiently long relaxation period [29]. If the particle is found in the left-hand (right-hand) well, the memory value is defined to be 1 (0).

The time evolution of the particle’s position  $x(t)$  is governed by Brownian dynamics

$$\dot{x} = -\frac{1}{\gamma} \partial_x U(x(t), t) + F(t) \quad (5.1)$$

for Gaussian white noise  $F(t)$  satisfying

$$\langle F(t) \rangle = 0, \quad \langle F(t)F(t') \rangle = \frac{2}{\beta\gamma} \delta(t - t'). \quad (5.2)$$

Here,  $\gamma$  is the Cartesian friction coefficient and  $U(x, t)$  is a generic double-well potential satisfying  $U(x, t) \rightarrow \infty$  as  $|x| \rightarrow \infty$ . We will find the equivalent statistical description in terms of the Fokker-Planck equation

$$\partial_t \rho = D [\partial_x (\beta U'(x, t) \rho) + \partial_x^2 \rho] \equiv -\partial_x G \quad (5.3)$$

convenient, where  $\rho(x, t)$  is the position probability density,  $G(x, t)$  is the probability current, and  $D$  is the diffusion coefficient.

Out of equilibrium, the system’s probability distribution over microstates fundamentally de-

depends on the history of the control parameters  $\lambda$ , which we denote by the control parameter protocol  $\Lambda$ ;  $\langle \cdot \rangle_\Lambda$  denotes the average over the nonequilibrium probability distribution arising from the parameter protocol  $\Lambda$ .

We are primarily interested in optimizing finite-time erasure efficiency over cyclic protocols for the classical single bit model described above. When classical information is being erased, the difference in Shannon entropies of the final and initial probability distributions must satisfy  $\Delta S \equiv S_f - S_i < 0$ , which would allow us to define the erasure efficiency  $\epsilon \equiv -\Delta S / (k_B \langle \beta Q \rangle_\Lambda)$  as the ratio of this decrease in Shannon entropy to the average heat  $\langle Q \rangle_\Lambda$  released into the thermal bath [33, 83]. However, in addition to erasure, we will also consider arbitrary initial and final spatial distributions for the particle, so we will state our results in terms of dissipation rather than efficiency.

Our goal will be to minimize the dissipated heat subject to constraints on the initial and final probability distributions, similar in some ways to “stochastic optimal control,” but for a different cost function [84]. Since we are constraining the initial and final probability distributions, we expect our optimal protocols to have jump discontinuities at the endpoints based on experience with optimization in the context of stochastic thermodynamics in general [42, 41, 43, 55, 85] and erasure efficiency in particular [34, 35]. These jump discontinuities warrant caution when defining thermodynamic quantities such as the average dissipated heat [85].

Suppose  $\lambda$  is smooth for  $t \in (0, t_f)$  but possesses jump discontinuities at  $t = 0, t_f$ . Then  $U(x, t) \equiv U(x, \lambda(t))$  is smooth on the interval  $(\delta, t_f - \delta)$  and it is possible to write

$$U(x(t_f - \delta), t_f - \delta) - U(x(\delta), \delta) = \int_\delta^{t_f - \delta} dt \left[ \frac{d\lambda}{dt} \right]^T \cdot \frac{\partial U}{\partial \lambda}(x(t), \lambda(t)) + \int_\delta^{t_f - \delta} \frac{\partial U}{\partial x}(x(t), \lambda(t)) \circ dx(t), \quad (5.4)$$

where the integral over the fluctuating quantity  $x$  is computed using Stratonovich integral calculus [35, 55, 58] and  $\delta$  is small and positive. We may also write this expression as

$$U(x(t_f), t_f) - U(x(0), 0) = [U(x(t_f), t_f) - U(x(t_f - \delta), t_f - \delta)] - [U(x(0), 0) - U(x(\delta), \delta)] + \int_\delta^{t_f - \delta} dt \left[ \frac{d\lambda}{dt} \right]^T \cdot \frac{\partial U}{\partial \lambda}(x(t), \lambda(t)) + \int_\delta^{t_f - \delta} \frac{\partial U}{\partial x}(x(t), \lambda(t)) \circ dx(t). \quad (5.5)$$

The nonequilibrium ensemble average of the left-hand side is given by

$$\langle U(x(t_f), t_f) - U(x(0), 0) \rangle_\Lambda = \int_{\mathbb{R}} dx [U(x, t_f)\rho(x, t_f) - U(x, 0)\rho(x, 0)]. \quad (5.6)$$

The last equality follows from the fact that  $\langle \delta(x(t) - x) \rangle_\Lambda = \rho(x, t)$  which is assumed continuously

differentiable on  $(0, t_f)$  and continuous on  $[0, t_f]$ . Note that  $\langle U(x(t_f), t_f) - U(x(0), 0) \rangle_{\Lambda} \equiv \Delta U$  depends only on  $\lambda(0)$ ,  $\lambda(t_f)$  and the initial and final probability distributions.

Taking the nonequilibrium ensemble average of both sides of Eq. (5.5) and then taking the limit  $\delta \rightarrow 0$ , we find the average work done on the system [58, 55]

$$\langle W \rangle_{\Lambda} = \int_{0^+}^{t_f^-} dt \left[ \frac{d\lambda}{dt} \right]^T \cdot \left\langle \frac{\partial U}{\partial \lambda}(\lambda(t)) \right\rangle_{\Lambda} + \int_{\mathbb{R}} \rho(x, t_f) U(x, t) \Big|_{t=t_f^-}^{t=t_f} - \int_{\mathbb{R}} \rho(x, 0) U(x, t) \Big|_{t=0^+}^{t=0} \quad (5.7)$$

where we abuse notation by defining

$$\int_{\mathbb{R}} dx \frac{\partial U}{\partial \lambda}(x, \lambda(t)) \rho(x, t) \equiv \left\langle \frac{\partial U}{\partial \lambda}(\lambda(t)) \right\rangle_{\Lambda}. \quad (5.8)$$

The last two terms of the average work explicitly take into account the cost of jump discontinuities at the beginning and end of the protocol [85].

By definition of average dissipation for nonequilibrium transitions from stationary states [6, 38],

$$\langle \beta W_{ex} \rangle_{\Lambda} = \int_{0^+}^{t_f^-} dt \left[ \frac{d\lambda}{dt} \right]^T \cdot \left\langle \frac{\partial \phi}{\partial \lambda}(\lambda(t)) \right\rangle_{\Lambda} + \int_{\mathbb{R}} \rho(x, t_f) \beta U(x, t) \Big|_{t=t_f^-}^{t=t_f} - \int_{\mathbb{R}} \rho(x, 0) \beta U(x, t) \Big|_{t=0^+}^{t=0} \quad (5.9)$$

where  $W_{ex} \equiv W - \Delta F$ ,  $\Delta F \equiv F(\lambda(t_f^-)) - F(\lambda(0^+))$ ,  $F(\lambda) \equiv -\beta^{-1} \ln \left( \int_{\mathbb{R}} \exp\{-\beta U(x; \lambda)\} \right)$ ,  $\phi(x, \lambda) \equiv -\ln \rho_{eq}(x; \lambda)$  and  $\rho_{eq}(x; \lambda) = \exp\{\beta (F(\lambda) - U(x, \lambda))\}$  is the Boltzmann distribution.

To simplify the mathematics, we consider a piecewise constant potential as illustrated in Fig. 5.1 and similar to the model considered in [86]. This model admits a reasonable “discretization” of the system, providing a means of calculating optimal protocols exactly. We use the discrete approximation of [87] to obtain transition rates for the master equations [54]

$$\frac{dp_i}{dt} = \sum_{j \neq i} r_{j \rightarrow i} p_j - \sum_{j \neq i} r_{i \rightarrow j} p_i \quad (5.10)$$

governing the time evolution of  $p_i$ . Here and throughout,  $p_l$  ( $p_r$ ) is the probability of the particle being on the left (right) of the barrier, corresponding to memory value 1 (0).

The double-well potential has a natural decomposition into “compartments”: we define the in-

terval  $(-l-w, -w)$  as compartment 1,  $(-w, w)$  as compartment 2 and  $(w, l+w)$  as compartment 3, which we will denote as  $l$  (left),  $b$  (barrier), and  $r$  (right) respectively. We may discretize the continuum dynamics a la [87] to obtain master equations governing the probability of finding the particle in compartment  $i$  at any given time.

If we define

$$q_i(x) = e^{-\beta U(x)} \int_{x_{i-1}}^x dx' e^{\beta U(x')} \quad (5.11)$$

for  $x_0 = -l-w$ ,  $x_1 = -w$ ,  $x_2 = w$ ,  $x_3 = l+w$  and

$$e^{-\beta G_i} = \int_{x_{i-1}}^{x_i} dx' e^{-\beta U(x')}, \quad (5.12)$$

then the transition rates are given by

$$r_{i \rightarrow i+1} = Dh_i^+[U], \quad r_{i+1 \rightarrow i} = Dh_i^-[U] \quad (5.13)$$

where

$$h_i^+[U] = \frac{e^{-\beta G_{i+1}}}{e^{-\beta G_i} \int_{x_i}^{x_{i+1}} dx q_i(x) - e^{-\beta G_{i+1}} \int_{x_{i-1}}^{x_i} dx q_i(x)} \quad (5.14)$$

and

$$h_i^-[U] = \frac{e^{-\beta G_i}}{e^{-\beta G_i} \int_{x_i}^{x_{i+1}} dx q_i(x) - e^{-\beta G_{i+1}} \int_{x_{i-1}}^{x_i} dx q_i(x)}. \quad (5.15)$$

A straightforward calculation determines the nontrivial transition rates for our model system:

$$\begin{aligned} r_{l \rightarrow b} &= \frac{2D}{l^2} \frac{1}{1 + 4 \left(\frac{w}{l}\right)^2 \frac{\eta}{\xi}} \\ r_{b \rightarrow l} &= \frac{2D}{l^2} \frac{\eta}{\xi} \frac{1}{1 + 4 \left(\frac{w}{l}\right)^2 \frac{\eta}{\xi}} \\ r_{b \rightarrow r} &= \frac{2D}{l^2} \frac{1}{4 \left(\frac{w}{l}\right)^2 + \frac{\xi}{1-\eta-\xi}} \\ r_{r \rightarrow b} &= \frac{2D}{l^2} \frac{1}{1 + 4 \left(\frac{w}{l}\right)^2 \left(\frac{1-\eta-\xi}{\xi}\right)}. \end{aligned} \quad (5.16)$$

The quantities  $\xi$  and  $\eta$  serve as an intermediate coordinate system in the geometric construction of optimal protocols. They are explicitly written in terms of physical coordinates  $(V_b, V_l)$  in Eq. (5.58). The transition rates simplify considerably if we assume terms proportional to  $(w/l)^2$  are negligible. Physically, this means that the width of the barrier is negligible compared to the

width of the wells. In that case,

$$r_{l \rightarrow b} \approx \frac{2D}{l^2}, \quad r_{b \rightarrow l} \approx \frac{2D}{l^2} \frac{\eta}{\xi}, \quad r_{b \rightarrow r} \approx \frac{2D}{l^2} \frac{1 - \eta - \xi}{\xi}, \quad r_{r \rightarrow b} \approx \frac{2D}{l^2}. \quad (5.17)$$

From the perspective of our optimization problem, minimizing the dissipation is equivalent to minimizing the average work done on the system, so in addition to yielding an exact solution, this problem is amenable to the geometric framework for calculating optimal protocols. The inverse diffusion tensor can be calculated directly from the Fokker-Planck equation Eq. (5.3), allowing us to compare the exact answer with this approximate solution.

We take as control parameters the “tilt”  $V_l$  and the potential barrier height  $V_b$  (see Fig. 5.1(c)), and we initially focus on the class of protocols resulting in (partial) erasure of the classical bit. For example, increasing  $V_l$  and decreasing  $V_b$  appropriately as in Fig. 5.1(c) ensures near unity probability of finding the particle in the right well (*i.e.*, memory value of 0) regardless of its initial state.

We consider protocols consisting of two stages. During the first (erasure) stage, the initial equilibrium distribution transitions to a final nonequilibrium distribution in which the system is overwhelmingly likely to have memory value 0. In the second (reset) stage, the control parameters are brought instantaneously back to their original values while keeping the particle probability distribution constant. We allow these protocols to have jump discontinuities at the endpoints of each stage, and the optimal protocols will indeed exhibit them.

### 5.3 Exact optimizer

For the discrete dynamics, Eq. (5.9) simplifies to

$$\langle \beta W_{ex} \rangle_{\Lambda} = \int_{0^+}^{\bar{t}_f^-} d\bar{t} \left[ \dot{\lambda} \cdot \sum_i p_i(\bar{t}) \frac{\partial \phi_i}{\partial \lambda} \right] + \sum_i p_i(t_f) \beta U_i(\bar{t}) \Big|_{\bar{t}=\bar{t}_f^-}^{\bar{t}=\bar{t}_f^-} - \sum_i p_i(0) \beta U_i(\bar{t}) \Big|_{\bar{t}=0^+}^{\bar{t}=0^+} \quad (5.18)$$

where

$$\beta \Delta F \equiv \ln \left[ \frac{l + l e^{-\beta V_l(0^+)} + 2w e^{-\beta V_b(0^+)}}{l + l e^{-\beta V_l(t_f^-)} + 2w e^{-\beta V_b(t_f^-)}} \right]. \quad (5.19)$$

Integrating by parts,  $\langle \beta W \rangle_{\Lambda}$  equals

$$\sum_i p_i(t_f) \beta U_i(\bar{t}_f) - \sum_i p_i(0) \beta U_i(0) - \int_{0^+}^{\bar{t}_f^-} d\bar{t} \sum_i \phi_i \dot{p}_i. \quad (5.20)$$

It follows immediately that the average work performed during the reset stage of the cycle is given by

$$\langle \beta W \rangle_{\Lambda_{reset}} = \sum_i p_i(t_f) \left[ \beta U_i(0) - \beta U_i(t_f) \right] \quad (5.21)$$

and so

$$\langle \beta W \rangle_{\Lambda_{cycle}} = \sum_i [p_i(t_f) - p_i(0)] \beta U_i(0) - \int_{0^+}^{\bar{t}_f^-} d\bar{t} \sum_i \phi_i \dot{p}_i. \quad (5.22)$$

By the first law of stochastic thermodynamics,

$$\langle \beta Q \rangle_{\Lambda_{cycle}} = - \int_{0^+}^{\bar{t}_f^-} d\bar{t} \sum_i \phi_i \dot{p}_i. \quad (5.23)$$

The average heat over the cycle is given explicitly by

$$\int_{0^+}^{\bar{t}_f^-} d\bar{t} \left[ \dot{p}_l \ln(\eta) + \dot{p}_b \ln(\xi) + \dot{p}_r \ln(1 - \xi - \eta) \right], \quad (5.24)$$

or, since  $p_b = 1 - p_l - p_r$ ,

$$\int_{0^+}^{\bar{t}_f^-} d\bar{t} \left[ \dot{p}_l \ln\left(\frac{\eta}{\xi}\right) + \dot{p}_r \ln\left(\frac{1 - \xi - \eta}{\xi}\right) \right]. \quad (5.25)$$

We may write this integral and hence the total average heat lost over the cycle explicitly as a functional of the probabilities and then optimize over these variables. We begin by rewriting Eq. (5.54) as

$$\frac{1}{1 - p_l - p_r} \begin{pmatrix} \dot{p}_l + p_l \\ \dot{p}_r + 1 - p_l \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\eta}{\xi} \\ \frac{1}{\xi} \end{pmatrix}, \quad (5.26)$$

which allows us to solve for  $\eta$  and  $\xi$  explicitly in terms of the probabilities and their time derivatives:

$$\xi = \frac{1 - p_l - p_r}{1 + \dot{p}_l + \dot{p}_r}, \quad \eta = \frac{\dot{p}_l + p_l}{1 + \dot{p}_l + \dot{p}_r}. \quad (5.27)$$

Using these expressions and performing another integration by parts,

$$\langle \beta Q \rangle_{\Lambda_{cycle}} = Q_b + \int_{0^+}^{\bar{t}_f^-} d\bar{t} \left[ \dot{p}_l \ln(\dot{p}_l + p_l) + \dot{p}_r \ln(\dot{p}_r + p_r) \right], \quad (5.28)$$

with a boundary term defined as

$$Q_b \equiv p_b(t_f) [\ln(p_b(t_f)) - 1] - p_b(0) [\ln(p_b(0)) - 1], \quad (5.29)$$

where  $\bar{t} \equiv \frac{2D}{l^2} t$ ,  $\dot{p} \equiv \partial p / \partial \bar{t}$ , and  $f(t^\pm) \equiv \lim_{\delta \rightarrow 0^+} f(t \pm \delta)$

The boundary term  $Q_b$  depends only on the probability distributions at the endpoints. Moreover, we see that the “bulk” term of the average heat functional is a sum  $I[p_l] + I[p_r]$ , where

$$I[z] \equiv \int_{0^+}^{\bar{t}_f^-} d\bar{t} [\dot{z} \ln(\dot{z} + z)]. \quad (5.30)$$

Therefore, to extremize the average heat functional, we can solve the Euler-Lagrange equations for  $I$ .

Suppose  $z(t)$  satisfies the Euler-Lagrange equation for the Lagrangian  $L[z, \dot{z}] \equiv \dot{z} \ln(\dot{z} + z)$ . Then it must be true that

$$\dot{z} \frac{\partial L}{\partial \dot{z}} - L = \frac{(\dot{z})^2}{\dot{z} + z} \quad (5.31)$$

is a constant.

Therefore, probability distributions extremizing  $\langle \beta Q \rangle_{\Lambda_{cycle}}$  satisfy  $\dot{p}_i^2 = K_i (\dot{p}_i + p_i)$  for positive constants  $K_i$  and  $i = l, r$ . Over the course of the erasure stage,  $p_l$  ( $p_r$ ) decreases (increases). The constants  $K_i$  may in turn be numerically fixed by imposing the constraints  $p_l(0) = \frac{1}{2(1+\gamma)} = p_r(0)$ ,  $p_l(t_f) = \delta$ ,  $p_r(t_f) = 1 - 2\delta$ , where  $\delta$  and  $\gamma$  are small and positive. The uniqueness of our solution combined with the Second Law guarantee that this is the minimum we sought.

In the long duration regime,  $K_i$  is a very small quantity. This observation is important in determining the optimal efficiency in the long time regime. Recall that

$$\langle \beta Q \rangle_{\Lambda_{cycle}} = Q_b + \int_{0^+}^{\bar{t}_f^-} d\bar{t} [\dot{p}_l \ln(\dot{p}_l + p_l) + \dot{p}_r \ln(\dot{p}_r + p_r)] \quad (5.32)$$

for boundary term

$$Q_b \equiv p_b(t_f) [\ln(p_b(t_f)) - 1] - p_b(0) [\ln(p_b(0)) - 1] \quad (5.33)$$

and  $\epsilon \equiv (-\Delta S) / (k_B \langle \beta Q \rangle_{\Lambda_{cycle}})$  for change in Shannon entropy

$$\Delta S \equiv S_f - S_i \equiv -k_B \sum_i p_i^f \ln p_i^f + k_B \sum_i p_i^o \ln p_i^o. \quad (5.34)$$

To find the approximate efficiency in the long duration limit, we must approximate the integrals  $\int_{0^+}^{\bar{t}_f^-} d\bar{t} \dot{p}_i \ln(\dot{p}_i + p_i)$ . We first consider  $i = l$ : since

$$\dot{p}_l = \frac{K_l}{2} \left( 1 - \sqrt{1 + \frac{4}{K_l} p_l} \right), \quad (5.35)$$

we may rewrite the integral as

$$\int_{p_l^o}^{p_l^f} dz \ln \left[ \frac{K_l}{2} \left( 1 - \sqrt{1 + \frac{4}{K_l} z} \right) + z \right] \quad (5.36)$$

after a change of variables. We may factorize the logarithm as

$$\ln \left[ \frac{K_l}{2z} \left( 1 - \sqrt{1 + \frac{4}{K_l} z} \right) + 1 \right] + \ln(z). \quad (5.37)$$

In the long duration limit, we have

$$\ln \left[ \frac{K_l}{2z} \left( 1 - \sqrt{1 + \frac{4}{K_l} z} \right) + 1 \right] \approx -\sqrt{\frac{K_l}{z}}. \quad (5.38)$$

Therefore,

$$\int_{p_l^o}^{p_l^f} dz \ln \left[ \frac{K_l}{2} \left( 1 - \sqrt{1 + \frac{4}{K_l} z} \right) + z \right] \approx p_l^f \ln p_l^f - p_l^f - p_l^o \ln p_l^o + p_l^o - 2\sqrt{K_l} \left( \sqrt{p_l^f} - \sqrt{p_l^o} \right). \quad (5.39)$$

A similar calculation demonstrates

$$\int_{p_r^o}^{p_r^f} dz \ln \left[ \frac{K_r}{2} \left( 1 + \sqrt{1 + \frac{4}{K_r} z} \right) + z \right] \approx p_r^f \ln p_r^f - p_r^f - p_r^o \ln p_r^o + p_r^o + 2\sqrt{K_r} \left( \sqrt{p_r^f} - \sqrt{p_r^o} \right). \quad (5.40)$$

The total average heat dissipated during the cycle is equal to the sum of these two integrals plus the boundary term Eq. (5.33). The first two terms of each integral approximation plus the boundary term simply yield the Landauer term  $-\Delta S$ . This follows from the observation that the sum over  $p_i$  for  $i = l, b, r$  must be 1 both at the beginning and end of the time course.

Therefore,

$$\langle \beta Q \rangle_{\Lambda_{cycle}} \approx -\Delta S/k_B + 2\sqrt{K_r} \left( \sqrt{p_r^f} - \sqrt{p_r^o} \right) - 2\sqrt{K_l} \left( \sqrt{p_l^f} - \sqrt{p_l^o} \right). \quad (5.41)$$

Furthermore, since

$$\frac{\dot{p}_l}{p_l} = \frac{K_l}{2p_l} \left( 1 - \sqrt{1 + \frac{4}{K_l} p_l} \right) \approx -\sqrt{\frac{K_l}{p_l}} \quad (5.42)$$

and

$$\frac{\dot{p}_r}{p_r} = \frac{K_r}{2p_r} \left( 1 + \sqrt{1 + \frac{4}{K_r} p_r} \right) \approx \sqrt{\frac{K_r}{p_r}}, \quad (5.43)$$

we find that

$$-\sqrt{K_l} \bar{t}_f \approx 2 \left( \sqrt{p_l^f} - \sqrt{p_l^o} \right), \quad \sqrt{K_r} \bar{t}_f \approx 2 \left( \sqrt{p_r^f} - \sqrt{p_r^o} \right). \quad (5.44)$$

In the long duration limit, we obtain the simple result

$$\langle \beta Q \rangle_{\Lambda_{opt}} \approx \frac{-\Delta S}{k_B} + \frac{4K}{t_f}, \quad (5.45)$$

where

$$K \equiv \left( \sqrt{p_r(t_f)} - \sqrt{p_r(0)} \right)^2 + \left( \sqrt{p_l(t_f)} - \sqrt{p_l(0)} \right)^2 \quad (5.46)$$

is twice the square of the Hellinger distance [88], a measure of similarity between pairs of probability distributions. Note that  $K$  contains no terms for  $p_b(0)$  and  $p_b(t_f)$ , which are both small for the cases we consider.

Consistent with two previous studies [35, 33], we find that the total dissipation for our optimized protocols consists of the sum of two terms: one given by the Landauer bound, which is proportional to the decrease in Shannon entropy resulting from the erasure of information, and a second term that falls as  $1/t_f$  (Eq. (5.45)). Fortunately, we have arrived at a simple closed-form expression for the total dissipation, valid for arbitrary temperature, that can be experimentally tested using existing setups [82, 28] by comparing our optimal path through the two dimensional parameter space with alternate protocols. Fig. 5.2(a) depicts an optimally efficient finite-time erasure cycle constructed based on the calculations in this section. For the parameter values selected to generate Fig. 5.2, the erasure efficiency  $\epsilon$  is about 94%.

## 5.4 Inverse diffusion tensor-based approximation

The components of the inverse diffusion matrix [38] are

$$\zeta_{ij}(\boldsymbol{\lambda}) \equiv \int_0^\infty dt' \left\langle \frac{\partial \phi}{\partial \lambda^i}(t') \frac{\partial \phi}{\partial \lambda^j}(0) \right\rangle_{\boldsymbol{\lambda}}. \quad (5.47)$$

This matrix is symmetric and positive semi-definite in general for systems relaxing to an equilibrium state [36] and it defines a Riemannian geometry on the space of parameters.

We calculate the components of the inverse diffusion tensor using both the continuum and

discrete dynamics. In the continuum case, the components are given by

$$\zeta_{ij}(\boldsymbol{\lambda}) = \frac{1}{D} \int_{-l-w}^{l+w} dx \left[ \frac{\partial_{\lambda^i} \Pi_{eq}(x, \boldsymbol{\lambda}) \partial_{\lambda^j} \Pi_{eq}(x, \boldsymbol{\lambda})}{\rho_{eq}(x, \boldsymbol{\lambda})} \right]. \quad (5.48)$$

which is a specific case of Eq. (6.18) constructed in Ch. 6. Here,  $Z(\boldsymbol{\lambda}) \equiv \int_{-(l+w)}^{l+w} dx e^{-\beta U(x, \boldsymbol{\lambda})}$  is the classical partition function,  $\rho_{eq}(x, \boldsymbol{\lambda})$  is the equilibrium probability distribution and  $\Pi_{eq}(x, \boldsymbol{\lambda}) = \int_{-l-w}^x dx' \rho_{eq}(x', \boldsymbol{\lambda})$  is the cumulative distribution function.

Since the potential is piecewise constant, it is possible to evaluate the iterated integrals in Eq. (5.48) explicitly. However, the resulting expressions are quite complicated when left in terms of the ‘‘physical’’ parameters  $V_b$  and  $V_l$ . It is mathematically advantageous at this stage to make a coordinate transformation in parameter space so that the metric tensor components are compact. Define

$$\eta \equiv \frac{l e^{-\beta V_l}}{Z}, \quad \xi \equiv \frac{2w e^{-\beta V_b}}{Z}, \quad (5.49)$$

where  $Z = l + l e^{-\beta V_l} + 2w e^{-\beta V_b}$ . The inverse diffusion tensor in this coordinate system is

$$\zeta^{cont}(\boldsymbol{\lambda}) = \frac{1}{D} \begin{pmatrix} \frac{l^2}{3} \frac{1-\xi}{\eta(1-\eta-\xi)} + \frac{4w^2}{\xi} & \frac{2w^2}{\xi} + \frac{l^2}{3} \frac{1}{1-\eta-\xi} \\ \frac{2w^2}{\xi} + \frac{l^2}{3} \frac{1}{1-\eta-\xi} & \frac{4w^2}{3\xi} + \frac{l^2}{3(1-\eta-\xi)} \end{pmatrix}. \quad (5.50)$$

Since we are ignoring  $(w/l)^2$  terms,

$$\zeta^{cont}(\boldsymbol{\lambda}) \approx \frac{l^2}{3D} \frac{1}{1-\eta-\xi} \begin{pmatrix} \frac{1-\xi}{\eta} & 1 \\ 1 & 1 \end{pmatrix}. \quad (5.51)$$

For the discrete dynamics, Eq. (5.47) may be written as

$$\zeta_{ij}(\boldsymbol{\lambda}) = \int_0^\infty dt \sum_{\sigma, \sigma'} p_{\sigma'}(t|\sigma) p_\sigma^* \partial_{\lambda^i} \phi_\sigma \partial_{\lambda^j} \phi_{\sigma'}. \quad (5.52)$$

Here,  $p_\sigma^*$  denotes the equilibrium probability distribution

$$p_l^* = \eta, \quad p_b^* = \xi, \quad p_r^* = 1 - \eta - \xi, \quad (5.53)$$

$\phi_\sigma \equiv -\ln p_\sigma^*$ , and  $p_{\sigma'}(t|\sigma)$  represents the solution to the master equations Eq. (5.10) satisfying the initial condition  $p_{\sigma'}(0|\sigma) = \delta_{\sigma, \sigma'}$  for fixed  $\eta, \xi$ .

To obtain  $p_{\sigma'}(t|\sigma)$ , we recognize that  $p_b(t|\sigma) = 1 - p_l(t|\sigma) - p_r(t|\sigma)$  and write the master

equations Eq. (5.10) with rates Eq. (5.17) as a linear system of equations:

$$\frac{d}{d\bar{t}} \begin{pmatrix} p_l \\ p_r \end{pmatrix} = - \begin{pmatrix} 1 + \frac{\eta}{\xi} & \frac{\eta}{\xi} \\ \frac{1-\eta-\xi}{\xi} & \frac{1-\eta}{\xi} \end{pmatrix} \begin{pmatrix} p_l \\ p_r \end{pmatrix} + \begin{pmatrix} \frac{\eta}{\xi} \\ \frac{1-\eta-\xi}{\xi} \end{pmatrix}, \quad (5.54)$$

where  $\bar{t} \equiv \frac{2D}{l^2}t$ . This system may be solved by standard methods [66]. We find

$$\begin{aligned} \begin{pmatrix} p_l(t|1) \\ p_r(t|1) \end{pmatrix} &= \begin{pmatrix} \eta \left( \frac{\xi}{1-\xi} e^{-\frac{\bar{t}}{\xi}} + 1 \right) + \frac{1-\eta-\xi}{1-\xi} e^{-\bar{t}} \\ (1-\eta-\xi) \left( \frac{\xi}{1-\xi} e^{-\frac{\bar{t}}{\xi}} + 1 \right) - \frac{1-\eta-\xi}{1-\xi} e^{-\bar{t}} \end{pmatrix}, \\ \begin{pmatrix} p_l(t|2) \\ p_r(t|2) \end{pmatrix} &= \left( 1 - e^{-\frac{\bar{t}}{\xi}} \right) \begin{pmatrix} \eta \\ 1-\eta-\xi \end{pmatrix}, \\ \begin{pmatrix} p_l(t|3) \\ p_r(t|3) \end{pmatrix} &= \begin{pmatrix} \eta \left( \frac{\xi}{1-\xi} e^{-\frac{\bar{t}}{\xi}} + 1 \right) - \frac{\eta}{1-\xi} e^{-\bar{t}} \\ (1-\eta-\xi) \left( \frac{\xi}{1-\xi} e^{-\frac{\bar{t}}{\xi}} + 1 \right) + \frac{\eta}{1-\xi} e^{-\bar{t}} \end{pmatrix}. \end{aligned} \quad (5.55)$$

Evaluation of Eq. (5.52) using Eq. (5.55) yields

$$\zeta^{disc}(\boldsymbol{\lambda}) = \frac{l^2}{2D} \frac{1}{1-\eta-\xi} \begin{pmatrix} \frac{1-\xi}{\eta} & 1 \\ 1 & 1 \end{pmatrix}. \quad (5.56)$$

It follows that  $\zeta^{cont}(\boldsymbol{\lambda}) = 2/3 \zeta^{disc}(\boldsymbol{\lambda})$  when  $(w/l)^2$  is negligible. Both dynamics then yield precisely the same geodesics since Christoffel symbols are invariant under constant scalings of the metric tensor [67].

Ignoring constant prefactors, (approximate) optimal protocols are precisely the geodesics of the line element

$$d\Sigma^2 = \frac{1}{1-\eta-\xi} \left[ \frac{(1-\xi)}{\eta} (d\eta)^2 + 2 d\eta d\xi + (d\xi)^2 \right]. \quad (5.57)$$

The Ricci scalar vanishes identically; therefore, there must exist a coordinate transformation in which the line element is Euclidean [67]. Indeed, if we define  $x = 2\sqrt{1-\eta-\xi}$ ,  $y = 2\sqrt{\eta}$ , then  $d\Sigma^2 = dx^2 + dy^2$ .

Geodesics are most conveniently calculated in  $(x, y)$ -coordinates. Physically, we should express quantities in terms of the  $(V_b, V_l)$ -coordinate system. For convenience, we list here the explicit formulae allowing us to transform between the two coordinate systems via the intermediate coordinates  $(\eta, \xi)$ :

$$\begin{aligned} \eta &= \frac{l e^{-\beta V_l}}{l + l e^{-\beta V_l} + 2w e^{-\beta V_b}}, \quad \xi = \frac{2w e^{-\beta V_b}}{l + l e^{-\beta V_l} + 2w e^{-\beta V_b}}, \\ x &= 2\sqrt{1-\eta-\xi}, \quad y = 2\sqrt{\eta}, \end{aligned} \quad (5.58)$$

$$\xi = 1 - \frac{1}{4}(x^2 + y^2), \quad \eta = \frac{y^2}{4},$$

$$\beta V_b = \ln \left[ \frac{2w}{l} \frac{1 - \eta - \xi}{\xi} \right], \quad \beta V_l = \ln \left[ \frac{1 - \eta - \xi}{\eta} \right]. \quad (5.59)$$

As originally constructed [36], the formulation of the inverse diffusion tensor assumes smooth protocols on the entire domain of definition. Here we constructed an approximation for the optimizer in the interior of the erasure stage.

The question of what endpoints should be selected for the approximate optimizer is answered by enforcing the probability constraints at the endpoints. This is achieved by using the derivative truncation method [37, 38] to obtain the approximate time evolution of the probability distributions when the parameters are adjusted according to the geodesic protocol. This in turn gives us approximate formulae for the probability distributions at the beginning and end of the erasure stage in terms of the endpoints of the geodesic. A numerical solution of these constraints is easily obtained.

Note the strong agreement between the exact solution and the approximate solution (Fig. 5.2(b)) based on the inverse diffusion tensor.

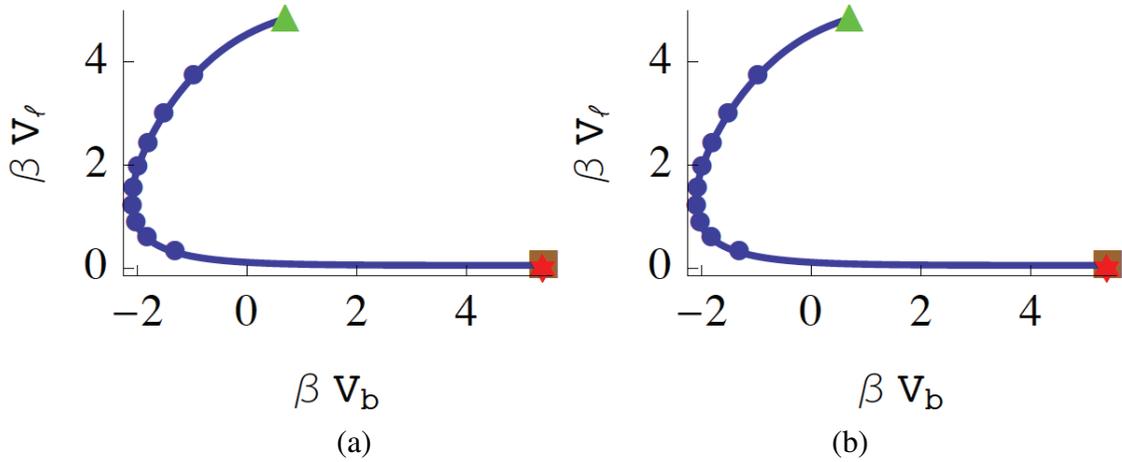


Figure 5.2: Optimally efficient finite-time erasure cycles. (a) The optimal cycle consists of two parts: the erasure stage and the reset stage. The erasure stage begins at  $(\ln\{w/(l\gamma)\}, 0)$  (red star) then jumps to the initial point of the erasure protocol (brown square). The erasure protocol (blue) proceeds for time  $t_f$  until reaching its terminus (green triangle). The reset stage consists of the jump from this terminus to the parameter values defining the original equilibrium state (red star). Blue dots indicate points separated by equal times along the erasure stage. For these parameters ( $\bar{t}_f \equiv (2D/l^2)t_f = 50$ ,  $\delta = 0.01$ ,  $\gamma = \exp(-10)$ ,  $w/l = 0.01$ ) the efficiency of the optimal cycle is 94.01%. (b) An approximate optimal efficiency erasure cycle determined by the inverse diffusion tensor framework is nearly identical to the exact solution shown in panel (a).

## 5.5 Beyond erasure

So far we have focused on the case of complete erasure of one bit, for which both the Landauer bound and the added dissipation necessary to achieve erasure in finite time are always positive. However, our formalism is completely general, valid for any change in the spatial distribution of the particle. Importantly, even in this broader setting, the second term is always nonnegative, as the Hellinger distance (and its square) is positive for any nonzero difference between the initial and final probability distributions, unlike the Landauer term, which can be zero or even negative.

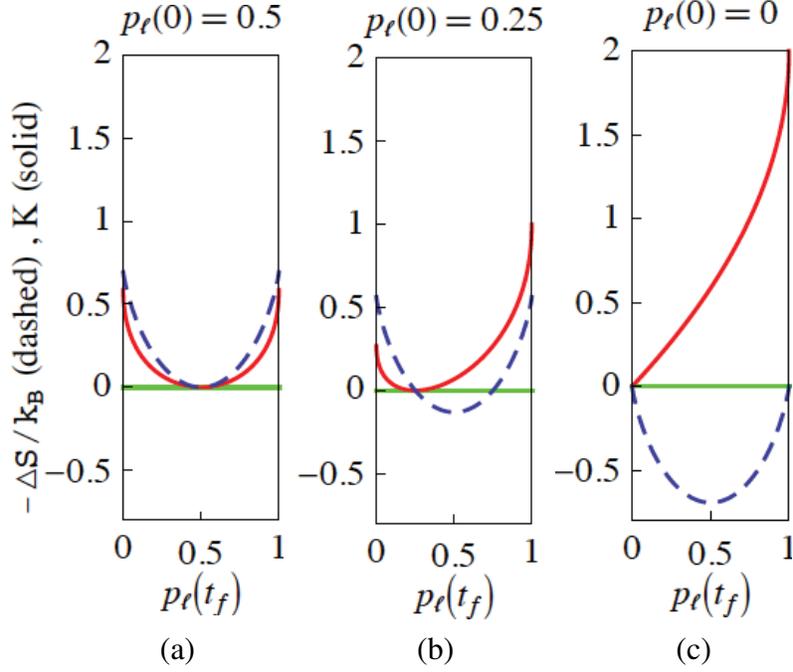


Figure 5.3: The Landauer bound and the added dissipation necessary for finite-time cycles exhibit different dependences on the spatial distribution of the particle. (a) If the particle is initially distributed equally between the two wells,  $p_l(0) = 0.5$ , where  $p_l(0)$  is the probability of being found in the left well at  $t = 0$ , then the Landauer bound, proportional to  $-\Delta S$  (dashed blue curve), is zero for no change in the likelihood of finding the particle on the left ( $p_l(t_f) = 0.5$ ) and positive for any other final distribution. This is always true for the second term in the full dissipation (solid red curve), which is proportional to  $K$  (Eq. (5.46)). In all panels,  $p_b(0) = p_b(t_f) = 0$ , where  $p_b(t)$  is the probability of finding the particle at the central barrier at time  $t$ . (b,c) For other initial conditions, the Landauer bound can be positive, zero, or negative depending on how the particle's spatial distribution changes.

## 5.6 Discussion

We have obtained a simple, closed-form expression for the dissipated heat of optimally efficient, finite-time erasure cycles, providing falsifiable predictions for currently achievable experiments. The solutions we have found are nontrivial, in that both of our control parameters are

continuously varying throughout the optimal protocol, yet our parametric solutions can be easily described. In addition to erasure, our solutions are valid for any initial and final particle distributions.

We find that the total dissipation for the optimal protocol consists of the Landauer bound plus a nonnegative second term proportional to the square of the Hellinger distance between the initial and final particle distributions (Eqs. (5.45,5.46)). Fittingly, one can think of this second term as a measure of how far out of equilibrium the system must be during the driving protocol, as it is the ratio of the “distance” between the initial and final probability distributions and the time allowed to make the transition. Indeed, one can show [88] that  $K$  (Eq. (5.46)) is a lower bound on the relative entropy  $D[\mathbf{p}(t_f) \parallel \mathbf{p}(0)] \equiv \sum_i p_i(t_f) \log[p_i(t_f)/p_i(0)]$  between the distributions, which is precisely the dissipation that would result from allowing the system to relax from the final distribution back to the initial equilibrium distribution with the control parameters held fixed to their initial values [89, 90, 91]. This is a tight bound in many cases; for example, for perfect erasure of one bit,  $K \approx 0.59$  and  $D[\mathbf{p}(t_f) \parallel \mathbf{p}(0)] \approx 0.69$  are comparable.

The exact and approximate optimal erasure cycles we found are nearly identical and both achieved high efficiencies for the finite cycle duration selected here, suggesting that the inverse diffusion tensor formalism could be an indispensable tool for predicting optimally efficient finite-time erasure cycles for complex model systems more relevant for biology or engineering.

## Chapter 6

# Inverse Diffusion Tensor for General Potentials

In Ch. 3 and Ch. 4, the method to calculate the inverse diffusion tensor components relied on the potential being harmonic. In this chapter, we demonstrate how the inverse diffusion tensor may be computed for a more general class of potentials in terms of the equilibrium probability distribution. Our starting point will be the Fokker-Planck equation which we assume fully describes the physics of the system. Furthermore, we show that the inverse diffusion tensor arises naturally through an expansion in temporal derivatives in this general setting.

We use this construction to compute optimal protocols for two model systems of physical interest. First, we consider a one-dimensional system modeling the storage and erasure of a single classical bit of information as in Ch. 5. There, we illustrated Landauer's principle that the erasure of information results in energy dissipation with a simple piece-wise constant potential. In this chapter, our erasure model consists of an overdamped colloidal particle diffusing under the influence of a continuous double-well potential with a large central barrier stabilizing the memory. If the particle is found to the left (right) of the origin, the memory value is 1 (0). We seek the most efficient protocols altering the shape of the confining potential so that the particle has overwhelmingly probability to be found to the right of the origin, thus setting the memory value to 0 and erasing the single bit of classical information originally encoded by the system.

The erasure cycle consists of a continuous stage in which the wells merge and the central barrier is lowered. A reset stage in which the potential returns instantaneously to its original state and leaves the final probability distribution undisturbed completes the erasure cycle. The inverse diffusion tensor predicts optimal erasure cycles in the long duration limit.

The second model system consists of an overdamped colloidal particle diffusing in one-dimension while coupled to a ratchet potential and an optical trap. In this simplified model, the coordinate of the diffusing particle may be identified with a mechanical state variable of a molecular motor [92] and the thermal bath consists of the huge number of irrelevant degrees of freedom of the liquid surrounding the motor as well as the internal degrees of freedom of the motor itself and the struc-

tures with which it interacts. The inverse diffusion tensor framework predicts the optimal time course for the optical trap center which represents an external driving of this simplistic molecular machine.

## 6.1 General construction

In this section, we present the general construction of the inverse diffusion tensor directly from the Fokker-Planck equation. We adapted the argument used below from [93]. We assume that the potential satisfies  $U(x, \boldsymbol{\lambda}) \rightarrow \infty$  as  $|x| \rightarrow \infty$ . However, the construction also applies for reflecting-wall boundary conditions.

For potential  $U(x, \boldsymbol{\lambda})$ , the Fokker-Planck equation governing the probability distribution  $\rho(x, t)$  in the overdamped case is

$$\frac{\partial \rho}{\partial t} = D \left[ \frac{\partial}{\partial x} \left( \frac{\partial (\beta U(x, \boldsymbol{\lambda}))}{\partial x} \rho \right) + \frac{\partial^2 \rho}{\partial x^2} \right] \quad (6.1)$$

where  $D$  is the diffusion coefficient. The inverse diffusion tensor components are given by

$$\zeta_{ij}(\boldsymbol{\lambda}) = \int_0^\infty dt' \left\langle \frac{\partial \phi}{\partial \lambda^i}(t') \frac{\partial \phi}{\partial \lambda^j}(0) \right\rangle_{eq, \boldsymbol{\lambda}} \quad (6.2)$$

where  $\phi(x, \boldsymbol{\lambda}) \equiv -\ln \rho_{eq}(x, \boldsymbol{\lambda})$  and  $\rho_{eq}(x, \boldsymbol{\lambda})$  is the equilibrium distribution.

We rewrite Eq. (6.2) as

$$\zeta_{ij}(\boldsymbol{\lambda}) = \int_0^\infty dt' \left[ \int_{-\infty}^\infty dx_0 \rho_{eq}(x_0, \boldsymbol{\lambda}) \partial_{\lambda^i} \phi(x_0, \boldsymbol{\lambda}) \times \left( \int_{-\infty}^\infty dx \rho(x, t'; x_0) \partial_{\lambda^j} \phi(x, \boldsymbol{\lambda}) \right) \right] \quad (6.3)$$

where  $\rho(x, t; x_0)$  satisfies Eq. (6.1) with initial condition  $\rho(x, t = 0; x_0) = \delta_{x, x_0}$  and  $\rho(x, t; x_0) \rightarrow 0$  for  $|x| \rightarrow \infty$ . For simplicity, define

$$m_i(t) \equiv \int_{-\infty}^\infty dx \rho(x, t; x_0) \partial_{\lambda^i} \phi(x, \boldsymbol{\lambda}) \quad (6.4)$$

so that

$$\zeta_{ij}(\boldsymbol{\lambda}) = \int_{-\infty}^\infty dx_0 \rho_{eq}(x_0, \boldsymbol{\lambda}) \partial_{\lambda^i} \phi(x_0, \boldsymbol{\lambda}) \int_0^\infty dt' m_i(t'). \quad (6.5)$$

Note that we have suppressed the dependence of  $m_i$  on  $x_0$  and  $\boldsymbol{\lambda}$  for convenience. We evaluate  $\int_0^\infty dt' m_i(t')$  by computing the Laplace transform

$$\hat{m}_i(s) \equiv \int_0^\infty dt' m_i(t') e^{-st'} \quad (6.6)$$

and taking the limit as  $s \rightarrow 0^+$ .

Integrating by parts,

$$\int_0^\infty dt' \frac{dm_i}{dt'}(t') e^{-st'} = s \hat{m}_i(s) - m_i(0). \quad (6.7)$$

Note that  $m_i(\infty)$  vanishes since  $\lim_{t \rightarrow \infty} \rho(x, t; x_0) = \rho_{eq}(x, \boldsymbol{\lambda})$ ; *i.e.* the system comes to a stationary state after a sufficiently long time has elapsed. By definition of  $m_i$ ,

$$m_i'(t) = \int_{-\infty}^\infty dx \partial_t \rho(x, t; x_0) \partial_{\lambda^i} \phi(x, \boldsymbol{\lambda}). \quad (6.8)$$

In terms of the probability current  $G(x, t)$ ,

$$\hat{m}_i(s) = \frac{m_i(0) - \int_{-\infty}^\infty dx \partial_x \hat{G}(x, s) \partial_{\lambda^i} \phi(x, \boldsymbol{\lambda})}{s}. \quad (6.9)$$

Therefore, to compute  $\hat{m}_i(s)$ , we need the Laplace transform of the probability current.

The Fokker-Planck equation may be used to derive an equation for the probability current:

$$\partial_t G(x, t) = D [\beta U'(x, \boldsymbol{\lambda}) \partial_x G(x, t) + \partial_x^2 G(x, t)]. \quad (6.10)$$

Taking the Laplace transform of both sides, we have

$$s \hat{G}(x, s) - G(x, 0) = D [\beta U'(x, \boldsymbol{\lambda}) \partial_x \hat{G}(x, s) + \partial_x^2 \hat{G}(x, s)] \quad (6.11)$$

which follows from  $\lim_{t \rightarrow \infty} G(x, t) = 0$ . Multiplying both sides by  $s$  and defining  $H(x, s) \equiv s \hat{G}(x, s)$ ,

$$s H(x, s) - s G(x, 0) = D [\beta U'(x, \boldsymbol{\lambda}) \partial_x H(x, s) + \partial_x^2 H(x, s)]. \quad (6.12)$$

We may obtain a solution to Eq. (6.12) by expanding  $H(x, s)$  as a series in  $s$ . If we define  $H(x, s) \equiv H_0(x) + s H_1(x) + \dots$ , then

$$0 = \beta U'(x, \boldsymbol{\lambda}) H_0'(x) + H_0''(x) \quad (6.13)$$

$$H_0(x) - G(x, 0) = D [\beta U'(x, \boldsymbol{\lambda}) \partial_x H_1(x) + \partial_x^2 H_1(x)] \quad (6.14)$$

$$H_{k-1}(x) = D [\beta U'(x, \boldsymbol{\lambda}) \partial_x H_k(x) + \partial_x^2 H_k(x)] \quad (6.15)$$

follow from substituting the expansion into Eq. (6.12) and comparing the coefficients of powers of  $s$  on both sides. The boundary conditions on the probability current must also be satisfied by  $H_k$  for each  $k$ .

We see that these differential equations may be solved iteratively. Fortunately, it turns out that

only  $H_2(x)$  is needed for our purposes as a short calculation using Eq. (6.9) shows that

$$\int_0^\infty dt' m_i(t') = - \int_{-\infty}^\infty dx \partial_x H_2(x) \partial_{\lambda^i} \phi(x, \boldsymbol{\lambda}). \quad (6.16)$$

For potentials that grow unbounded as  $|x| \rightarrow \infty$ , the probability current must vanish in the limit of large  $|x|$ . With these boundary conditions it is not difficult to show

$$H_2(x) = \frac{1}{D} \left[ - \left( \int_{-\infty}^\infty dx e^{-\beta U(x, \boldsymbol{\lambda})} \int_a^x dx' e^{\beta U(x', \boldsymbol{\lambda})} (\theta(x' - x_0) - \Pi_{eq}(x', \boldsymbol{\lambda})) \right) \Pi_{eq}(x, \boldsymbol{\lambda}) + \int_{-\infty}^x dx' e^{-\beta U(x', \boldsymbol{\lambda})} \int_a^{x'} dx'' e^{\beta U(x'', \boldsymbol{\lambda})} (\theta(x'' - x_0) - \Pi_{eq}(x'', \boldsymbol{\lambda})) \right] \quad (6.17)$$

where  $\theta$  denotes the Heaviside function and  $\Pi_{eq}(x, \boldsymbol{\lambda}) = \int_{-\infty}^x dx' \rho_{eq}(x', \boldsymbol{\lambda})$  is the equilibrium cumulative distribution function. Here,  $a$  is an arbitrary real constant. Surprisingly,  $H_2(x)$  is independent of  $a$  and we will have occasion to choose different convenient values for computational purposes.

From this result we see that the inverse diffusion tensor has the compact form

$$\zeta_{ij}(\boldsymbol{\lambda}) = \int d\nu(x, x', x'') \left[ e^{\beta(U(x', \boldsymbol{\lambda}) - U(x, \boldsymbol{\lambda}) - U(x'', \boldsymbol{\lambda}))} \partial_{\lambda^i} \phi(x, \boldsymbol{\lambda}) \partial_{\lambda^j} \phi(x'', \boldsymbol{\lambda}) \right] \quad (6.18)$$

where we have used the shorthand

$$\int d\nu(x, x', x'') \rightarrow -\frac{1}{DZ(\boldsymbol{\lambda})} \int_{-\infty}^\infty dx \int_a^x dx' \int_{-\infty}^{x'} dx''. \quad (6.19)$$

This expression may be further simplified by observing that

$$\int_{-\infty}^{x'} dx'' \partial_{\lambda^j} \phi(x'', \boldsymbol{\lambda}) e^{-\beta U(x'', \boldsymbol{\lambda})} = -Z(\boldsymbol{\lambda}) \partial_{\lambda^j} \Pi_{eq}(x', \boldsymbol{\lambda}) \quad (6.20)$$

and

$$\partial_{\lambda^i} \phi(x, \boldsymbol{\lambda}) e^{-\beta U(x, \boldsymbol{\lambda})} = -Z(\boldsymbol{\lambda}) \partial_{\lambda^i} \rho_{eq}(x, \boldsymbol{\lambda}) \quad (6.21)$$

which follow from the definition of the nonequilibrium potential  $\phi$ . These expressions may be used to rewrite Eq. (6.18) as

$$\zeta_{ij}(\boldsymbol{\lambda}) = -\frac{Z(\boldsymbol{\lambda})}{D} \int_{-\infty}^\infty dx \left[ \partial_{\lambda^i} \rho_{eq}(x, \boldsymbol{\lambda}) \int_a^x dx' e^{\beta U(x', \boldsymbol{\lambda})} \partial_{\lambda^j} \Pi_{eq}(x', \boldsymbol{\lambda}) \right] \quad (6.22)$$

or

$$\zeta_{ij}(\boldsymbol{\lambda}) = -\frac{Z(\boldsymbol{\lambda})}{D} \int_{-\infty}^{\infty} dx \left[ \partial_x \left( \partial_{\lambda^i} \Pi_{eq}(x, \boldsymbol{\lambda}) \right) \int_a^x dx' e^{\beta U(x', \boldsymbol{\lambda})} \partial_{\lambda^j} \Pi_{eq}(x', \boldsymbol{\lambda}) \right]. \quad (6.23)$$

If

$$\lim_{x \rightarrow \pm\infty} \partial_{\lambda^i} \Pi_{eq}(x, \boldsymbol{\lambda}) \int_a^x dx' e^{\beta U(x', \boldsymbol{\lambda})} \partial_{\lambda^j} \Pi_{eq}(x', \boldsymbol{\lambda}) = 0, \quad (6.24)$$

then

$$\zeta_{ij}(\boldsymbol{\lambda}) = \frac{1}{D} \int_{-\infty}^{\infty} dx \left[ \frac{\partial_{\lambda^i} \Pi_{eq}(x, \boldsymbol{\lambda}) \partial_{\lambda^j} \Pi_{eq}(x, \boldsymbol{\lambda})}{\rho_{eq}(x, \boldsymbol{\lambda})} \right]. \quad (6.25)$$

As an example, this equation may be used to construct the components of the inverse diffusion tensor arising in Ch. 5 for the classical bit system with continuum dynamics. These components are recorded in Eq. (5.50). The potential is piecewise constant in the interval  $(-l - w, l + w)$  and  $+\infty$  elsewhere. Specifically,

$$U(x, \boldsymbol{\lambda}) = \begin{cases} V_l & -(l + w) < x < -w \\ V_b & -w < x < w \\ 0 & w < x < l + w \\ \infty & \text{otherwise} \end{cases} \quad (6.26)$$

where (for now)  $\boldsymbol{\lambda} = \{V_l, V_b\}$ . The integral appearing in Eq. (6.25) may be computed through a straightforward calculation.

Before we proceed, we note that computations using the physical coordinates  $\boldsymbol{\lambda} = \{V_l, V_b\}$  very rapidly become unwieldy. We define

$$\eta \equiv \frac{l e^{-\beta V_l}}{Z}, \quad \xi \equiv \frac{2w e^{-\beta V_b}}{Z} \quad (6.27)$$

where  $Z = l + l e^{-\beta V_l} + 2w e^{-\beta V_b}$  to facilitate the following calculations. In other words, we use the equilibrium probabilities of locating the particle in each compartment as control parameters. In terms of these parameters, the equilibrium distribution is given by

$$\rho_{eq}(x, \boldsymbol{\lambda}) = \begin{cases} \frac{\eta}{l} & -(l + w) < x < -w \\ \frac{\xi}{2w} & -w < x < w \\ \frac{1 - \eta - \xi}{l} & w < x < l + w \end{cases}. \quad (6.28)$$

The equilibrium distribution has a relatively compact form when expressed in terms of  $\eta$  and  $\xi$ .

The corresponding cumulative distribution function  $\Pi_{eq}(x, \boldsymbol{\lambda})$  is given by

$$\Pi_{eq}(x, \boldsymbol{\lambda}) = \begin{cases} \eta\left(\frac{x}{l} + 1 + \frac{w}{l}\right) & -(l+w) < x < -w \\ \xi\left(\frac{x}{2w} + \frac{1}{2}\right) + \eta & -w < x < w \\ \xi + \eta + (1 - \eta - \xi)\frac{x-w}{l} & w < x < l+w \end{cases}. \quad (6.29)$$

We see that

$$e^{\beta U(x', \boldsymbol{\lambda})} \partial_{\xi} \Pi_{eq}(x', \boldsymbol{\lambda}) = \begin{cases} 0 & -(l+w) < x' < -w \\ e^{\beta V_b}(x'/(2w) + 1/2) & -w < x' < w \\ 1 - (x' - w)/l & w < x' < l+w \end{cases} \quad (6.30)$$

and

$$e^{\beta U(x', \boldsymbol{\lambda})} \partial_{\eta} \Pi_{eq}(x', \boldsymbol{\lambda}) = \begin{cases} e^{\beta V_l}(x'/l + 1 + w/l) & -(l+w) < x' < -w \\ e^{\beta V_b} & -w < x' < w \\ 1 - (x' - w)/l & w < x' < l+w \end{cases} \quad (6.31)$$

and so

$$\begin{aligned} \zeta_{\eta\eta} &= \frac{1}{D} \left[ \frac{l^2}{3} \frac{1 - \xi}{\eta(1 - \eta - \xi)} + \frac{4w^2}{\xi} \right] \\ \zeta_{\eta\xi} &= \frac{1}{D} \left[ \frac{l^2}{3(1 - \eta - \xi)} + \frac{2w^2}{\xi} \right] \\ \zeta_{\xi\xi} &= \frac{1}{D} \left[ \frac{l^2}{3} \frac{1}{1 - \eta - \xi} + \frac{4w^2}{3\xi} \right] \end{aligned} \quad (6.32)$$

precisely as recorded in Eq. (5.50).

### 6.1.1 Example: Erasure Model

We consider the following model to represent a single classical bit of information: an overdamped Brownian colloidal particle diffusing in a one-dimensional double-well potential in contact with a thermal bath of temperature  $T$  [29, 28]. The wells are initially separated by a potential barrier whose height is much larger than the energy scale  $\beta^{-1} \equiv k_B T$  set by thermal fluctuations, ensuring stability of memory. Explicitly,

$$U(x, \lambda) \equiv -\frac{1}{\beta} \log \left[ \frac{\alpha e^{-\alpha(x-1+\lambda)}}{(1 + e^{-\alpha(x-1+\lambda)})^2} + \frac{\alpha e^{-\alpha(x-1)}}{(1 + e^{-\alpha(x-1)})^2} \right] \quad (6.33)$$

where  $x$  is a dimensionless spatial coordinate and  $\alpha \gg 1$ . Initially,  $\lambda = 2$  and there are two distinct wells and a central barrier with height governed by  $\alpha$ . As  $\lambda$  decreases to 0, the barrier height diminishes and the left-hand well shifts to merge with the right-hand well.

The system is prepared so that the particle has equal probability of being found in either well. This may be achieved, for example, by selecting the initial position of the particle to be at the midpoint of the potential barrier and waiting a sufficiently long relaxation period [29]. If the particle is found to the left (right) of the origin, the memory value is defined to be 1 (0).

We are primarily interested in optimizing finite-time erasure efficiency over cyclic protocols for the classical single bit model described above. When classical information is being erased, the difference in Shannon entropies of the final and initial probability distributions must satisfy  $\Delta S \equiv S_f - S_i < 0$ , which would allow us to define the erasure efficiency  $\epsilon \equiv -\Delta S / (k_B \langle \beta Q \rangle_\Lambda)$  as the ratio of this decrease in Shannon entropy to the average heat  $\langle Q \rangle_\Lambda$  released into the thermal bath [33, 83].

Our goal will be to minimize the dissipated heat over protocols in which  $\lambda$  decreases from 2 to 0. According to the first law of stochastic thermodynamics [55],

$$\langle \beta W_{ex} \rangle_\Lambda + \frac{-\Delta S}{k_B} = \langle \beta Q \rangle_\Lambda \quad (6.34)$$

and so

$$\epsilon = \frac{1}{1 + \frac{\langle \beta W_{ex} \rangle_\Lambda}{(-\Delta S)/k_B}}, \quad (6.35)$$

Therefore, minimizing the dissipation is equivalent to minimizing the average work done on the system by the first law of stochastic thermodynamics. Furthermore, minimizing the average excess work done on the system is equivalent to maximizing the efficiency of the erasure cycle.

The inverse diffusion tensor formalism allows us to compute an approximation to the optimal time course for  $\lambda$ . We consider protocols consisting of two stages. During the first (erasure) stage, the initial equilibrium distribution transitions to a final nonequilibrium distribution in which the system is overwhelmingly likely to have memory value 0. Equivalently, the control parameter  $\lambda$  decreases from 2 to 0 during this erasure stage. In the second (reset) stage,  $\lambda$  is brought instantaneously back to its original value while keeping the particle probability distribution constant. No heat is generated during this stage.

From Eq. (6.25), we have

$$\zeta(\lambda) = \frac{1}{2D} \left\{ 1 - \frac{\frac{\pi}{2} - \tan^{-1} \left( \sqrt{2 / [\cosh(\alpha\lambda) - 1]} \right)}{\sqrt{2 [\cosh(\alpha\lambda) - 1]}} \right\}. \quad (6.36)$$

The equilibrium distribution for the potential Eq. (6.33) has the form

$$\rho_{eq}(x, \lambda) = \frac{1}{2} [f(x + \lambda) + f(x)] \quad (6.37)$$

for

$$f(x) = \frac{\alpha e^{-\alpha(x-1)}}{(1 + e^{-\alpha(x-1)})^2} = \frac{d}{dx} \left[ \frac{1}{1 + e^{-\alpha(x-1)}} \right] \quad (6.38)$$

by construction.

It is straightforward to show that

$$\partial_\lambda \Pi_{eq}(x, \lambda) = \frac{1}{2} f(x + \lambda) \quad (6.39)$$

and

$$\frac{[\partial_\lambda \Pi_{eq}]^2}{\rho_{eq}} = \frac{1}{2} \frac{f(x + \lambda)^2}{f(x + \lambda) + f(x)} = \frac{1}{2} \left\{ f(x + \lambda) - \frac{1}{\frac{1}{f(x+\lambda)} + \frac{1}{f(x)}} \right\}. \quad (6.40)$$

Therefore,

$$\begin{aligned} \int_{-\infty}^{\infty} dx \frac{[\partial_\lambda \Pi_{eq}]^2}{\rho_{eq}} &= \frac{1}{2} \left\{ 1 - \int_{-\infty}^{\infty} dx \frac{1}{\frac{1}{f(x+\lambda)} + \frac{1}{f(x)}} \right\} \\ &= \frac{1}{2} \left\{ 1 - \int_{-\infty}^{\infty} dx \frac{\alpha e^{-\alpha(x-1)}}{(1 + e^{-\alpha(x-1)})^2 + e^{\alpha\lambda} (1 + e^{-\alpha\lambda} e^{-\alpha(x-1)})^2} \right\} \end{aligned} \quad (6.41)$$

Making the change of variables  $u = e^{-\alpha(x-1)}$ , we have that

$$\begin{aligned} \int_{-\infty}^{\infty} dx \frac{[\partial_\lambda \Pi_{eq}]^2}{\rho_{eq}} &= \frac{1}{2} \left\{ 1 - \int_0^{\infty} \frac{du}{(1+u)^2 + e^{\alpha\lambda} (1 + e^{-\alpha\lambda} u)^2} \right\} \\ &= \frac{1}{2} \left\{ 1 - \frac{1}{1 + e^{-\alpha\lambda}} \int_{\frac{2}{(1+e^{-\alpha\lambda})}}^{\infty} \frac{dv}{v^2 + \left( \frac{e^{-\alpha\lambda} + e^{\alpha\lambda} - 2}{(1+e^{-\alpha\lambda})^2} \right)} \right\} \\ &= \frac{1}{2} \left\{ 1 - \frac{1}{\sqrt{2} [\cosh(\alpha\lambda) - 1]} \left( \frac{\pi}{2} - \tan^{-1} \left( \sqrt{\frac{2}{\cosh(\alpha\lambda) - 1}} \right) \right) \right\} \end{aligned} \quad (6.42)$$

which implies Eq. (6.36).

Since

$$\langle \beta W_{ex} \rangle_{\mathbf{\Lambda}} \approx \int_0^\tau dt \zeta(\lambda(t)) \left( \frac{d\lambda}{dt} \right)^2, \quad (6.43)$$

the Euler-Lagrange equation for the optimal time course is

$$\frac{d\lambda}{dt} = - \frac{\int_0^2 dz \sqrt{\zeta(z)}/\tau}{\sqrt{\zeta(\lambda(t))}}. \quad (6.44)$$

Eq. (6.44) yields an implicit expression for the optimal protocol:

$$\int_{\lambda}^2 dz \sqrt{\zeta(z)} / \int_0^2 dz \sqrt{\zeta(z)} = \frac{t}{\tau}. \quad (6.45)$$

### 6.1.2 Example: Ratchet model with optical trap coupling

We consider an overdamped colloidal particle diffusing in one dimension subject to optical trap confinement and a tilted ratchet potential [92]. Mathematically,

$$\beta U(x, \boldsymbol{\lambda}) = \frac{1}{2} \beta k (x - x_0)^2 - \beta F x + \beta V_0 \varphi(x/l). \quad (6.46)$$

Here,  $l$  is some characteristic length scale,  $F$  is the “tilt” of the ratchet, and  $V_0$  is the magnitude of the ratchet potential. Furthermore, we choose a single control parameter, namely  $x_0$ , the position of the center of the optical trap.

The model possesses relative mathematical simplicity and captures essential physics of chemical processes relevant to the operation of cellular machinery. We view the following as the first step towards applying the inverse diffusion tensor framework to optimization of nonequilibrium processes underlying the functionality of nanoscale biological motors and machines.

If we consider an isothermal chemical reaction in the presence of a catalyst protein (*i.e.* an enzyme), then the reaction can be described by a single reaction coordinate, cycling through a number of chemical states in the simplest case [92]. A suitable working model is then an overdamped Brownian particle (reaction coordinate) in the presence of thermal fluctuations in a periodic potential. For this reason, we select  $\varphi(y) = \sin(y)$ .

If the concentrations of the reactants and products are away from their equilibrium ratio, then the catalyst molecule will loop through the chemical reaction cycle preferably in one direction [92]. In the corresponding ratchet model, the periodic potential must be supplemented by a constant tilt, *i.e.*  $F$ .

Define

$$\begin{aligned} \alpha &\equiv \beta k l^2, \quad y \equiv x/l, \quad f \equiv F/(kl), \\ u_0(y, y_0) &\equiv \frac{1}{2} (y - y_0)^2 - f y, \quad \epsilon \equiv V_0/(kl^2) \end{aligned} \quad (6.47)$$

so that  $\beta U(x, x_0) = \alpha [u_0(y, y_0) + \epsilon \varphi(y)]$ . Assuming the strength of the optical trap far exceeds the strength of the ratchet potential,  $\epsilon$  is a small parameter and we may apply perturbation theory; *i.e.* we expand all quantities to first order in  $\epsilon$  and discard higher order terms.

It is straightforward to show that

$$\rho_{eq}(y, y_0) \approx \frac{1}{l} \sqrt{\frac{\alpha}{2\pi}} e^{-\frac{1}{2}\alpha(y-y_0-f)^2} \left[ 1 - \alpha \epsilon (\varphi(y) - \langle \varphi \rangle_0) \right] \quad (6.48)$$

where the subscript “0” indicates an average with  $\epsilon = 0$ .

Using Eq. (6.25), we compute

$$\zeta(y_0) \approx \frac{l^2}{D} (1 - 2\alpha\epsilon\langle\varphi''\rangle_0). \quad (6.49)$$

The equilibrium distribution for the ratchet model (Eq. (6.48)) is

$$\rho_{eq}(y, y_0) \approx \frac{1}{l} \sqrt{\frac{\alpha}{2\pi}} e^{-\frac{1}{2}\alpha(y-y_0-f)^2} \left[ 1 - \alpha\epsilon(\varphi(y) - \langle\varphi\rangle_0) \right] \quad (6.50)$$

with cumulative distribution function

$$\Pi_{eq}(y, y_0) = \int_{-\infty}^y dy' \rho_{eq}(y', y_0). \quad (6.51)$$

Since

$$\partial_{y_0}\langle\varphi\rangle_0 = \langle\varphi'\rangle_0, \quad (6.52)$$

we have that

$$\begin{aligned} \partial_{y_0}\Pi_{eq}(y, y_0) = & \sqrt{\frac{\alpha}{2\pi}} \left\{ -e^{-\frac{1}{2}\alpha(y-y_0-f)^2} + \alpha\epsilon \left( (\varphi(y) - \langle\varphi\rangle_0) e^{-\frac{1}{2}\alpha(y-y_0-f)^2} - \right. \right. \\ & \left. \left. \int_{-\infty}^y dy' e^{-\frac{1}{2}\alpha(y'-y_0-f)^2} (\varphi'(y') - \langle\varphi'\rangle_0) \right) \right\} \end{aligned} \quad (6.53)$$

and so

$$\begin{aligned} \partial_{y_0}\Pi_{eq}(y, y_0) e^{\beta U(y, y_0)} = & \sqrt{\frac{\alpha}{2\pi}} e^{-\alpha(fy_0 + \frac{1}{2}f^2)} \left\{ -1 + \alpha\epsilon \left( \varphi(y) - \langle\varphi\rangle_0 - e^{\frac{1}{2}\alpha(y-y_0-f)^2} \int_y^\infty dy' e^{-\frac{1}{2}\alpha(y'-y_0-f)^2} \times \right. \right. \\ & \left. \left. (\varphi'(y') - \langle\varphi'\rangle_0) \right) \right\}. \end{aligned} \quad (6.54)$$

According to [94],

$$e^{w^2} \int_w^\infty e^{-z^2} dz \leq \frac{1}{w + \sqrt{w^2 + \frac{4}{\pi}}} \quad (6.55)$$

for  $w \geq 0$ . Since  $\varphi$  and all of its derivatives are bounded on  $\mathbb{R}$ ,

$$\lim_{y \rightarrow \infty} \partial_{y_0}\Pi_{eq}(y, y_0) \int_0^y dy' e^{\beta U(y', y_0)} \partial_{y_0}\Pi_{eq}(y', y_0) = 0. \quad (6.56)$$

A similar argument shows that

$$\lim_{y \rightarrow -\infty} \partial_{y_0} \Pi_{eq}(y, y_0) \int_0^y dy' e^{\beta U(y', y_0)} \partial_{y_0} \Pi_{eq}(y', y_0) = 0, \quad (6.57)$$

and so we may apply Eq. (6.25) to compute the inverse diffusion tensor.

Using Eqs. (6.50) and (6.53), we see that

$$\frac{[\partial_{y_0} \Pi_{eq}]^2}{\rho_{eq}} \approx l \sqrt{\frac{\alpha}{2\pi}} \left\{ e^{-\frac{\alpha}{2}(y-y_0-f)^2} + \alpha \epsilon \left( -(\varphi(y) - \langle \varphi \rangle_0) \times \right. \right. \\ \left. \left. e^{-\frac{\alpha}{2}(y-y_0-f)^2} + 2 \int_{-\infty}^y dy' (\varphi'(y') - \langle \varphi' \rangle_0) e^{-\frac{\alpha}{2}(y'-y_0-f)^2} \right) \right\}. \quad (6.58)$$

From Eq. (6.25) we have

$$\zeta(y_0) \approx \frac{l^2}{D} \left\{ 1 + 2\alpha \epsilon \int_{-\infty}^{\infty} dy \int_{-\infty}^y dy' (\varphi'(y') - \langle \varphi' \rangle_0) \times \right. \\ \left. e^{-\frac{\alpha}{2}(y'-y_0-f)^2} / \sqrt{\frac{2\pi}{\alpha}} \right\}. \quad (6.59)$$

The integral may be evaluated using integration by parts:

$$\int_{-\infty}^{\infty} dy \int_{-\infty}^y dy' (\varphi'(y') - \langle \varphi' \rangle_0) e^{-\frac{\alpha}{2}(y'-y_0-f)^2} = \\ \int_{-\infty}^{\infty} dy \left\{ \frac{d}{dy} (y - y_0 - f) \int_{-\infty}^y dy' (\varphi'(y') - \langle \varphi' \rangle_0) e^{-\frac{\alpha}{2}(y'-y_0-f)^2} \right\} = \\ \int_{-\infty}^{\infty} dy (y - y_0 - f) (\langle \varphi' \rangle_0 - \varphi'(y)) e^{-\frac{\alpha}{2}(y-y_0-f)^2} = \\ \int_{-\infty}^{\infty} dy (\varphi'(y) - \langle \varphi' \rangle_0) \frac{d}{dy} \left( \frac{1}{\alpha} e^{-\frac{\alpha}{2}(y-y_0-f)^2} \right) = \\ -\frac{1}{\alpha} \int_{-\infty}^{\infty} dy \varphi''(y) e^{-\frac{\alpha}{2}(y-y_0-f)^2} = -\sqrt{\frac{2\pi}{\alpha}} \langle \varphi'' \rangle_0. \quad (6.60)$$

Therefore

$$\zeta(y_0) \approx \frac{l^2}{D} (1 - 2\alpha \epsilon \langle \varphi'' \rangle_0). \quad (6.61)$$

For  $\varphi(y) = \sin(y)$ ,

$$\zeta(y_0) \approx \frac{l^2}{D} \left( 1 + 2\epsilon e^{-\frac{1}{2\alpha}} \sin(y_0 + f) \right). \quad (6.62)$$

Since

$$\langle \beta W_{ex} \rangle_{\Lambda} \approx \int_0^{\tau} dt \zeta(y_0) \left( \frac{dy_0}{dt} \right)^2, \quad (6.63)$$

the optimal time course for the trap center is given by the Euler-Lagrange equation

$$\frac{dy_0}{dt} = \frac{C}{\sqrt{\frac{D}{l^2}} \sqrt{\zeta(y_0)}}. \quad (6.64)$$

Here,  $C$  is a constant of integration.

Using perturbation theory, an approximate solution is given by

$$y_0(t) \approx Ct + B + \epsilon e^{-\frac{1}{2\alpha}} \cos(Ct + f + A) \quad (6.65)$$

where the constants  $A$ ,  $B$  and  $C$  are determined by the endpoints of the protocol and the duration  $\tau$ . Note that the temporal periodicity of the optimal protocol driving the center of the harmonic trap reflects the spatial periodicity of the ratchet.

## 6.2 Derivative truncation

We conclude this chapter with a discussion of the derivative truncation approach to deriving the inverse diffusion tensor for general potentials in the overdamped case. Recall that in Ch. 3 and Ch. 4 we presented an alternative means of computing the inverse diffusion tensor which obviated the need to compute time integrals of stationary state correlators. A first-order derivative truncation expansion forms the basis of this alternative construction. In this concluding section, we demonstrate how the derivative truncation method reproduces the inverse diffusion tensor for general potentials.

Since we assume general, non-harmonic potentials, the moments appearing in the mean Y-value functional will not satisfy a closed, finite set of ordinary differential equations. Therefore, the construction used in Ch. 3 and Ch. 4 does not obviously carry over to the more general situation. Instead, we base our proof on the mathematics developed in this chapter.

Referring to Eq. (6.1), we assume the nonequilibrium probability density has the approximate form

$$\rho(x, t) \approx \rho_{eq}(x, \boldsymbol{\lambda}(t)) + \frac{d\lambda^i}{dt} \mathcal{G}_i(x, \boldsymbol{\lambda}(t)) \quad (6.66)$$

where  $\mathcal{G}_i(x, \boldsymbol{\lambda})$  is to be determined. Substituting this expression into Eq. (6.1) and neglecting

higher-order derivatives, we see

$$\frac{\partial \rho_{eq}(x, \boldsymbol{\lambda})}{\partial \lambda^i} = D \left[ \frac{\partial}{\partial x} \left( \frac{\partial (\beta U(x, \boldsymbol{\lambda}))}{\partial x} \mathcal{G}_i(x, \boldsymbol{\lambda}) \right) + \frac{\partial^2}{\partial x^2} \mathcal{G}_i(x, \boldsymbol{\lambda}) \right]. \quad (6.67)$$

Furthermore, since both  $\rho(x, t)$  and  $\rho_{eq}(x, \boldsymbol{\lambda})$  are normalized probability distributions, we have the constraint

$$\int_{-\infty}^{\infty} dx \mathcal{G}_i(x, \boldsymbol{\lambda}) = 0. \quad (6.68)$$

We may systematically integrate Eq. (6.67) to obtain a solution which also satisfies Eq. (6.68) and appropriate boundary conditions. For our purposes, a more expedient way of arriving at the solution is to simply state a candidate and demonstrate that it satisfies the necessary requirements.

Our candidate is

$$\mathcal{G}_j(x, \boldsymbol{\lambda}) \equiv -\langle \partial_x H_2(x; \boldsymbol{\lambda}, x_0) \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}) \rangle_{eq, \boldsymbol{\lambda}}. \quad (6.69)$$

Here,  $H_2$  is defined by Eqs. (6.13), (6.14) and (6.15) and appropriate boundary conditions as dictated by the physics of the problem. The average  $\langle \cdot \rangle_{eq, \boldsymbol{\lambda}}$  applies to the variable  $x_0$  and is defined in terms of the stationary state probability distribution characterized by  $\boldsymbol{\lambda}$ .

We can quickly establish Eq. (6.68). Bringing the integral  $\int_{-\infty}^{\infty} dx$  inside of the stationary state average in Eq. (6.69), we see that  $\int_{-\infty}^{\infty} dx \partial_x H_2(x; \boldsymbol{\lambda}, x_0)$  vanishes by the fundamental theorem of calculus in the case of an unbounded potential at  $\pm\infty$ .

Substituting Eq. (6.69) into Eq. (6.67), we see that we must establish

$$\frac{\partial \rho_{eq}(x, \boldsymbol{\lambda})}{\partial \lambda^j} = -D \partial_x \left[ \langle (\beta U'(x, \boldsymbol{\lambda}) \partial_x H_2(x; \boldsymbol{\lambda}, x_0) + \partial_x H_2(x; \boldsymbol{\lambda}, x_0)) \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}) \rangle_{eq, \boldsymbol{\lambda}} \right]. \quad (6.70)$$

From Eq. (6.15) follows

$$D [\beta U'(x, \boldsymbol{\lambda}) \partial_x H_2(x; \boldsymbol{\lambda}, x_0) + \partial_x H_2(x; \boldsymbol{\lambda}, x_0)] = H_1(x; \boldsymbol{\lambda}, x_0) \quad (6.71)$$

and so we must show

$$\frac{\partial \rho_{eq}(x, \boldsymbol{\lambda})}{\partial \lambda^j} = -\langle \partial_x H_1(x; \boldsymbol{\lambda}, x_0) \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}) \rangle_{eq, \boldsymbol{\lambda}}. \quad (6.72)$$

For unbounded-potential boundary conditions,  $H_1(x; \boldsymbol{\lambda}, x_0) = \theta(x - x_0) + \Pi_{eq}(x, \boldsymbol{\lambda})$ . Therefore,

$$\langle \partial_x H_1(x; \boldsymbol{\lambda}, x_0) \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}) \rangle_{eq, \boldsymbol{\lambda}} = \langle \delta(x - x_0) \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}) \rangle_{eq, \boldsymbol{\lambda}}. \quad (6.73)$$

The second term vanishes since  $\rho_{eq}(x, \boldsymbol{\lambda})$  is independent of  $x_0$  and  $\langle \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}) \rangle_{eq, \boldsymbol{\lambda}} = 0$ . More-

over,

$$\langle \delta(x - x_0) \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}) \rangle_{eq, \boldsymbol{\lambda}} = \rho_{eq}(x, \boldsymbol{\lambda}) \partial_{\lambda^j} \phi(x, \boldsymbol{\lambda}) = -\partial_{\lambda^j} \rho_{eq}(x, \boldsymbol{\lambda}), \quad (6.74)$$

establishing the claim.

We are now in position to relate this derivative truncation approximation to the inverse diffusion tensor approximation. Recall

$$\langle Y \rangle_{\Lambda} \equiv \int_0^{\tau} dt \left[ \frac{d\boldsymbol{\lambda}^T}{dt} \right] \cdot \left\langle \frac{\partial \phi}{\partial \boldsymbol{\lambda}}(\boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} \quad (6.75)$$

where

$$\left\langle \frac{\partial \phi}{\partial \lambda^i}(\boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} \equiv \int_{-\infty}^{\infty} dx \rho(x, t) \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)). \quad (6.76)$$

Using the derivative truncation approximation,

$$\begin{aligned} \left\langle \frac{\partial \phi}{\partial \lambda^i}(\boldsymbol{\lambda}(t)) \right\rangle_{\Lambda} &\approx \frac{d\lambda^j}{dt} \int_{-\infty}^{\infty} dx \mathcal{G}_j(x, \boldsymbol{\lambda}(t)) \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \\ &= -\frac{d\lambda^j}{dt} \int_{-\infty}^{\infty} dx \langle \partial_x H_2(x; \boldsymbol{\lambda}(t), x_0) \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}(t)) \rangle_{eq, \boldsymbol{\lambda}(t)} \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \\ &= -\frac{d\lambda^j}{dt} \left\langle \left[ \int_{-\infty}^{\infty} dx \partial_x H_2(x; \boldsymbol{\lambda}(t), x_0) \frac{\partial \phi}{\partial \lambda^i}(x, \boldsymbol{\lambda}(t)) \right] \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}(t)) \right\rangle_{eq, \boldsymbol{\lambda}(t)} \\ &= \frac{d\lambda^j}{dt} \left\langle \left[ \int_0^{\infty} dt' \int_{-\infty}^{\infty} dx \rho(x, t'; x_0) \frac{\partial \phi}{\partial \lambda^i}(x; \boldsymbol{\lambda}(t)) \right] \partial_{\lambda^j} \phi(x_0, \boldsymbol{\lambda}(t)) \right\rangle_{eq, \boldsymbol{\lambda}(t)} \\ &= \frac{d\lambda^j}{dt} \int_0^{\infty} dt' \langle \partial_{\lambda^i} \phi(t') \partial_{\lambda^j} \phi(0) \rangle_{eq, \boldsymbol{\lambda}(t)}. \end{aligned} \quad (6.77)$$

Therefore, we see that the derivative truncation approximation reproduces the inverse diffusion tensor for general potentials in the overdamped regime.

# Bibliography

- [1] Trepagnier EH, Jarzynski C, Ritort F, Crooks GE, Bustamante CJ, et al. (2004) Experimental test of hatano and sasa's nonequilibrium steady-state equality. *Proc Natl Acad Sci USA* 101: 15038-15041.
- [2] Evans DJ, Cohen EGD, Morriss GP (1993) Probability of 2nd law violations in shearing steady-states. *Phys Rev Lett* 71: 2401–2404.
- [3] Evans DJ, Searles DJ (1994) Equilibrium microstates which generate 2nd law violating steady-states. *Phys Rev E* 50: 1645–1648.
- [4] Gallavotti G, Cohen EGD (1995) Dynamical ensembles in nonequilibrium statistical-mechanics. *Phys Rev Lett* 74: 2694-2697.
- [5] Crooks GE (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E* 60: 2721–2726.
- [6] Hatano T, Sasa S (2001) Steady-state thermodynamics of Langevin systems. *Phys Rev Lett* 86: 3463–3466.
- [7] Wang GM, Sevick EM, Mittag E, Searles DJ, Evans DJ (2002) Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Phys Rev Lett* 89: 050601.
- [8] Carberry DM, Reid JC, Wang GM, Sevick EM, Searles DJ, et al. (2004) Fluctuations and irreversibility: An experimental demonstration of a second-law-like theorem using a colloidal particle held in an optical trap. *Phys Rev Lett* 92: 140601.
- [9] Garnier N, Ciliberto S (2005) Nonequilibrium fluctuations in a resistor. *Phys Rev E* 71: 060101.
- [10] Toyabe S, Sagawa T, Ueda M, Muneyuki E, Sano M (2010) Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat Phys* 6: 988-992.
- [11] Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Phys Rev Lett* 78: 2690–2693.

- [12] Liphardt JT, Dumont S, Smith SB, Tinoco Jr I, Bustamante C (2002) Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* 296: 1832–1835.
- [13] Seifert U (2005) Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys Rev Lett* 95: 040602.
- [14] Sagawa T, Ueda M (2010) Generalized Jarzynski equality under nonequilibrium feedback control. *Phys Rev Lett* 104: 090602.
- [15] Neal RM (2001) Annealed importance sampling. *Stat Comput* 11: 125–139.
- [16] Nilmeier JP, Crooks GE, Minh DDL, Chodera JD (2011) Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *P Natl Acad Sci USA* 108: E1009–18.
- [17] Still S, Sivak DA, Bell AJ, Crooks GE (2012) Thermodynamics of prediction. *Phys Rev Lett* 109: 120604.
- [18] Schrödinger E (1992) *What is Life?* Cambridge: Cambridge University Press.
- [19] Howard J (2001) *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, Massachusetts: Sinauer.
- [20] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, et al. (2002) *Molecular Biology of the Cell*. New York: Garland Science.
- [21] Andresen B (2011) Current trends in finite-time thermodynamics. *Angew Chem, Int Ed* 50: 2690-2704.
- [22] Chen L, Sun F, editors (2004) *Advances in finite time thermodynamics: analysis and optimization*. New York: Nova Science Publishers.
- [23] Frank M (2002) The physical limits of computing. *Comput Sci Eng* 4: 16-26.
- [24] Maxwell J (1871) *Theory of Heat*. London: Appleton.
- [25] Szilard L (1929) On the minimization of entropy in a thermodynamic system with interferences of intelligent beings. *ZPhys* 53: 840-856.
- [26] Landauer R (1961) Irreversibility and heat generation in the computing process. *IBM J Res Dev* 5: 183-191.
- [27] Bennett C (1982) The thermodynamics of computation—a review. *Int J Theor Phys* 21: 905-940.
- [28] Beacuterut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R, et al. (2012) Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* 483: 187-9.

- [29] Dillenschneider R, Lutz E (2009) Memory Erasure in Small Systems. *Phys Rev Lett* 102: 210601.
- [30] Klages R, Just W, Jarzynski C, editors (2013) Nonequilibrium statistical physics of small systems. Weinheim, Germany: Wiley-VCH Verlag and Co. KGaA.
- [31] Sagawa T Thermodynamics of information processing in small systems. Ph.D. thesis, The University of Tokyo, Japan.
- [32] Sagawa T, Ueda M (2012) Nonequilibrium thermodynamics of feedback control. *Phys Rev E* 85: 021104.
- [33] Diana G, Bagci GB, Esposito M (2013) Finite-time erasing of information stored in fermionic bits. *Phys Rev E* 87: 012111.
- [34] Esposito M, Kawai R, Lindenberg K, Van den Broeck C (2010) Finite-time thermodynamics for a single-level quantum dot. *Europhys Lett* 89: 20003.
- [35] Aurell E, Gawedzki K, Mejia-Monasterio C, Mohayaei R, Muratore-Ginanneschi P (2012) Refined Second Law of Thermodynamics for Fast Random Processes. *J Stat Phys* 147: 487-505.
- [36] Sivak DA, Crooks GE (2012) Thermodynamic Metrics and Optimal Paths. *Phys Rev Lett* 108: 190602.
- [37] Zulkowski PR, Sivak DA, Crooks GE, DeWeese MR (2012) Geometry of thermodynamic control. *Phys Rev E* 86: 041148.
- [38] Zulkowski PR, Sivak DA, DeWeese MR (2013) Optimal control of transitions between nonequilibrium steady states. arXiv:1303.6596.
- [39] Shenfeld DK, Xu H, Eastwood MP, Dror RO, Shaw DE (2009) Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. *Phys Rev E* 80: 046705.
- [40] Brody DC, Hook DW (2009) Information geometry in vapour-liquid equilibrium. *J Phys A* 42: 023001.
- [41] Gomez-Marin A, Schmiedl T, Seifert U (2008) Optimal protocols for minimal work processes in underdamped stochastic thermodynamics. *J Chem Phys* 129: 024114 (8).
- [42] Schmiedl T, Seifert U (2007) Optimal finite-time processes in stochastic thermodynamics. *Phys Rev Lett* 98: 108301.
- [43] Aurell E, Mejía-Monasterio C, Muratore-Ginanneschi P (2011) Optimal protocols and optimal transport in stochastic thermodynamics. *Phys Rev Lett* 106: 250601 (4).

- [44] Weinhold F (1975) Metric geometry of equilibrium thermodynamics. *J Chem Phys* 63: 2479-2483.
- [45] Ruppeiner G (1979) Thermodynamics: A riemannian geometric model. *Phys Rev A* 20: 1608-1613.
- [46] Schlögl F (1985) Thermodynamic metric and stochastic measures. *Z Phys B* 59: 449-454.
- [47] Salamon P, Nulton J, Ihrig E (1984) On the relation between entropy and energy versions of thermodynamic length. *J Chem Phys* 80: 436-437.
- [48] Salamon P, Berry RS (1983) Thermodynamic length and dissipated availability. *Phys Rev Lett* 51: 1127-1130.
- [49] Brody D, Rivier N (1995) Geometrical aspects of statistical mechanics. *Phys Rev E* 51: 1006-1011.
- [50] Crooks GE (2007) Measuring thermodynamic length. *Phys Rev Lett* 99: 100602.
- [51] Burbea J, Rao CR (1982) Entropy differential metric, distance and divergence measures in probability spaces - a unified approach. *J Multivariate Anal* 12: 575-596.
- [52] Sivak DA, Crooks GE (2012) Thermodynamic metrics and optimal paths. *Phys Rev Lett* 108: 190602.
- [53] Mazonka O, Jarzynski C (1999) Exactly solvable model illustrating far-from-equilibrium predictions. *ArXiv:9912.121*.
- [54] Zwanzig R (2001) Nonequilibrium statistical mechanics. New York: Oxford University Press.
- [55] Seifert U (2012) Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics* 75: 126001.
- [56] Sivak DA, Crooks GE (2012) Near-equilibrium measurements of nonequilibrium free energy. *Phys Rev Lett* 108: 150601.
- [57] Crooks GE (1998) Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J Stat Phys* 90: 1481-1487.
- [58] Sekimoto K (1998) Langevin equation and thermodynamics. *Prog Theor Phys Suppl* 130: 17-27.
- [59] Jarzynski C (1998) Equilibrium free energies from nonequilibrium processes. *Acta Phys Pol B* 29: 1609-1622.
- [60] Imperato A, Peliti L, Pesce G, Rusciano G, Sasso A (2007) Work and heat probability distribution of an optically driven brownian particle: theory and experiments. *Phys Rev E* 76: 050101-1-4.

- [61] Crooks GE (1999) Excursions in Statistical Dynamics. Ph.D. thesis, University of California at Berkeley.
- [62] Jarzynski C (1999) Microscopic analysis of clausius-duhem processes. *Journal of Statistical Physics* 96: 415-427.
- [63] Douarche F, Ciliberto S, Petrosyan A, Rabbiosi I (2005) An experimental test of the Jarzynski equality in a mechanical experiment. *Europhys Lett* 70: 593–599.
- [64] Ciliberto S, Joubaud S, Petrosyan A (2010) Fluctuations in out-of-equilibrium systems: from theory to experiment. *J Stat Mech: Theor Exp* : P12003.
- [65] Vaikuntanathan S, Jarzynski C (2009) Dissipation and lag in irreversible processes. *Europhys Lett* 87: 60005.
- [66] Boyce W, DiPrima R (2000) Elementary differential equations and boundary value problems. Wiley, New York, seventh edition.
- [67] do Carmo MP (1992) Riemannian geometry. Boston: Birkhäuser.
- [68] Jost J (2005) Riemannian geometry and geometric analysis. Berlin: Springer, fourth edition.
- [69] Carroll SM (2004) Spacetime and geometry: An introduction to general relativity. San Francisco: Addison Wesley, first edition.
- [70] Ruppeiner G (2010) Thermodynamic curvature measures interactions. *Am J Phys* 78: 1170-80.
- [71] Shenfeld DK, Xu H, Eastwood MP, Dror RO, Shaw DE (2009) Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. *Phys Rev E* 80: 046705.
- [72] Aurell E, Mejía-Monasterio C, Muratore-Ginanneschi P (2011) Optimal protocols and optimal transport in stochastic thermodynamics. *Phys Rev Lett* 106: 250601.
- [73] Kardar M (2007) Statistical physics of particles. Cambridge: Cambridge University Press. URL <http://dx.doi.org/10.1017/CBO9780511815898>.
- [74] Peterman EJ, Gittes F, Schmidt CF (2003) Laser-induced heating in optical traps. *Biophysical Journal* 84: 1308 - 1316.
- [75] Risken H (1996) The Fokker-Planck equation. Berlin: Springer-Verlag, 3rd edition.
- [76] Neuman K, Block S (2004) Optical trapping. *Review of Scientific Instruments* 75: 2787-2809.
- [77] Joykuty J, Mathur V, Venkataraman V, Natarajan V (2005) Direct Measurement of the Oscillation Frequency in an Optical-Tweezers Trap by Parametric Excitation. *Phys Rev Lett* 95: 193902.

- [78] Lambson B, Carlton D, Bokor J (2011) Exploring the thermodynamic limits of computation in integrated systems: Magnetic memory, nanomagnetic logic, and the landauer limit. *Phys Rev Lett* 107: 010604.
- [79] Mehta P, Schwab D (2012) Energetic costs of cellular computation. *Proceedings of the National Academy of Sciences* 109: 17978-17982.
- [80] Zulkowski PR, Sivak DA, DeWeese MR (2013) Optimal control of transitions between nonequilibrium steady states. *Public Library of Science* (in press).
- [81] Maroney O (2005) The (absence of a) relationship between thermodynamic and logical reversibility. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 36: 355-374.
- [82] Jun Y, Bechhoefer J (2012) Virtual potentials for feedback traps. *Phys Rev E* 86: 061106.
- [83] Plenio M, Vitelli V (2001) The physics of forgetting: Landauer's erasure principle and information theory. *Contemp Phys* 42: 25-60.
- [84] Chernyaka V, Chertkov M, Bierkens J, Kappen H (2013) Stochastic optimal control as non-equilibrium statistical mechanics: Calculus of variations over density and current. *ArXiv:1306.6572v1*.
- [85] Then H, Engel A (2008) Computing the optimal protocol for finite-time processes in stochastic thermodynamics. *Phys Rev E* 77: 041105.
- [86] Barkeshli MM (2006) Dissipationless information erasure and landauer's principle. *ArXiv:0504323v3*.
- [87] Xing J, Wang H, Oster G (2005) From continuum fokker-planck models to discrete kinetic models. *Biophys J* 89: 1551-1563.
- [88] Reiss RD (1989) *Approximate Distributions of Order Statistics with Applications to Non-Parametric Statistics*. Springer Series in Statistics. Springer-Verlag.
- [89] Shaw R (1984) *The Dripping Faucet as a Model Chaotic System*. Santa Cruz, CA: Aerial.
- [90] Gaveau B, Schulman LS (1997) *Phys Lett A* 229: 347.
- [91] Sivak DA, Crooks GE (2012) Near-equilibrium measurements of nonequilibrium free energy. *Phys Rev Lett* 108: 150601.
- [92] Reimann P (2002) Brownian motors: noisy transport far from equilibrium. *Physics Reports* 361: 57 - 265.
- [93] Pankratov A (1999) Time evolution of averages in dynamical systems driven by noise. *Phys Lett A* 255: 17-22.

- [94] Abramowitz M, Stegun I (1972) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York.
- [95] Weinberg S (1972) Gravitation and cosmology: principles and applications of the general theory of relativity. Wiley, New York, first edition.

# Appendix A

## Riemannian Geometry

A careful definition of a maximal atlas of coordinate charts on a topological space would typically initiate any complete introduction to the subject of Riemannian geometry. However, since the parameter spaces of the applications presented in this thesis are open subsets of  $\mathbb{R}^n$ , it will be most convenient and expedient to present the relevant geometric quantities in terms of the familiar concept of vectors defined on Euclidean space. This way, we may avoid being bogged down by unnecessary exposition or lost in irrelevant mathematical technicalities. The interested reader may refer to [67, 68] for details related to topics covered below and for a complete treatment of the fascinating field of Riemannian geometry. A strong introduction written by a physicist may be found in [69].

### A.1 Metric Tensor

We begin our treatment with a discussion of geometry in  $\mathbb{R}^3$  which is assumed familiar to the reader. The Euclidean inner product of vectors (or first fundamental form for the more sophisticated reader) may be defined as follows: given the standard basis of vectors for this space, it is possible to uniquely expand any two vectors as  $\mathbf{v} = \sum_{i=1}^3 v^i \mathbf{e}_i$ ,  $\mathbf{w} = \sum_{i=1}^3 w^i \mathbf{e}_i$  and to define their inner product as  $\langle \mathbf{v}, \mathbf{w} \rangle \equiv \sum_{i=1}^3 v^i w^i$ . Note that the inner product is symmetric in its arguments and that  $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$  always with equality holding if and only if  $\mathbf{v} = 0$ . Since  $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ , it is natural to use this quantity as a measure of length. Furthermore, it is possible to show that  $\langle \mathbf{v}, \mathbf{w} \rangle = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \sqrt{\langle \mathbf{w}, \mathbf{w} \rangle} \cos \theta$  where  $\theta$  is the angle between the two vectors. In this way the inner product encodes the geometry of three-dimensional Euclidean space.

We may rewrite the inner product in a more illuminating way:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i,j=1}^3 \delta_{ij} v^i w^j \tag{A.1}$$

where  $\delta_{ij}$  is the Kronecker delta. Note that  $\delta_{ij}$  is trivially a symmetric, positive-definite matrix. It is possible to generalize the inner product by making the substitution  $\delta_{ij} \rightarrow g_{ij}(\mathbf{x})$  where  $g_{ij}(\mathbf{x})$  is a symmetric, positive-definite matrix depending on position in Euclidean space  $\mathbf{x}$ :

$$\langle \mathbf{v}, \mathbf{w} \rangle_{g,\mathbf{x}} \equiv \sum_{ij} g_{ij}(\mathbf{x}) v^i w^j. \quad (\text{A.2})$$

To be consistent with the notation of the thesis, we will henceforth adhere to the Einstein summation convention and suppress summation symbols for the remainder of this appendix. Unless stated otherwise, if an index symbol is duplicated in both a superscript and a subscript, we assume the index is summed over. For instance, the generalized inner product may be written as

$$\langle \mathbf{v}, \mathbf{w} \rangle_{g,\mathbf{x}} \equiv g_{ij}(\mathbf{x}) v^i w^j. \quad (\text{A.3})$$

where  $i$  and  $j$  are implicitly summed over. The matrix  $g_{ij}(\mathbf{x})$  is referred to as the metric tensor and generally encodes a richer geometry than the standard Euclidean inner product defined in terms of the Kronecker tensor. Note that Eq. (A.2) may be extended to spaces of dimension  $n \neq 3$  trivially.

The term ‘‘tensor’’ may invoke trepidation in the newcomer. However, for our purposes, it is convenient and sufficient to think of tensors as objects that transform a specific way under coordinate transformations. A hallmark of geometric quantities is their coordinate invariance: things like length, angle, area of a surface, etc. should not depend on the particular coordinate system we humans employ to calculate them. Students familiar with the basic concepts of gravitation will recognize that the physical importance of coordinate invariance is a mantra often repeated by general relativists. We will give an argument for the behavior of the metric tensor under coordinate transformations by thinking of the coordinate invariant notion of curve length.

Recall how one may compute the length of a parameterized curve in  $\mathbb{R}^3$  endowed with the standard inner product: one divides the curve into very small segments and approximates each segment by the tangent vector to the curve. The length of each tangent vector is measured using the inner product and the total length is approximated by a sum over all segments. In the limit where the segment length approaches 0, the approximation converges to the definition of the curve length. This process is not restricted to the standard inner product and may be extended to a more general metric tensor  $g_{ij}(\mathbf{x})$ . A most convenient condensation of this process is given by the metric line element which encodes the distance between two infinitesimally separated points:

$$ds^2 \equiv g_{ij}(\mathbf{x}) dx^i dx^j. \quad (\text{A.4})$$

Here,  $x^j$  is a coordinate system chosen *a priori* and does not necessarily have to coincide with Cartesian coordinates and  $dx^j$  is an increment of the  $j$ -th coordinate.

Since length must be a coordinate invariant concept,  $ds^2$  will not change under a redefinition of

coordinates. Not only is this a convenient way of demonstrating the metric tensor's behavior under coordinate redefinitions but it is a fact we use extensively throughout this thesis to find analytic expressions for geodesics which we define shortly. Suppose a new set of coordinates  $\tilde{x}^J(x^i)$  is defined in terms of the old where we assume a smooth inverse transformation exists. Here, we use upper case Latin symbols for the indices of the new coordinate system. Then, by the chain rule of calculus,

$$dx^i = \frac{\partial x^i}{\partial \tilde{x}^I} d\tilde{x}^I \quad (\text{A.5})$$

and so

$$ds^2 = g_{ij}(\mathbf{x}) dx^i dx^j = g_{ij}(\mathbf{x}(\tilde{\mathbf{x}})) \frac{\partial x^i}{\partial \tilde{x}^I} \frac{\partial x^j}{\partial \tilde{x}^J} d\tilde{x}^I d\tilde{x}^J = \tilde{g}_{IJ}(\tilde{\mathbf{x}}) d\tilde{x}^I d\tilde{x}^J. \quad (\text{A.6})$$

From this expression, we see that the coordinate invariance of length is ensured if

$$\tilde{g}_{IJ}(\tilde{\mathbf{x}}) \equiv g_{ij}(\mathbf{x}(\tilde{\mathbf{x}})) \frac{\partial x^i}{\partial \tilde{x}^I} \frac{\partial x^j}{\partial \tilde{x}^J}. \quad (\text{A.7})$$

This is the tensorial nature of the metric.

It is furthermore assumed in the expression Eq. (A.2) that the vectors  $\mathbf{v}$ ,  $\mathbf{w}$  are in some sense “attached” to the point  $\mathbf{x}$ . More concretely,  $\mathbf{v}$ ,  $\mathbf{w}$  are assumed to be elements of the tangent space attached to  $\mathbf{x}$ . The collection of tangent spaces (each a vector space copy of  $\mathbb{R}^3$ ) is referred to as the tangent bundle. It is necessary to distinguish tangent spaces in the presence of a non-trivial metric since it is no longer possible to parallel transport vectors in the standard way.

In a standard calculus course, calculations often call for the comparison of two vectors in some sense. For instance, addition and taking the inner product of two vectors are two operations that come immediately to mind. When  $\mathbb{R}^3$  is endowed with the standard inner product, it is acceptable to move vectors anywhere in space without rotation in order to make these comparisons. In this circumstance it was not necessary to distinguish the vector  $\mathbf{v}$  with base at  $\mathbf{x}$  from the vector with the same components but based at  $\mathbf{x}' \neq \mathbf{x}$ . In the presence of curvature this way of moving vectors about is no longer consistent and so it becomes fundamentally important to distinguish tangent spaces at different points. This issue of parallel transport will arise again during our brief discussion of curvature.

## A.2 Curvature Tensor

The intrinsic curvature of a surface can be readily understood in terms of a simple thought experiment. Suppose a two-dimensional being inhabits the surface of a sphere which we embed into  $\mathbb{R}^3$ . We emphasize that this embedding is not necessary or even practical for the general definition of intrinsic curvature; the example used here is a convenient visual heuristic. The two-dimensional

being is unaware of any other spatial dimensions but notices something very strange during an expedition on the sphere. With an arrow fixed in her frame of reference, the explorer sets out from the north pole and travels down a meridian. She then cuts across a latitude and returns to the north pole along another meridian. Though she is careful not to disturb the arrow's direction in her reference frame, she finds that the arrow points in a different direction upon her return. The situation is illustrated in Fig. A.1.

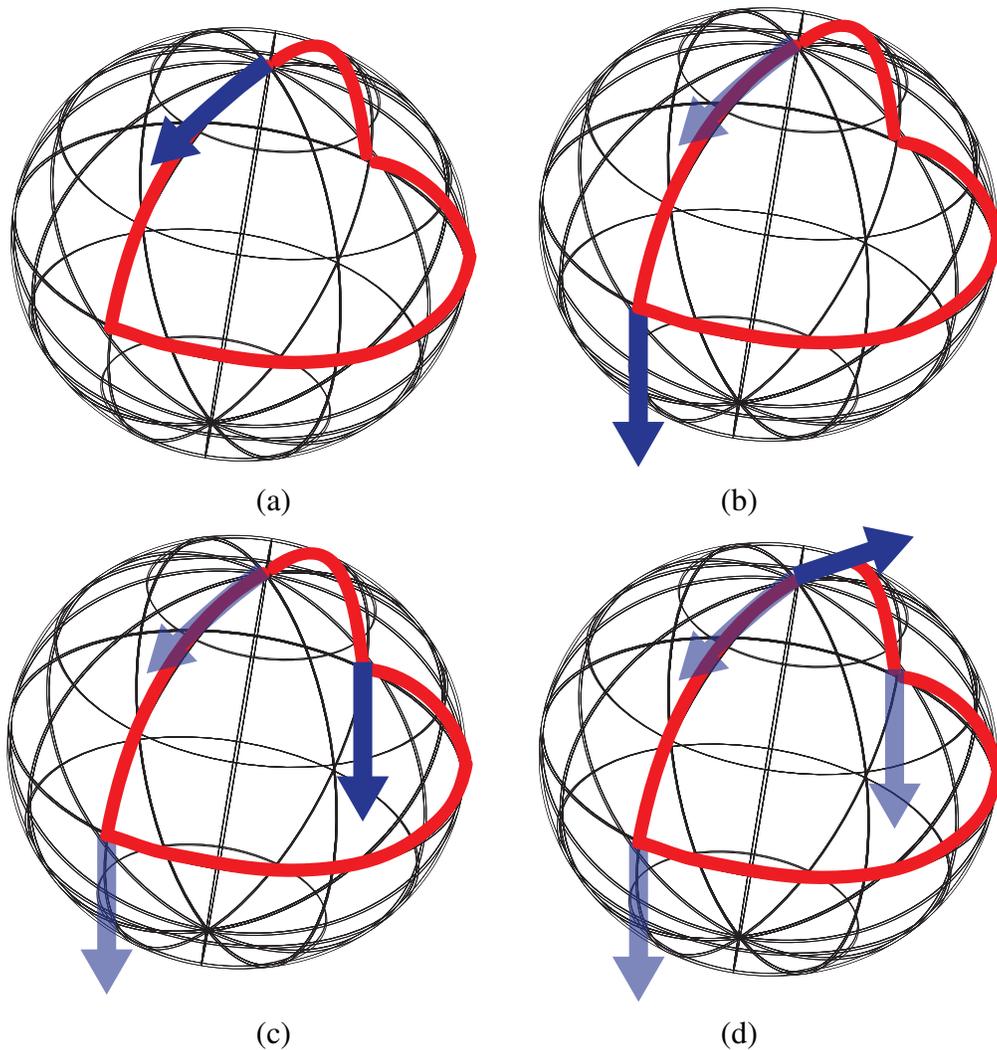


Figure A.1: A two-dimensional explorer traverses a closed curve on the surface of a sphere (red) while carrying an arrow (blue) in a fixed direction in her frame of reference. In (a) the explorer begins at the north pole with the arrow tangent to a meridian, moves down the meridian in (b), along the equator in (c) and returns to the north pole along the meridian in (d) which is part of the same great circle as the meridian in (b). The result is a rotation of the vector due to the intrinsic curvature of the sphere.

The explorer rightfully comes to the conclusion that her two-dimensional universe must be curved. In Riemannian geometry, the Riemann tensor encodes information about the intrinsic

geometry of space and is defined in terms of parallel transport of vectors around small closed loops [95]. We need not go into the details of the construction of this tensor nor many of its properties. We hope the reader will find the above narrative a useful visualization of the idea of intrinsic curvature. For our purposes, it is sufficient to give the explicit mathematical definition of the tensor and a few key characteristics which we use extensively throughout this thesis. In the following, observe that the Riemann tensor may be expressed entirely in terms of the metric tensor.

To begin, we need to define the so-called Christoffel symbols (of the second kind) which arise naturally in the context of vector parallel transport. We assume a coordinate system has been chosen from the outset.

$$\Gamma_{jk}^i \equiv \frac{1}{2} g^{il} (\partial_k g_{jl} + \partial_j g_{kl} - \partial_l g_{kj}) \quad (\text{A.8})$$

where  $g^{ij}$  are the components of the matrix inverse of  $g_{ij}$ . Please note that the Christoffel symbols *do not* transform tensorially under changes of coordinates.

The Christoffel symbols are used to construct the Riemann tensor:

$$R^i_{jkl} \equiv \partial_k \Gamma_{jl}^i - \partial_l \Gamma_{jk}^i + \Gamma_{jl}^m \Gamma_{mk}^i - \Gamma_{jk}^m \Gamma_{ml}^i. \quad (\text{A.9})$$

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

$$R_{ij} = R^l_{ilj}, \quad R = g^{ij} R_{ij}, \quad (\text{A.10})$$

which are useful for determining the curvature content of the space. For our purposes, the Ricci scalar is the most useful quantity recorded here. In every application presented in this thesis we construct a metric tensor initially in physical coordinates; *i.e.* those under the direct control of an experimentalist. Geometrically, one coordinate system is not singled out over another, though some coordinate systems prove to be more *convenient* than others depending on the situation as any multi-variable calculus student can attest. The utility of the Ricci scalar is due to its coordinate invariance; *i.e.* the Ricci scalar can tell us something about the underlying geometry independently of our coordinate system.

We use the Ricci scalar extensively to seek out coordinate transformations making our computations as simplistic as possible. For instance, in Ch. 3, we demonstrate that the Ricci scalar of the  $(\beta, k)$ -submanifold is constant and negative implying that the underlying geometry is that of hyperbolic space, a well-known and thoroughly understood geometry. In Ch. 4 and Ch. 5, computation of the Ricci scalar reveals to us that the underlying geometry in both instances is that of Euclidean space with the standard inner product. In all cases considered in this thesis, the physical coordinates obscure the often-times simplistic geometry describing our models.

### A.3 Geodesics

Loosely speaking, a geodesic is a curve of minimal length joining two fixed points in space. From Eq. (A.4), the length of the parameterized curve  $\gamma(t)$  is given by

$$L[\gamma] \equiv \int_a^b \sqrt{g_{ij}(\boldsymbol{\lambda}(t)) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt}} dt \quad (\text{A.11})$$

where the parameter  $t$  takes on values from  $[a, b]$ . Using standard arguments from the calculus of variations [67, 68, 95], extrema of this functional must satisfy the geodesic equation

$$\frac{d^2\gamma^i}{dt^2} + \Gamma_{jk}^i \frac{d\gamma^j}{dt} \frac{d\gamma^k}{dt} = 0 \quad (\text{A.12})$$

where  $\Gamma_{jk}^i$  are the Christoffel symbols. The geodesic equation Eq. (A.12) holds regardless of the coordinate system used. (For a proof, see [95].) Therefore, obtaining useful coordinate systems in which the metric tensor and Christoffel symbols simplify expedites the solution of the geodesic equation.

A quick calculation demonstrates that the norm of the tangent vector to a geodesic must necessarily be constant; *i.e.*  $g_{ij}(\boldsymbol{\lambda}(t)) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} = \text{const.}$ :

$$\frac{d}{dt} \left( g_{ij}(\boldsymbol{\lambda}(t)) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} \right) = \frac{dg_{ij}(\boldsymbol{\lambda}(t))}{dt} \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} + 2g_{ij}(\boldsymbol{\lambda}(t)) \frac{d\gamma^i}{dt} \frac{d^2\gamma^j}{dt^2} = \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} \frac{d\gamma^k}{dt} \left( \partial_k g_{ij} - 2g_{il} \Gamma_{jk}^l \right) \quad (\text{A.13})$$

which vanishes by definition of the Christoffel symbols in terms of the metric (Eq. (A.8)). This fact proves useful in finding geodesics in Ch. 3.

### A.4 Killing Vectors

We conclude this appendix with a brief account of Killing vector fields. In general, Killing vector fields generate isometries of the metric tensor [68, 69]. In words, if we follow the flow generated by a Killing vector field, distances are preserved. Mathematically, a Killing vector field  $\mathbf{K}$  satisfies the Killing equation [68, 69]:

$$\partial_i K_j + \partial_j K_i - 2\Gamma_{ij}^k K_k = 0 \quad (\text{A.14})$$

where  $K_i \equiv g_{ij} K^j$ . Killing vector fields may be very useful in computing closed-form expressions of geodesics since they determine conserved quantities along geodesics. Indeed, we shall now

prove that

$$K_i \frac{d\gamma^i}{dt} \tag{A.15}$$

is parameter-independent when  $\gamma(t)$  satisfies the geodesic equation Eq. (A.12).

The proof is elementary: from basic calculus, we see that

$$\frac{d}{dt} \left( K_i \frac{d\gamma^i}{dt} \right) = K_i \frac{d^2\gamma^i}{dt^2} + \partial_j K_i \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt}. \tag{A.16}$$

Since  $\mathbf{K}$  is a Killing field,

$$\partial_j K_i \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} = \Gamma_{ij}^k K_k \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt}. \tag{A.17}$$

Furthermore, since  $\gamma$  is a geodesic, we have

$$\partial_j K_i \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} = -K_k \frac{d^2\gamma^k}{dt^2} \tag{A.18}$$

and so

$$\frac{d}{dt} \left( K_i \frac{d\gamma^i}{dt} \right) = 0. \tag{A.19}$$

With a Killing vector field in hand, it is therefore possible to obtain a first integral of the geodesic equation, simplifying the calculation of closed-form expressions for geodesics. However, it may seem we have traded one difficulty (calculating solutions to the second order geodesic equation) for another (solving the Killing equation). Though true in general, it is sometimes possible to write down Killing vector fields based on some symmetry of the metric tensor. In particular, if it is possible to find a coordinate system in which the metric components do not depend on one of the coordinates, say  $x^j$ , then the vector pointing in the  $j$ -th direction is a Killing vector field [69]. This observation is used to great effect in Ch. 3.

# Appendix B

## Fokker-Planck Equations

Fokker-Planck equations form the basis for many of the results discussed in this thesis and are useful descriptions in nonequilibrium statistical mechanics when the underlying dynamical system has Markovian friction and Gaussian white noise. They are often used as an alternative to Langevin equations which are difficult to solve when nonlinearities are present. Therefore, for sake of completeness and for the benefit of the newcomer to nonequilibrium statistical mechanics, a chapter devoted to the derivation of the Fokker-Planck equation seems appropriate. The reader should not mistakenly believe that the material presented in this appendix originated with the author: the exposition given here is largely drawn from Ch. 2.2 of [54].

### B.1 Derivation of a Fokker-Planck Equation

Let us suppose that the dynamical system of interest is described by a set of variables  $\{a_1, a_2 \dots\}$  which is more conveniently denoted by a vector  $\mathbf{a}$ . For the reader familiar with the basics of classical mechanics, the vector  $\mathbf{a}$  could for instance represent a point in the phase space of a system consisting of  $N$  particles. In that case, the components of  $\mathbf{a}$  would consist of the three position coordinates and the three components of momentum for each of the  $N$  particles under consideration. We suppose at the outset that the noise-free part of the dynamics is Markovian; *i.e.* has no memory. Furthermore, we require that the noise is white and has a Gaussian distribution.

Specifically, we assume equations of motion

$$\frac{d\mathbf{a}}{dt} = \mathbf{v}(\mathbf{a}) + \mathbf{F}(t), \quad (\text{B.1})$$

where  $\mathbf{v}(\mathbf{a})$  is a given function and the noise  $\mathbf{F}(t)$  is Gaussian with

$$\langle \mathbf{F}(t) \rangle = 0, \quad \langle \mathbf{F}(t) \mathbf{F}(t') \rangle = 2\mathbf{B}\delta(t - t'). \quad (\text{B.2})$$

In general,  $\mathbf{B}$  is a symmetric matrix which may depend upon  $\mathbf{a}$ . However, for the applications found in this thesis, the matrix  $\mathbf{B}$  will be fixed by enforcing the Einstein relation between the diffusion constant and the ambient temperature.

As an alternative to solving the equations of motion directly, we seek a probability distribution  $f(\mathbf{a}, t)$  of the values of  $\mathbf{a}$  at time  $t$ . More to the point, we want the average of such a probability distribution over the noise. First, we recognize that the probability distribution is a conserved quantity:

$$\int d\mathbf{a} f(\mathbf{a}, t) = 1 \quad (\text{B.3})$$

for all time. Therefore, we expect the  $f(\mathbf{a}, t)$  to satisfy a conservation law:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \mathbf{a}} \cdot \left( \frac{\partial \mathbf{a}}{\partial t} f \right) = 0. \quad (\text{B.4})$$

If we use Eq. (B.1) in the conservation law Eq. (B.4), then we obtain

$$\frac{\partial f}{\partial t} = - \frac{\partial}{\partial \mathbf{a}} \cdot (\mathbf{v}(\mathbf{a})f + \mathbf{F}(t)f) \quad (\text{B.5})$$

which contains a random term.

Following the notation set by [54], we define the operator  $L$  by

$$L\Phi \equiv \frac{\partial}{\partial \mathbf{a}} \cdot (\mathbf{v}(\mathbf{a})\Phi) \quad (\text{B.6})$$

so that formal solutions to Eq. (B.4) with  $\mathbf{F}(t) = 0$  are given by

$$f(\mathbf{a}, t) = e^{-tL} f(\mathbf{a}, 0). \quad (\text{B.7})$$

With the noise term, the conservation equation reads

$$\frac{\partial f}{\partial t} = -Lf - \frac{\partial}{\partial \mathbf{a}} \cdot \mathbf{F}(t)f. \quad (\text{B.8})$$

An integration over time leads to the operator equation

$$f(\mathbf{a}, t) = e^{-tL} f(\mathbf{a}, 0) - \int_0^t ds e^{-(t-s)L} \frac{\partial}{\partial \mathbf{a}} \cdot \mathbf{F}(s)f(\mathbf{a}, s). \quad (\text{B.9})$$

Substituting this back into the conservation law yields

$$\frac{\partial}{\partial t} f(\mathbf{a}, t) = -Lf(\mathbf{a}, t) - \frac{\partial}{\partial \mathbf{a}} \cdot \mathbf{F}(t)f(\mathbf{a}, 0) + \frac{\partial}{\partial \mathbf{a}} \cdot \mathbf{F}(t) \int_0^t ds e^{-(t-s)L} \frac{\partial}{\partial \mathbf{a}} \cdot \mathbf{F}(s)f(\mathbf{a}, s). \quad (\text{B.10})$$

Upon averaging over the noise, terms with a single  $F$  vanish and we obtain

$$\frac{\partial}{\partial t} \langle f(\mathbf{a}, t) \rangle = -\frac{\partial}{\partial \mathbf{a}} \cdot \mathbf{v}(\mathbf{a}) \langle f(\mathbf{a}, t) \rangle + \frac{\partial}{\partial \mathbf{a}} \cdot \mathbf{B} \cdot \frac{\partial}{\partial \mathbf{a}} \langle f(\mathbf{a}, t) \rangle. \quad (\text{B.11})$$

This is the Fokker-Planck equation. In this thesis, we abuse notation and identify  $f(\mathbf{a}, t)$  as the probability distribution obtained after averaging over noise.