Title
Contraction and Reaction in Generalized Schrödinger Bridges

ISBN
9798273396425

Author
Teter, Alexis Mikayla Horne

Publication Date
2025-10-31

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**CONTRACTION AND REACTION
IN GENERALIZED SCHRÖDINGER BRIDGES**

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

APPLIED MATHEMATICS

by

Alexis Teter

December 2025

The dissertation of Alexis Teter
is approved:

Professor Qi Gong, Chair

Associate Professor Abhishek Halder

Professor Hongyun Wang

Associate Professor Yongxin Chen

Peter Biehl
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

List of Notations

In this dissertation, the following symbols are used.

Notation	**Definition**
$\preceq, \prec, \succeq, \succ$	Conic inequalities
$\mathbb{R}^n$	Euclidean space of dimension n
$\overline{\mathbb{R}}$	Set of extended reals
$\mathbb{S}^{n-1}$	Euclidean unit sphere in $\mathbb{R}^n$
$\mathbb{N}$	The set of natural numbers $\{1, 2, \cdots\}$
$\mathbb{N}_0$	The set of whole numbers $\mathbb{N} \cup \{0\}$
$\mathrm{Skew}\,(n, \mathbb{R})$	The set of $n \times n$ skew-symmetric matrices with real entries
$H(\boldsymbol{X}, \boldsymbol{D})$	Weyl operator
$h(\boldsymbol{x}, \boldsymbol{\xi})$	Weyl symbol
$\mathbb{E}_{\mathbb{P}}\,[\cdot]$	Expectation operator with respect to the probability measure $\mathbb{P}$
$\\|\cdot\\|_2$	Euclidean magnitude
$\langle \cdot, \cdot \rangle$	Hilbert-Schmidt inner product
$\propto$	Proportional to
$\sim$	Follows the statistical law
$\sharp$	Pushforward operator
$^\top$	Transpose
†	Complex conjugate
t	Time
$\boldsymbol{x}$	State vector
$\boldsymbol{u}$	Control input vector
$\mathcal{U}$	Set of finite energy Markovian control policies
ρ	Joint probability density function
$\mathcal{P}_2\,(\mathbb{R}^n)$	Set of probability density functions supported over $\mathbb{R}^n$ with finite second moments
$\nabla_{\boldsymbol{x}}, \nabla_{\boldsymbol{x}}\cdot, \Delta_{\boldsymbol{x}}$	Euclidean gradient, divergence and Laplacian operators with respect to the vector $\boldsymbol{x}$
$\Delta_{\boldsymbol{\Sigma}}$	Weighted Laplacian operator for $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$
$\mathrm{Hess}_{\boldsymbol{x}}$	Euclidean Hessian operator with respect to the vector $\boldsymbol{x}$, i.e., the composite operator $\nabla_{\boldsymbol{x}} \circ \nabla_{\boldsymbol{x}} = \nabla^2_{\boldsymbol{x}}$
$\\|\cdot\\|^2_{\boldsymbol{\Sigma}}$	The squared weighted norm given by $(\cdot)^\top \boldsymbol{\Sigma}(\cdot)$
$\{\cdot, \cdot\}$	Poisson bracket
$\mathrm{conv}(\cdot)$	Convex hull
ε	Diffusive regularization parameter
S	Value function
$\widehat{\varphi}, \varphi$	Schrödinger factors

Notation	**Definition**
$C^{m,n}(\mathcal{T};\mathcal{S})$	Space of functions which are m and n times continuously differentiable on $\mathcal{T}$ and $\mathcal{S}$, respectively
$D_{\mathrm{KL}}(\mathbb{P} \parallel \mathbb{Q})$	Relative entropy a.k.a. the Kullback-Leibler divergence between probability measures $\mathbb{P}$ and $\mathbb{Q}$
$\mathbb{P} \ll \mathbb{Q}$	Measure $\mathbb{P}$ is absolutely continuous with respect to measure $\mathbb{Q}$
$\mathcal{F}[\cdot], \mathcal{F}^{-1}[\cdot]$	Fourier and inverse Fourier transform
$\mathcal{L}[\cdot], \mathcal{L}^{-1}[\cdot]$	Laplace and inverse Laplace transform
$(f * g)(\boldsymbol{x})$	Convolution of functions f and g, defined as $\int_{\mathbb{R}^d} f(\boldsymbol{x}-\boldsymbol{y})g(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \int_{\mathbb{R}^d} f(\boldsymbol{y})g(\boldsymbol{x}-\boldsymbol{y})\mathrm{d}\boldsymbol{y}$, for $f,g:\mathbb{R}^d \to \mathbb{R}$
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian probability density function with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathrm{diag}(\boldsymbol{x})$	Diagonal matrix with main diagonal entries being the elements of the vector $\boldsymbol{x}$
$\det(\cdot)$	Determinant
$\mathrm{eig}(\cdot)$	Eigenvalue operator
$\mathbf{1}_{\{\cdot\}}$	Indicator function
$q(\boldsymbol{x})$	State-dependent cost function
$k(\cdot,\cdot,\cdot,\cdot)$	Markov kernel
$\mathrm{L}(\cdot,\cdot,\cdot)$	Lagrangian
$\boldsymbol{I}_n$	Identity matrix of size $n \times n$
i	The imaginary number $\sqrt{-1}$
$\oslash$	Elementwise (Hadamard) division
$d_H(\cdot,\cdot)$	Hilbert's projective metric
κ	Contraction coefficient
γ	Worst-case contraction coefficient
$h_{\mathcal{K}}(\cdot)$	The support function for a closed convex set $\mathcal{K}$
$\mathbb{W}$	Wiener measure
H_n	Physicist's Hermite polynomial of degree $n \in \mathbb{N}_0$
$\psi, \psi^{\dagger}$	Wave function and its conjugate
$\mathfrak{R}, \mathfrak{I}$	Real and imaginary parts of a complex number
$\hbar$	Reduced Planck's constant

List of Acronyms

In this dissertation, the following acronyms are used.

Acronyms	**Expansion**
AI	Artificial Intelligence
a.k.a.	Also Known As
a.s.	Almost Sure
ECI	Earth Centered Inertial
FPK	Fokker-Planck-Kolmogorov
HJB	Hamilton-Jacobi-Bellman
i.i.d.	Independent and Identically Distributed
IVP	Initial Value Problem
LQ	Linear Quadratic
LTV	Linear Time Varying
OCP	Optimal Control Problem
ODE	Ordinary Differential Equation
OMT	Optimal Mass Transport
PDE	Partial Differential Equation
PDF	Probability Density Function
QFT	Quantum Field Theory
SBP	Schrödinger Bridge Problem
SDE	Stochastic Differential Equation
w.r.t.	With Respect To

Abstract

Contraction and Reaction in Generalized Schrödinger Bridges

Alexis Teter

A bridge is a diffusion process that connects two given points in a finite dimensional vector space that are pinned at two given times. A Schrödinger bridge is a diffusion process that lifts this concept to the infinite dimensional manifold of probability measures. This dissertation contributes to the rapidly growing research area of Schrödinger bridge that is undergoing an explosion of mathematical and algorithmic developments across several disciplines: stochastic control, non-equilibrium statistical mechanics, and generative AI. This burgeoning interest is partly due to the fact that a Schrödinger bridge comes with a large deviation guarantee: it is the most-likely measure-valued path connecting the given endpoint measures. In this sense, the Schrödinger bridge is the most parsimonious model consistent with observed snapshots at two times. Part of its modern popularity is also because both the theory and algorithm for Schrödinger bridge are nonparametric, i.e., dispense the form or even the existence of finite dimensional sufficient statistic for the underlying measures. The computation is done directly on samples, without gridding or parameterizing the underlying state space. Starting from the classical Schrödinger bridge–a diffusive version of the optimal mass transport–this work considers several generalizations, and make progress in two fronts: contraction and reaction.

The issue of *contraction* concerns with the convergent numerical solution in the generalized Schrödinger bridge problems. The standard approach for solving such

problems is to deploy dynamic Sinkhorn recursions that are contractive with respect to Hilbert's projective metric. We quantify the worst-case contraction coefficient for the linear Schrödinger bridge in terms of the problem data. The issue of *reaction* concerns with the presence of an additive state cost in the objective that regularizes the controlled sample paths for all times, in addition to enforcing the exact steering between the endpoint statistics. Solving such problems via dynamic Sinkhorn recursions require Markov kernels associated with certain reaction-advection-diffusion partial differential equations, where the state costs play the role of the reaction rates. Such Markov kernels are not transition probability kernels since the reaction terms contribute to the non-conservation of probability mass, making direct application of dynamic Sinkhorn recursion challenging. For both classical and linear quadratic Schrödinger bridge problems, we overcome this challenge by deriving the associated Markov kernels in closed form, and showing how the dynamic Sinkhorn recursions with the derived kernels still apply. We show that the probabilistic Lambert problem–a problem of current interest in astrodynamics–is an instance of the generalized Schrödinger bridge where the state cost, and thus the reaction rate arises not from regularization, but from the nonlinear gravitational potential. This connects two hitherto disparate research areas, and enables nonparametric solution of the probabilistic Lambert problem. We conclude with several ongoing and future directions of research, including new equivalence between the Schrödinger bridge problems and the two point boundary value problems involving the quantum mechanical Schrödinger equation with generalized Bohm potential that we derive. Surprisingly, we show that such generalized Bohm potentials are necessarily complex-valued wherein the real part of the potential encodes elastic scattering (transmission of wave function), and the imaginary part encodes inelastic scattering (absorption of wave function).

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Abhishek Halder. His guidance and insights have been essential at every step in my journey as a graduate student. I am profoundly grateful to him for his mentorship. I will be grateful to him for the support he provided throughout this program for the rest of my career.

I would also like to thank all of my co-authors, whose insights were critical to the developments of my research.

During my Ph.D., I was supported in part by the National Science Foundation grant 2112755. The work in Chapter 6 was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract DEAC52-07NA27344. Partial funding for this work was provided by LLNL Laboratory Directed Research and Development grant GS 25-ERD-044. These supports are gratefully acknowledged.

During my fourth year of doctoral study, I was supported by the Achievement Rewards for College Scientists (ARCS) Foundation Fellowship. This support was critical to my success, as was meeting my fellow ARCS scholars and witnessing their inspirational research.

I'm incredibly grateful for the Teaching Assistantships from the Department of Applied Mathematics, and the opportunity to serve as an instructor for the course "Feedback Control" in 2022 California State Summer School for Mathematics and Science (COSMOS) at UC Santa Cruz.

I extend my gratitude to my colleagues at LLNL this past summer. I would particularly like to thank Alexx Perloff, who provided invaluable advice and support on navigating national laboratories and research institutions.

Finally, I thank my friends and family, who have been incredibly patient and supportive throughout my journey in this program. My sister, Charlynn, and my cat, Treble, deserve particular thanks.

1 | Introduction

1.1 Overview

In this introductory chapter, I define two problems key to this dissertation: the problem of optimal mass transport (OMT) and the Schrödinger Bridge Problem (SBP). I begin with the classical formulation of each, before introducing a generalized formulation. This dissertation will explore specific generalizations of the OMT and the SBP, with a particular focus on the latter.

It turns out that the classical SBP itself is a generalization of the classical OMT. While defining each problem's formulation, I highlight this link between the two. Identifying this connection between the classical formulations lays the groundwork for developments presented in Chapter 5.

After establishing the formulation of the OMT and the SBP, I provide a high-level description of the approach used to solve the SBP throughout this dissertation. I initially define this approach for the special case of the classical SBP. To clarify any restrictions on this approach as applied to generalizations of the Schrödinger Bridge Problem, I present the results of applying the relevant transform to a generalized SBP. In doing so, I lay the groundwork for the generalized cases pursued later in this dissertation.

This introductory chapter concludes with an outline of the structure of the ensuing chapters and an explanation of my publication history.

1.2 Optimal Mass Transport

The first problem of interest is that of optimal mass transport, henceforth referred to as OMT. In this section, I provide a brief overview of the history, formulation, and interpretation of OMT. I begin with the classical case of OMT, before providing the generalized form of OMT.

1.2.1 Static Formulation of OMT

Historically, the first formulation of the classical OMT was by Gaspard Monge in 1781 [1]. Monge's formulation of the classical OMT sought a mass-preserving measurable map $\boldsymbol{\tau} : \mathbb{R}^n \mapsto \mathbb{R}^n$ that pushforwards a given probability density function (PDF) ρ_0 to another given PDF ρ_1, while minimizing the transportation cost. In other words, the Monge formulation of OMT is the static optimization problem:

$$\begin{aligned} &\underset{\text{measurable } \boldsymbol{\tau}:\mathbb{R}^n \mapsto \mathbb{R}^n}{\arg\inf} \quad \mathbb{E}_{\rho_0} \frac{1}{2} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2 \\ &\text{subject to} \qquad \boldsymbol{x}_0 \sim \rho_0, \quad \boldsymbol{x}_1 \sim \rho_1, \quad \boldsymbol{\tau}(\boldsymbol{x}_0) = \boldsymbol{x}_1, \end{aligned} \tag{1.1}$$

where the random vector $\boldsymbol{x}_0 \in \mathbb{R}^n$ has the PDF ρ_0, and the random vector $\boldsymbol{x}_1 \in \mathbb{R}^n$ has the PDF ρ_1, and $\mathbb{E}_{\rho_0}$ denotes the expectation with respect to the PDF ρ_0. Throughout this dissertation, $\| \cdot \|_2$ represents the Euclidean norm. Problem (1.1) is an infinite dimensional optimization problem that is nonlinear and nonconvex in the decision variable $\boldsymbol{\tau}$, since the constraint can be equivalently written as $\rho_1 = \boldsymbol{\tau}_\sharp \rho_0$, where $\sharp$ denotes the pushforward.

In 1942, Leonid Kantorovich reformulated [2] the static OMT by shifting focus from finding the optimal transport map $\boldsymbol{\tau}^{\text{opt}}$ to instead computing an optimal coupling, π^{opt}, between the given ρ_0 and ρ_1. This optimal coupling solves the following infinite-dimensional *linear* optimization problem:

$$\underset{\pi\in\Pi(\rho_0,\rho_1)}{\arg\inf}\ \mathbb{E}_\pi \frac{1}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2, \tag{1.2}$$

where $\mathbb{E}_\pi[\cdot] := \int[\cdot]\mathrm{d}\pi$, and $\Pi(\rho_0,\rho_1)$ denotes the set of all joint probability measures π supported over the product space $\mathbb{R}^n \times \mathbb{R}^n$ such that $\int_{\mathbb{R}^n}\pi(\boldsymbol{x}_0,\boldsymbol{x}_1)\mathrm{d}\boldsymbol{x}_1 = \rho_0$ and $\int_{\mathbb{R}^n}\pi(\boldsymbol{x}_0,\boldsymbol{x}_1)\mathrm{d}\boldsymbol{x}_0 = \rho_1$. Under mild conditions, the optimal transport map $\boldsymbol{\tau}^{\mathrm{opt}}$ in (1.1) corresponds exactly to the support of the optimal coupling π^{opt} in (1.2). From a computational point of view, Kantorovich's reformulation (1.2) of the classical OMT as an infinite-dimensional linear program is more manageable than Monge's original formulation (1.1) of the classical OMT as an infinite-dimensional nonlinear nonconvex problem over measurable maps $\boldsymbol{\tau}(\cdot)$. See e.g., [3].

In contrast to the static formulations mentioned above, in this dissertation, we will focus on equivalent dynamic formulations mentioned next. These formulations are atypical stochastic optimal control problems.

1.2.2 Dynamic Formulation of OMT

Let $\mathcal{P}_2(\mathbb{R}^n)$ denote the set of PDFs supported over $\mathbb{R}^n$, whose elements have finite second raw moments, i.e.,

$$\mathcal{P}_2(\mathbb{R}^n) = \left\{\rho : \mathbb{R}^n \mapsto \mathbb{R}_{\geq 0} \mid \int_{\mathbb{R}^n}\rho\,\mathrm{d}\boldsymbol{x} = 1,\ \int_{\mathbb{R}^n}\|\boldsymbol{x}\|_2^2\,\rho\,\mathrm{d}\boldsymbol{x} < \infty\right\}.$$

For $\rho_0,\rho_1 \in \mathcal{P}_2(\mathbb{R}^n)$, let $\mathcal{P}_{01}$ denote the collection of all PDF-valued trajectories $\rho(t,\boldsymbol{x})$ that are continuous in $t\in[t_0,t_1]$, and supported over $\mathbb{R}^n$ for each $t\in[t_0,t_1]$, such that $\rho(t_0,\boldsymbol{x}) = \rho_0(\boldsymbol{x})$ and $\rho(t_1,\boldsymbol{x}) = \rho_1(\boldsymbol{x})$. Let $\mathcal{U}$ be the set comprising of all Markovian finite energy control policies, i.e.,

$$\mathcal{U} := \{\boldsymbol{u} : \mathbb{R}^n \times [t_0,t_1] \mapsto \mathbb{R}^n \mid \|\boldsymbol{u}\|_2^2 < \infty\}. \tag{1.3}$$

For a given time horizon $[t_0, t_1]$, the dynamic formulation of OMT, introduced by Benamou and Brenier [4], is the variational problem:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{u}\|_2^2\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{1.4a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u}) = 0, \tag{1.4b}$$

$$\rho(t_0,\boldsymbol{x}) = \rho_0, \quad \rho(t_1,\boldsymbol{x}) = \rho_1, \tag{1.4c}$$

where $\nabla_{\boldsymbol{x}}\cdot$ is the divergence operator with respect to vector $\boldsymbol{x}$. For details, refer to [5, Thm. 8.1], [6, Ch. 8]. The existence-uniqueness of the solution for (1.4) is guaranteed [7] for $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^n)$.

In (1.4), the vector $\boldsymbol{x} \in \mathbb{R}^n$ can be seen as the state of a controlled dynamical system, and the $\boldsymbol{u} \in \mathcal{U}$ as a feasible control policy. From a control perspective, (1.4) amounts to designing a minimum effort stochastic optimal control policy that is guaranteed to steer the controlled state statistics from the given ρ_0 at t_0 to the given ρ_1 at t_1.

1.2.3 Generalized OMT

A generalized version of OMT takes the form

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n} \left(q(\boldsymbol{x}) + r(\boldsymbol{u})\right)\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{1.5a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u})) = 0, \tag{1.5b}$$

$$\rho(t=t_0,\boldsymbol{x}) = \rho_0, \quad \rho(t=t_1,\boldsymbol{x}) = \rho_1. \tag{1.5c}$$

Assume that $r(\boldsymbol{u})$ is strictly convex and coercive, and that the cost-to-go $q(\boldsymbol{x})+r(\boldsymbol{u})$ is bounded.

The problem (1.5) generalizes the classical OMT (1.4) in two distinct ways. *First,* the cost-to-go in the objective function is now $q(\boldsymbol{x}) + r(\boldsymbol{u})$, the sum of a state-dependent cost and a control-dependent cost. This contrasts with the classical case (1.4), since there $q(\cdot) \equiv 0$, $r(\cdot) \equiv \frac{1}{2}\|\cdot\|_2^2$. *Second*, the vector field in the PDE constraint (1.5b) is allowed to be generic Lipschitz in time t, state $\boldsymbol{x}$, and control $\boldsymbol{u}$. In contrast, (1.4) concerns with the special case $\boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{u}) \equiv \boldsymbol{u}$.

In the subsequent chapters, we will consider specific choices for the state cost $q(\boldsymbol{x})$, the control cost $r(\boldsymbol{u})$, and the drift vector field $\boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{u})$.

1.3 Schrödinger Bridge Problem

The Schrödinger Bridge Problem, henceforth referred to as the SBP, is the primary focus of this dissertation. As in the case of OMT above, I begin with the classical formulation of this problem. I then provide the generalized form of the SBP, in preparation for the variants of the SBP considered in future chapters.

1.3.1 Static Formulation of SBP

The SBP originated from the works of Erwin Schrödinger in 1931-2, with the motivation of finding a probabilistic interpretation of quantum mechanics [8–10]. For the static SBP formulation, fix a (not necessarily small) regularization parameter $\varepsilon > 0$. Then consider the entropy-regularized version of the static OMT (1.2), which is an infinite dimensional *nonlinear* optimization problem:

$$\underset{\pi \in \Pi(\rho_0, \rho_1)}{\arg\inf} \; \mathbb{E}_\pi \left\{ \frac{1}{2} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2 + \varepsilon \log \pi \right\}. \tag{1.6}$$

The problem (1.6) is the static SBP. In contrast to (1.2), problem (1.6) is *strictly* convex in the decision variable $\pi \in \Pi(\rho_0, \rho_1)$.

Schrödinger's motivation behind (1.6) was not really to formulate a regularization of (1.2). Instead, the intrinsic motivation behind problem (1.6) was that it can be re-written as a *relative entropy* a.k.a. *Kullback-Leibler divergence* $\mathrm{D}_{\mathrm{KL}}(\cdot \parallel \cdot)$ minimization problem:

$$\underset{\pi\in\Pi(\rho_0,\rho_1)}{\arg\inf}\ \varepsilon\mathrm{D}_{\mathrm{KL}}\left(\pi(\boldsymbol{x}_0,\boldsymbol{x}_1) \parallel k(\boldsymbol{x}_0,\boldsymbol{x}_1)\left(\pi_0(\boldsymbol{x}_0)\otimes\pi_1(\boldsymbol{x}_1)\right)\right), \tag{1.7}$$

where π_0, π_1 are the marginal measures associated with the PDFs ρ_0, ρ_1, respectively, $\pi_0(\boldsymbol{x}_0)\otimes\pi_1(\boldsymbol{x}_1)$ denotes the product measure, and

$$k(\boldsymbol{x}_0,\boldsymbol{x}_1) := \exp\left(-\frac{1}{2\varepsilon}\|\boldsymbol{x}_0-\boldsymbol{x}_1\|_2^2\right), \tag{1.8}$$

is the *Gibbs kernel.* The Kullback-Leibler divergence $\mathrm{D}_{\mathrm{KL}}(\cdot \parallel \cdot)$ between two probability measures $\pi, \widetilde{\pi}$ is given by

$$\mathrm{D}_{\mathrm{KL}}\left(\pi \parallel \widetilde{\pi}\right) := \mathbb{E}_{\pi}\left\{\log\frac{\mathrm{d}\pi}{\mathrm{d}\widetilde{\pi}}\right\} = \begin{cases}\displaystyle\int \log\frac{\mathrm{d}\pi}{\mathrm{d}\widetilde{\pi}}\mathrm{d}\pi, & \text{if } \pi \ll \widetilde{\pi},\\ +\infty, & \text{otherwise,}\end{cases} \tag{1.9}$$

where $\pi \ll \widetilde{\pi}$ denotes that the measure π is absolutely continuous with respect to $\widetilde{\pi}$, and $\dfrac{\mathrm{d}\pi}{\mathrm{d}\widetilde{\pi}}$ denotes the Radon-Nikodym derivative.

Re-writing (1.6) in the form (1.7) has three merits. The *first merit* is that it clarifies the maximum-likelihood interpretation for the minimizer π^{opt} in the sense the coupling π^{opt} is the most-likely joint consistent with the prescribed marginal measures π_0, π_1 (or equivalently the marginal PDFs ρ_0, ρ_1). The *second merit* is that the existence-uniqueness of the minimizer π^{opt} becomes clear since problem (1.7) is a strictly convex program: the Kullback-Leibler projection of the Gibbs-weighted product measure $k(\boldsymbol{x}_0,\boldsymbol{x}_1)\left(\pi_0(\boldsymbol{x}_0)\otimes\pi_1(\boldsymbol{x}_1)\right)$ onto the convex polyhedron $\Pi(\rho_0,\rho_1)$. The *third merit* is that (1.7) can be numerically solved via the celebrated Sinkhorn algorithm [11–13], which is an recursive algorithm to solve the diagonal matrix scaling problem: given elementwise nonnegative square matrix $\boldsymbol{M}$, find positive diagonal matrices $\boldsymbol{D}_1, \boldsymbol{D}_2$ such that the scaled matrix $\boldsymbol{D}_1\boldsymbol{M}\boldsymbol{D}_2$ has a prescribed row sum and a prescribed column sum. The Sinkhorn recursions are contractive w.r.t. Hilbert's projective metric, and hence enjoy existence-uniqueness and worst-case linear rate of convergence guarantees.

1.3.2 Dynamic Formulation of SBP

While Schrödinger's original formulation is most naturally described as a static variational desribed above, equivalent stochastic control formulation for the same has been understood since the late 20th century, which is what is most relevant in this dissertation. The problem is to compute the minimum effort control required to steer a given PDF to another given PDF over a given time horizon, subject to a controlled Brownian motion constraint. This leads to a stochastic optimal control problem:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{u}\|_2^2\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{1.10a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u}) = \varepsilon\Delta_{\boldsymbol{x}}\rho, \quad \varepsilon > 0, \tag{1.10b}$$

$$\rho(t_0,\boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \rho(t_1,\boldsymbol{x}) = \rho_1(\boldsymbol{x}), \tag{1.10c}$$

where $\Delta_{\boldsymbol{x}}$ is the Laplacian operator with respect to the vector $\boldsymbol{x}$, and as before, $\mathcal{P}_{01}$ is the collection of PDF-valued trajectories $\rho(t,\cdot)$ that are continuous in $t \in [t_0, t_1]$ and satisfy $\rho(t_0,\cdot) = \rho_0$, $\rho(t_1,\cdot) = \rho_1$.

Note that the macroscopic dynamics (1.10b) corresponds to the microscopic controlled sample path dynamics given by the Itô stochastic differential equation

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{u}(t,\boldsymbol{x})\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}\boldsymbol{w}(t), \tag{1.11}$$

where $\boldsymbol{w} \in \mathbb{R}^n$ denotes the standard Wiener process. The Itô diffusion (1.11) is a controlled Brownian motion.

From inspection, the classical SBP (1.10) is a generalization of (1.4) in which the first order advection PDE, a.k.a. the Liouville PDE (1.4b) (see e.g., [14]) is replaced by the second order adevtion-diffusion PDE a.k.a. the Fokker-Planck-Kolmogorov (FPK) PDE (1.10b). When $\varepsilon \downarrow 0$, the SBP (1.10) reduces to the OMT problem (1.4), and the solution of (1.10) converges to the solution of (1.4) [15–17]. Chapter 5 includes discussion of the similar connection between a particular generalization of OMT and a corresponding generalization of the SBP.

1.3.3 Generalized SBP

Since this dissertation will explore a variety of generalizations of (1.10), I now define a generalized (dynamical) SBP as follows:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n} \left(q(\boldsymbol{x}) + r(\boldsymbol{u})\right)\, \rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{1.12a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot\left(\rho\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u})\right) = \Delta_{\boldsymbol{\Sigma}(t,\boldsymbol{x},\boldsymbol{u})}\rho \tag{1.12b}$$

$$\rho(t_0,\boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \rho(t_1,\boldsymbol{x}) = \rho_1(\boldsymbol{x}), \tag{1.12c}$$

where the weighted Laplacian

$$\Delta_{\boldsymbol{\Sigma}(t,\boldsymbol{x},\boldsymbol{u})}\rho := \sum_{i,j=1}^{n} \frac{\partial^2}{\partial_{x_i}\partial_{x_j}}\left((\boldsymbol{\Sigma})_{ij}\rho\right), \tag{1.13}$$

and the *diffusion tensor* $\boldsymbol{\Sigma} := \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$ is an $n\times n$ symmetric positive semi-definite matrix field, where $\boldsymbol{\sigma} = \boldsymbol{\sigma}(t,\boldsymbol{x},\boldsymbol{u}) \in \mathbb{R}^{n\times p}$ is the *diffusion coefficient.* Note that the diffusion coefficient may be time-, state-, and control-dependent.

As in (1.5), this problem formulation generalizes both the cost-to-go function in the objective and the controlled dynamics arising in the PDE constraint. The FPK

PDE (1.12b) is induced by the controlled sample path dynamics

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{u})\,\mathrm{d}t + \sqrt{2}\boldsymbol{\sigma}(t, \boldsymbol{x}, \boldsymbol{u})\,\mathrm{d}\boldsymbol{w}(t), \tag{1.14}$$

a generalization of (1.11). Specifically, the Itô stochastic differential equation (1.14) has controlled drift coefficient $\boldsymbol{f}$, and controlled diffusion coefficient $\boldsymbol{\sigma}$. Note that by choosing $q(\boldsymbol{x}) \equiv 0$, $r(\boldsymbol{u}) \equiv \frac{1}{2}\|\boldsymbol{u}\|_2^2$, and $\boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{u}) \equiv \boldsymbol{u}$, as in the OMT case, and additionally letting $\boldsymbol{\sigma}(t, \boldsymbol{x}, \boldsymbol{u}) \equiv \sqrt{\varepsilon}\boldsymbol{I}_n$, the microscopic (i.e., sample path) dynamics (1.14) reduces to (1.11), and the macroscopic (i.e., PDF) dynamics (1.12) reduces to (1.10).

Generalizations of the SBP considered in future chapters can likewise be recovered from (1.12) by choosing appropriate state cost $q(\boldsymbol{x})$, control cost $r(\boldsymbol{u})$, controlled drift coefficient $\boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{u})$, and controlled diffusion coefficient $\boldsymbol{\sigma}(t, \boldsymbol{x}, \boldsymbol{u})$.

1.4 Hopf-Cole Transform and Schrödinger Factors

I now outline the process of deriving the known solution to the classical SBP (1.10) in Ch. 1.4.1, and the same for generic control-affine SBP in Ch. 1.4.2.

1.4.1 Hopf-Cole Transform for Classical SBP

We begin by deriving the first order conditions for optimality for the classical SBP. By determining the Lagrangian of (1.10) and applying pointwise minimization, the optimal solution pair $(\rho_{\mathrm{opt}}, \boldsymbol{u}_{\mathrm{opt}})$ for the classical SBP (1.10) is found to solve the system of coupled, nonlinear PDEs:

$$\frac{\partial S}{\partial t} + \frac{1}{2}\|\nabla_{\boldsymbol{x}} S\|_2^2 + \varepsilon\Delta_{\boldsymbol{x}} S = 0, \tag{1.15a}$$

$$\frac{\partial \rho_{\mathrm{opt}}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\rho_{\mathrm{opt}} \nabla_{\boldsymbol{x}} S\right) = \varepsilon\Delta_{\boldsymbol{x}} \rho_{\mathrm{opt}}, \tag{1.15b}$$

with boundary conditions

$$\rho_{\text{opt}}(t_0, \boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \rho_{\text{opt}}(t_1, \boldsymbol{x}) = \rho_1(\boldsymbol{x}), \tag{1.16}$$

where $S \in C^{1,2}(\mathbb{R}^n; [t_0, t_1])$, and the optimal control $\boldsymbol{u}_\varepsilon^{\text{opt}} = \nabla_{\boldsymbol{x}} S(t, \boldsymbol{x})$. Specifically, (1.15a) is the Hamilton-Jacobi-Bellman (HJB) PDE, and (1.15b) is a controlled FPK PDE with control $\nabla_{\boldsymbol{x}} S(t, \boldsymbol{x})$. See Appendix A.1 for a general derivation of first order conditions; the derivation of (1.15) can be easily recovered.

The system of PDEs (1.15) in unknowns $\rho_{\text{opt}}(t, \boldsymbol{x}), S(t, \boldsymbol{x})$ are coupled and nonlinear. To transform this to a system of linear PDEs, the *Hopf-Cole transform* [18, 19] introduces the so-called *Schrödinger factors* φ_ε and $\widehat{\varphi}_\varepsilon$, as a change-of-variables $(\rho_{\text{opt}}, S) \mapsto (\widehat{\varphi}_\varepsilon, \varphi_\varepsilon)$ given by

$$\varphi_\varepsilon := \exp\left(\frac{S}{2\varepsilon}\right), \tag{1.17a}$$

$$\widehat{\varphi}_\varepsilon := \rho_{\text{opt}} \exp\left(-\frac{S}{2\varepsilon}\right). \tag{1.17b}$$

In the case of the classical SBP, this transformation allows the reformulation of the system (1.15)-(1.16) into a set of forward and backward heat equations

$$\frac{\partial \widehat{\varphi}_\varepsilon}{\partial t} = \varepsilon \Delta_{\boldsymbol{x}} \widehat{\varphi}_\varepsilon, \tag{1.18a}$$

$$\frac{\partial \varphi_\varepsilon}{\partial t} = -\varepsilon \Delta_{\boldsymbol{x}} \varphi_\varepsilon, \tag{1.18b}$$

with coupled boundary conditions

$$\begin{aligned} \widehat{\varphi}_\varepsilon(t_0, \boldsymbol{x}) \varphi_\varepsilon(t_0, \boldsymbol{x}) &= \rho_0, \\ \widehat{\varphi}_\varepsilon(t_1, \boldsymbol{x}) \varphi_\varepsilon(t_1, \boldsymbol{x}) &= \rho_1. \end{aligned} \tag{1.19}$$

Note that (1.18) can be trivially recovered from the derivation in Appendix A.2. Here, the subscript highlights the dependence of (1.17) on the choice of the parameter ε.

$$\begin{tikzcd}
\widehat{\varphi}_{\varepsilon}(t_0,\boldsymbol{x}) \arrow[rr, "\int"] & & \widehat{\varphi}_{\varepsilon}(t_1,\boldsymbol{x}) \arrow[d, "\rho_1(\boldsymbol{x})/\widehat{\varphi}_{\varepsilon}(t_1,\boldsymbol{x})"] \\
\varphi_{\varepsilon}(t_0,\boldsymbol{x}) \arrow[u, "\rho_0(\boldsymbol{x})/\varphi_{\varepsilon}(t_0,\boldsymbol{x})"] & & \varphi_{\varepsilon}(t_1,\boldsymbol{x}) \arrow[ll, "\int"]
\end{tikzcd}$$

Figure 1.1: Fixed point recursion over the endpoint Schrödinger factor pair $(\widehat{\varphi}_{\varepsilon,0}, \varphi_{\varepsilon,1})$ to solve the boundary-coupled system of PDEs (1.18)-(1.19). The arrows with integrals denote marching the solutions for the respective initial value problem (IVP) over time.

As is well-known [20, Sec. 8], [21, Sec. II], the SBP can thus be solved by computing the function pair $(\widehat{\varphi}_{\varepsilon}(t_0,\boldsymbol{x}), \varphi_{\varepsilon}(t_1,\boldsymbol{x}))$ satisfying a system of nonlinear integral equations, referred to as the *Schrödinger system*:

$$\rho_0(\boldsymbol{x}) = \widehat{\varphi}_{\varepsilon}(t_0,\boldsymbol{x}) \int_{\mathbb{R}^n} k(t_0,\boldsymbol{x},t_1,\boldsymbol{y})\varphi_{\varepsilon}(t_1,\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \tag{1.20a}$$

$$\rho_1(\boldsymbol{x}) = \varphi_{\varepsilon}(t_1,\boldsymbol{x}) \int_{\mathbb{R}^n} k(t_0,\boldsymbol{y},t_1,\boldsymbol{x})\widehat{\varphi}_{\varepsilon}(t_0,\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \tag{1.20b}$$

where k is the appropriate *uncontrolled* (i.e., $\boldsymbol{u} \equiv \boldsymbol{0}$) Markov transition probability kernel. In the case of the classical SBP (1.10), k is the well-known heat kernel, henceforth referred to as k_{B}, where the subscript B stands for the Brownian noise, a.k.a. the standard Wiener process in (1.11).

The Schrödinger system (1.20) can in turn be solved via the fixed point recursion over $(\widehat{\varphi}_{\varepsilon}(t_0,\boldsymbol{x}), \varphi_{\varepsilon}(t_1,\boldsymbol{x}))$ shown in Figure 1.1. This recursion proceeds as follows:

1. Make an everywhere positive initial guess for $\widehat{\varphi}_{\varepsilon}(t_0,\boldsymbol{x})$.

2. Using $\widehat{\varphi}_{\varepsilon}(t_0,\boldsymbol{x})$, integrate (1.18a) from $t = t_0$ to $t = t_1$ to determine $\widehat{\varphi}_{\varepsilon}(t_1,\boldsymbol{x})$.

3. Define

$$\varphi_\varepsilon(t_1, \boldsymbol{x}) = \frac{\rho_1}{\widehat{\varphi}_\varepsilon(t_1, \boldsymbol{x})}.$$

4. Integrate from $t = t_1$ to $t = t_0$ to determine $\varphi_\varepsilon(t_0, \boldsymbol{x})$.

5. Define

$$\widehat{\varphi}_\varepsilon(t_0, \boldsymbol{x}) = \frac{\rho_0}{\varphi_\varepsilon(t_0, \boldsymbol{x})}.$$

6. Repeat (2)-(5) until convergence.

The function pair $(\widehat{\varphi}_\varepsilon(t_0, \boldsymbol{x}), \varphi_\varepsilon(t_1, \boldsymbol{x}))$, thus computed, are used to find the Schrödinger factors

$$\widehat{\varphi}_\varepsilon(t, \boldsymbol{x}) := \int_{\mathbb{R}^n} k(t_0, \boldsymbol{y}, t, \boldsymbol{x})\widehat{\varphi}_{\varepsilon,0}(\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \quad t \geq t_0, \tag{1.21a}$$

$$\varphi_\varepsilon(t, \boldsymbol{x}) := \int_{\mathbb{R}^n} k(t, \boldsymbol{x}, t_1, \boldsymbol{y})\varphi_{\varepsilon,1}(\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \quad t \leq t_1. \tag{1.21b}$$

The optimally controlled state PDF and the optimal control for all $t \in [t_0, t_1]$ are then recovered from the Schrödinger factors as

$$\rho_{\text{opt}}(t, \boldsymbol{x}) = \widehat{\varphi}_\varepsilon(t, \boldsymbol{x})\varphi_\varepsilon(t, \boldsymbol{x}), \tag{1.22a}$$

$$\boldsymbol{u}_{\text{opt}}(t, \boldsymbol{x}) = 2\varepsilon\nabla_{\boldsymbol{x}} \log \varphi_\varepsilon(t, \boldsymbol{x}). \tag{1.22b}$$

Throughout this dissertation, an SBP will be referred to as *exactly solvable* if the corresponding (uncontrolled) Markov kernel κ is known in closed form, i.e., without integrals or series sum. This is the case for classical SBP discussed above, as the heat kernel k_{B} can be expressed as

$$k_{\text{B}}\left(t_0, \boldsymbol{x}_0, t_1, \boldsymbol{x}_1\right) = \left(4\pi\varepsilon(t_1 - t_0)\right)^{-n/2} \exp\left(-\frac{\|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2}{4\varepsilon(t_1 - t_0)}\right). \tag{1.23}$$

A closed form of the Markov kernel κ such as (1.23) allows for convenient numerical implementation of the forward and backward pass in dynamic Sinkhorn recursion (Fig. 1.1) via simple matrix-vector multiplications. This obviates randomized computation with Feynman-Kac path integrals or other nontrivial solvers [21,22] which incur function approximation errors. A key focus of this dissertation will therefore concern different approaches to determining a closed form Markov kernel for generalizations of the SBP.

1.4.2 Hopf-Cole Transform for Control-Affine SBP

In the case of the classical SBP, the Hopf-Cole transform, as discussed in Section 1.4.1, moves the coupling from the PDEs to the boundary conditions. In our pursuit of a closed form κ for generalizations of the SBP, we naturally wish to determine the conditions under which the Hopf-Cole transform remains beneficial.

To determine when the Hopf-Cole transform may be beneficially applied, consider a *control-affine* SBP:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(q\left(t,\boldsymbol{x}\right)+\frac{1}{2}\|\boldsymbol{u}\|_2^2\right)\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{1.24a}$$

$$\frac{\partial\rho}{\partial t}+\nabla_{\boldsymbol{x}}\cdot\left(\rho\left(\boldsymbol{A}\left(t,\boldsymbol{x}\right)+\boldsymbol{B}\left(t,\boldsymbol{x}\right)\boldsymbol{u}\right)\right)=\Delta_{\boldsymbol{\Sigma}(t,\boldsymbol{x},\boldsymbol{u})}\,\rho, \tag{1.24b}$$

$$\rho(t=t_0,\boldsymbol{x})=\rho_0,\quad \rho(t=t_1,\boldsymbol{x})=\rho_1. \tag{1.24c}$$

Note that (1.24) is an instance of the generalized SBP (1.12) for control cost $r(\boldsymbol{u})\equiv\frac{1}{2}\|\boldsymbol{u}\|_2^2$ and drift $\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u})\equiv\boldsymbol{A}\left(t,\boldsymbol{x}\right)+\boldsymbol{B}\left(t,\boldsymbol{x}\right)\boldsymbol{u}$. This choice of $\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u})$ implies that the controlled sample path dynamics is governed by the Itô stochastic differential equation

$$\mathrm{d}\boldsymbol{x}=\left(\boldsymbol{A}\left(t,\boldsymbol{x}\right)+\boldsymbol{B}\left(t,\boldsymbol{x}\right)\boldsymbol{u}\right)\mathrm{d}t+\boldsymbol{\sigma}(t,\boldsymbol{x},\boldsymbol{u})\mathrm{d}\boldsymbol{w}_t, \tag{1.25}$$

where the diffusion tensor in (1.24) is $\boldsymbol{\Sigma}=\boldsymbol{\sigma}\boldsymbol{\sigma}^{\top}$.

Our work [23] investigates solving (1.24) via the algorithm introduced in Section 1.4.1. We begin by making the following assumptions to ensure the strong solution [24, p. 66] for the uncontrolled ($\boldsymbol{u} \equiv \boldsymbol{0}$) version of (6.2), and to ensure that the corresponding transition density is everywhere continuous and positive [25, p. 364].

A1. (**Non-explosion and Lipschitz coefficients**) There exist constants $c_1, c_2 > 0$ such that $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\forall t \in [t_0, t_1]$,

$$\|\boldsymbol{A}(t,\boldsymbol{x})\|_2 + \|\boldsymbol{\sigma}(t,\boldsymbol{x})\|_2 \leq c_1 (1 + \|\boldsymbol{x}\|_2),$$

$$\|\boldsymbol{A}(t,\boldsymbol{x}) - \boldsymbol{A}(t,\boldsymbol{y})\|_2 \leq c_2 \|\boldsymbol{x} - \boldsymbol{y}\|_2.$$

A2. (**Uniformly lower bounded diffusion tensor**) There exists a constant $c_3 > 0$ such that $\forall \boldsymbol{x} \in \mathbb{R}^n$, $\forall t \in [t_0, t_1]$,

$$\langle \boldsymbol{x}, \boldsymbol{\Sigma}(t,\boldsymbol{x})\boldsymbol{x} \rangle \geq c_3 \|\boldsymbol{x}\|_2^2.$$

As in the classical case, we start by deriving the first order conditions for optimality for (1.24), arriving at the following theorem.

Theorem 1.1 (Coupled Nonlinear PDEs with linear boundary conditions)**.** *Let* ***A1****-****A2*** *hold. Consider two PDFs ρ_0, ρ_1 having finite second moments, a given time horizon $[t_0, t_1]$, and the state cost q suitably smooth. For $S \in \mathcal{C}^{1,2}([t_0,t_1];\mathbb{R}^n)$, the necessary conditions for optimality for problem* (1.24) *are*

Primal PDE:

$$\frac{\partial \rho_{\text{opt}}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\rho_{\text{opt}} \left(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{B}^\top \nabla_{\boldsymbol{x}} S\right)\right) = \Delta_{\boldsymbol{\Sigma}}\, \rho_{\text{opt}}, \tag{1.26a}$$

Dual PDE:

$$\frac{\partial S}{\partial t} + \langle \nabla_{\boldsymbol{x}} S, \boldsymbol{A} \rangle + \frac{1}{2} \langle \nabla_{\boldsymbol{x}} S, \boldsymbol{B}\boldsymbol{B}^\top \nabla_{\boldsymbol{x}} S \rangle + \langle \boldsymbol{\Sigma}, \mathit{Hess}_{\boldsymbol{x}} S \rangle = q, \tag{1.26b}$$

Primal Boundary Conditions :

$$\rho_{\text{opt}}\left(t_0, \boldsymbol{x}\right) = \rho_0(\boldsymbol{x}), \quad \rho_{\text{opt}}\left(t_1, \boldsymbol{x}\right) = \rho_1(\boldsymbol{x}) \tag{1.26c}$$

The optimal control is

$$\boldsymbol{u}_{\text{opt}} = \boldsymbol{B}^{\top} \nabla_{\boldsymbol{x}} S. \tag{1.27}$$

For a derivation of this Theorem, see Appendix A.1.

The derived necessary conditions (1.26) comprise of two coupled nonlinear PDEs (1.26a)-(1.26b) in the primal-dual variable pair $\left(\rho_{\text{opt}}, S\right)$ with two linear boundary conditions (1.26c). This system is atypical since both of these boundary conditions are available for the single variable ρ_{opt}.

Consider a generalized version of the Hopf-Cole transform 1.17 which introduces Schrödinger factors φ_λ and $\widehat{\varphi}_\lambda$ (see [18, 19]), as

$$\varphi_\lambda \coloneqq \exp\left(\frac{S}{\lambda}\right), \tag{1.28a}$$

$$\widehat{\varphi}_\lambda \coloneqq \rho_{\text{opt}} \exp\left(-\frac{S}{\lambda}\right). \tag{1.28b}$$

The constant $\lambda > 0$ is a scaling parameter; the dependence of this generalized Hopf-Cole transform on the choice of λ is emphasized by the subscript.

The appeal of (1.28) is that, *under certain conditions*, it converts (1.26), a system of coupled PDEs with linear boundary conditions, to a system of decoupled PDEs with bilinear boundary conditions.

Theorem 1.2 (Boundary-coupled nonlinear PDEs)**.** *Consider the setting as in Theorem 1.1. Let the pair* $\left(\rho_{\text{opt}}, S\right)$ *solve* (1.26)*. The pair* $\left(\widehat{\varphi}_\lambda, \varphi_\lambda\right)$ *given by* (1.28) *solves*

$$\frac{\partial \widehat{\varphi}_\lambda}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\widehat{\varphi}\left(\boldsymbol{A} + \boldsymbol{A}_\varphi\right)\right) - \Delta_{\boldsymbol{\Sigma}}\, \widehat{\varphi}_\lambda + \left(\frac{q}{\lambda} + q_\varphi\right) \widehat{\varphi}_\lambda = 0, \tag{1.29a}$$

$$\frac{\partial \varphi_\lambda}{\partial t} + \left\langle \nabla_{\boldsymbol{x}} \varphi_\lambda, \boldsymbol{A} + \boldsymbol{A}_\varphi \right\rangle + \left\langle \boldsymbol{\Sigma}, \mathit{Hess}_{\boldsymbol{x}} \varphi_\lambda \right\rangle - \left(\frac{q}{\lambda} + q_\varphi\right) \varphi_\lambda = 0, \tag{1.29b}$$

$$\widehat{\varphi}_\lambda(t_0,\boldsymbol{x})\,\varphi_\lambda(t_0,\boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \widehat{\varphi}_\lambda(t_1,\boldsymbol{x})\,\varphi_\lambda(t_1,\boldsymbol{x}) = \rho_1(\boldsymbol{x}), \tag{1.29c}$$

where

$$\boldsymbol{A}_\varphi := \left(\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}} \log\varphi_\lambda, \tag{1.30}$$

$$q_\varphi := \frac{1}{2}\left(\nabla_{\boldsymbol{x}}\log\varphi_\lambda\right)^\top \left(\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}} \log\varphi_\lambda. \tag{1.31}$$

The minimizing pair $(\rho_{\text{opt}}, \boldsymbol{u}_{\text{opt}})$ *for problem* (1.24) *is then recovered from*

$$\rho_{\text{opt}}(t,\boldsymbol{x}) = \widehat{\varphi}_\lambda(t,\boldsymbol{x})\varphi_\lambda(t,\boldsymbol{x}), \tag{1.32}$$

$$\boldsymbol{u}_{\text{opt}}(t,\boldsymbol{x}) = \lambda\,\boldsymbol{B}^\top\nabla_{\boldsymbol{x}}\log\varphi_\lambda(t,\boldsymbol{x}), \tag{1.33}$$

for all $t \in [t_0, t_1]$.

The proof of Theorem 1.2 is given in Appendix A.2.

Notice that the PDEs (1.29a)-(1.29b) are still *coupled and nonlinear*. Similar to (1.26a)-(1.26b), the equation-level coupling in (1.29a)-(1.29b) is one way in the sense (1.29b) has no dependence on $\widehat{\varphi}_\lambda$ but (1.29a) depends on both $\varphi_\lambda, \widehat{\varphi}_\lambda$. Unlike (1.26c), the boundary conditions (1.29c) are now bilinear. For these reasons, the utility of applying this generalized Hopf-Cole transform (1.28) to (1.24) is not immediately apparent.

However, a closer inspection of (1.29a)-(1.29b) reveals that the coupling and the nonlinearities emanate precisely from the φ_λ-dependent vector field $\boldsymbol{A}_\varphi$ and the the φ_λ-dependent reaction term q_φ defined in (1.30) and (1.31), respectively. The $\boldsymbol{A}_\varphi$ and q_φ, are both linear in the matrix $\lambda\boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}$. So, these PDEs become *decoupled and linear* provided the control and noise channels coincide. This is summarized in the following Corollary 1.3.

Corollary 1.3 (Boundary-coupled linear PDEs)**.** *Consider the same setting as in The-*

orem 1.2. If $\boldsymbol{B}\boldsymbol{B}^\top \propto \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$ or equivalently, $\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma} = \boldsymbol{0}$ for some $\lambda > 0$, then (1.29) *reduces to the following system of linear advection-diffusion-reaction PDEs that are only coupled through bilinear boundary conditions:*

$$\frac{\partial \widehat{\varphi}_\lambda}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\widehat{\varphi}_\lambda \boldsymbol{A}\right) - \Delta_{\boldsymbol{\Sigma}}\, \widehat{\varphi}_\lambda + \frac{q\widehat{\varphi}_\lambda}{\lambda} = 0, \tag{1.34a}$$

$$\frac{\partial \varphi_\lambda}{\partial t} + \langle \nabla_{\boldsymbol{x}}\varphi, \boldsymbol{A}\rangle + \langle \boldsymbol{\Sigma}, \mathit{Hess}_{\boldsymbol{x}}\varphi_\lambda \rangle - \frac{q\varphi_\lambda}{\lambda} = 0, \tag{1.34b}$$

$$\widehat{\varphi}_\lambda\left(t_0, \boldsymbol{x}\right) \varphi_\lambda\left(t_0, \boldsymbol{x}\right) = \rho_0(\cdot), \quad \widehat{\varphi}_\lambda\left(t_1, \boldsymbol{x}\right) \varphi_\lambda\left(t_1, \boldsymbol{x}\right) = \rho_1(\boldsymbol{x}). \tag{1.34c}$$

Proof. From (1.30)-(1.31), both $\boldsymbol{A}_\varphi, q_\varphi$ vanish for $\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma} = \boldsymbol{0}$. □

When $\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma} \neq \boldsymbol{0}$ for all $\lambda > 0$, the difference

$$\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}$$

quantifies the imbalance between the control and the noise fluctuation, causing an excess additive drift (1.30) and an excess additive reaction rate (1.31), to ensure the conservation of optimally controlled probability mass.

In a different context, Kappen [26, Sec. 3.1] provided a physical interpretation of $\boldsymbol{B}\boldsymbol{B}^\top \propto \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$: in the directions of low noise, optimal control (1.33) is expensive, and vice versa. Existing works [20, 21] assume *a priori* that $\boldsymbol{B}\boldsymbol{B}^\top \propto \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$. This assumption is practically motivated, as identical input and noise channels indeed occur for stochastic forcing and noisy actuators.

In this dissertation, the generalizations of the SBP under consideration will likewise concern cases in which $\boldsymbol{B}\boldsymbol{B}^\top \propto \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$, and therefore the Hopf-Cole transform (1.28) will prove beneficial.

1.5 Existing Literature and Original Contributions

This dissertation explores generalizations of the SBP. Prior works have generalized the classical SBP in multiple directions. One direction generalizes the SBP by considering more general drifts and diffusions [21, 27–35]. Specific versions of generalized OMTs have also appeared before; see e.g., [35, 36]. Another direction considers additional bounded domain constraints [37, 38] where the controlled sample paths are required to deterministically stay within the domain for all times.

However, works on the SBP with state cost-to-go, one of the main generalizations explored in this dissertation, are limited [10, 39], [40, Ch. 2.6, 2.7], [41]. On the other hand, deriving the worst-case contraction coefficient for a generalized SBP in terms of the problem data, as done in this dissertation, is novel.

In this dissertation, we deduce closed form kernels for the SBP with quadratic state cost, a novel result. By finding the Markov kernel k in closed form, our results enable the numerical solution of the SBP with quadratic state cost for *generic non-Gaussian endpoint distributions* ρ_0, ρ_1 *with finite second moments*. The existence and uniqueness of this problem was discussed in [42], wherein the kernel was constructed by solving Lyapunov equations. However, the solution with quadratic state cost presented in [42], while allowing for $\boldsymbol{Q}$ to be time-varying, only concerns with the special case of *Gaussian* ρ_0, ρ_1. In contrast, the closed form Markov kernel we derive makes *no parametric assumptions on* ρ_0, ρ_1.

We present two different approaches to analytically recovering the kernel for the SBP problem with quadratic state cost. In the process, we make an additional pedagogical contribution by explaining the general steps to derive the Green's functions or kernels of certain linear PDE initial value problems (IVPs) using *Weyl calculus* *tools* in a systematic manner. The mathematical development of Weyl calculus originated in quantum mechanics, but its relevance in solving problems outside quantum mechanics, especially in diffusion and control problems, are less known. Common references [43], [44, Ch. 2-3], [45–47] on Weyl calculus appeal to quantum physics and field-theoretic ideas to introduce the core mathematical concepts, albeit for good reasons. However, this style of exposition, even when mathematically rigorous, can become a barrier to the broader systems-control audience. To make Weyl calculus tools more accessible to a broad audience across different disciplines, we take a purposefully less rigorous recipe-style introduction of selective Weyl calculus tools with illustrative calculations.

We consider two ways to further generalize the SBP with quadratic state cost. First, we consider the case in which we steer a more general *linear time-varying diffusion*. We still allow for non-Gaussian initial and final distributions. We demonstrate that this *linear quadratic SBP* is still exactly solvable. Moreover, we present a novel approach that relies on finding a state-time dependent distance-like functional given by the solution of a deterministic optimal control problem. This technique breaks away from existing methods, uncovering a new connection between Markov kernels, distances, and optimal control. This connection is of interest beyond its immediate application in solving the linear quadratic SBP.

Our second method of generalizing the SBP with quadratic state cost is to instead consider the SBP with *bounded, positive* state cost that is *not necessarily quadratic*. In the process, we draw connections to a problem of interest to the astrodynamics community, the probabilistic Lambert Problem.

The deterministic Lambert problem has a substantial literature in the guidance-control community – a non-exhaustive list of references includes [48–52], [53, Ch.

13.4]. The Lambert problem with probabilistic uncertainty has been investigated before using Taylor expansion [54], using linearization followed by second order statistics (i.e., covariance) mapping [55,56], using the pushforward mapping of probability measures [57], using polynomial approximation [58], and using neural network and Gaussian process regression [59]. While these works consider the endpoint probabilistic uncertainties or some approximants thereof, they do not explicitly account for the stochastic dynamics.

While the importance of OMT in dynamics-control problems with probabilistic uncertainties is being recognized across different research communities at a very rapid pace [3, 6, 60–67], we uncover a hitherto unknown link between OMT and the probabilistic Lambert problem. Making a precise connection with OMT allows us to theoretically guarantee existence-uniqueness for the solution of the probabilistic Lambert problem.

By showing that the additive state cost in this generalized SBP corresponds to the probabilistic Lambert problem with process noise, we numerically solve the probabilistic Lambert problem using *nonparametric* computation. In the process, we propose a novel algorithm to numerically solve such a generalized SBP. While [68] also considers an SBP with killing or creation, [68] considers *unbalanced* endpoint marginals, which is not the case for us. While it is possible to convert that unbalanced problem to an equivalent balanced problem, doing so changes the mathematical structure and meaning of the stochastic optimal control problem, see [68, eq. (4.3) and Sec. 5].

In addition to exploring methods of exactly solving the generalizations discussed above, we also provide substantial contributions to the understanding of the convergence rate of the algorithm defined in Fig. 1.1. Thanks to the Banach contraction mapping theorem, this recursion has guaranteed *linear* convergence with a *contraction coefficient* $\kappa \in (0, 1)$. The smaller the κ, the faster the convergence. In practice, the contraction coefficient κ is numerically observed to be small (i.e., fast convergence) even in more general settings [69, Fig. 4], [21, 34, 70] than classical or linear SBPs. The recent work [71] investigates classical SBP convergence from a sample complexity perspective. [72] provides a *worst-case* contraction coefficient $\gamma \geq \kappa$ for the classical SBP. We provide novel control-theoretic as well as geometric interpretations for this result. We additionally derive a new formula for the worst-case contraction coefficient for the *linear SBP* in terms of the problem data, and provide control-theoretic and geometric interpretations of the same. We also provide a practical application of these formula, by highlighting how pre-conditioning the supports can help reduce the worst-case contraction coefficient, thereby improving the convergence of the linear SBP.

1.6 Organization of Dissertation

The remaining of this dissertation is organized as follows.

In Chapter 2, I summarize existing results on the contraction coefficient for the classical SBP. I then define the first generalization of the SBP discussed in this dissertation: the linear SBP. I identify the corresponding changes to the recursive algorithm introduced above (see Figure 1.1) when solving the linear SBP rather than the classical SBP. I then present a derivation of the contraction coefficient for this generalized case, as published in a joint work with Abhishek Halder and Yongxin Chen (see [73]). Our resulting geometric and control-theoretic interpretations, for both the classical and the linear SBP, are thoroughly explained. Chapter 2 concludes with a discussion of the potential applications we identified of the contraction coefficient, and an example of one such application.

Chapter 3 concerns a different generalization of the SBP; one which includes a fixed quadratic state cost. The primary focus of this chapter is the derivation of the Markov kernel k for such a SBP. I present two different approaches for deriving such a kernel; one which we employ Hermite polynomials (see [74]) and one which we use Weyl calculus (see [75]). Our successful derivation of the Markov kernel via each approach both demonstrates that this generalization of the SBP is *exactly solvable*, and provides the solution of this SBP.

In Chapter 4, we combine the generalizations discussed in Chapters 2 and 3 to consider a linear SBP with (not necessarily fixed) quadratic state cost. In pursuit of the Markov kernel k corresponding to this new generalization, we introduce an entirely new approach, which is designed from observations about the structure of known Markov kernels (see [76]). We demonstrate the success of this approach for known kernels, and for this new generalization.

In Chapter 5, we further generalize the SBP studied in Chapter 3, by considering an additive state cost of any structure, rather than restricting our developments to a *quadratic* state cost. I present the existence-uniqueness and first-order conditions of optimality of the solution for this SBP with state cost. I then present our work on a particular case study of the SBP with additive state cost: the probabilistic Lambert's problem (see [22]). After a brief overview of the historical and modern applications enabled by the solution of this problem to motivate the rest of the chapter, we leverage properties of generalized OMT and the SBP to prove the existence and uniqueness of the solution to the probabilistic Lambert's problem. To enable the solution of this problem via the typical Sinkhorn recursions, we then use the

Fourier and Laplace transforms to determine an approach for the necessary forward and backward integration. We then construct an algorithm to numerically solve the probabilistic Lambert's problem. The corresponding numerical simulation results highlight the practical use of the proposed algorithm.

I conclude this dissertation with a summary of the key results, as well as a discussion of our results from applying an alternative transform to the SBP, which suggests intriguing routes for further study.

1.7 Publications

The following table, Table 1.1, gives a list of my published works and works currently under review, as well as their correspondence with the chapters of this dissertation.

Table 1.1: Publications and related chapters in this dissertation.

Publication	Chapter	Ref.
"On the Hopf-Cole Transform for Control-affine Schrödinger Bridge", under review in *IEEE Trans. Automatic Control.* A.M.H. Teter and A. Halder.	Ch. 1	[23]
"On the Contraction Coefficient of the Schrödinger Bridge for Stochastic Linear Systems", *IEEE Control Systems Letters, 7:3325–3330, 2023.* A.M.H. Teter, Y. Chen, and A. Halder	Ch. 2	[73]

"Schrödinger Bridge with Quadratic State Cost is Exactly Solvable", accepted in *IEEE Trans. Automatic Control.* A.M.H. Teter, W. Wang and A. Halder	Ch. 3, 5	[74]
"Weyl Calculus and Exactly Solvable Schrödinger Bridges with Quadratic State Cost", *60th Annual Allerton Conference , 1–8, 2024, Invited paper in Session 'Duality in Control, Inference, and Learning'.* A.M.H. Teter, W. Wang and A. Halder	Ch. 3	[75]
"Markov Kernels, Distances and Optimal Control: A Parable of Linear Quadratic Non-Gaussian Distribution Steering", under review in *IEEE Trans. Automatic Control.* A.M.H. Teter, W. Wang, S. Shivakumar and A. Halder	Ch. 4	[76]
"Probabilistic Lambert Problem: Connections with Optimal Mass Transport, Schrödinger Bridge, and Reaction-Diffusion PDEs", *SIAM Journal on Applied Dynamical Systems, 24(1):16-43, 2025.* A.M.H. Teter, I. Nodozi and A. Halder	Ch. 5	[77]

"Solution of the Probabilistic Lambert's Problem: Optimal Transport Approach", *Mathematical Theory of Networks and Systems, 2024, Invited paper in Session 'Optimal Transport'.* A.M.H. Teter, I. Nodozi and A. Halder	Ch. 5	[22]
"Control-affine Schrödinger Bridge and Generalized Bohm Potential", under review in *IEEE Control Systems Letters.* A.M.H. Teter, A. Halder, M.D. Schneider, A.S. Perloff, J. Pratt, C.M. Artman, and M. Demireva	Ch. 6	[78]
"Proximal Mean Field Learning in Shallow Neural Networks", *Transactions on Machine Learning Research, 2024.* A.M.H. Teter, I. Nodozi, and A. Halder	N/A	[79]
"Set Invariance with Probability One for Controlled Diffusion: Score-based Approach", under review in *IEEE Trans. Automatic Control.* W. Wang, A.M.H. Teter, M. Arcak and A. Halder	N/A	[80]

2 | The Contraction Coefficient in Linear SBPs

2.1 Overview

As discussed in Chapter 1.4, the SBP is solved using fixed point recursions, repeated to convergence. Thanks to the Banach contraction mapping theorem, the fixed point recursion illustrated in Figure 1.1 of Chapter 1 is known to enjoy guaranteed *linear* convergence with a *contraction coefficient* $\kappa \in (0,1)$.

This contraction coefficient, κ, is understood as follows. Let Φ represent the mapping of one full recursion in Figure 1.1, i.e.,

$$\widehat{\varphi}^{(k+1)}(t_0,\boldsymbol{x}) = \Phi\left(\widehat{\varphi}^{(k)}(t_0,\boldsymbol{x})\right), \quad \text{recursion index } k \in \mathbb{N}_0,$$

where $\widehat{\varphi}^{(0)}(t_0,\boldsymbol{x})$ represents our initial, everywhere positive guess of the true $\widehat{\varphi}(t_0,\boldsymbol{x})$, and $\widehat{\varphi}^{(j)}(t_0,\boldsymbol{x})$ represents the updated approximation of $\widehat{\varphi}(t_0,\boldsymbol{x})$ after j full recursions.

Then the contraction coefficient κ is defined as

$$\begin{aligned}\kappa = \min_{0<c<1}\ & c\\ \text{such that}\quad & d_H(\Phi(x),\Phi(y)) \leq c\, d_H(x,y),\end{aligned}$$

where $d_H(\cdot,\cdot)$ is the Hilbert's projective metric between any two elements of a pointed convex cone in a real vector space, given by

$$d_H(x,y) := \log\left(\frac{\inf\{\lambda \geq 0 | x \leq \lambda y\}}{\sup\{\lambda \geq 0 | \lambda y \leq x\}}\right). \tag{2.1}$$

The $\leq$ in (2.1) is understood as the conic inequality. In particular, for $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n_{>0}$

(positive orthant), we have

$$d_H(\boldsymbol{x}, \boldsymbol{y}) = \log\left(\frac{\max_{i=1,\cdots,n} x_i/y_i}{\min_{i=1,\cdots,n} x_i/y_i}\right). \tag{2.2}$$

Notice that the smaller (resp., larger) the contraction coefficient κ, the faster (resp., slower) the convergence .

The contraction coefficient κ depends on the specific choices of ρ_0, ρ_1. This chapter thus concerns with the *worst-case contraction coefficient* γ satisfying $\kappa \le \gamma$, where the "worst-case" refers to all possible ρ_0, ρ_1 supported on given compact sets $\mathcal{X}_0, \mathcal{X}_1 \subset \mathbb{R}^n$. An understanding of γ is needed to evaluate the efficiency of the solution to the SBP.

The contributions detailed in this chapter are a joint work with Abhishek Halder and Yongxin Chen, and presented in [73]. This chapter contains our derivation of the worst-case contraction rate for a linear SBP in terms of the particular problem data. To lay the groundwork for these results, I first provide existing results on the contraction coefficient for the classical SBP, before detailing the generalization to a linear SBP. I then detail our geometric and optimal control-theoretic interpretations of the derived worst-case contraction rate for both the classical and linear SBP. This chapter concludes with our numerical example illustrating how the worst-case contraction coefficient can be employed in preconditioning the endpoint supports.

2.2 The Contraction Coefficient for the Classical SBP

To lay the groundwork for analyzing the contraction coefficient for the linear SBP, we must first collect corresponding ideas for the classical SBP.

Recall from Chapter 1 that, for a given $\varepsilon > 0$, the (scaled) standard Wiener process

$\sqrt{2\varepsilon}\,\mathrm{d}\boldsymbol{w}(t)$ in $\mathbb{R}^n$ has an associated Markov kernel

$$k_{\mathrm{B}}\left(t_0,\boldsymbol{x}_0,t_1,\boldsymbol{x}_1\right) := \left(4\pi\varepsilon(t_1-t_0)\right)^{-n/2}\exp\left(-\frac{\|\boldsymbol{x}_0-\boldsymbol{x}_1\|_2^2}{4\varepsilon(t_1-t_0)}\right), \tag{2.3}$$

where the subscript B stands for the Brownian a.k.a. standard Wiener process. Therefore, for the classical SBP, the Markov kernel $k \equiv k_{\mathrm{B}}$ in the Schrödinger system

$$\rho_0(\boldsymbol{x}) = \widehat{\varphi}_\varepsilon(t_0,\boldsymbol{x})\int_{\mathbb{R}^n} k(t_0,\boldsymbol{x},t_1,\boldsymbol{y})\varphi_\varepsilon(t_1,\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \tag{2.4a}$$

$$\rho_1(\boldsymbol{x}) = \varphi_\varepsilon(t_1,\boldsymbol{x})\int_{\mathbb{R}^n} k(t_0,\boldsymbol{y},t_1,\boldsymbol{x})\widehat{\varphi}_\varepsilon(t_0,\boldsymbol{y})\mathrm{d}\boldsymbol{y}. \tag{2.4b}$$

Note that k_{B} is continuous for all $(\boldsymbol{x}_0,\boldsymbol{x}_1) \in \mathbb{R}^n \times \mathbb{R}^n$. Furthermore, for $\mathcal{X}_0,\mathcal{X}_1$ compact, there exist constants $\alpha_{\mathrm{B}},\beta_{\mathrm{B}}$ such that $0 < \alpha_{\mathrm{B}} \le \beta_{\mathrm{B}} < \infty$, and

$$\alpha_{\mathrm{B}} \le k_{\mathrm{B}}\left(t_0,\boldsymbol{x}_0,t_1,\boldsymbol{x}_1\right) \le \beta_{\mathrm{B}} \quad \forall\,(\boldsymbol{x}_0,\boldsymbol{x}_1) \in \mathcal{X}_0 \times \mathcal{X}_1. \tag{2.5}$$

In [72, equation (17)], the rate of convergence for the Schrödinger system (2.4) associated with the classical SBP was related to the quantity

$$\gamma_{\mathrm{B}} := \tanh^2\left(\frac{1}{2}\log\left(\frac{\beta_{\mathrm{B}}}{\alpha_{\mathrm{B}}}\right)\right) \in (0,1). \tag{2.6}$$

Specifically, γ_{B} was shown [72, Lemma 5] be to the *worst-case contraction coefficient* for a single pass of the recursion shown in Figure 1.1.

Define

$$\widetilde{\alpha}_{\mathrm{B}} := \max_{\boldsymbol{x}_0\in\mathcal{X}_0,\boldsymbol{x}_1\in\mathcal{X}_1} \|\boldsymbol{x}_0-\boldsymbol{x}_1\|_2^2, \tag{2.7a}$$

$$\widetilde{\beta}_{\mathrm{B}} := \min_{\boldsymbol{x}_0\in\mathcal{X}_0,\boldsymbol{x}_1\in\mathcal{X}_1} \|\boldsymbol{x}_0-\boldsymbol{x}_1\|_2^2, \tag{2.7b}$$

wherein the maximum and minimum are guaranteed to exist due to the compactness of $\mathcal{X}_0,\mathcal{X}_1$. From (2.3) and (2.5), it then follows that the $\alpha_{\mathrm{B}},\beta_{\mathrm{B}}$ in (2.6) can be expressed as

$$\alpha_{\rm B} = \frac{\exp\left(-\widetilde{\alpha}_{\rm B}/(4\varepsilon)\right)}{\sqrt{(4\pi\varepsilon)^n}}, \quad \beta_{\rm B} = \frac{\exp\left(-\widetilde{\beta}_{\rm B}/(4\varepsilon)\right)}{\sqrt{(4\pi\varepsilon)^n}}. \tag{2.8}$$

Consequently, (2.6) may be rewritten

$$\gamma_{\rm B} = \tanh^2\left(\frac{\widetilde{\alpha}_{\rm B} - \widetilde{\beta}_{\rm B}}{8\varepsilon}\right) \in (0,1). \tag{2.9}$$

2.3 The Linear SBP

We now turn our attention to the SBP for a stochastic linear time-varying system. This variant concerns with minimum effort steering of a controlled stochastic process $\boldsymbol{x}(t)$ satisfying Itô diffusion

$$\mathrm{d}\boldsymbol{x}(t) = \left(\boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t,\boldsymbol{x})\right)\mathrm{d}t + \sqrt{2\varepsilon}\boldsymbol{B}(t)\,\mathrm{d}\boldsymbol{w}(t), \tag{2.10}$$

over a fixed time horizon $[t_0, t_1]$, from a given initial state PDF $\rho_0(\boldsymbol{x}) := \rho(t = t_0, \boldsymbol{x})$ to another given terminal state PDF $\rho_1(\boldsymbol{x}) := \rho(t = t_1, \boldsymbol{x})$. In (2.10), the controlled state $\boldsymbol{x} \in \mathbb{R}^n$, the control input $\boldsymbol{u} \in \mathbb{R}^m$ and the standard Wiener process noise $\boldsymbol{w} \in \mathbb{R}^m$.

The constant $\varepsilon > 0$ denotes the strength of the process noise, and is not necessarily small. In (2.10), the process noise enters through the same "channels" as the input. This is true of many practical applications, such as in noisy actuators, in external disturbances including wind gust affecting the system state via forcing, and in unmodeled dynamics.

The minimum effort objective translates to the following problem:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{x}\|_2^2\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{2.11a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot\left(\rho\ \left(\boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t,\boldsymbol{x})\right)\right) = \varepsilon\Delta_{\boldsymbol{x}}\rho, \qquad \varepsilon > 0, \tag{2.11b}$$

$$\rho(t_0, \boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \rho(t_1, \boldsymbol{x}) = \rho_1(\boldsymbol{x}), \tag{2.11c}$$

where $\mathcal{U}$ of finite energy Markovian control policies, as defined in (1.3). For convenience, problem (2.11) is referred to as the linear SBP.

Note that the classical SBP [8–10, 16] is a special case of problem (2.11) in which $\boldsymbol{A}(t) \equiv \boldsymbol{0}, \boldsymbol{B}(t) \equiv \boldsymbol{I}_n$.

Recall that the generalized SBP defined in Chapter 1 took the form

$$\begin{aligned}
&\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n} \left(q(\boldsymbol{x}) + r(\boldsymbol{u})\right)\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t\\
&\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot\left(\rho\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u})\right) = \Delta_{\boldsymbol{\Sigma}(t,\boldsymbol{x},\boldsymbol{u})}\rho\\
&\rho(t_0,\boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \rho(t_1,\boldsymbol{x}) = \rho_1(\boldsymbol{x}).
\end{aligned}$$

So (2.11) can be viewed as a specific case of this generalized SBP in which $\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u}) \equiv \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t,\boldsymbol{x})$ and $\boldsymbol{\sigma}(t,\boldsymbol{x},\boldsymbol{u}) \equiv \sqrt{2\varepsilon}\boldsymbol{B}(t)$, with costs $q(\boldsymbol{x}) \coloneqq \frac{1}{2}\|\boldsymbol{x}\|_2^2$ and $r(\boldsymbol{u}) \equiv 0$, as in the classical SBP.

Throughout this chapter, we assume the following:

A1. The matricial trajectory pair $(\boldsymbol{A}(t), \boldsymbol{B}(t))$ is continuous and bounded for all $t \in [t_0, t_1]$,

A2. $(\boldsymbol{A}(t), \boldsymbol{B}(t))$ is a controllable pair in the sense that the finite horizon *controllability Gramian*

$$\boldsymbol{M}_{10} \coloneqq \int_{t_0}^{t_1} \boldsymbol{\Phi}_{t\tau}\boldsymbol{B}(\tau)\boldsymbol{B}^\top(\tau)\boldsymbol{\Phi}_{t\tau}^\top \mathrm{d}\tau \tag{2.13}$$

is symmetric positive definite, where $\boldsymbol{\Phi}_{t\tau} \coloneqq \boldsymbol{\Phi}(t,\tau)$ for $t_0 \le \tau \le t \le t_1$ denotes the state transition matrix associated with the state matrix $\boldsymbol{A}(t)$,

A3. The given endpoint PDFs ρ_0, ρ_1 have compact supports $\mathcal{X}_0, \mathcal{X}_1 \subset \mathbb{R}^n$, respectively, satisfying $\int_{\mathcal{X}_0}\rho_0 = \int_{\mathcal{X}_1}\rho_1 = 1$.

Under the stated assumptions **A1**-**A3**, the existence-uniqueness for the solution of this problem are guaranteed, and the solution can be computed by solving the associated Schrödinger system (2.4), as introduced in Chapter 1, with Markov kernel $k := k_{\rm L}$ where

$$k_{\rm L}\left(t_0, \boldsymbol{x}_0, t_1, \boldsymbol{x}_1\right) := \det\left(\boldsymbol{M}_{10}\right)^{-1/2} k_{\rm B}\left(t_0, \boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\boldsymbol{x}_0, t_1, \boldsymbol{M}_{10}^{-1/2}\boldsymbol{x}_1\right). \tag{2.14}$$

The fixed point recursion introduced in Figure 1.1, for this definition of Markov kernel, is provably contractive [72, Sec. III] in Hilbert's projective metric [81,82]. See also [83]. For recent works elaborating connections between the SBP and stochastic optimal control, see [20, 84].

Note that the optimally controlled state PDF $\rho_{\rm opt}(t, \boldsymbol{x})$ and the optimal control $\boldsymbol{u}_{\rm opt}(t, \boldsymbol{x})$ for a linear SBP can be recovered from the Schroödinger factors as

$$\begin{aligned}
\rho_{\rm opt}(t, \boldsymbol{x}) &= \hat{\varphi}_{\varepsilon}(t, \boldsymbol{x})\varphi_{\varepsilon}(t, \boldsymbol{x})\\
\boldsymbol{u}_{\rm opt}(t, \boldsymbol{x}) &= 2\varepsilon\boldsymbol{B}(t)^{\top}\nabla_{\boldsymbol{x}}\log\varphi_{\varepsilon}(t, \boldsymbol{x})
\end{aligned}$$

for all $t \in [t_0, t_1]$, where φ_{ϵ} and $\hat{\varphi}_{\varepsilon}$ are defined as in (1.17a). This determines the solution for the linear SBP.

2.4 The Contraction Coefficient for the Linear SBP

Formula (2.9) provides an explicit relation between the worst-case contraction coefficient $\gamma_{\rm B}$ for the classical Schrödinger system and the problem data given by the tuple $(\mathcal{X}_0, \mathcal{X}_1, \varepsilon)$. Here, we report results from our investigation into how the worst-case contraction coefficient $\gamma_{\rm L}$ for the problem (2.11) depends on the problem data given by the tuple $(\mathcal{X}_0, \mathcal{X}_1, \varepsilon, \boldsymbol{A}(t), \boldsymbol{B}(t))$, involving both the endpoint PDFs (with their compact supports) as well as the dynamical coefficients $\boldsymbol{A}(t), \boldsymbol{B}(t), \varepsilon$ in (2.10).

Intuition suggests that, keeping $(\mathcal{X}_0, \mathcal{X}_1, \varepsilon)$ fixed, the rate of convergence will differ for different choices of controllable pair $(\boldsymbol{A}(t), \boldsymbol{B}(t))$. Faster (resp. slower) convergence is expected for linear systems which are "easier" (resp. "harder") to control than others. Thus, the worst-case contraction coefficient $\gamma_{\rm L}$ is expected to depend on the controllability Gramian (2.13).

We have the following result.

Theorem 2.1. *Consider the linear SBP* (2.11) *with assumptions* ***A1-A3****. The associated Schrödinger system* (2.4) *with* $q \equiv k_L$ *has worst-case contraction coefficient* $\gamma_L \in (0,1)$ *given by*

$$\gamma_L = \tanh^2\left(\frac{\widetilde{\alpha}_L - \widetilde{\beta}_L}{8\varepsilon}\right), \tag{2.15}$$

where

$$\widetilde{\alpha}_L := \max_{\boldsymbol{x}_0 \in \mathcal{X}_0, \boldsymbol{x}_1 \in \mathcal{X}_1} \left(\boldsymbol{\Phi}_{t_1 t_0}\boldsymbol{x}_0 - \boldsymbol{x}_1\right)^\top \boldsymbol{M}_{10}^{-1}\left(\boldsymbol{\Phi}_{t_1 t_0}\boldsymbol{x}_0 - \boldsymbol{x}_1\right), \tag{2.16a}$$

$$\widetilde{\beta}_L := \min_{\boldsymbol{x}_0 \in \mathcal{X}_0, \boldsymbol{x}_1 \in \mathcal{X}_1} \left(\boldsymbol{\Phi}_{t_1 t_0}\boldsymbol{x}_0 - \boldsymbol{x}_1\right)^\top \boldsymbol{M}_{10}^{-1}\left(\boldsymbol{\Phi}_{t_1 t_0}\boldsymbol{x}_0 - \boldsymbol{x}_1\right). \tag{2.16b}$$

For a proof of Theorem 2.1, see Appendix B.1.

2.5 Control-Theoretic Interpretation

The objective in (2.16) is precisely the minimum cost for the deterministic optimal control problem:

$$\min_{\boldsymbol{v}} \quad \int_{t_0}^{t_1} \frac{1}{2}\|\boldsymbol{u}\|_2^2 \,\mathrm{d}t \tag{2.17a}$$

$$\text{subject to} \quad \dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{B}(t)\boldsymbol{u}, \tag{2.17b}$$

$$\boldsymbol{x}(t = t_0) = \boldsymbol{x}_0, \quad \boldsymbol{x}(t = t_1) = \boldsymbol{x}_1, \tag{2.17c}$$

i.e., the cost for minimum effort steering of a controllable linear time-varying system from a fixed initial state $\boldsymbol{x}_0$ to a fixed terminal state $\boldsymbol{x}_1$ over the given time horizon $[t_0, t_1]$. See e.g., [85, p. 194].

For fixed $(\boldsymbol{A}(t), \boldsymbol{B}(t))$, and therefore for fixed $\boldsymbol{\Phi}_{t_1 t_0}, \boldsymbol{M}_{10}$, the optimal cost (2.17a) varies with the variation in endpoints $\boldsymbol{x}_0 \in \mathcal{X}_0, \boldsymbol{x}_1 \in \mathcal{X}_1$. Thus, $\widetilde{\alpha}_{\mathrm{L}}$ (resp. $\widetilde{\beta}_{\mathrm{L}}$) equals the worst-case (resp. best-case) optimal state transfer cost for the source and target supports $\mathcal{X}_0, \mathcal{X}_1$. As $\tanh^2(\cdot)$ increases over positive real values, and $\widetilde{\alpha}_{\mathrm{L}} - \widetilde{\beta}_{\mathrm{L}} \geq 0$, γ_{L} in (2.15) is an increasing function of the *range* of optimal state transfer cost: $\widetilde{\alpha}_{\mathrm{L}} - \widetilde{\beta}_{\mathrm{L}}$.

Likewise, note that γ_{L} is a decreasing function of process noise ε due to the presence of process noise ε in the denominator in (2.15). When other parameters are held fixed, a stronger process noise reduces γ_{L}. A stronger process noise improving the contraction coefficient matches intuition.

2.6 Geometric Interpretation

In the following, I provide our geometric insights for (2.7) and (2.16) and discuss the computation of γ_{L}.

In (2.7) and (2.16), the $\widetilde{\alpha}_{\mathrm{B}}, \widetilde{\alpha}_{\mathrm{L}}$ (resp. $\widetilde{\beta}_{\mathrm{B}}, \widetilde{\beta}_{\mathrm{L}}$) can be seen as the maximal (resp. minimal) separation between the sets $\mathcal{X}_0, \mathcal{X}_1$ or their linear transforms. While (2.7a) and (2.16a) are invariant under convexification of the supports, (2.7b) and (2.16b) can decrease or stay the same under convexification. Therefore, the contraction rate (2.9) (resp. (2.15)) under convexification is greater than or equal to (meaning slower or same rate) the corresponding rate without convexification of the supports.

When the compact sets $\mathcal{X}_0, \mathcal{X}_1$ are also convex, computing the minimum values in

(2.7b) and (2.16b) reduces to solving the "best approximation pair" problem; e.g., [86]. Then, (2.7b) and (2.16b) can be numerically computed using the Gilbert-Johnson-Keerthi (GJK) algorithm or its improved variants [87–89].

On the other hand, the maximum values in (2.7a) and (2.16a) correspond to the *squared diameters* of the Cartesian products $\mathcal{X}_0 \times \mathcal{X}_1$ and $\boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\mathcal{X}_0 \times \boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1$, respectively. When $\mathcal{X}_0, \mathcal{X}_1$ are compact and convex, so are these Cartesian products. Therefore, the maximum values in (2.7a) and (2.16a) must be attained at the boundaries of the sets $\mathcal{X}_0 \times \mathcal{X}_1$ and $\boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\mathcal{X}_0 \times \boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1$, respectively. However, numerical computation of these maximum values may be cumbersome depending on what kind of descriptions for the convex sets $\mathcal{X}_0, \mathcal{X}_1$ are available.

To offer key geometric insights on (2.7) and (2.16) for the case of convex $\mathcal{X}_0, \mathcal{X}_1$, we express $\widetilde{\alpha}_{\mathrm{B}}$ and $\widetilde{\alpha}_{\mathrm{B}}$ in terms of support functions.

Definition *The support function $h_{\mathcal{K}}(\cdot)$ for a closed convex set $\mathcal{K}$, defined as*

$$h_{\mathcal{K}}(\boldsymbol{y}) \coloneqq \sup_{\boldsymbol{x}\in\mathcal{K}} \langle \boldsymbol{y}, \boldsymbol{x} \rangle, \quad \boldsymbol{y} \in \mathbb{R}^n, \tag{2.18}$$

measures the distance from the origin to a supporting hyperplane of $\mathcal{K}$ with normal along $\boldsymbol{y}$. Note that, given convex set $\mathcal{K}$ and $\boldsymbol{T} \in \mathbb{R}^{n\times n}$, $\boldsymbol{\tau} \in \mathbb{R}^n$, the support function of the affine transformed set $\boldsymbol{T}\mathcal{K} + \boldsymbol{\tau}$ is

$$h_{\boldsymbol{T}\mathcal{K}+\boldsymbol{\tau}}(\boldsymbol{y}) = h_{\mathcal{K}}\left(\boldsymbol{T}^{\top}\boldsymbol{y}\right) + \langle \boldsymbol{\tau}, \boldsymbol{y} \rangle. \tag{2.19}$$

Using this definition, (2.7) and (2.16) can be expressed in terms of the support functions of $\mathcal{X}_0, \mathcal{X}_1$, as follows.

Theorem 2.2. *Consider compact convex $\mathcal{X}_0, \mathcal{X}_1$ with respective support functions $h_{\mathcal{X}_0}(\cdot), h_{\mathcal{X}_1}(\cdot)$. Let $\mathbb{S}^{n-1}$ denote the Euclidean unit sphere in $\mathbb{R}^n$. Then* (2.7) *can be* *expressed as*

$$\widetilde{\alpha}_B = \left\{ \max_{\boldsymbol{y}\in\mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}(\boldsymbol{y}) + h_{\mathcal{X}_1}(-\boldsymbol{y}) \right) \right\}^2, \tag{2.20a}$$

$$\widetilde{\beta}_B = \left\{ \min_{\boldsymbol{y}\in\mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}(\boldsymbol{y}) + h_{\mathcal{X}_1}(-\boldsymbol{y}) \right) \right\}^2. \tag{2.20b}$$

Furthermore, (2.16) *can be expressed as*

$$\widetilde{\alpha}_L = \left\{ \max_{\boldsymbol{y}\in\mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}\left(\boldsymbol{\Phi}_{t_1 t_0}^{\top} \boldsymbol{M}_{10}^{-1/2} \boldsymbol{y} \right) + h_{\mathcal{X}_1}\left(-\boldsymbol{M}_{10}^{-1/2} \boldsymbol{y} \right) \right) \right\}^2, \tag{2.21a}$$

$$\widetilde{\beta}_L = \left\{ \min_{\boldsymbol{y}\in\mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}\left(\boldsymbol{\Phi}_{t_1 t_0}^{\top} \boldsymbol{M}_{10}^{-1/2} \boldsymbol{y} \right) + h_{\mathcal{X}_1}\left(-\boldsymbol{M}_{10}^{-1/2} \boldsymbol{y} \right) \right) \right\}^2. \tag{2.21b}$$

Proof. Consider the set difference

$$\mathcal{X}_0 - \mathcal{X}_1 := \mathcal{X}_0 + (-\mathcal{X}_1) = \{ \boldsymbol{x}_0 - \boldsymbol{x}_1 \mid \boldsymbol{x}_0 \in \mathcal{X}_0, \boldsymbol{x}_1 \in \mathcal{X}_1 \}.$$

Let $\mathbb{B}^n := \mathrm{conv}\left(\mathbb{S}^{n-1}\right)$, the convex hull of $\mathbb{S}^{n-1}$, i.e., the n dimensional unit Euclidean ball.

From (2.7a), we have

$$\begin{aligned}
\widetilde{\alpha}_{\mathrm{B}} &= \left\{ \max_{\boldsymbol{x}_0\in\mathcal{X}_0, \boldsymbol{x}_1\in\mathcal{X}_1} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2 \right\}^2 \\
&= \left\{ \max_{\boldsymbol{x}\in\mathcal{X}_0-\mathcal{X}_1} \|\boldsymbol{x}\|_2 \right\}^2 \\
&= \left\{ \max_{\boldsymbol{x}\in\mathcal{X}_0-\mathcal{X}_1} \left\langle \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}, \boldsymbol{x} \right\rangle \right\}^2 \\
&= \left\{ \max_{\boldsymbol{x}\in\mathcal{X}_0-\mathcal{X}_1} \max_{\boldsymbol{y}\in\mathbb{S}^{n-1}} \langle \boldsymbol{y}, \boldsymbol{x} \rangle \right\}^2 \\
&= \left\{ \max_{\boldsymbol{y}\in\mathbb{S}^{n-1}} \max_{\boldsymbol{x}\in\mathcal{X}_0-\mathcal{X}_1} \langle \boldsymbol{y}, \boldsymbol{x} \rangle \right\}^2 \\
&= \left\{ \max_{\boldsymbol{y}\in\mathbb{S}^{n-1}} h_{\mathcal{X}_0-\mathcal{X}_1}(\boldsymbol{y}) \right\}^2
\end{aligned} \tag{2.22}$$

$$= \left\{ \max_{\boldsymbol{y} \in \mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}(\boldsymbol{y}) + h_{-\mathcal{X}_1}(\boldsymbol{y}) \right) \right\}^2, \tag{2.23}$$

where (2.22) follows from the definition (2.18), and (2.23) holds because the support function is distributive over Minkowski sum. Using (2.19), we get $h_{-\mathcal{X}_1}(\boldsymbol{y}) = h_{\mathcal{X}_1}(-\boldsymbol{y})$, and therefore (2.23) equals (2.20a).

Likewise, from (2.7b), we have

$$\begin{aligned}
\widetilde{\beta}_{\mathrm{B}} &= \left\{ \min_{\boldsymbol{x}_0 \in \mathcal{X}_0, \boldsymbol{x}_1 \in \mathcal{X}_1} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2 \right\}^2 \\
&= \left\{ -\max_{\boldsymbol{x} \in \mathcal{X}_0 - \mathcal{X}_1} \left(-\|\boldsymbol{x}\|_2 \right) \right\}^2 \\
&= \left\{ \max_{\boldsymbol{x} \in \mathcal{X}_0 - \mathcal{X}_1} \left\langle \frac{-\boldsymbol{x}}{\|\boldsymbol{x}\|_2}, \boldsymbol{x} \right\rangle \right\}^2 \\
&= \left\{ \max_{\boldsymbol{x} \in \mathcal{X}_0 - \mathcal{X}_1} \min_{\boldsymbol{y} \in \mathbb{S}^{n-1}} \langle \boldsymbol{y}, \boldsymbol{x} \rangle \right\}^2 \qquad (2.24) \\
&= \left\{ \max_{\boldsymbol{x} \in \mathcal{X}_0 - \mathcal{X}_1} \min_{\boldsymbol{y} \in \mathbb{B}^{n}} \langle \boldsymbol{y}, \boldsymbol{x} \rangle \right\}^2, \qquad (2.25)
\end{aligned}$$

where the last line is due to the linear objective which allows lossless convexification for the inner minimization in (2.24). This can be seen explicitly from the Cauchy-Schwarz inequality: $-\|\boldsymbol{y}\|_2 \ \|\boldsymbol{x}\|_2 \leq \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ where the equality is achieved when $\boldsymbol{y}$ is an unit vector pointing opposite to $\boldsymbol{x}$.

Since the sets $\mathcal{X}_0 - \mathcal{X}_1, \mathbb{B}^n$ are both compact convex, applying the Von Neumann minimax theorem [90, 91], we rewrite (2.25) as

$$\widetilde{\beta}_{\mathrm{B}} = \left\{ \min_{\boldsymbol{y} \in \mathbb{B}^n} \max_{\boldsymbol{x} \in \mathcal{X}_0 - \mathcal{X}_1} \langle \boldsymbol{y}, \boldsymbol{x} \rangle \right\}^2 = \left\{ \min_{\boldsymbol{y} \in \mathbb{B}^n} h_{\mathcal{X}_0 - \mathcal{X}_1}(\boldsymbol{y}) \right\}^2. \tag{2.26}$$

We next revert back the feasible set of the minimization in (2.26) to $\mathbb{S}^{n-1}$. To justify this, note that since the origin is within $\mathbb{B}^n$, the minimum in (2.26) cannot be positive. If this minimum value is zero, we can scale the $\arg\min$ to lie on the unit sphere. So it remains to consider the case when the minimum value

$h_{\mathcal{X}_0-\mathcal{X}_1}(\boldsymbol{y}^{\text{opt}}) < 0$, achieved by $\boldsymbol{y}^{\text{opt}}$ with $0 < \|\boldsymbol{y}^{\text{opt}}\| = \delta < 1$. In other words, $\boldsymbol{y}^{\text{opt}}$ is in the interior of $\mathbb{B}^n$. Now consider a vector $\widetilde{\boldsymbol{y}} := \boldsymbol{y}^{\text{opt}}/\delta \in \mathbb{S}^{n-1}$, which is feasible. Thanks to the positive homogeneity of the support function, we have $h_{\mathcal{X}_0-\mathcal{X}_1}(\widetilde{\boldsymbol{y}}) = \frac{1}{\delta} h_{\mathcal{X}_0-\mathcal{X}_1}(\boldsymbol{y}^{\text{opt}})$, which yields

$$h_{\mathcal{X}_0-\mathcal{X}_1}(\widetilde{\boldsymbol{y}}) = \underbrace{\frac{1}{\delta}}_{>0} \underbrace{h_{\mathcal{X}_0-\mathcal{X}_1}(\boldsymbol{y}^{\text{opt}})}_{<0} < h_{\mathcal{X}_0-\mathcal{X}_1}(\boldsymbol{y}^{\text{opt}}),$$

contradicting the supposition that $\boldsymbol{y}^{\text{opt}}$ is a minimizer. So the minimizer must lie on the boundary of the feasible set $\mathbb{B}^n$, i.e., on $\mathbb{S}^{n-1}$. Therefore, we can express (2.26) as

$$\widetilde{\beta}_{\text{B}} = \left\{\min_{\boldsymbol{y}\in\mathbb{S}^{n-1}} h_{\mathcal{X}_0-\mathcal{X}_1}(\boldsymbol{y})\right\}^2 = \left\{\min_{\boldsymbol{y}\in\mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}(\boldsymbol{y}) + h_{-\mathcal{X}_1}(\boldsymbol{y})\right)\right\}^2$$

which is indeed (2.20b) since $h_{-\mathcal{X}_1}(\boldsymbol{y}) = h_{\mathcal{X}_1}(-\boldsymbol{y})$.

To derive (2.21), we start by rewriting (2.16a) as

$$\begin{aligned}\widetilde{\alpha}_{\text{L}} &= \max_{\boldsymbol{x}_0\in \boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\mathcal{X}_0,\, \boldsymbol{x}_1\in\boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2 \\ &= \max_{\boldsymbol{x}\in \boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\mathcal{X}_0 - \boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1} \|\boldsymbol{x}\|_2^2,\end{aligned} \tag{2.27}$$

and (2.16b) as

$$\begin{aligned}\widetilde{\beta}_{\text{L}} &= \min_{\boldsymbol{x}_0\in \boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\mathcal{X}_0,\, \boldsymbol{x}_1\in\boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2 \\ &= \min_{\boldsymbol{x}\in \boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\mathcal{X}_0 - \boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1} \|\boldsymbol{x}\|_2^2.\end{aligned} \tag{2.28}$$

We then follow the same steps as before with the additional usage of the formula (2.19) relating the support functions of affine transformed sets in terms of the support functions of their pre-images. This completes the proof. □

The equalities (2.27) and (2.28) highlight that $\widetilde{\alpha}_{\text{L}}, \widetilde{\beta}_{\text{L}}$ are respectively the maximal and minimal separation between the linear transformed sets $\boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1 t_0}\mathcal{X}_0$ and $\boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1$. Specializing (2.27)-(2.28) for the classical SBP with $\boldsymbol{A}(t) \equiv \boldsymbol{0}$, $\boldsymbol{B}(t) \equiv \boldsymbol{I}_n$, and thus with $\boldsymbol{M}_{10} = \boldsymbol{\Phi}_{t_1 t_0} = \boldsymbol{I}_n$, recovers (2.7), which is the maximal and minimal separation between the original supports $\mathcal{X}_0, \mathcal{X}_1$, as expected.

2.7 Improved Computation via Preconditioning

Previous works such as [92] have explored the use of preconditioning to improve the performance of optimal transport algorithms. The preconditioning procedure described in [92] transforms the measures and corresponding support sets through a deterministic map such that the preconditioned measures are moved closer together, by creating new measures with the same (zero) mean and (diagonal) covariance matrix. The relation between solutions to the optimal transport problem, before and after preconditioning, can be determined according to the preconditioning. Such a strategy can be extended to the SBP because the SBP is an entropy-regularized optimal transport problem [15, 93].

We explore the application of such a preconditioning procedure for improved computation of γ_{L}. The following example illustrates the main idea.

Example 2.1: Consider an instance of the linear SBP (2.11) with time-invariant coefficients

$$\boldsymbol{A}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{B}(t) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \varepsilon = 0.5,$$

i.e., (2.10) is noisy double integrator

$$\mathrm{d}x = y\,\mathrm{d}t, \quad \mathrm{d}y = u\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}w.$$

In this case,

$$\boldsymbol{\Phi}_{t_1 t_0} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{M}_{10}^{-1} = \begin{bmatrix} 12 & -6 \\ -6 & 4 \end{bmatrix}.$$

We consider ellipsoidal supports

$$\mathcal{X}_i = \mathcal{E}_i\left(\boldsymbol{c}_i, \boldsymbol{S}_i\right) := \{\boldsymbol{x} \in \mathbb{R}^2 \mid (\boldsymbol{x} - \boldsymbol{c}_i)^\top \boldsymbol{S}_i^{-1}(\boldsymbol{x} - \boldsymbol{c}_i) \leq 1\}$$

for $i \in \{0, 1\}$, with respective center vectors

$$\boldsymbol{c}_0 := \boldsymbol{\Phi}_{t_1 t_0}^{-1} \boldsymbol{M}_{10}^{1/2} \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \quad \boldsymbol{c}_1 := \boldsymbol{M}_{10}^{1/2} \begin{pmatrix} 3 \\ 0 \end{pmatrix},$$

and respective positive definite shape matrices

$$\boldsymbol{S}_0 := \boldsymbol{\Phi}_{t_1 t_0}^{-1} \boldsymbol{M}_{10} \Phi_{10}^{-\top}, \quad \boldsymbol{S}_1 := \boldsymbol{M}_{10} = \begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix}.$$

Then

$$\boldsymbol{M}_{10}^{-1/2} \boldsymbol{\Phi}_{t_1 t_0} \mathcal{X}_0 = \{(x, y) \in \mathbb{R}^2 | x^2 + (y - 3)^2 \leq 1\}, \tag{2.29a}$$

$$\boldsymbol{M}_{10}^{-1/2} \mathcal{X}_1 = \{(x, y) \in \mathbb{R}^2 | (x - 3)^2 + y^2 \leq 1\}. \tag{2.29b}$$

Without the use of preconditioning, we determine γ_{L} from Theorem 2.1 by considering the maximum and minimum separation between the sets $\boldsymbol{M}_{10}^{-1/2} \boldsymbol{\Phi}_{t_1 t_0} \mathcal{X}_0$ and $\boldsymbol{M}_{10}^{-1/2} \mathcal{X}_1$, which in **Example 2.1**, are two disjoint circular disks such that $\widetilde{\alpha}_{\mathrm{L}} = 2 + 2\sqrt{3}$ and $\widetilde{\beta}_{\mathrm{L}} = -2 + 2\sqrt{3}$. Theorem 2.1 yields $\gamma_{\mathrm{L}} = \tanh^2(1) \approx 0.580$.

If the pushforwards $\left(\boldsymbol{M}_{10}^{-1/2} \boldsymbol{\Phi}_{t_1 t_0}\right)_\sharp \rho_0$, $\left(\boldsymbol{M}_{10}^{-1/2}\right)_\sharp \rho_1$, i.e.,

$$\frac{\sqrt{\det(\boldsymbol{M}_{10})}}{\det\left(\boldsymbol{\Phi}_{t_1 t_0}\right)} \rho_0(\boldsymbol{\Phi}_{t_1 t_0}^{-1} \boldsymbol{M}_{10}^{1/2}(\cdot)), \ \sqrt{\det(\boldsymbol{M}_{10})} \rho_1(\boldsymbol{M}_{10}^{1/2}(\cdot))$$

have identical, diagonal covariance matrices, then applying the preconditioning pro-

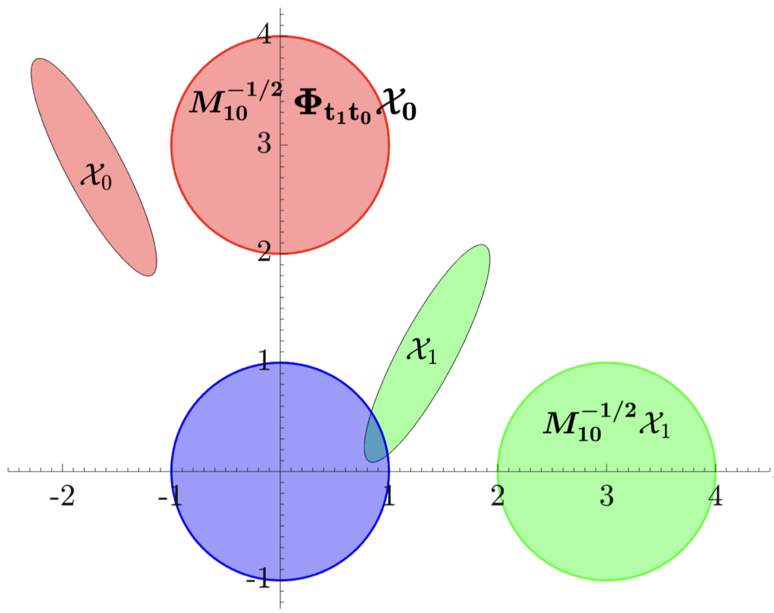

Figure 2.1: The elliptical support sets $\mathcal{X}_0, \mathcal{X}_1$ and circular disks $\boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1t_0}\mathcal{X}_0, \boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1$ in **Example 2.1**. The preconditioned supports coincide with the origin-centered unit circular disk (in *blue*).

cedure as in [92, Sec. 5] amounts to translating the means of the supports $\boldsymbol{M}_{10}^{-1/2}\boldsymbol{\Phi}_{t_1t_0}\mathcal{X}_0$ and $\boldsymbol{M}_{10}^{-1/2}\mathcal{X}_1$ to the origin. In **Example 2.1**, the preconditioned supports (2.29) are both origin-centered unit disks (Fig. 2.1). Consequently, $\widetilde{\alpha}_{\mathrm{L}}^{\text{precond}} = 2$, $\widetilde{\beta}_{\mathrm{L}}^{\text{precond}} = 0$, and thus $\gamma_{\mathrm{L}}^{\text{precond}} = \tanh^2(0.5) = 0.214$, which is an improvement on the original $\gamma_{\mathrm{L}} \approx 0.580$.

Theorem 2.1 can be applied in two distinct fashions. First, as evidenced by **Example 2.1**, Theorem 2.1 can allow for improved ease of calculation of $\gamma_{\mathrm{L}}^{\text{precond}}$, via preconditioning procedures to transform the supports into more geometrically convenient forms. More significantly, Theorem 2.1 can demonstrate the effectiveness of proposed preconditioning procedures in reducing the upper bound of the contraction coefficient, γ_{L}. This suggests a potential application of Theorem 2.1 in which we leverage the results of this chapter to optimally construct a preconditioning procedure for a given SBP.

3 | Classical SBP with Quadratic State Cost

3.1 Overview

This chapter concerns a second generalization of the SBP, given by

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \quad \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\frac{1}{2}\|\boldsymbol{u}\|_2^2+\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x}\right)\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{3.1a}$$

$$\frac{\partial\rho}{\partial t}+\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})=\Delta_{\boldsymbol{x}}\rho, \tag{3.1b}$$

$$\boldsymbol{x}(t=t_0)\sim\rho_0\,\text{(given)},\quad \boldsymbol{x}(t=t_1)\sim\rho_1\,\text{(given)}. \tag{3.1c}$$

This variant (3.1) will henceforth be referred to as the classical SBP with quadratic state cost.

We note that (3.1b) corresponds to the Itô diffusion process

$$\mathrm{d}\boldsymbol{x}=\boldsymbol{u}\,(t,\boldsymbol{x})\;\mathrm{d}t+\sqrt{2}\,\mathrm{d}\boldsymbol{w}. \tag{3.2}$$

Comparing (3.1) to the classical SBP introduced in Chapter 1 as

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf}\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\frac{1}{2}\|\boldsymbol{u}\|_2^2\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t,$$
$$\frac{\partial\rho}{\partial t}+\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})=\varepsilon\Delta_{\boldsymbol{x}}\rho,\quad \varepsilon>0,$$
$$\rho(t_0,\boldsymbol{x})=\rho_0(\boldsymbol{x}),\quad \rho(t_1,\boldsymbol{x})=\rho_1(\boldsymbol{x}),$$

we note that (3.1) both includes an additional term in the objective function (hence, we refer to (3.1) as classical SBP *with quadratic state cost*) and, for convenience, fixes $\varepsilon=1$.

Recall that the generalized SBP first defined in Chapter 1 took the form

$$\begin{aligned}
&\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n} \left(q(\boldsymbol{x}) + r(\boldsymbol{u})\right)\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t\\
&\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot\left(\rho\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u})\right) = \Delta_{\boldsymbol{\Sigma}(t,\boldsymbol{x},\boldsymbol{u})}\rho\\
&\rho(t_0,\boldsymbol{x}) = \rho_0(\boldsymbol{x}),\quad \rho(t_1,\boldsymbol{x}) = \rho_1(\boldsymbol{x}).
\end{aligned}$$

In the above, setting $q(\boldsymbol{x}) \equiv \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x}$, with fixed $\boldsymbol{Q} \succeq \boldsymbol{0}$, $r(\boldsymbol{u}) \equiv \frac{1}{2}\|\boldsymbol{u}\|_2^2$, $\boldsymbol{\Sigma} \equiv \boldsymbol{I}$, and $\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{u}$, we recover (3.1).

The quadratic state cost regularization may be viewed as a mechanism to bias the trajectory toward some desired or nominal constant (w.l.o.g. zero) vector for all $t \in [t_0, t_1]$ in addition to fulfilling the endpoint statistics constraints. Fig. 3.1 highlights this use of quadratic state cost; notice that the sample paths generated by the classical SBP (blue) are much more prone to deviations from 0 than the sample paths generated by the classical SBP with quadratic state cost (red). From a control-theoretic viewpoint, therefore, the proposed quadratic regularization is natural since penalizing state excursions away from nominal can be a soft way to promote safety.

This formulation is in similar spirit as classical linear quadratic optimal control [94] where the stage cost indeed comprises of two summands: a positive semi-definite state cost-to-go (here $\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x}$) and a positive definite control cost-to-go (here $\frac{1}{2}\|\boldsymbol{u}\|_2^2$). The relative modeling importance of one summand over another is determined by $\boldsymbol{Q} \succeq 0$.

Unlike the conventional Schrödinger bridge, the additive state-cost induces a state-dependent rate of killing and creation of probability mass. Determining the solution of the SBP with quadratic state cost therefore requires determining the Markov kernel of a reaction-diffusion partial differential equation.

Specifically, solving SBP with quadratic state cost requires solving the following

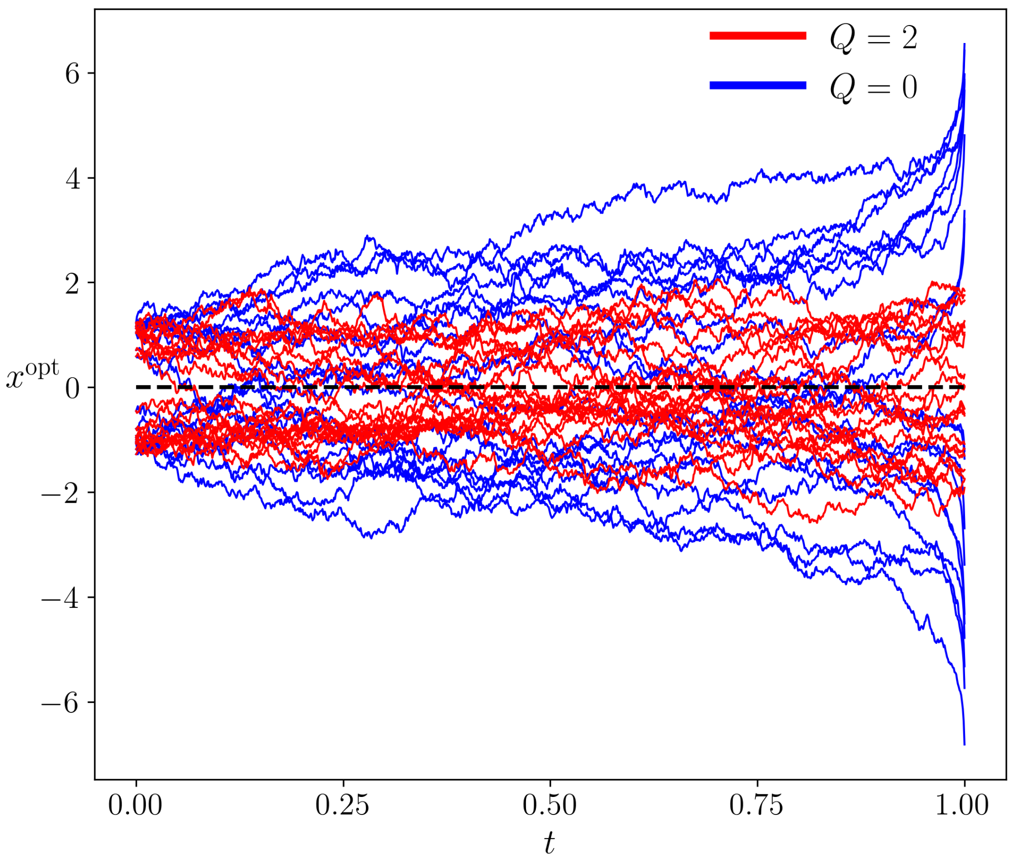

Figure 3.1: Sample paths for 1D SBP with quadratic state cost $\frac{1}{2}Qx^2$ for fixed $Q \geq 0$, time horizon $[t_0, t_1] = [0, 1]$, and endpoint PDFs $\rho_0 = \sum_{\mu_0 \in \{-1,1\}} \frac{1}{2}\mathcal{N}(\mu_0, 0.05^2)$, $\rho_1 = \mathcal{N}(0, 0.5^2)$.

PDEs:

$$\frac{\partial \widehat{\varphi}}{\partial t} = \left(\Delta_{\boldsymbol{x}} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}\right)\widehat{\varphi}, \tag{3.5a}$$

$$\frac{\partial \varphi}{\partial t} = \left(-\Delta_{\boldsymbol{x}} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}\right)\varphi. \tag{3.5b}$$

Fortunately, it suffices to derive the Markov kernel for the reaction-diffusion PDE (3.5a), since the same result may be applied to determine a solution to the PDE (3.5b) under the change-of-variable $t \mapsto \bar{t} := t_1 - (t - t_0)$. Without loss of generality, the developments in the remainder of this chapter are therefore restricted to the derivation of the kernel to (3.5a).

Deriving this Markov kernel in closed form demonstrates that the SBP with quadratic state cost is exactly solvable, *even for non-Gaussian endpoints.* In other words, finding the Markov kernel k in closed form enables the numerical solution of the SBP with quadratic state cost for *generic non-Gaussian endpoint distributions μ_0, μ_1 with finite second moments.* The solution of SBP with quadratic state cost in the special case of Gaussian μ_0, μ_1 has appeared before in [42, 95]. The closed-form solution of the SBP without state cost reported in [96] was also for Gaussian μ_0, μ_1.

Unlike these prior works, the closed form Markov kernel derived here makes no parametric assumptions on μ_0, μ_1, and therefore represents an advancement in the state-of-the-art. The idea of exactly solving the SBP by finding associated Green's function has been independently pursued in the recent work [97].

3.2 Deriving k_+ Via Hermite Polynomials

In [74], we derive the Markov kernel in closed form for the SBP with quadratic state cost using Hermite polynomials. In this section, I present our results. I begin with our deduction of this Markov kernel where $\boldsymbol{Q} \succ \boldsymbol{0}$ (Theorem 3.1), denoted as k_{++}. I then generalize our results for $\boldsymbol{Q} \succeq \boldsymbol{0}$ (Theorem 3.3) to derive the corresponding kernel k_+.

This derivation uses the *physicist's Hermite polynomials* [98, Ch. 22] of degree $n \in \mathbb{N}_0$, given by

$$H_n(x) := (-1)^n e^{x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(e^{-x^2}\right), \tag{3.6}$$

which are orthogonal in the sense that

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2}\mathrm{d}x = \begin{cases} 0 & \text{if} \quad m \neq n, \\ \sqrt{\pi}2^n n! & \text{otherwise.} \end{cases} \tag{3.7}$$

The closing section of this chapter will explore an alternate derivation of k_{++} and k_+.

The following Lemma will be useful in our development.

Lemma 3.1 (A series sum for exponential of negative quadratic). *For α, β, t_0, x fixed, and $0 \leq t_0 < t < \infty$,*

$$\sum_{j=0}^{\infty} \frac{(\alpha\,(t-t_0))^j}{j!\beta^j} \frac{\mathrm{d}^j}{\mathrm{d}x^j} e^{-(\beta x)^2} = e^{-(\beta x + \alpha(t-t_0))^2}. \tag{3.8}$$

For a proof of Lemma 3.1, see Appendix C.1.

We will use the following result, which is sometimes referred to as the "central identity" of quantum field theory (QFT) [99, Appendix A].

Lemma 3.2 (Central identity of QFT). *[100, p. 2] [99, p. 15] For a suitably smooth function $f : \mathbb{R}^n \mapsto \mathbb{R}$, and matrix $\boldsymbol{A} \in \mathrm{Sym}_{++}(n)$,*

$$\int_{\mathbb{R}^n} \exp\!\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}\right) f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \sqrt{\frac{(2\pi)^n}{\det(\boldsymbol{A})}} \exp\left(\frac{1}{2}\nabla_{\boldsymbol{x}}^\top \boldsymbol{A}^{-1}\nabla_{\boldsymbol{x}}\right) f(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\boldsymbol{0}}, \tag{3.9}$$

where the exponential of a differential operator is understood as a power series. In particular, for $f(\boldsymbol{x}) = \exp\langle \boldsymbol{b}, \boldsymbol{x}\rangle$, $\boldsymbol{b} \in \mathbb{R}^n$, we have $\int_{\mathbb{R}^n} \exp\!\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \langle \boldsymbol{b}, \boldsymbol{x}\rangle\right)\mathrm{d}\boldsymbol{x} = \sqrt{\frac{(2\pi)^n}{\det(\boldsymbol{A})}} \exp\left(\frac{1}{2}\boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}\right)$.

For diagonalizable $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ with $\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$, and function $g(\cdot)$ well-defined on the spectrum of $\boldsymbol{X}$, the matrix function

$$g\,(\boldsymbol{X}) = \boldsymbol{V} g(\boldsymbol{\Lambda})\boldsymbol{V}^{-1}, \tag{3.10}$$

see e.g., [101, Ch. 5], [102, Ch. 1].

3.2.1 Kernel for $\boldsymbol{Q} \succ 0$

For $\boldsymbol{Q} \succ \boldsymbol{0}$, recall the eigen-decomposition $\frac{1}{2}\boldsymbol{Q} = \boldsymbol{V}^\top \boldsymbol{\Lambda} \boldsymbol{V}$ where the positive diagonal matrix $\boldsymbol{\Lambda}$ has the eigenvalues of $\frac{1}{2}\boldsymbol{Q}$ along its main diagonal, and the columns of the orthogonal matrix $\boldsymbol{V}$ are the eigenvectors of $\frac{1}{2}\boldsymbol{Q}$. For convenience, let $\lambda_i > 0$ be the ith diagonal entry of $\boldsymbol{\Lambda}$, let $\boldsymbol{y} := \boldsymbol{V}\boldsymbol{x}$, and

$$\widehat{\eta}(t, \boldsymbol{y}) := \widehat{\varphi}\left(t, \boldsymbol{x} = \boldsymbol{V}^\top \boldsymbol{y}\right). \tag{3.11}$$

The change-of-variable $\boldsymbol{x} \mapsto \boldsymbol{y}$ gives $\Delta_{\boldsymbol{x}}\widehat{\varphi}(t, \boldsymbol{x}) = \Delta_{\boldsymbol{y}}\widehat{\eta}(t, \boldsymbol{y})$, and (3.5a) becomes

$$\begin{aligned} \frac{\partial \widehat{\eta}}{\partial t} &= \Delta_{\boldsymbol{y}}\widehat{\eta} - \left(\boldsymbol{y}^\top \boldsymbol{\Lambda} \boldsymbol{y}\right)\widehat{\eta} \\ &= \sum_{i=1}^{n} \left(\frac{\partial^2}{\partial y_i^2} - \lambda_i y_i^2\right)\widehat{\eta}. \end{aligned} \tag{3.12}$$

We seek a separation-of-variables solution ansatz $\widehat{\eta}(t, \boldsymbol{y}) = T(t)\prod_{i=1}^{n} Y_i(y_i)$ for (3.12), which requires us to satisfy $\frac{1}{T}\frac{\mathrm{d}T}{\mathrm{d}t} = \sum_{i=1}^{n}\left(\frac{1}{Y_i}\frac{\mathrm{d}^2 Y_i}{\mathrm{d}y_i^2} - \lambda_i y_i^2\right) = c$ for some constant $c \in \mathbb{R}$. Equivalently, we need to satisfy the $n+1$ ODEs

$$\begin{aligned} &\frac{\mathrm{d}T}{\mathrm{d}t} = cT, \\ &\frac{\mathrm{d}^2 Y_1}{\mathrm{d}y_1^2} - \left(\lambda_1 y_1^2 + c_1\right) Y_1 = 0, \\ &\quad \vdots \\ &\frac{\mathrm{d}^2 Y_n}{\mathrm{d}y_n^2} - \left(\lambda_n y_n^2 + c_n\right) Y_n = 0, \end{aligned}$$

for constants $c, c_1, \cdots, c_n \in \mathbb{R}$ subject to the constraint $c_1 + \ldots + c_n = c$.

The time ODE has solution $T(t) = k e^{c(t-t_0)}$ for some constant $k \in \mathbb{R}$. To determine each $Y_k(y_k)$, we use the following Lemma, proved in Appendix C.2.

Lemma 3.3 (Solution of a parametric second order nonlinear ODE). *For pa-*

rameters $c \in \mathbb{R}$ and $\lambda > 0$, the solutions to the parametric ODE

$$\frac{\mathrm{d}^2 Y}{\mathrm{d}y^2} - (\lambda y^2 + c)Y = 0, \tag{3.13}$$

are given by

$$Y = a \exp\left(-\frac{\sqrt{\lambda} y^2}{2}\right) H_n\left(\lambda^{1/4} y\right), \tag{3.14}$$

for any $n = -\frac{c}{2\sqrt{\lambda}} - \frac{1}{2} \in \mathbb{N}_0$, $a \in \mathbb{R}$.

Applying (3.14) to the ODEs for $Y_1, Y_2, \cdots, Y_n$, we find that the general solution to (3.12) is of the form

$$\begin{aligned}
&\widehat{\eta}(t, \boldsymbol{y}) \\
&= \sum_{m_n=0}^{\infty} \cdots \sum_{m_1=0}^{\infty} a_{\boldsymbol{m}} \prod_{i=1}^{n} \exp\left\{-\left(t - t_0\right)\sqrt{\lambda_i}\left(2m_i + 1\right) - \frac{y_i^2 \sqrt{\lambda_i}}{2}\right\} H_{m_i}\left(\lambda_i^{1/4} y_i\right),
\end{aligned} \tag{3.15}$$

where $\boldsymbol{m} := (m_1, m_2, \ldots, m_n) \in \mathbb{N}_0^n$, and the coefficients $a_{\boldsymbol{m}}$ remain to be determined. Using (3.15) together with the orthogonality of Hermite polynomials, we express the IVP solution to (3.12) with known initial condition $\widehat{\eta}_0(\boldsymbol{y}) := \widehat{\eta}(t_0, \boldsymbol{y}) = \widehat{\varphi}_0\left(\boldsymbol{V}^\top \boldsymbol{y}\right)$ as follows.

Proposition 3.1 (IVP solution for (3.12) **with positive diagonal $\boldsymbol{\Lambda}$ as series sum).** *For $0 \leq t_0 < t < \infty$ and initial condition $\widehat{\eta}_0(\boldsymbol{y}) := \widehat{\eta}(t_0, \boldsymbol{y})$, the unique solution to* (3.12) *with positive diagonal matrix $\boldsymbol{\Lambda}$ is given by*

$$\widehat{\eta}(t, \boldsymbol{y}) = \sum_{m_n=0}^{\infty} \cdots \sum_{m_1=0}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (*)\, \widehat{\eta}_0(\boldsymbol{z})\, \mathrm{d}z_1 \ldots \mathrm{d}z_n \tag{3.16}$$

where

$$\begin{aligned}
(*) := \prod_{i=1}^{n} \frac{\lambda_i^{1/4}}{\sqrt{\pi} 2^{m_i} (m_i!)} \exp\left\{-\left(t - t_0\right)\sqrt{\lambda_i}(2m_i + 1) - \frac{(y_i^2 + z_i^2)\sqrt{\lambda_i}}{2}\right\} \\
\times H_{m_i}\left(\lambda_i^{1/4} y_i\right) H_{m_i}\left(\lambda_i^{1/4} z_i\right).
\end{aligned} \tag{3.17}$$

See Appendix C.3 for a proof of Proposition 3.1.

We then resolve the series sum in (3.16)-(3.17) to derive the *closed form* Markov kernel $k \equiv k_{++}$ for (3.12). The calculations of this derivation can be found in Appendix C.4. Note that functions such as the hyperbolic sinusoidal function are understood to be applied elementwise for matricial arguments.

Theorem 3.1 (Closed form kernel k_{++} for IVP solution of (3.12) with positive diagonal $\boldsymbol{\Lambda}$). *Consider the same setting as in Proposition 3.1. The solution* (3.16) *simplifies to*

$$\widehat{\eta}(t,\boldsymbol{y}) = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} k_{++}(t_0,\boldsymbol{y},t,\boldsymbol{z})\,\widehat{\eta}_0(\boldsymbol{z})\,\mathrm{d}z_1\dots\mathrm{d}z_n \tag{3.18}$$

where the kernel

$$\begin{aligned} &k_{++}(t_0,\boldsymbol{y},t,\boldsymbol{z})\\ &= \frac{(\det(\boldsymbol{M}_{tt_0}))^{1/4}}{(2\pi)^{n/2}\sqrt{\det(\sinh(2\,(t-t_0)\sqrt{\boldsymbol{\Lambda}}))}} \times \exp\left(-\frac{1}{2}\begin{pmatrix}\boldsymbol{y}^\top & \boldsymbol{z}^\top\end{pmatrix}\boldsymbol{M}_{tt_0}\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix}\right), \end{aligned} \tag{3.19}$$

and the matrix

$$\boldsymbol{M}_{tt_0} = \begin{bmatrix}\sqrt{\boldsymbol{\Lambda}}\coth\left(2\,(t-t_0)\sqrt{\boldsymbol{\Lambda}}\right) & -\sqrt{\boldsymbol{\Lambda}}\,\mathrm{csch}\left(2\,(t-t_0)\sqrt{\boldsymbol{\Lambda}}\right)\\ -\sqrt{\boldsymbol{\Lambda}}\,\mathrm{csch}\left(2\,(t-t_0)\sqrt{\boldsymbol{\Lambda}}\right) & \sqrt{\boldsymbol{\Lambda}}\coth\left(2\,(t-t_0)\sqrt{\boldsymbol{\Lambda}}\right)\end{bmatrix}. \tag{3.20}$$

We note that unlike the heat kernel k_{B}, defined in (1.23), k_{++} is not radial in the sense that (3.19) is not a function of $\|\boldsymbol{y}-\boldsymbol{z}\|_2$ alone.

The next Proposition (proof in Appendix C.5) states that the matrix $\boldsymbol{M}_{tt_0}$ in (3.20) can be seen as positive diagonal scaling of a symplectic positive definite matrix, and is therefore positive definite. This will be useful in establishing properties of $\widehat{\varphi}(t,\boldsymbol{x})$

in Proposition 3.4. Note that, for any $n \in \mathbb{N}$, the set of *symplectic matrices*

$$\mathrm{Sp}\left(2n,\mathbb{R}\right) := \left\{ \boldsymbol{S} \in \mathbb{R}^{2n\times 2n} \mid \boldsymbol{S}^\top \boldsymbol{J}\boldsymbol{S} = \boldsymbol{J} := \begin{pmatrix} 0 & \boldsymbol{I} \\ -\boldsymbol{I} & 0 \end{pmatrix} \right\} \tag{3.21}$$

is a connected Lie group that is a subgroup of $\mathrm{SL}(2n,\mathbb{R})$, the set of volume and orientation preserving linear maps in $\mathbb{R}^{2n}$. We denote the set of $n\times n$ symmetric positive definite matrices as $\mathrm{Sym}_{++}(n)$, i.e., $\boldsymbol{A} \in \mathrm{Sym}_{++}(n)$ means $\boldsymbol{A} \succ \boldsymbol{0}$. Let $\mathrm{Sp}_{++}\left(2n,\mathbb{R}\right) := \mathrm{Sp}\left(2n,\mathbb{R}\right) \cap \mathrm{Sym}_{++}(2n)$.

Proposition 3.2 (Properties of matrix $\boldsymbol{M}_{tt_0}$). *For $0 \le t_0 < t < \infty$, and given positive diagonal matrix $\boldsymbol{\Lambda}$, the matrix $\boldsymbol{M}_{tt_0}$ in* (3.20) *satisfies*

(i) $\boldsymbol{M}_{tt_0} = \begin{bmatrix} \boldsymbol{\Lambda}^{1/4} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}^{1/4} \end{bmatrix} \boldsymbol{M}^{(1)}_{tt_0}\boldsymbol{M}^{(2)}_{tt_0} \begin{bmatrix} \boldsymbol{\Lambda}^{1/4} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}^{1/4} \end{bmatrix}$ *where*

$$\boldsymbol{M}^{(1)}_{tt_0} = \begin{bmatrix} \cosh(2(t-t_0)\sqrt{\boldsymbol{\Lambda}}) & -\boldsymbol{I} \\ -\boldsymbol{I} & \cosh(2(t-t_0)\sqrt{\boldsymbol{\Lambda}}) \end{bmatrix},$$

$$\boldsymbol{M}^{(2)}_{tt_0} = \begin{bmatrix} \mathrm{csch}(2(t-t_0)\sqrt{\boldsymbol{\Lambda}}) & \boldsymbol{0} \\ \boldsymbol{0} & \mathrm{csch}(2(t-t_0)\sqrt{\boldsymbol{\Lambda}}) \end{bmatrix},$$

(ii) $\boldsymbol{M}^{(1)}_{tt_0}\boldsymbol{M}^{(2)}_{tt_0} \in \mathrm{Sp}_{++}\left(2n,\mathbb{R}\right)$,

(iii) $\boldsymbol{M}_{tt_0} \in \mathrm{Sym}_{++}\left(2n,\mathbb{R}\right)$.

Since $\boldsymbol{M}_{tt_0} \succ \boldsymbol{0}$, the exponential term in (3.19) is of the form $\exp\left(-\frac{1}{2}\mathrm{dist}^2_{tt_0}(\boldsymbol{y},\boldsymbol{z})\right)$ where $\mathrm{dist}_{tt_0}(\boldsymbol{y},\boldsymbol{z})$ is a distance metric. Our next result provides geometric insight that $\mathrm{dist}^2_{tt_0}(\boldsymbol{y},\boldsymbol{z})$ is, in fact, the minimal value of the action integral induced by the Lagrangian $L : \mathcal{T}\mathbb{R}^n \mapsto \mathbb{R}$ given by[1]

$$L(\boldsymbol{q},\boldsymbol{u}) := \frac{1}{2}\|\boldsymbol{u}\|_2^2 + \boldsymbol{q}^\top\left(2\boldsymbol{\Lambda}\right)\boldsymbol{q}, \quad (\boldsymbol{q},\boldsymbol{u}) \in \mathcal{T}\mathbb{R}^n, \tag{3.22}$$

[1]Recall that the diagonal matrix associated with the eigen-decomposition of $\boldsymbol{Q}$ is $2\boldsymbol{\Lambda}$.

where $\mathcal{T}$ denotes the tangent bundle.

Proposition 3.3 (Induced distance in k_{++}). *Given $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$, the minimum value of the action integral associated with the Lagrangian* (3.22)*:*

$$dist^2_{tt_0}(\boldsymbol{y}, \boldsymbol{z}) := \inf_{\gamma(\cdot)\in\left\{\mathcal{C}^2([t_0,t]) \mid \gamma(t_0)=\boldsymbol{y}, \gamma(t)=\boldsymbol{z}\right\}} \int_{t_0}^{t} L\left(\gamma(\tau), \dot{\gamma}(\tau)\right) \mathrm{d}\tau \tag{3.23}$$

equals to $\begin{pmatrix}\boldsymbol{y}^\top & \boldsymbol{z}^\top\end{pmatrix} \boldsymbol{M}_{tt_0} \begin{pmatrix}\boldsymbol{y} \\ \boldsymbol{z}\end{pmatrix}$*, where $\boldsymbol{M}_{tt_0}$ is given by* (3.20)*.*

For the proof of Proposition 3.3, see Appendix C.6.

As a point of comparison, recall that the scaled heat kernel in $\mathbb{R}^n$ associated with the diffusion process $\mathrm{d}\boldsymbol{x} = \sqrt{2}\mathrm{d}\boldsymbol{w}$ is denoted as k_0, i.e.,

$$k_0(t_0, \boldsymbol{y}, t, \boldsymbol{z}) = (4\pi(t - t_0))^{-n/2} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{z}\|_2^2}{4(t - t_0)}\right), \tag{3.24}$$

where $0 \le s < t < \infty, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Note that this is a specialized version of k_{B} in which $\varepsilon = 1$.

Additionally, note that the linear kernel associated with the diffusion process $\mathrm{d}\boldsymbol{x} = \boldsymbol{A}(t)\boldsymbol{x}(t)\mathrm{d}t + \sqrt{2}\boldsymbol{B}(t)\mathrm{d}\boldsymbol{w}$, denoted as k_{L}, is given by

$$\begin{aligned} k_{\mathrm{L}}\left(t_0, \boldsymbol{y}, t, \boldsymbol{z}\right) = (4\pi(t - t_0))^{-n/2} \det\left(\boldsymbol{\Gamma}_{tt_0}\right)^{-1/2} \\ \times \exp\left(-\frac{\left(\boldsymbol{\Phi}_{tt_0}\boldsymbol{y} - \boldsymbol{z}\right)^\top \boldsymbol{\Gamma}_{tt_0}^{-1} \left(\boldsymbol{\Phi}_{tt_0}\boldsymbol{y} - \boldsymbol{z}\right)}{4(t - t_0)}\right) \end{aligned} \tag{3.25}$$

where the controllability Gramian

$$\boldsymbol{\Gamma}_{t_1 t_0} := \int_{t_0}^{t_1} \boldsymbol{\Phi}_{t_1\tau} \boldsymbol{B}(\tau) \boldsymbol{B}^\top(\tau) \boldsymbol{\Phi}_{t_1\tau}^\top \mathrm{d}\tau \succ \boldsymbol{0}.$$

Remark 3.1. *Unlike the Markov kernels* (3.24) *and* (3.25)*, the kernel k_{++} in* (3.19) *is not a transition probability density because of the reaction term contributing to state-dependent creation or killing of mass. This can be seen by applying the Central Identity* *of QFT* (3.9) *to evaluate*

$$\int_{\mathbb{R}^{2n}} k_{++}(t_0,\boldsymbol{y},t,\boldsymbol{z})\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z} = \det\left(\sqrt{\boldsymbol{\Lambda}}\sinh\left(2(t-t_0)\sqrt{\boldsymbol{\Lambda}}\right)\right)^{-1/2} \neq 1$$

.

3.2.2 Known Kernels as Special Cases of k_{++}

In the special case $\frac{1}{2}\boldsymbol{Q}$, and hence $\boldsymbol{\Lambda}$, equals $\boldsymbol{I}$, direct substitution reduces (3.19) to the *multivariate Mehler kernel* [103], [44, Thm. 1]

$$k_{\text{Mehler}}(t_0,\boldsymbol{y},t,\boldsymbol{z}) = \frac{1}{(2\pi\sinh(2(t-t_0)))^{n/2}} \times\exp\left(\operatorname{csch}(2(t-t_0))\langle\boldsymbol{y},\boldsymbol{z}\rangle - \frac{1}{2}\coth\left(2(t-t_0)\right)\left(\|\boldsymbol{y}\|_2^2+\|\boldsymbol{z}\|_2^2\right)\right),$$

which is the propagator or Green's function for the *isotropic quantum harmonic oscillator*. Well-known extensions of the Mehler kernel include the *Kibble-Slepian formula* [104, 105] and variants thereof [106–108].

Another special case of interest for k_{++} is the limit $\boldsymbol{\Lambda}\downarrow\boldsymbol{0}$ understood as $\lambda_i\downarrow 0\ \forall i=1,\cdots,n$. In this limit, the kernel (3.19) recovers the heat kernel (3.24), as stated in the following theorem.

Theorem 3.2 (Heat kernel as limiting case of k_{++}). *Consider the kernel k_{++} in* (3.19)-(3.20) *and the heat kernel k_0 in* (3.24). *Then,*

$$\lim_{(\lambda_1,\cdots,\lambda_n)\downarrow\boldsymbol{0}} k_{++}(t_0,\boldsymbol{y},t,\boldsymbol{z}) = k_0(t_0,\boldsymbol{y},t,\boldsymbol{z}).$$

Proof. Notice that the kernel (3.19) is a separable product in tuple $(t-t_0,d_i,y_i,z_i)$, cf. (C.10). Letting $\delta_i := \sqrt{d_i}\ \forall i=1,\cdots,n$, we then have

$$\lim_{(d_1,\cdots,d_n)\downarrow\boldsymbol{0}} \kappa_{++}(t_0,\boldsymbol{y},t,\boldsymbol{z})$$

$$= \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n} \sqrt{\ell_1} \exp\left(-\frac{1}{2}\begin{pmatrix} y_i & z_i \end{pmatrix}\begin{bmatrix} \ell_2 & -\ell_3 \\ -\ell_3 & \ell_2 \end{bmatrix}\begin{pmatrix} y_i \\ z_i \end{pmatrix}\right), \tag{3.26}$$

where

$$\ell_1 := \lim_{\delta_i \downarrow 0} \frac{\delta_i}{\sinh(2\,(t-t_0)\,\delta_i)},$$

$$\ell_2 := \lim_{\delta_i \downarrow 0} \delta_i \coth(2(t-t_0)\delta_i),$$

$$\ell_3 := \lim_{\delta_i \downarrow 0} \delta_i \operatorname{csch}(2(t-t_0)\delta_i).$$

We evaluate the limits ℓ_1, ℓ_2, ℓ_3 by applying L'Hôpital's rule: $\ell_1 = \ell_2 = \ell_3 = 1/(2(t - t_0))$. Substituting these back in (3.26) yields

$$\lim_{(d_1,\cdots,d_n)\downarrow \mathbf{0}} \kappa_{++}(t_0, \boldsymbol{y}, t, \boldsymbol{z}) = (4\pi(t-t_0))^{-n/2} \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{z}\|^2}{4(t-t_0)}\right) = \kappa_0(t_0, \boldsymbol{y}, t, \boldsymbol{z}),$$

as claimed. □

That k_0 arises as a limiting case of k_{++} is consistent with the principle of least action in Proposition 3.3 since in the limit $\boldsymbol{\Lambda} \downarrow \mathbf{0}$, the Lagrangian in (3.22) equals $\frac{1}{2}\|\boldsymbol{u}\|_2^2$. Then, (3.23) yields the squared Euclidean distance $\|\boldsymbol{y} - \boldsymbol{z}\|_2^2$, which is what appears inside the exponential in (3.24) up to the usual scaling.

3.2.3 Kernel for $\boldsymbol{Q} \succeq 0$

In Theorem 3.3 that follows, we generalize the closed form Markov kernel for the case $\boldsymbol{Q} \succeq \mathbf{0}$. We denote this generalized kernel as k_+. To account for the fact that zero eigenvalues of $\frac{1}{2}\boldsymbol{Q}$ can be interlaced with its positive eigenvalues, we re-index the spatial variables. In the following theorem, the subscripted index vector notation $\boldsymbol{x}_{[i_1:i_r]}$ stands for the appropriate length r sub-vector of the vector $\boldsymbol{x}$. Let

$\mathbb{N} := \{1, 2, \cdots\}$, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, and the n-fold tensor product $\mathbb{N}_0^n := \underbrace{\mathbb{N}_0 \times \mathbb{N}_0 \times \cdots \mathbb{N}_0}_{n \text{ times}}$.

Theorem 3.3 (Closed form kernel k_+ for IVP solution of (3.12) with nonnegative diagonal $\boldsymbol{\Lambda}$). *Consider diagonal matrix $\boldsymbol{\Lambda}$ having $p \in \mathbb{N}_0$ zero diagonal entries where $p < n$, and $n - p$ positive diagonal entries. Let $i_1, i_2, \ldots, i_{n-p}$ be the indices corresponding to positive diagonal entries of $\boldsymbol{\Lambda}$. Likewise, let $i_{n-p+1}, i_{n-p+2}, \ldots, i_n$ be the indices corresponding to zero diagonal entries of $\boldsymbol{\Lambda}$.*

For $0 \leq t_0 < t < \infty$ and initial condition $\widehat{\eta}_0(\boldsymbol{y}) := \widehat{\eta}(t_0, \boldsymbol{y})$, the unique solution to (3.12) *with such diagonal $\boldsymbol{\Lambda}$, is*

$$\widehat{\eta}(t, \boldsymbol{y}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} k_+(t_0, \boldsymbol{y}, t, \boldsymbol{z})\, \widehat{\eta}_0(\boldsymbol{z})\, \mathrm{d}z_1 \ldots \mathrm{d}z_n \tag{3.27}$$

where the kernel

$$\begin{aligned} k_+(t_0, \boldsymbol{y}, t, \boldsymbol{z}) = & k_{++}\left(t_0, \boldsymbol{y}_{[i_1:i_{n-p}]}, t, \boldsymbol{z}_{[i_1:i_{n-p}]}\right) \\ & \times k_0\left(t_0, \boldsymbol{y}_{[i_{n-p+1}:i_n]}, t, \boldsymbol{z}_{[i_{n-p+1}:i_n]}\right), \end{aligned} \tag{3.28}$$

and k_{++}, k_0 are as in Theorem 3.1 and (3.24)*, respectively.*

For a proof of 3.3, see Appendix C.7.

Remark 3.2. *For a given $\boldsymbol{Q} \geq \boldsymbol{0}$, the solution* (3.18) *or* (3.27)*, can be brought back to the original $\boldsymbol{x}$ coordinates to determine the Schrödinger factor*

$$\widehat{\varphi}(t, \boldsymbol{x}) = \widehat{\eta}(t, \boldsymbol{y} = \boldsymbol{V}\boldsymbol{x}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} k_+(t_0, \boldsymbol{V}\boldsymbol{x}, t, \boldsymbol{z})\, \widehat{\varphi}_0\left(\boldsymbol{V}^\top \boldsymbol{z}\right) \mathrm{d}z_1 \ldots \mathrm{d}z_n, \tag{3.29}$$

wherein k_+ specializes to k_{++} when $\boldsymbol{Q} \succ \boldsymbol{0}$. As mentioned earlier, the PDE (3.5b) *and its solution can be mapped to that of* (3.5a) *via transformation $t \mapsto t_0 + t_1 - t$; therefore, the other Schrödinger factor $\varphi(t, \boldsymbol{x}) = \widehat{\eta}(t_0 + t_1 - t, \boldsymbol{y} = \boldsymbol{V}\boldsymbol{x})$.*

From (3.28) and (3.29), we determine the following properties of the kernel k_+ and Schrödinger factor $\widehat{\varphi}(t, \boldsymbol{x})$.

Proposition 3.4 (Properties of the kernel and the solution). *For* $0 \leq t_0 < t < \infty$, *given* $\boldsymbol{Q} \succeq \boldsymbol{0}$, $\widehat{\varphi}_0(\cdot)$ *Lebesgue measurable and everywhere positive, the kernel* k_+ *and the solution* (3.29) *satisfy:*

(i) ***(spatial symmetry)*** $k_+(t_0, \boldsymbol{y}, t, \boldsymbol{z}) = k_+(t_0, \boldsymbol{z}, t, \boldsymbol{y})$,

(ii) ***(positivity)*** $\widehat{\varphi}(t, \boldsymbol{x}) > 0$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$,

(iii) ***(decay)*** $\lim_{|\boldsymbol{x}| \to \infty} \widehat{\varphi}(t, \boldsymbol{x}) = \lim_{t \to \infty} \widehat{\varphi}(t, \boldsymbol{x}) = 0$.

For a proof of Proposition 3.4, see Appendix C.8.

3.3 Numerical Simulations

In Fig. 3.2, Fig. 3.3 and Fig. 3.4, we illustrate the solution $\widehat{\varphi}(t, \boldsymbol{x})$ in (3.29) for $n = 2$ dimensions with $t_0 = 0$, initial conditions $\widehat{\varphi}_0(\cdot) = 1$, $\widehat{\varphi}_0(\cdot) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, and $\widehat{\varphi}_0(\cdot) \propto \exp\left(-f(x_1, x_2)/35\right)$ where

$$f(x_1, x_2) := \left((10x_1 - 5)^2 + 10x_2 - 16\right)^2 + \left(10x_1 - 12 + (10x_2 - 5)^2\right)^2,$$
$$\text{and } (x_1, x_2) \in [0, 1]^2,$$

respectively. We point out that the last $\widehat{\varphi}_0$ is of non-Gaussian Gibbs-type [109, Section V-C], and f there is a scaled Himmelblau function, often used as a benchmark in nonconvex optimization [110].

In these figures, the columns correspond to different times and the rows correspond to different problem data: $\boldsymbol{Q} \succeq \boldsymbol{0}$. These results agree with the properties in Proposition 3.4, and were obtained by using the derived kernels in Theorem 3.1 and Theorem 3.3. For the three cases of $\widehat{\varphi}_0$ reported here, we analytically performed the integration in (3.29) using Lemma 3.2.

Specifically, for $t_0 = 0$, $\widehat{\varphi}_0(\cdot) = 1$, and $\boldsymbol{Q} \succ \boldsymbol{0}$, (3.18)-(3.19) give

$$\begin{aligned}\widehat{\eta}(t,\boldsymbol{y})\big|_{\widehat{\varphi}_0=1} &= \frac{(\det(\boldsymbol{\Lambda}))^{1/4}\exp\!\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\coth\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\boldsymbol{y}\right)}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\sinh(2t\sqrt{\boldsymbol{\Lambda}}))}}\\ &\qquad\int_{\mathbb{R}^n}\exp\!\left(-\frac{1}{2}\boldsymbol{z}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\coth\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\boldsymbol{z}+\boldsymbol{y}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\operatorname{csch}\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\boldsymbol{z}\right)\mathrm{d}\boldsymbol{z}\\ &= \frac{\exp\!\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\tanh\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\boldsymbol{y}\right)}{\sqrt{\det\left(\cosh\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\right)}},\end{aligned}\tag{3.30}$$

where the last equality follows from invoking Lemma 3.2 for the integral in the preceding line. The corresponding

$$\begin{aligned}\widehat{\varphi}(t,\boldsymbol{x})\big|_{\widehat{\varphi}_0=1} &= \widehat{\eta}\left(t,\boldsymbol{y}=\boldsymbol{V}^\top\boldsymbol{x}\right)\|_{\widehat{\varphi}_0=1}\\ &= \frac{\exp\!\left(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\tanh\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\boldsymbol{V}^{-1}\boldsymbol{x}\right)}{\sqrt{\det\left(\cosh\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\right)}}\\ &= \frac{\exp\!\left(-\frac{1}{4}\boldsymbol{x}^\top\sqrt{2\boldsymbol{Q}}\tanh\left(t\sqrt{2\boldsymbol{Q}}\right)\boldsymbol{x}\right)}{\sqrt{\det\left(\cosh\left(t\sqrt{2\boldsymbol{Q}}\right)\right)}},\end{aligned}\tag{3.31}$$

where the last equality used (3.10). For sanity check, we verify from (3.31) that $\widehat{\varphi}(t=0,\boldsymbol{x})\big|_{\widehat{\varphi}_0=1} = 1$, as expected. Also, for $\boldsymbol{\Lambda}\downarrow\boldsymbol{0}$ (heat kernel limit), (3.31) returns unity which is intuitive since the heat kernel leaves constant function invariant. For the results in Fig. 3.2, we used (3.31) to illustrate the effect of varying $\boldsymbol{Q}\succeq\boldsymbol{0}$. In particular, the last row of Fig. 3.2 used Theorem 3.3 to account for a zero eigenvalue. That row shows the joint effect of one eigen-direction evolving toward uniform (diffusion due to k_0) and another eigen-direction evolving as per k_{++}.

Likewise, for $t_0 = 0$, $\widehat{\varphi}_0(\cdot) = \mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}\right)$ and $\boldsymbol{Q}\succ\boldsymbol{0}$, (3.18)-(3.19) give

$$\begin{aligned}\widehat{\eta}(t,\boldsymbol{y})\big|_{\widehat{\varphi}_0=\mathcal{N}(\boldsymbol{0},\boldsymbol{I})} &= \frac{(\det(\boldsymbol{\Lambda}))^{\frac{1}{4}}\exp\!\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\coth\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\boldsymbol{y}\right)}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\sinh(2t\sqrt{\boldsymbol{\Lambda}}))}}\\ &\qquad\int_{\mathbb{R}^n}\exp\!\left(-\frac{1}{2}\boldsymbol{z}^\top\left\{\boldsymbol{\Lambda}^{\frac{1}{2}}\coth\left(2t\sqrt{\boldsymbol{\Lambda}}\right)+\boldsymbol{I}\right\}\boldsymbol{z}+\boldsymbol{y}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\operatorname{csch}\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\boldsymbol{z}\right)\mathrm{d}\boldsymbol{z}\end{aligned}$$

$$= \frac{(\det\boldsymbol{\Lambda})^{\frac{1}{4}}}{\sqrt{\det\left(\sqrt{\boldsymbol{\Lambda}}\cosh\left(2t\sqrt{\boldsymbol{\Lambda}}\right)+\sinh\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\right)}}$$
$$\exp\Bigg(-\frac{1}{2}\boldsymbol{y}^{\top}\boldsymbol{\Lambda}^{\frac{1}{2}}\Big\{\coth\left(2t\sqrt{\boldsymbol{\Lambda}}\right)-\operatorname{csch}\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\times$$
$$\left(\sqrt{\boldsymbol{\Lambda}}\coth\left(2t\sqrt{\boldsymbol{\Lambda}}\right)+\boldsymbol{I}\right)^{-1}\sqrt{\boldsymbol{\Lambda}}\operatorname{csch}\left(2t\sqrt{\boldsymbol{\Lambda}}\right)\Big\}\boldsymbol{y}\Bigg) \tag{3.32}$$

where, once again, the last equality follows from invoking Lemma 3.2 for the integral in the preceding line. The corresponding

$$\widehat{\varphi}(t,\boldsymbol{x})\big|_{\widehat{\varphi}_0=\mathcal{N}(\boldsymbol{0},\boldsymbol{I})} = \widehat{\eta}\left(t,\boldsymbol{y}=\boldsymbol{V}^{\top}\boldsymbol{x}\right)\|_{\widehat{\varphi}_0=\mathcal{N}(\boldsymbol{0},\boldsymbol{I})}$$
$$= \frac{2^{-n/4}(\det\boldsymbol{Q})^{\frac{1}{4}}}{\sqrt{\det\left(\sqrt{2\boldsymbol{Q}}\cosh\left(t\sqrt{2\boldsymbol{Q}}\right)/2+\sinh\left(t\sqrt{2\boldsymbol{Q}}\right)\right)}}$$
$$\exp\Bigg(-\frac{1}{4}\boldsymbol{x}^{\top}\sqrt{2\boldsymbol{Q}}\Big\{\coth\left(t\sqrt{2\boldsymbol{Q}}\right)-\operatorname{csch}\left(t\sqrt{2\boldsymbol{Q}}\right)\times$$
$$\left(\frac{\sqrt{2\boldsymbol{Q}}}{4}\coth\left(t\sqrt{2\boldsymbol{Q}}\right)+\boldsymbol{I}\right)^{-1}\sqrt{2\boldsymbol{Q}}\operatorname{csch}\left(t\sqrt{2\boldsymbol{Q}}\right)\Big\}\boldsymbol{x}\Bigg), \tag{3.33}$$

where the last equality is due to (3.10). For the results in Fig. 3.3, we used (3.33) to illustrate the effect of varying $\boldsymbol{Q} \geq \boldsymbol{0}$. As before, the last row of Fig. 3.3 used Theorem 3.3 to account for a zero eigenvalue.

Likewise, the results in Fig. 3.4 show the transient action of the kernel on non-Gaussian $\widehat{\varphi}_0$.

Our final simulation brings together the ideas to solve the SBP with state cost-to-go. In Fig. 3.5, we show the closed loop optimally controlled sample paths together with endpoint PDFs $\rho_0 = \sum_{\mu_0\in\{-1,1\}}\frac{1}{2}\mathcal{N}(\mu_0, 0.05^2)$, $\rho_1 = \mathcal{N}(0, 0.5^2)$ for an 1D SB with state cost $\frac{1}{2}Qx^2$ with $Q = 2$ and time horizon $[t_0, t_1] = [0, 1]$. This result is obtained by performing the dynamic Sinkhorn recursions shown in Fig. 1.1 wherein we solved the forward-backward IVPs for the PDEs (3.5) as in Remark 3.2, which was in turn made possible by the kernel k_{++} deduced in Theorem 3.1.

This numerical simulation demonstrates that our results can indeed be leveraged to solve (3.1), including with non-Gaussian ρ_0. However, this approach does not naturally lend itself to extending our results for time-varying weights in quadratic state-cost-to-go or with prior dynamics

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{A}(t)\boldsymbol{x}(t)\mathrm{d}t + \sqrt{2}\boldsymbol{B}(t)\mathrm{d}\boldsymbol{w}$$

as opposed to the case $(\boldsymbol{A}(t), \boldsymbol{B}(t)) = (\boldsymbol{0}, \boldsymbol{I})$ considered in this chapter. Thus motivated, we next investigate an alternative route to deriving kernels k_{++} and k_+.

3.4 Deriving k_+ via Weyl Calculus

This section explores how ideas from Weyl calculus in quantum mechanics, specifically the Weyl operator and the Weyl symbol, can help determine Markov kernels, as demonstrated in [75]. We illustrate these ideas by explicitly finding the Markov kernel for the case of convex quadratic state cost via Weyl calculus, recovering as well as generalizing our earlier results but avoiding the tedious computation with Hermite polynomials.

The mathematical development of Weyl calculus originated in quantum mechanics, but its relevance in solving problems outside quantum mechanics, especially in diffusion and control problems, is less well known. Common references [43], [44, Ch. 2-3], [45–47] on Weyl calculus appeal to quantum physics and field-theoretic ideas to introduce the core mathematical concepts.In contrast, the reminder of this chapter takes a different, expository approach: a purposefully less rigorous recipe-style introduction of selective Weyl calculus tools with illustrative calculations. The recipe for finding the kernel follows the path

$$\text{PDE} \longrightarrow \text{Weyl operator} \longrightarrow \text{Weyl symbol} \longrightarrow \text{kernel}.$$

This section explores the two middle ingredients, the Weyl operator and the Weyl symbol, and their connections with the familiar PDE and kernel.

3.4.1 Key Notations and Background for Weyl Calculus

Before presenting our results from [75], I fix key notations and background that find use in the following developments.

Poisson bracket. Let $\mathrm{Skew}\,(2n,\mathbb{R})$ denote the set of $2n \times 2n$ skew-symmetric matrices with real entries. Given sufficiently smooth functions $f, g : \mathbb{R}^{2n} \mapsto \mathbb{R}$, define their *Poisson bracket* $\{\cdot,\cdot\}$ as

$$\{f,g\} := \langle \nabla f, \boldsymbol{\Omega}\nabla g\rangle,\ \boldsymbol{\Omega} := \begin{pmatrix} \mathbf{0} & \boldsymbol{I}_n \\ -\boldsymbol{I}_n & \mathbf{0}\end{pmatrix} \in \mathrm{Skew}\,(2n,\mathbb{R})\,, \tag{3.34}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard Euclidean inner product. Note that $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Omega}^{\top} = -\boldsymbol{\Omega}$, and $\det(\boldsymbol{\Omega}) = 1$. As is well-known [111, Prop. 2.23], the Poisson bracket is bilinear, skew symmetric, and satisfies the Jacobi identity.

For $\boldsymbol{x}, \boldsymbol{\xi} \in \mathbb{R}^n$, we think of f, g as mappings which act on the tuple $(\boldsymbol{x},\boldsymbol{\xi})$, i.e., $f, g : (\boldsymbol{x},\boldsymbol{\xi}) \in \mathbb{R}^{2n} \mapsto \mathbb{R}$. Following [112, p. 128-9], let

$$\{f,g\}_1(\boldsymbol{x},\boldsymbol{\xi}) := \frac{1}{2\mathrm{i}}\{f,g\}(\boldsymbol{x},\boldsymbol{\xi}), \quad \mathrm{i} := \sqrt{-1}.$$

More generally, for any $j \in \mathbb{N}_0 := \{0,1,2,\cdots\}$, let

$$\{f,g\}_j\,(\boldsymbol{x},\boldsymbol{\xi}) := \left(\frac{1}{2\mathrm{i}}\right)^j \left(\sum_{k=1}^{n}\left(\frac{\partial^2}{\partial y_k \partial \xi_k} - \frac{\partial^2}{\partial x_k \partial \nu_k}\right)\right)^j f(\boldsymbol{x},\boldsymbol{\xi})g(\boldsymbol{y},\boldsymbol{\eta})\Big|_{\boldsymbol{y}=\boldsymbol{x},\boldsymbol{\eta}=\boldsymbol{\xi}}. \tag{3.35}$$

In particular, for $j = 0$, definition (3.35) gives

$$\{f,g\}_0\,(\boldsymbol{x},\boldsymbol{\xi}) = f\,(\boldsymbol{x},\boldsymbol{\xi})\,g\,(\boldsymbol{x},\boldsymbol{\xi})\,,$$

i.e., the product of the two functions.

Standard symplectic form. For $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^n$, the standard symplectic form associated with the matrix $\boldsymbol{\Omega}$ in (3.34) is a bilinear mapping[2] $\sigma : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \mapsto \mathbb{R}$, given by

$$\sigma\left((\boldsymbol{x}, \boldsymbol{\xi})\,;(\boldsymbol{y}, \boldsymbol{\eta})\right) \coloneqq \sum_{k=1}^{n} \left(\xi_k y_k - x_k \eta_k\right). \tag{3.36}$$

The symplectic form σ is skew-symmetric ($\sigma(\boldsymbol{v};\boldsymbol{w}) = \boldsymbol{v}^\top \boldsymbol{\Omega} \boldsymbol{w} = -\sigma(\boldsymbol{w};\boldsymbol{v})\ \forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^{2n}$), and non-degenerate (for any $\boldsymbol{v} \in \mathbb{R}^{2n}$, $\sigma(\boldsymbol{v};\boldsymbol{w}) = 0\ \forall \boldsymbol{w} \in \mathbb{R}^{2n} \Rightarrow \boldsymbol{v} = \boldsymbol{0}$).

3.4.2 Weyl Calculus

Armed with the initial Weyl calculus background above, I can now explain how we apply Weyl calculus to determine the Markov kernel associated with a linear PDE IVP of interest. While we are primarily motivated for the case when the PDE is reaction-diffusion type, this section proceeds formally in terms of more generic operators. Additional necessary ideas and nomenclatures from Weyl calculus will appear throughout this chapter.

For $t_0 \geq 0$ and $\mathcal{X} \subseteq \mathbb{R}^n$, we begin by considering a linear PDE IVP

$$\frac{\partial}{\partial t}\widehat{\varphi} = -\mathcal{L}\widehat{\varphi}, \quad \widehat{\varphi}(t_0, \boldsymbol{x}) = \widehat{\varphi}_0(\boldsymbol{x}) \text{ (given)}, \tag{3.37}$$

where the unknown $\widehat{\varphi} : [t_0, \infty) \times \mathcal{X} \mapsto \mathbb{R}$, and the spatial operator $\mathcal{L}$ is time-independent. Formally, the solution of (3.37) can be expressed via the semigroup $\exp\left(-(t - t_0)\mathcal{L}\right)$ as

$$\widehat{\varphi} = \exp\left(-\left(t - t_0\right)\mathcal{L}\right)\widehat{\varphi}_0. \tag{3.38}$$

We seek a kernel representation of the solution $\widehat{\varphi}$, i.e., for $0 \leq t_0 \leq s \leq t < \infty$,

[2] More generally, the mapping $\sigma : \mathbb{C}^{2n} \times \mathbb{C}^{2n} \mapsto \mathbb{R}$ remains well-defined as stated here but we will not need this fact.

we seek to explicitly determine $k(s, \boldsymbol{x}, t, \boldsymbol{y})$ with suitable domain of definition, such that (3.38) is expressible as

$$\widehat{\varphi}(t, \boldsymbol{x}) = \int_{\mathcal{X}} k(t_0, \boldsymbol{x}, t, \boldsymbol{y}) \widehat{\varphi}_0 (\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}. \tag{3.39}$$

The initial condition $\widehat{\varphi}(t_0, \boldsymbol{x}) = \widehat{\varphi}_0(\boldsymbol{x})$ necessitates that

$$k(t_0, \boldsymbol{x}, t_0, \boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y}), \text{ the Dirac delta.} \tag{3.40}$$

For convenience, fix $\mathcal{X} \equiv \mathbb{R}^n$ for ease of exposition; the development goes through for general subsets of $\mathbb{R}^n$.

When the PDE in (3.37) is the forward Kolmogorov a.k.a. Fokker-Planck PDE, then $\mathcal{L}$ is an advection-diffusion operator and under mild regularity assumptions on the underlying drift and diffusion coefficients, the kernel k is a transition probability density; see e.g., [113, Ch. 1]. However, if $\mathcal{L}$ has an additional reaction term, then (3.37) is non-conservative and k no longer has the interpretation of transition probability density. In such more generic settings, an analytical handle on k can still help to numerically compute the solution of (3.37) via (3.39). In the absence of an analytic handle on k, numerically computing $\widehat{\varphi}$ becomes particularly non-trivial in the presence of (state dependent) reaction [22, Section 5.2, 5.3].

The relevant instance of (3.37) in this dissertation is the reaction-diffusion PDE IVP with $\mathcal{L} \equiv -\Delta_{\boldsymbol{x}} + q(\boldsymbol{x})$, wherein $\Delta_{\boldsymbol{x}}$ denotes the Euclidean Laplacian operator, and $q(\boldsymbol{x})$ is a given bounded continuous reaction rate. If q is constant, then a change of variable $\widehat{\varphi} \mapsto \widehat{\vartheta} := e^{tq}\widehat{\varphi}$ transforms the problem to the heat PDE IVP:

$$\frac{\partial}{\partial t}\widehat{\vartheta} = \Delta_{\boldsymbol{x}}\widehat{\vartheta}, \quad \widehat{\vartheta}(t = t_0, \boldsymbol{x}) = \widehat{\varphi}_0(\boldsymbol{x}),$$

for which k is well-known [114, p. 44-47]. However, for state-dependent $q(\boldsymbol{x})$ new ideas are needed. Using the Weyl calculus tools discussed in Section 3.4.3, we ultimately find k for quadratic $q(\boldsymbol{x})$, as it appears in (3.1).

3.4.3 From PDE to Weyl Operator

In pursuit of getting an analytic handle on k in (3.39), we start by defining the following operators as per the quantum mechanics convention [43, p. 1]:

$$X_k := x_k \qquad \forall k \in [n], \tag{3.41a}$$

$$D_k := \frac{1}{\mathrm{i}} \frac{\partial}{\partial x_k} \qquad \forall k \in [n], \tag{3.41b}$$

and let

$$\boldsymbol{X} := \begin{pmatrix} X_1 & \cdots & X_n \end{pmatrix}^\top, \quad \boldsymbol{D} := \begin{pmatrix} D_1 & \cdots & D_n \end{pmatrix}^\top.$$

We then determine a representation of $\exp(-(t-t_0)\mathcal{L})$ in terms of the operators $\boldsymbol{X}$ and $\boldsymbol{D}$. This representation is referred to as the *Weyl Operator* $H(\boldsymbol{X}, \boldsymbol{D})$.

Example 3.1. *[Weyl operator for the heat PDE] Let* (3.37) *be the heat PDE IVP with* $\mathcal{L} \equiv -\Delta_{\boldsymbol{x}}$. *Letting* $\nabla_{\boldsymbol{x}} := \begin{pmatrix} \frac{\partial}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} \end{pmatrix}^\top$, *note that*

$$\|\boldsymbol{D}\|_2^2 := \langle \boldsymbol{D}, \boldsymbol{D} \rangle = (-\mathrm{i})^2 \left(\nabla_{\boldsymbol{x}} \cdot \nabla_{\boldsymbol{x}}\right) = -\Delta_{\boldsymbol{x}},$$

i.e., $\Delta_{\boldsymbol{x}} = -\|\boldsymbol{D}\|_2^2$. *So the corresponding Weyl operator is*

$$H_{\text{heat}}(\boldsymbol{X}, \boldsymbol{D}) = \exp\left(-(t-t_0)\|\boldsymbol{D}\|_2^2\right).$$

3.4.4 From Weyl Operator to Weyl Symbol

After identifying the Weyl operator associated with our PDE of interest, we seek the corresponding *Weyl symbol* denoted as $h(\boldsymbol{x}, \boldsymbol{\xi})$.

Note that while the Weyl operator depends on $\boldsymbol{X}$ and $\boldsymbol{D}$, the Weyl symbol is dependent on variables $\boldsymbol{x}$ and $\boldsymbol{\xi}$. An approach for determining $h(\boldsymbol{x}, \boldsymbol{\xi})$ discussed in [43, p. 32-36] proceeds as follows:

1. Rewrite the Weyl operator $H(\boldsymbol{X}, \boldsymbol{D})$ such that all of the D_k operators are multiplied on the right of the X_k operators. As the operators in general are non-commutative, this may require the use of the commutation relation

$$[X_k, D_k] := X_k D_k - D_k X_k = \mathrm{i}, \quad k \in [n].$$

2. Define $R(\boldsymbol{x}, \boldsymbol{\xi})$ by replacing the operator $\boldsymbol{X}$ with $\boldsymbol{x}$ and the operator $\boldsymbol{D}$ with $\boldsymbol{\xi}$ in the version of the Weyl operator $H(\boldsymbol{X}, \boldsymbol{D})$ satisfying Step 1 above.

3. Calculate the Weyl symbol $h(\boldsymbol{x}, \boldsymbol{\xi})$ via one of the following formulas:

$$h(\boldsymbol{x}, \boldsymbol{\xi}) = \frac{1}{\pi^n} \int_{\mathbb{R}^{2n}} R(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\xi}}) \exp\left(2\mathrm{i}\langle \tilde{\boldsymbol{x}} - \boldsymbol{x}, \tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}\rangle\right) d\tilde{\boldsymbol{x}}\, d\tilde{\boldsymbol{\xi}}, \tag{3.42a}$$

$$= \sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{\mathrm{i}}{2}\right)^m \left(\frac{\partial^m}{\partial \boldsymbol{x}^m} \cdot \frac{\partial^m}{\partial \boldsymbol{\xi}^m}\right) R(\boldsymbol{x}, \boldsymbol{\xi}). \tag{3.42b}$$

The differential operators in (3.42b) are understood as the m^{th} order mixed partial derivatives. In particular, (3.42b) can be seen as the series expansion of

$$\exp\left(-\frac{1}{2\mathrm{i}} \frac{\partial}{\partial \boldsymbol{x}} \cdot \frac{\partial}{\partial \boldsymbol{\xi}}\right) R(\boldsymbol{x}, \boldsymbol{\xi}).$$

Example 3.2. *[Weyl symbol for the heat PDE] The Weyl operator $H_{\text{heat}}(\boldsymbol{X}, \boldsymbol{D})$ found in Example 3.1, does not involve any multiplication of $\boldsymbol{X}$ with $\boldsymbol{D}$, or indeed any dependence on $\boldsymbol{X}$. Therefore, Steps 1 and 2 mentioned above are reduced to a direct replacement of $\boldsymbol{D}$ with $\boldsymbol{\xi}$. Because the resulting*

$$R_{\text{heat}}(\boldsymbol{x}, \boldsymbol{\xi}) = R_{\text{heat}}(\boldsymbol{\xi}) = \exp\left(-(t - t_0)\|\boldsymbol{\xi}\|_2^2\right)$$

is independent of $\boldsymbol{x}$, (3.42b) *immediately yields the Weyl symbol*

$$h_{\text{heat}}(\boldsymbol{x}, \boldsymbol{\xi}) = e^{-(t-t_0)\|\boldsymbol{\xi}\|_2^2}. \tag{3.43}$$

This example illustrates that finding the Weyl symbol associated with (3.37) *is straightforward when the relevant Weyl operator depends on either $\boldsymbol{X}$ or $\boldsymbol{D}$ alone.*

3.4.5 From Weyl Symbol to Kernel

We follow the development in Hörmander [115, p. 161], to obtain the kernel k in (3.39) from the associated Weyl symbol $h(\boldsymbol{x},\boldsymbol{\xi})$ as

$$k(t_0,\boldsymbol{x},t,\boldsymbol{y}) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} h\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{2},\boldsymbol{\xi}\right)e^{\mathrm{i}\langle\boldsymbol{x}-\boldsymbol{y},\boldsymbol{\xi}\rangle}\mathrm{d}\boldsymbol{\xi}. \tag{3.44}$$

The RHS of (3.44) can be seen as the non-unitary inverse Fourier transform of

$$\boldsymbol{\xi}\mapsto h\left(\frac{(\boldsymbol{x}+\boldsymbol{y})}{2},\boldsymbol{\xi}\right).$$

Thus, for (3.40) to hold, the Weyl symbol h evaluated at t_0 must be equal to unity (see e.g., the heat symbol (3.43)). We will make use of this observation in our subsequent computation.

Example 3.3. *[Kernel for the heat PDE] Applying* (3.44) *to the Weyl symbol $h\equiv h_{\text{heat}}$ derived in Example 3.2, we obtain*

$$\begin{aligned} k_{\text{heat}}(t_0,\boldsymbol{x},t,\boldsymbol{y}) &= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\exp\left(-(t-t_0)\|\boldsymbol{\xi}\|_2^2\right)\exp\left(\mathrm{i}\langle\boldsymbol{x}-\boldsymbol{y},\boldsymbol{\xi}\rangle\right)\mathrm{d}\boldsymbol{\xi} \\ &= \frac{1}{(2\pi)^n}\prod_{k=1}^{n}\left(\int_{-\infty}^{\infty}\exp\left(-\xi_k^2(t-t_0)+\mathrm{i}(x_k-y_k)\xi_k\right)\mathrm{d}\xi_k\right). \end{aligned} \tag{3.45}$$

To evaluate (3.45)*, we specialize the Gaussian-like integral*[3]

$$\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}ax^2+\mathrm{i}Jx\right)\mathrm{d}x = \left(\frac{2\pi}{a}\right)^{1/2}\exp\left(-\frac{J^2}{2a}\right) \tag{3.46}$$

with $a=2(t-t_0)$, $J=x_k-y_k$. This yields

$$k_{\text{heat}}(t_0,\boldsymbol{x},t,\boldsymbol{y}) = \frac{1}{(2\pi)^n}\prod_{k=1}^{n}\left(\frac{\pi}{t-t_0}\right)^{1/2}\exp\left(-\frac{(x_k-y_k)^2}{4(t-t_0)}\right)$$

[3]This integral is simply the Fourier transform of the Gaussian with appropriate scaling.

$$= \frac{1}{(4\pi(t-t_0))^{n/2}} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{4(t-t_0)}\right),$$

which is indeed the well-known heat kernel [114, p. 44-47].

In summary, we may derive the solution of a PDE (3.37) in the kernel form (3.39) by first identifying the associated Weyl operator $H(\boldsymbol{X},\boldsymbol{D}) = \exp\left(-(t-t_0)\mathcal{L}(\boldsymbol{X},\boldsymbol{D})\right)$ as in Section 3.4.3. This is then followed up with a computation of the corresponding Weyl symbol $h(\boldsymbol{x},\boldsymbol{\xi})$ as in Section 3.4.4. Using this Weyl symbol h, we then arrive at an expression for the kernel k via (3.44).

For pedagogical clarity, Examples 3.1-3.3 illustrate the aforementioned process for the classical heat kernel. Now the question arises: Can these computations be performed for reaction-diffusion PDEs with quadratic state-dependent reaction rate that emerge from the SBP with quadratic state cost, (3.1)?

It turns out that for such problems, the Weyl symbol is most conveniently determined via a PDE approach [112, p. 130-132]. To prepare ground for this PDE approach, we need the product rule of Weyl calculus.

Product Rule of Weyl Calculus: Consider a Weyl operator $C(\boldsymbol{X},\boldsymbol{D})$ which admits a decomposition

$$C(\boldsymbol{X},\boldsymbol{D}) = A(\boldsymbol{X},\boldsymbol{D})B(\boldsymbol{X},\boldsymbol{D}) \tag{3.47}$$

in terms of two other Weyl operators $A(\boldsymbol{X},\boldsymbol{D}), B(\boldsymbol{X},\boldsymbol{D})$.

Consider the corresponding Weyl operator-Weyl symbol pairs (A,a), (B,b) and (C,c). Given the operator decomposition (3.47), the product rule is a recipe to determine the Weyl symbol c from the Weyl symbols a, b.

It is known that [115, p. 161] if either A or B is a polynomial, then the Weyl symbol $c(\boldsymbol{x},\boldsymbol{\xi})$ can be exactly determined via Taylor expansion of[4]

$$\exp\left(\frac{\mathrm{i}}{2}\sigma\left(\boldsymbol{D}_{\boldsymbol{x}},\boldsymbol{D}_{\boldsymbol{\xi}};\boldsymbol{D}_{\boldsymbol{y}},\boldsymbol{D}_{\boldsymbol{\eta}}\right)\right)(a(\boldsymbol{x},\boldsymbol{\xi})b(\boldsymbol{y},\boldsymbol{\eta}))\big|_{\boldsymbol{y}=\boldsymbol{x},\boldsymbol{\eta}=\boldsymbol{\xi}}$$

with *finitely many terms*, where σ is the standard symplectic form (3.36). In this case, we can express $c(\boldsymbol{x},\boldsymbol{\xi})$ as [112, p. 128-129]

$$c(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{j=0}^{d_A\wedge d_B}\frac{1}{j!}\{a,b\}_j(\boldsymbol{x},\boldsymbol{\xi}), \tag{3.48}$$

where $\{a,b\}_j(\boldsymbol{x},\boldsymbol{\xi})$ is defined as in (3.35), and $d_A\wedge d_B$ is the minimum of the degrees d_A, d_B for the polynomials A, B.

When only one of A, B is a polynomial, the corresponding polynomial degree is to be used as the upper limit of the summation index j in (3.48). When neither A nor B is polynomial, the equality in (3.48) is replaced with asymptotic equivalence: $c(\boldsymbol{x},\boldsymbol{\xi}) \sim \sum_{j\in\mathbb{N}_0}\frac{1}{j!}\{a,b\}_j(\boldsymbol{x},\boldsymbol{\xi})$.

3.4.6 Kernel for the SBP with Quadratic State Cost

As shown earlier in this chapter, solving the classical SBP with quadratic state cost and generic endpoint distributions (with finite second moments) leads to finding the Markov kernel of the reaction-diffusion PDE IVP

$$\frac{\partial\widehat{\varphi}}{\partial t} = \left(\Delta_{\boldsymbol{z}} - \frac{1}{2}\boldsymbol{z}^\top\boldsymbol{Q}\boldsymbol{z}\right)\widehat{\varphi}, \quad \widehat{\varphi}(t_0,\cdot) = \widehat{\varphi}_0, \tag{3.49}$$

for given $\boldsymbol{Q}\geq\boldsymbol{0}$, wherein $\boldsymbol{z}\in\mathbb{R}^n$ and the unknown $\widehat{\varphi}$ is a Schrödinger factor. Here and throughout, $0\leq t_0\leq t<\infty$.

Problem (3.49) is indeed of the form (3.37). The associated kernel in (3.39) was derived in Section 3.2 via tedious computation with Hermite polynomials. Since that computation is difficult to generalize for other state costs (thus for other state-

[4] Here $\boldsymbol{D}_{\boldsymbol{x}} := \begin{pmatrix}D_{x_1} & \cdots & D_{x_n}\end{pmatrix}^\top$, and likewise for $\boldsymbol{D}_{\boldsymbol{\xi}},\boldsymbol{D}_{\boldsymbol{y}},\boldsymbol{D}_{\boldsymbol{\eta}}$.

dependent reaction rates), we explore applying the Weyl calculus ideas in Section 3.4.2 for finding such kernels. Here we show that the steps outlined in Section 3.4.2 indeed recovers the kernel for (3.49) in a systematic manner without invoking Hermite polynomials.

For $\boldsymbol{Q} \geq \boldsymbol{0}$, consider again the eigen-decomposition $\frac{1}{2}\boldsymbol{Q} = \boldsymbol{V}^\top \boldsymbol{\Lambda} \boldsymbol{V}$, where $\boldsymbol{\Lambda}$ is an $n \times n$ diagonal matrix with its main diagonal comprising the eigenvalues $\{\lambda_k\}_{k=1}^n \in \mathbb{R}^n_{\geq 0}$ of $\frac{1}{2}\boldsymbol{Q}$. Now consider a linear change-of-variable $\boldsymbol{z} \mapsto \boldsymbol{x}$ as

$$\boldsymbol{x} := \boldsymbol{V}\boldsymbol{z}. \tag{3.50}$$

Then,

$$\widehat{\nu}(t, \boldsymbol{x}) := \widehat{\varphi}\left(t, \boldsymbol{z} = \boldsymbol{V}^\top \boldsymbol{x}\right), \tag{3.51}$$

and $\Delta_{\boldsymbol{z}} \widehat{\varphi}(t, \boldsymbol{z}) = \Delta_{\boldsymbol{x}} \widehat{\nu}(t, \boldsymbol{x})$. Therefore, in the $\boldsymbol{x}$ coordinates, (3.49) takes the form

$$\begin{aligned} \frac{\partial \widehat{\nu}}{\partial t} &= \Delta_{\boldsymbol{x}} \widehat{\nu} - \left(\boldsymbol{x}^\top \boldsymbol{\Lambda} \boldsymbol{x}\right) \widehat{\nu} \\ &= \sum_{k=1}^n \left(\frac{\partial^2}{\partial x_k^2} - \lambda_k x_k^2 \right) \widehat{\nu}. \end{aligned} \tag{3.52}$$

For notational convenience, we define the Weyl operator

$$Q_{\boldsymbol{\Lambda}}(\boldsymbol{X}, \boldsymbol{D}) := \|\boldsymbol{D}\|_2^2 + \sum_{k=1}^n \lambda_k X_k^2. \tag{3.53}$$

Following Section 3.4.3, the Weyl operator of (3.52) is

$$H_{\boldsymbol{\Lambda}}(\boldsymbol{X}, \boldsymbol{D}) = \exp\left(-(t - t_0) Q_{\boldsymbol{\Lambda}}(\boldsymbol{X}, \boldsymbol{D})\right), \tag{3.54}$$

i.e., the Weyl operator of (3.52), $H_{\boldsymbol{\Lambda}}(\boldsymbol{X}, \boldsymbol{D})$, can be seen as a composite operator.

Note that for the Weyl operator $Q_{\boldsymbol{\Lambda}}$ in (3.53), we can readily find its Weyl symbol $q_{\boldsymbol{\Lambda}}$ using (3.42b) as

$$q_{\boldsymbol{\Lambda}}(\boldsymbol{x}, \boldsymbol{\xi}) = \|\boldsymbol{\xi}\|_2^2 + \sum_{k=1}^n \lambda_k x_k^2. \tag{3.55}$$

However, for the composite Weyl operator $H_{\mathbf{\Lambda}}$, determining the corresponding Weyl symbol $h_{\mathbf{\Lambda}}$ via (3.42b) becomes challenging since partial derivatives of $H_{\mathbf{\Lambda}}(\boldsymbol{X},\boldsymbol{D})$ of any order are nonzero. To circumvent this issue, we adopt a PDE approach from [112, p. 130-132] explained next. The main idea is to derive a PDE IVP for Weyl symbol $h_{\mathbf{\Lambda}}$ and to solve the same in closed form.

Since the Weyl operator $H_{\mathbf{\Lambda}}$ must satisfy (3.52), we have

$$\frac{\partial}{\partial t}H_{\mathbf{\Lambda}}(\boldsymbol{X},\boldsymbol{D}) = -Q_{\mathbf{\Lambda}}(\boldsymbol{X},\boldsymbol{D})H_{\mathbf{\Lambda}}(\boldsymbol{X},\boldsymbol{D}).$$

Then, applying (3.48), we note that the Weyl symbol $h_{\mathbf{\Lambda}}$ satisfies

$$\frac{\partial}{\partial t}h_{\mathbf{\Lambda}}(\boldsymbol{x},\boldsymbol{\xi}) = -\sum_{j=0}^{2}\frac{1}{j!}\{q_{\mathbf{\Lambda}},h_{\mathbf{\Lambda}}\}_j(\boldsymbol{x},\boldsymbol{\xi}). \tag{3.56}$$

Applying (3.35) and (3.55), we find that

$$\{q_{\mathbf{\Lambda}},h_{\mathbf{\Lambda}}\}_1 = 0, \tag{3.57}$$

and

$$\begin{aligned}\{q_{\mathbf{\Lambda}},h_{\mathbf{\Lambda}}\}_2 &= \left(\sum_{k=1}^{n}\left(2\lambda_k(t-t_0) - 2\lambda_k^2(t-t_0)^2x_k^2 - 2\lambda_k(t-t_0)^2\xi_k^2\right)\right)\\ &\qquad\times\exp\left(-(t-t_0)\left(\|\boldsymbol{\xi}\|_2^2 + \sum_{k=1}^{n}\lambda_k x_k^2\right)\right)\\ &= \left(2(t-t_0)\,\mathrm{trace}(\mathbf{\Lambda}) - 2(t-t_0)^2\left(\boldsymbol{x}^\top\mathbf{\Lambda}^2\boldsymbol{x} + \boldsymbol{\xi}^\top\mathbf{\Lambda}\boldsymbol{\xi}\right)\right)\\ &\qquad\times\exp\left(-(t-t_0)q_{\mathbf{\Lambda}}(\boldsymbol{x},\boldsymbol{\xi})\right).\end{aligned} \tag{3.58}$$

Since the second order derivative of $h_{\mathbf{\Lambda}}(\boldsymbol{x},\boldsymbol{\xi})$ w.r.t. $x_k \forall\, k\in[n]$ is

$$\frac{\partial^2}{\partial x_k^2}h(x,\xi) = \left(-2\lambda_k(t-t_0) + 4\lambda_k^2x_k^2(t-t_0)^2\right)\exp\left(-(t-t_0)q_{\mathbf{\Lambda}}(\boldsymbol{x},\boldsymbol{\xi})\right),$$

and likewise

$$\frac{\partial^2}{\partial\xi_k^2}h_{\mathbf{\Lambda}}(\boldsymbol{x},\boldsymbol{\xi}) = \left(-2(t-t_0) + 4\xi_k^2(t-t_0)^2\right)\exp\left(-(t-t_0)q_{\mathbf{\Lambda}}(\boldsymbol{x},\boldsymbol{\xi})\right),$$

(3.58) can be expressed as

$$\{q_{\mathbf{\Lambda}}, h_{\mathbf{\Lambda}}\}_2 = -\frac{1}{2}\left(\sum_{k=1}^{n} \lambda_k \frac{\partial^2 h_{\mathbf{\Lambda}}}{\partial \xi_k^2}(\boldsymbol{x}, \boldsymbol{\xi}) + \frac{\partial^2 h_{\mathbf{\Lambda}}}{\partial x_k^2}(\boldsymbol{x}, \boldsymbol{\xi})\right). \tag{3.59}$$

Combining (3.56), (3.57) and (3.59), we then have

$$\begin{aligned}\frac{\partial}{\partial t} h_{\mathbf{\Lambda}}(\boldsymbol{x}, \boldsymbol{\xi}) &= -q_{\mathbf{\Lambda}}(\boldsymbol{x}, \boldsymbol{\xi}) h_{\mathbf{\Lambda}}(\boldsymbol{x}, \boldsymbol{\xi}) - \frac{1}{2}\{q_{\mathbf{\Lambda}}, h_{\mathbf{\Lambda}}\}_2(\boldsymbol{x}, \boldsymbol{\xi}) \\ &= -q_{\mathbf{\Lambda}} h_{\mathbf{\Lambda}} + \frac{1}{4}\left(\sum_{k=1}^{n} \lambda_k \frac{\partial^2 h_{\mathbf{\Lambda}}}{\partial \xi_k^2} + \frac{\partial^2 h_{\mathbf{\Lambda}}}{\partial x_k^2}\right). \end{aligned} \tag{3.60}$$

We next focus on solving for $h_{\mathbf{\Lambda}}$ from (3.60) subject to the initial condition $h_{\mathbf{\Lambda}}|_{t=t_0} = 1$.

We note that the Weyl symbol $q_{\mathbf{\Lambda}}$ in (3.55) is time-invariant. Let $q_k := \lambda_k x_k^2 + \xi_k^2$ $\forall k \in [n]$, so that $q_{\mathbf{\Lambda}} = \sum_{k=1}^{n} q_k$.

We seek to solve (3.60) in the form

$$h_{\mathbf{\Lambda}}(\boldsymbol{x}, \boldsymbol{\xi}) = g(t, q_1, \cdots, q_n). \tag{3.61}$$

Substituting (3.61) into (3.60) yields

$$\frac{\partial g}{\partial t} = -\left(\sum_{k=1}^{n} q_k\right) g + \frac{1}{4}\sum_{k=1}^{n}\left(\frac{\partial}{\partial x_k}\left(\frac{\partial g}{\partial q_k}\frac{\partial q_k}{\partial x_k}\right) + \lambda_k \frac{\partial}{\partial \xi_k}\left(\frac{\partial g}{\partial q_k}\frac{\partial q_k}{\partial \xi_k}\right)\right). \tag{3.62}$$

As

$$\begin{aligned}\frac{\partial}{\partial x_k}\left(\frac{\partial g}{\partial q_k}\frac{\partial q_k}{\partial x_k}\right) &= \frac{\partial}{\partial x_k}\left(\frac{\partial g}{\partial q_k}\left(2\lambda_k x_k\right)\right) \\ &= \frac{\partial^2 g}{\partial q_k^2}\left(4\lambda_k^2 x_k^2\right) + 2\lambda_k \frac{\partial g}{\partial q_k},\end{aligned}$$

and likewise

$$\begin{aligned}\frac{\partial}{\partial \xi_k}\left(\frac{\partial g}{\partial q_k}\frac{\partial q_k}{\partial \xi_k}\right) &= \frac{\partial}{\partial \xi_k}\left(\frac{\partial g}{\partial q_k}(2\xi_k)\right) \\ &= 4\xi_k^2 \frac{\partial^2 g}{\partial q_k^2} + 2\frac{\partial g}{\partial q_k},\end{aligned}$$

the PDE (3.62) simplifies to

$$\frac{\partial g}{\partial t} = -\left(\sum_{k=1}^{n} q_k\right) g + \sum_{k=1}^{n}\left(\lambda_k q_k \frac{\partial^2 g}{\partial q_k^2} + \lambda_k \frac{\partial g}{\partial q_k}\right). \tag{3.63}$$

To solve (3.63), we consider the ansatz

$$g(t, q_1, \cdots, q_n) = \alpha(t)\exp\left(-\sum_{k=1}^{n}\beta_k(t) q_k\right) \tag{3.64}$$

for suitably smooth $\alpha(t), \beta_1(t), \cdots, \beta_n(t)$ that are not identically zero for any $t_0 < t$. Recall from Section 3.4.5 that $h_{\mathbf{\Lambda}}|_{t=t_0} = 1$, and hence from (3.61) and (3.64), the initial conditions for $\alpha(\cdot), \beta_1(\cdot), \cdots, \beta_n(\cdot)$ are

$$\alpha\left(t_0\right) = 1, \quad \beta_k\left(t_0\right) = 0 \quad \forall\, k \in [n]. \tag{3.65}$$

Next, notice that

$$\frac{\partial g}{\partial t} = \left(\dot{\alpha} - \alpha\sum_{k=1}^{n}\dot{\beta}_k q_k\right)\exp\left(-\sum_{k=1}^{n}\beta_k(t) q_k\right), \tag{3.66a}$$

$$\frac{\partial g}{\partial q_k} = -\alpha\beta_k \exp\left(-\sum_{k=1}^{n}\beta_k q_k\right), \tag{3.66b}$$

$$\frac{\partial^2 g}{\partial q_k^2} = \alpha\beta_k^2 \exp\left(-\sum_{k=1}^{n}\beta_k q_k\right), \tag{3.66c}$$

where $\dot{\alpha} := \frac{\mathrm{d}\alpha}{\mathrm{d}t}, \dot{\beta} := \frac{\mathrm{d}\beta}{\mathrm{d}t}$. Substituting (3.64) into (3.63), using (3.66), and dividing through by $\exp\left(-\sum_{k=1}^{n}\beta_k(t)q_k\right)$ yields

$$\dot{\alpha} - \alpha\sum_{k=1}^{n}\dot{\beta}_k q_k = \sum_{k=1}^{n}\left(-\alpha q_k + \lambda_k\alpha\beta_k^2 q_k + \lambda_k\left(-\alpha\beta_k\right)\right).$$

Rearranging the above to collect terms involving q_k in the RHS, we obtain

$$\dot{\alpha} + \sum_{k=1}^{n}\lambda_k\alpha\beta_k = \sum_{k=1}^{n}\left(\alpha\dot{\beta}_k - \alpha + \lambda_k\alpha\beta_k^2\right)q_k. \tag{3.67}$$

As the LHS of (3.67) is independent of $q_k\ \forall k \in [n]$, we note that the coefficients of q_k in the RHS must be zero $\forall k \in [n]$. This observation, together with the fact that

$\alpha(t) \neq 0$ (otherwise (3.64) gives trivial solution), results in the nonlinear ODE IVPs

$$\dot{\beta}_k = 1 - \lambda_k \beta_k^2, \quad \beta_k(t_0) = 0 \quad \forall k \in [n], \tag{3.68}$$

where we have imposed the $\beta_k(\cdot)$ initial conditions from (3.65).

Separation-of-variables and u substitution with $\tanh(u) = \sqrt{\lambda_k}\beta_k$, determine the solution to (3.68) as

$$\beta_k(t) = \frac{1}{\sqrt{\lambda_k}} \tanh\left(\sqrt{\lambda_k}(t - t_0)\right), \quad \forall k \in [n]. \tag{3.69}$$

Since all coefficients in the RHS of (3.67) are zero, combining (3.67) and (3.69) gives

$$\frac{\dot{\alpha}}{\alpha} = -\sum_{k=1}^{n} \sqrt{\lambda_k} \tanh\left(\sqrt{\lambda_k}(t - t_0)\right), \quad \alpha(t_0) = 1, \tag{3.70}$$

where we have imposed the $\alpha(\cdot)$ initial condition from (3.65). Direct integration yields the solution of the ODE IVP (3.70):

$$\alpha(t) = \prod_{k=1}^{n} \frac{1}{\cosh\left(\sqrt{\lambda_k}(t - t_0)\right)}. \tag{3.71}$$

Combining (3.61), (3.64), (3.69), (3.71), we have thus determined that the Weyl symbol $h_{\boldsymbol{\Lambda}}$ associated with the Weyl operator $H_{\boldsymbol{\Lambda}}$, is

$$h_{\boldsymbol{\Lambda}}(\boldsymbol{x}, \boldsymbol{\xi}) \tag{3.72}$$

$$= \left(\prod_{k=1}^{n} \frac{1}{\cosh\left(\sqrt{\lambda_k}(t - t_0)\right)}\right) \times \exp\left(-\sum_{k=1}^{n} \frac{\lambda_k x_k^2 + \xi_k^2}{\sqrt{\lambda_k}} \tanh\left(\sqrt{\lambda_k}(t - t_0)\right)\right). \tag{3.73}$$

We now apply (3.44) to determine the kernel $k_{\boldsymbol{\Lambda}}$ corresponding to the Weyl symbol $h_{\boldsymbol{\Lambda}}$ given by (3.73).

Since (3.73) is a product of n terms, each with its unique subscript $k \in [n]$, so (3.44) can be written as the product of n univariate integrals, each being

$$\frac{1}{2\pi} \frac{1}{\cosh\left(\sqrt{\lambda_k}(t - t_0)\right)}$$

$$\times \exp\left(-\frac{\sqrt{\lambda_k}\left(x_k^2 + 2x_k y_k + y_k^2\right)}{4}\tanh\left(\sqrt{\lambda_k}(t-t_0)\right)\right)$$
$$\underbrace{\times\int_{-\infty}^{+\infty}\exp\left(-\frac{\xi_k^2}{\sqrt{\lambda_k}}\tanh\left(\sqrt{\lambda_k}(t-t_0)\right)+\mathrm{i}(x_k-y_k)\xi_k\right)\mathrm{d}\xi_k}_{I_k}, \tag{3.74}$$

where $k \in [n]$.

Invoking (3.46) with

$$a = \frac{2}{\sqrt{\lambda_k}}\tanh\left(\sqrt{\lambda_k}(t-t_0)\right), \qquad J = x_k - y_k,$$

the integral I_k in (3.74) evaluates to

$$\left(\frac{\pi\sqrt{\lambda_k}}{\tanh\left(\sqrt{\lambda_k}(t-t_0)\right)}\right)^{1/2}\exp\left(-\frac{\sqrt{\lambda_k}(x_k-y_k)^2}{4\tanh\left(\sqrt{\lambda_k}(t-t_0)\right)}\right).$$

Thus, (3.74) is the product of a pre-factor and an exponential term, wherein the pre-factor is

$$\begin{aligned}\frac{1}{2\pi}\frac{1}{\cosh\left(\sqrt{\lambda_k}(t-t_0)\right)}&\left(\frac{\pi\sqrt{\lambda_k}}{\tanh\left(\sqrt{\lambda_k}(t-t_0)\right)}\right)^{1/2}\\ &= \frac{\lambda_k^{1/4}}{\sqrt{2\pi\sinh\left(2\sqrt{\lambda_k}(t-t_0)\right)}},\end{aligned} \tag{3.75}$$

and letting $\theta_k := \sqrt{\lambda_k}(t-t_0)$, the argument of the exponential term is

$$\begin{aligned}&-\frac{\sqrt{\lambda_k}(x_k^2+2x_ky_k+y_k^2)}{4}\tanh\theta_k - \frac{\sqrt{\lambda_k}(x_k-y_k)^2}{4\tanh\theta_k}\\ &= -\frac{\sqrt{\lambda_k}}{4}\left\{\left(x_k^2+y_k^2\right)\left(\frac{\sinh\theta_k}{\cosh\theta_k}+\frac{\cosh\theta_k}{\sinh\theta_k}\right)+2x_ky_k\left(\frac{\sinh\theta_k}{\cosh\theta_k}-\frac{\cosh\theta_k}{\sinh\theta_k}\right)\right\}\\ &= -\frac{\sqrt{\lambda_k}}{2}(x_k^2+y_k^2)\frac{\cosh(2\theta_k)}{\sinh(2\theta_k)}+\frac{\sqrt{\lambda_k}x_ky_k}{\sinh(2\theta_k)}.\end{aligned} \tag{3.76}$$

Therefore, taking the product of the n integrals of the form (3.74) from $k=1$ to $k=n$, and substituting back $\theta_k = \sqrt{\lambda_k}(t-t_0)$, we arrive at the kernel

$$k_{\mathbf{\Lambda}}(t_0,\boldsymbol{x},t,\boldsymbol{y}) = \left(\prod_{k=1}^{n}\frac{\lambda_k^{1/4}}{\sqrt{2\pi\sinh\left(2\sqrt{\lambda_k}(t-t_0)\right)}}\right)$$

$$\times \exp\left(-\sum_{k=1}^{n}\frac{\sqrt{\lambda_k}}{2}(x_k^2+y_k^2)\frac{\cosh\left(2\sqrt{\lambda_k}(t-t_0)\right)}{\sinh\left(2\sqrt{\lambda_k}(t-t_0)\right)}\right)$$
$$\times \exp\left(\sum_{k=1}^{n}\sqrt{\lambda_k}x_k y_k\left(\frac{1}{\sinh\left(2\sqrt{\lambda_k}(t-t_0)\right)}\right)\right). \tag{3.77}$$

The expression (3.77) for the kernel matches with that derived in Section 3.2. Herein, by leveraging the Weyl calculus, we circumvent the use of Hermite polynomial-based tedious computation in 3.2.

3.4.7 Further Generalization Via Weyl Calculus

Formula (3.77) serves as the kernel or Green's function for (3.52), and helps write the solution for the PDE IVP (3.49) as

$$\widehat{\varphi}(t,\boldsymbol{z}) = \widehat{\nu}(t,\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{V}\boldsymbol{z}} = \int_{\mathbb{R}^n} k_{\boldsymbol{\Lambda}}(t_0,\boldsymbol{V}\boldsymbol{z},t,\boldsymbol{y})\,\varphi_0\left(\boldsymbol{V}^\top\boldsymbol{y}\right)\mathrm{d}\boldsymbol{y}. \tag{3.78}$$

We now extend this result for a modified version of (3.49) given by

$$\frac{\partial\widehat{\varphi}}{\partial t} = \left(\Delta_{\boldsymbol{z}} - \left(\frac{1}{2}\boldsymbol{z}^\top\boldsymbol{Q}\boldsymbol{z} + \boldsymbol{r}^\top\boldsymbol{z} + s\right)\right)\widehat{\varphi}, \quad \widehat{\varphi}(t_0,\cdot) = \widehat{\varphi}_0, \tag{3.79}$$

where $\boldsymbol{Q} \succeq \boldsymbol{0}$, $\boldsymbol{r} \in \mathbb{R}^n$, $s \in \mathbb{R}$. In other words, we generalize the reaction rate to have additional affine terms, i.e., *generic convex quadratic reaction rate*: $\frac{1}{2}\boldsymbol{z}^\top\boldsymbol{Q}\boldsymbol{z} + \boldsymbol{r}^\top\boldsymbol{z} + s$.

Theorem 3.4. *Given $\boldsymbol{Q} \succeq \boldsymbol{0}$, $\boldsymbol{r} \in \mathbb{R}^n$, $s \in \mathbb{R}$, consider the eigen-decomposition $\frac{1}{2}\boldsymbol{Q} = \boldsymbol{V}^\top\boldsymbol{\Lambda}\boldsymbol{V}$, where the main diagonal entries of the diagonal matrix Λ are $\{\lambda_k\}_{k=1}^n$. Let $\boldsymbol{v}_k$ be the kth column of $\boldsymbol{V}^\top$ for all $k \in [n]$.*

The solution for the PDE IVP (3.79) *is*

$$\widehat{\varphi}(t,\boldsymbol{z}) = \int_{\mathbb{R}^n} k_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(t_0,\boldsymbol{V}\boldsymbol{z},t,\boldsymbol{y})\,\varphi_0(\boldsymbol{V}^\top\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}, \tag{3.80}$$

where the kernel

$$k_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(t_0,\boldsymbol{x},t,\boldsymbol{y}) \tag{3.81}$$

$$
\begin{aligned}
&= \left(\prod_{k=1}^{n} \frac{\lambda_k^{1/4} \exp\left(-c_k(t-t_0)\right)}{\sqrt{2\pi \sinh\left(2\sqrt{\lambda_k}(t-t_0)\right)}} \right) \\
&\quad \times \exp\left(\sum_{k=1}^{n} -\frac{\sqrt{\lambda_k}}{2}(x_k^2+y_k^2)\frac{\cosh\left(2\sqrt{\lambda_k}(t-t_0)\right)}{\sinh\left(2\sqrt{\lambda_k}(t-t_0)\right)} + \frac{\sqrt{\lambda_k}x_k y_k}{\sinh\left(2\sqrt{\lambda_k}(t-t_0)\right)} \right. \\
&\qquad \left. - \frac{\left(\frac{1}{2}\boldsymbol{r}^\top \boldsymbol{v}_k\,(x_k+y_k) + \frac{1}{4\lambda_k}\left(\boldsymbol{r}^\top\boldsymbol{v}_k\right)^2\right)\tanh(\sqrt{\lambda_k}(t-t_0))}{\sqrt{\lambda_k}} \right),
\end{aligned}
\tag{3.82}
$$

and the constants

$$
c_k := \frac{1}{4\lambda_k}(\boldsymbol{r}^\top\boldsymbol{v}_k)^2 - \frac{s}{n}, \quad \forall k \in [n]. \tag{3.83}
$$

For a proof of Theorem 3.4, see Appendix C.9

Our calculations via Weyl calculus recovered the result (3.28), initially determined via computations with Hermite polynomials that were not only more tedious but also difficult to generalize for other variants of such PDEs. While Weyl calculus tools were invented to address problems in quantum mechanics, our developments highlight the potential of Weyl calculus in explicitly recovering the kernels of PDEs arising from diffusion and control problems.

This chapter focused on a generalization of the classical SBP in which there is an additional state-dependent quadratic cost. Throughout this chapter, we assume that the quadratic state-cost has a *fixed* $\boldsymbol{Q}$. The following chapter removes this assumption, to consider a time-dependent $\boldsymbol{Q}(t)$, and linear time-varying dynamics.

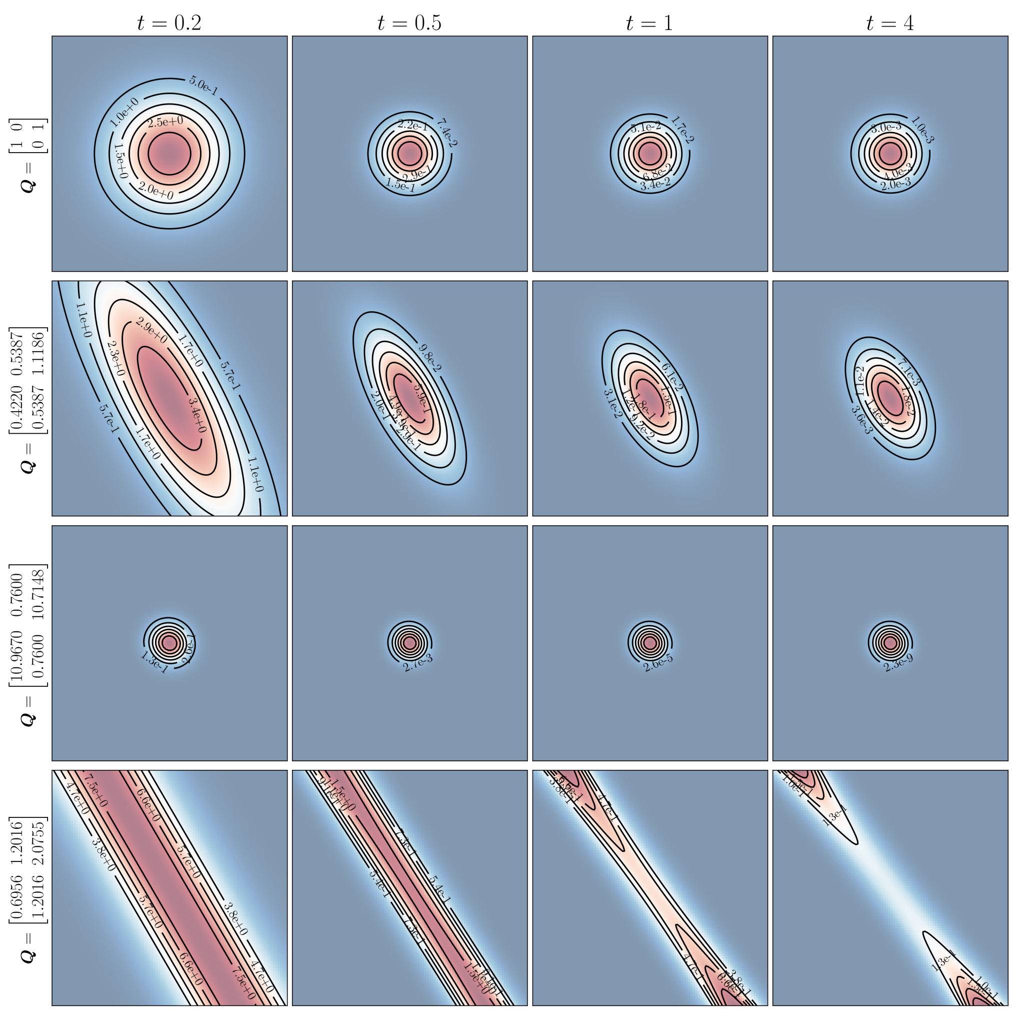

Figure 3.2: Solution to PDE (5.7a) in $n = 2$ dimensions with $t_0 = 0$ and initial condition $\widehat{\varphi}_0(\cdot) = 1$. All subfigures are over the spatial domain $[-7, 7]^2$.

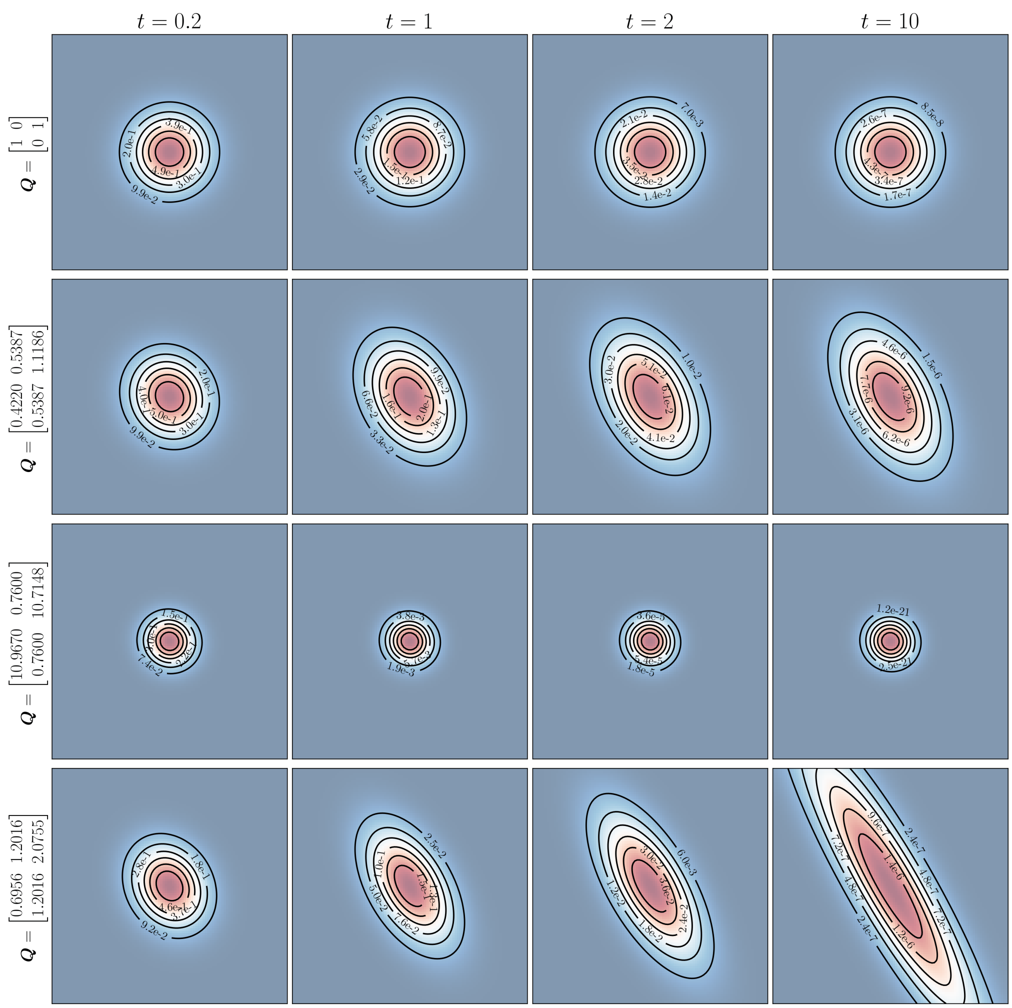

Figure 3.3: Solution to PDE (5.7a) in $n = 2$ dimensions with $t_0 = 0$ and initial condition $\widehat{\varphi}_0(\cdot) = \frac{1}{2\pi} e^{-\frac{1}{2}\|\cdot\|^2}$. All subfigures are over the spatial domain $[-5, 5]^2$.

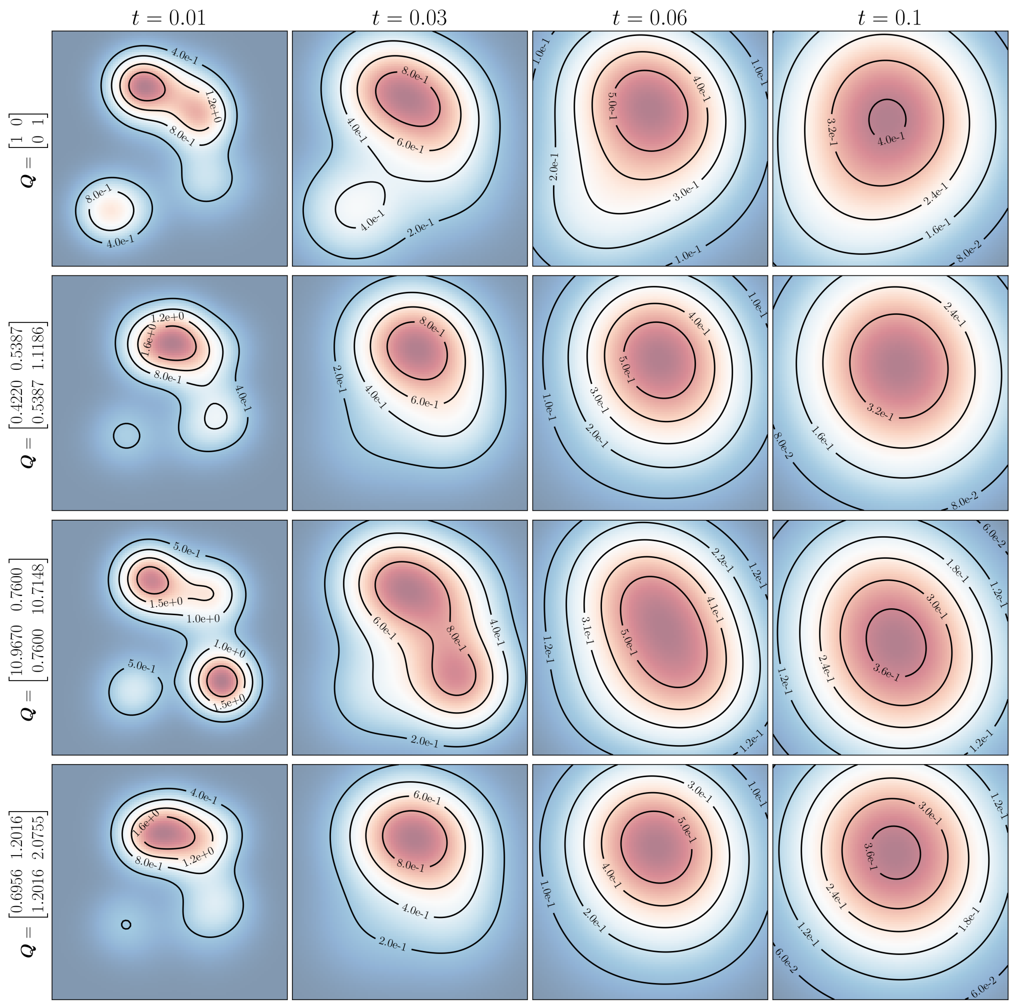

Figure 3.4: Solution to PDE (5.7a) in $n = 2$ dimensions with $t_0 = 0$ and initial condition $\widehat{\varphi}_0(\cdot) = \rho_\infty$. All subfigures are over the spatial domain $[-0.25, 1.25]^2$.

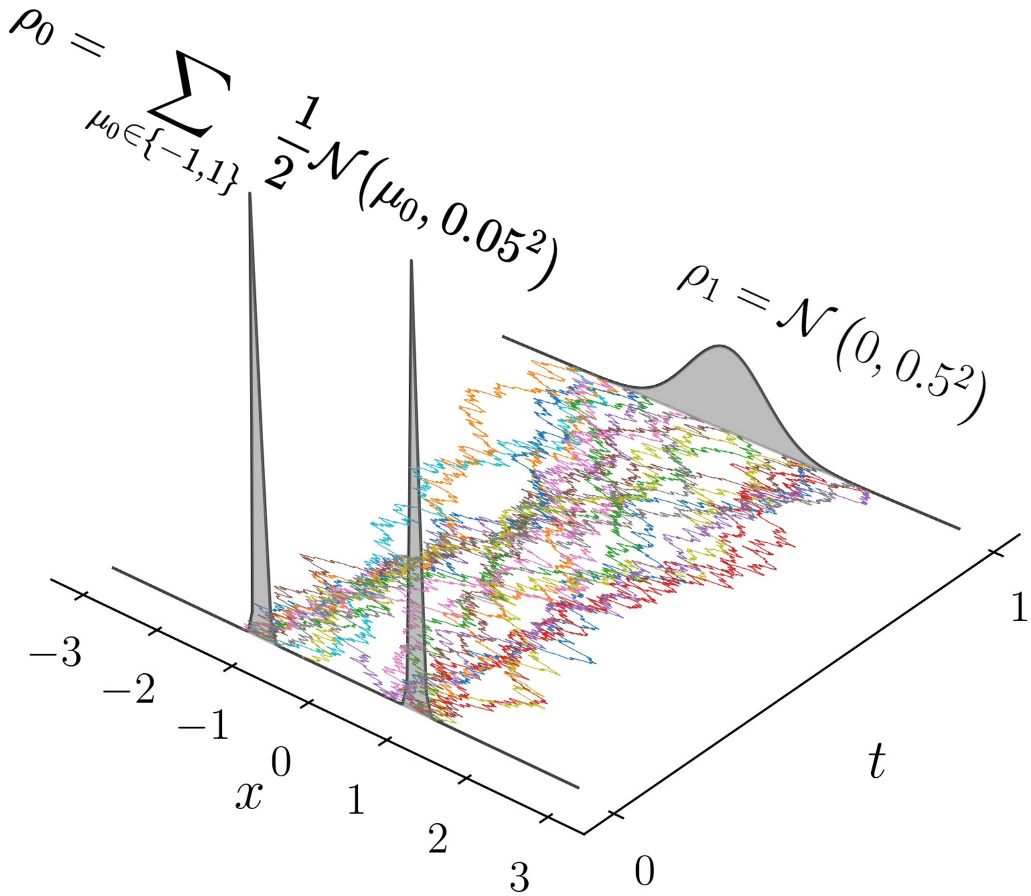

Figure 3.5: Optimally controlled sample paths for 1D SB with state cost $\frac{1}{2}Qx^2$, $Q = 2$, shown with endpoint endpoint PDFs $\rho_0 = \sum_{\mu_0 \in \{-1,1\}} \frac{1}{2}\mathcal{N}(\mu_0, 0.05^2)$, $\rho_1 = \mathcal{N}(0, 0.5^2)$ and time horizon $[t_0, t_1] = [0, 1]$.

4 | Linear SBP with Quadratic State Cost

4.1 Overview

The chapter investigates a further generalization of the SBP with quadratic state cost explored in Chapter 3. I now replace (3.2) with the controlled linear time-varying (LTV) dynamics:

$$\mathrm{d}\boldsymbol{x} = \left(\boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{B}(t)\boldsymbol{u}\right)\,\mathrm{d}t + \sqrt{2}\boldsymbol{B}(t)\,\mathrm{d}\boldsymbol{w}, \tag{4.1}$$

wherein $\left(\boldsymbol{A}(t), \boldsymbol{B}(t)\right)$ is controllable matrix-valued trajectory pair in $\left(\mathbb{R}^{n\times n}, \mathbb{R}^{n\times m}\right)$ that is bounded and continuous in $t \in [t_0, t_1]$.

The resulting generalized SBP is

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\frac{1}{2}\|\boldsymbol{u}\|_2^2 + \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}(t)\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{4.2a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot\left(\rho\left(\boldsymbol{A}(t)\,\boldsymbol{x} + \boldsymbol{B}(t)\,\boldsymbol{u}\right)\right) = \Delta_{\boldsymbol{B}\boldsymbol{B}^\top}\,\rho, \tag{4.2b}$$

$$\rho(t = t_0, \boldsymbol{x}) = \rho_0, \quad \rho(t = t_1, \boldsymbol{x}) = \rho_1, \tag{4.2c}$$

henceforth referred to as the LQ SBP.

Comparing this problem formulation to that of the generalized SBP introduced in Chapter 1,

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(q(\boldsymbol{x}) + r(\boldsymbol{u})\right)\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot\left(\rho\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u})\right) = \Delta_{\boldsymbol{\Sigma}(t,\boldsymbol{x},\boldsymbol{u})}\rho$$

$$\rho(t_0,\boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \rho(t_1,\boldsymbol{x}) = \rho_1(\boldsymbol{x}),$$

note that the LQ SBP can be recovered from the generalized SBP by setting $q(\boldsymbol{x}) \equiv$

$\frac{1}{2}\boldsymbol{x}\boldsymbol{Q}(t)\boldsymbol{x}$, $r(\boldsymbol{u}) \equiv \frac{1}{2}\|\boldsymbol{u}\|_2^2$, $\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u}) \equiv \boldsymbol{A}(t)\,\boldsymbol{x} + \boldsymbol{B}(t)\,\boldsymbol{u}$, and $\boldsymbol{\Sigma} \equiv \boldsymbol{B}\boldsymbol{B}^\top$ (or equivalently, $\boldsymbol{\sigma} \equiv \boldsymbol{B}$).

A natural question is whether a closed-form formula for the kernel k can be derived when (3.2) is replaced by (4.1). Finding such k would enable solving *LQ SBPs with generic non-Gaussian* μ_0, μ_1 having finite second moments. From a probabilistic point of view, this k is the Markov kernel of the Itô process

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{A}(t)\boldsymbol{x}\,\mathrm{d}t + \sqrt{2}\,\boldsymbol{B}(t)\mathrm{d}\boldsymbol{w}, \quad \boldsymbol{w} \in \mathbb{R}^m, \tag{4.4}$$

with creation or killing of probability mass with rate $q(\boldsymbol{x}) \coloneqq \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}(t)\boldsymbol{x}$, $\boldsymbol{Q}(t) \succeq \boldsymbol{0}$.

This Markov kernel is the Green's function for the linear reaction-advection-diffusion partial differential equation

$$\frac{\partial k}{\partial t} = -\langle\nabla_{\boldsymbol{x}}, k\boldsymbol{A}(t)\boldsymbol{x}\rangle + \langle\boldsymbol{B}(t)\boldsymbol{B}(t)^\top, \nabla_{\boldsymbol{x}}^2 k\rangle - \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}(t)\boldsymbol{x}\,k. \tag{4.5}$$

This chapter contains the results of joint work, [76], with Wenqing Wang, Sachin Shivakumar, and Abhishek Halder. In our derivation for the $(\boldsymbol{A}(t), \boldsymbol{B}(t), \boldsymbol{Q}(t))$-parametrized kernel, we introduce a new approach relying on finding a state-time dependent distance-like functional given by the solution of a deterministic optimal control problem. In the process, we uncover a new connection between Markov kernels, distances, and optimal control. This connection is of interest beyond its immediate application in solving the LQ SBP.

Note that Markov kernels often can be interpreted as *transition probability*, which are measurable maps sending *Borel probability measures*[1] supported on subsets of $\mathbb{R}^n$ to itself. The well-known (Euclidean) *heat kernel* [114, p. 44-47]

$$k_{\mathrm{Heat}}(t_0, \boldsymbol{x}, t, \boldsymbol{y}) \coloneqq \frac{1}{(4\pi\,(t - t_0))^{n/2}} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2}{4\,(t - t_0)}\right), \tag{4.6}$$

[1] endowed with the topology of weak convergence

is one such example, as the transition probability for the Itô process

$$\mathrm{d}\boldsymbol{x}(t) = \sqrt{2}\,\mathrm{d}\boldsymbol{w}(t), \quad \boldsymbol{w}(t) \in \mathbb{R}^n. \tag{4.7}$$

Likewise, the Markov kernel

$$\begin{aligned} &k_{\text{Linear}}\left(t_0, \boldsymbol{x}, t, \boldsymbol{y}\right) \\ &\quad= \left(4\pi\left(t-t_0\right)\right)^{-n/2} \det\left(\boldsymbol{\Gamma}_{tt_0}\right)^{-1/2} \exp\left(-\frac{\left(\boldsymbol{\Phi}_{tt_0}\boldsymbol{x}-\boldsymbol{y}\right)^{\top}\boldsymbol{\Gamma}_{tt_0}^{-1}\left(\boldsymbol{\Phi}_{tt_0}\boldsymbol{x}-\boldsymbol{y}\right)}{4\left(t-t_0\right)}\right), \end{aligned} \tag{4.8}$$

which is the transition probability for the Itô process (4.4), where $(\boldsymbol{A}(t), \boldsymbol{B}(t))$ is a uniformly controllable matrix-valued trajectory pair in $(\mathbb{R}^{n\times n}, \mathbb{R}^{n\times m})$ that is bounded and continuous in $t \in [t_0, \infty)$, and the state transition matrix $\boldsymbol{\Phi}_{t\tau} := \boldsymbol{\Phi}(t,\tau)\,\forall t_0 \leq \tau \leq t$.

Both (4.6) and (4.8) satisfy $k \geq 0$, $\int_{\mathbb{R}^n} k\mathrm{d}\boldsymbol{y} = 1$. They solve Kolmogorov's forward PDE initial value problem with Dirac delta initial condition,

$$\frac{\partial k}{\partial t} = Lk, \quad k(t_0, \boldsymbol{x}, t_0, \boldsymbol{y}) = \delta(\boldsymbol{x}-\boldsymbol{y}), \tag{4.9}$$

where L is an *advection-diffusion* spatial operator induced by the drift and diffusion coefficients of the underlying Itô process. For example, for the Itô process (4.7), $Lk \equiv \Delta_{\boldsymbol{x}} k$ (the standard Laplacian). For (4.4), $Lk \equiv -\langle \nabla_{\boldsymbol{x}}, k\boldsymbol{A}(t)\boldsymbol{x}\rangle + \boldsymbol{B}(t)\boldsymbol{B}(t)^{\top}\Delta_{\boldsymbol{x}} k$.

However, the Markov kernel k is *not always* a transition probability. Consider cases like (4.5), in which the underlying Itô process, in addition to drift and diffusion, allows the creation or killing of probability mass at a rate $q(\boldsymbol{x})$ for some bounded measurable q. In such cases, k is a measurable map that sends *nonnegative Borel measures*[2] supported on subsets of $\mathbb{R}^n$ to itself. Then, (4.9) gets replaced with

$$\frac{\partial k}{\partial t} = (L-q)\,k, \quad k(t_0, \boldsymbol{x}, t_0, \boldsymbol{y}) = \delta(\boldsymbol{x}-\boldsymbol{y}), \tag{4.10}$$

[2] still endowed with the topology of weak convergence

Ref.	$(\boldsymbol{A}_t, \boldsymbol{B}_t)$	$\boldsymbol{Q}_t \geq \boldsymbol{0}$	Distributions μ_0, μ_1	Markov kernel κ
[121, 122]	$(\boldsymbol{0}, \boldsymbol{I})$	$\boldsymbol{0}$	generic	(4.6)
[123]	generic	$\boldsymbol{0}$	Gaussian	(4.8)
[116, 117]	generic	generic	Gaussian	n/a
[73], Ch. 2	generic	$\boldsymbol{0}$	generic	(4.8)
[74, 75], Ch. 3	$(\boldsymbol{0}, \boldsymbol{I})$	fixed $\boldsymbol{Q} \geq \boldsymbol{0}$	generic	(3.28)
This work	generic	generic	generic	(4.37)

Table 4.1: Comparison of related works on the LQ SBP.

where L is an advection-diffusion operator as before ($q = 0$ case) and $L - q$ becomes a *reaction-advection-diffusion* operator. We say that q is the reaction rate.

Compared to (4.9), explicit formula for k in (4.10) are less known. In Chapter 3, I presented the derivation of a closed-form formula for k in the special case of $L \equiv \Delta_{\boldsymbol{x}}$ and convex quadratic $q(\boldsymbol{x})$, first by using Hermite polynomials and then again via Weyl calculus. However, our methods derived the kernel only for $(\boldsymbol{A}(t), \boldsymbol{B}(t)) = (\boldsymbol{0}, \boldsymbol{I})$ and constant $\boldsymbol{Q} \geq 0$.

Of particular relevance to the results presented in this chapter are the works [74, 75, 116, 117], all of which consider the quadratic state cost in the cost-to-go. The developments in [116, 117] focused on *Gaussian* ρ_0, ρ_1, and therefore did not derive the kernel, since linear dynamics preserve Gaussianity. Recently, the idea of using the Green's functions to solve SBPs for the purpose of sampling generative artificial intelligence models has appeared in [118, 119]. Also related is the work [120] that proposed a feasible bridge between non-Gaussian marginals with linear time-invariant dynamics, but with no guarantee for optimality. For reference, Table 4.1 contrasts the technical contribution of this Chapter vis-à-vis the related works in the literature.

Neither the Hermite polynomial approach nor the Weyl calculus computation discussed in Chapter 3 generalize in any obvious way to derive the kernel k in the LQ setting of our interest in this Chapter. Therefore, we chart a new path motivated by a basic observation on the structure of the Markov kernels, distances and certain deterministic optimal control problems.

We propose a new technique that involves identifying a deterministic optimal control problem from the Itô diffusion, solving the same to find a distance function, and then to identify the Markov kernel in a structure defined by the same. We provide computational details to demonstrate that the proposed technique systematically recovers Markov kernels of interest.

4.2 Markov Kernels and Distances

Both (4.6) and (4.8) are of the form

$$k = c\,(t, t_0) \exp\left(-\frac{1}{2}\mathrm{dist}^2_{tt_0}\,(\boldsymbol{x}, \boldsymbol{y})\right) \tag{4.11}$$

for some suitable distance function

$$\mathrm{dist}_{tt_0} : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}_{\geq 0}.$$

The subscript tt_0 signifies the distance function's parametric dependence on t, t_0.

Because (4.6) and (4.8) are both transition probabilities, the distance function dist_{tt_0} uniquely determines k. Once dist_{tt_0} is identified, the pre-factor $c\,(t, t_0)$, and hence k is completely determined by the normalization condition

$$\int_{\mathbb{R}^n} k \mathrm{d}\boldsymbol{y} = 1.$$

Interestingly, the k associated with problem (3.1) derived in Chapter 3 is also of the form (4.11), even though that k is not a transition probability, and $c(t, t_0)$ thus does not follow from normalization.

These exemplars hint that for controlled dynamics over $\mathbb{R}^n$, the distance function might be induced by some principle of least action.

That Markov kernels are related to distances is in itself not a new observation. The most well-known result relating kernels and distances is *Varadhan's formula* [124–126] which says that the heat kernel $k_{\text{Heat}}^{\mathcal{M}}$ on a complete Riemannian manifold $\mathcal{M}$ satisfies

$$\lim_{t \downarrow t_0} t \log k_{\text{Heat}}^{\mathcal{M}}(t_0, \boldsymbol{x}, t, \boldsymbol{y}) = -\frac{1}{2}\text{dist}^2(\boldsymbol{x}, \boldsymbol{y}) \tag{4.12}$$

uniformly on every compact subsets of $\mathcal{M} \times \mathcal{M}$, where dist is the minimal geodesic distance connecting $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{M}$. In other words, dist can be recovered as the short time asymptotic of the heat kernel on $\mathcal{M}$. In the context of Varadhan's formula (4.12), the L in (4.9) is the Laplace-Beltrami operator on $\mathcal{M}$.

For specific manifolds, a few exact or asymptotic relations are also known [127, Ch. VI], [128], [129, Ch. 5]:

$$k_{\text{Heat}}^{\mathbb{H}^3}(t_0, \boldsymbol{x}, t, \boldsymbol{y}) = \frac{1}{(4\pi(t-t_0))^{3/2}} \frac{\text{dist}(\boldsymbol{x}, \boldsymbol{y})}{\sinh \text{dist}(\boldsymbol{x}, \boldsymbol{y})} \times \exp\left(-t - \frac{\text{dist}^2(\boldsymbol{x}, \boldsymbol{y})}{4(t-t_0)}\right), \tag{4.13a}$$

$$k_{\text{Heat}}^{\mathbb{H}^n}(t_0, \boldsymbol{x}, t, \boldsymbol{y}) \sim \frac{1}{(4\pi(t-t_0))^{n/2}} \left(\frac{\text{dist}(\boldsymbol{x}, \boldsymbol{y})}{\sinh \text{dist}(\boldsymbol{x}, \boldsymbol{y})}\right)^{(n-1)/2} \times \exp\left(-\frac{\text{dist}^2(\boldsymbol{x}, \boldsymbol{y})}{4(t-t_0)}\right), \tag{4.13b}$$

$$k_{\text{Heat}}^{\mathbb{S}^n}(t_0, \boldsymbol{x}, t, \boldsymbol{y}) \sim \frac{1}{(4\pi(t-t_0))^{n/2}} \left(\frac{\text{dist}(\boldsymbol{x}, \boldsymbol{y})}{\sin \text{dist}(\boldsymbol{x}, \boldsymbol{y})}\right)^{(n-1)/2} \times \exp\left(-\frac{\text{dist}^2(\boldsymbol{x}, \boldsymbol{y})}{4(t-t_0)}\right), \tag{4.13c}$$

where $\mathbb{H}^n, \mathbb{S}^n$ denote the n dimensional hyperbolic manifold and the sphere, respectively. The symbol $\sim$ in (4.13) denotes asymptotic equivalence.

In contrast, the form (4.11) that we focus on is rather specific, and is for a flat geometry. It is then natural to speculate that dist_{tt_0} may arise from the finite horizon reachability constraint over $[t_0, t]$ imposed by the controlled dynamics, i.e., dist_{tt_0} is of *sub-Riemannian* or *Carnot-Carathéodory* type [130–132]. This motivates us to formulate dist_{tt_0} as the minimal value of an action integral explained in the following section.

4.3 Distances from Optimal Control

We postulate that for k of the form (4.11), the dist_{tt_0} is induced by a *deterministic optimal control problem (OCP)* of particular structure. The objective for this OCP is

$$\frac{1}{2}\mathrm{dist}^2_{tt_0}(\boldsymbol{x},\boldsymbol{y}) = \min_{\boldsymbol{u}_\tau} \int_{t_0}^{t} \left(\frac{1}{2}\|\boldsymbol{u}(\tau)\|_2^2 + q(\boldsymbol{z}(\tau))\right)\mathrm{d}\tau. \tag{4.14}$$

The constraint for this OCP is a *controlled ODE* obtained by replacing $\mathrm{d}\boldsymbol{w}(\tau)$ in the underlying Itô diffusion with a controlled drift $\boldsymbol{u}(\tau)\,\mathrm{d}\tau$, and boundary conditions $\boldsymbol{z}(\tau = t_0) = \boldsymbol{x}$, $\boldsymbol{z}(\tau = t) = \boldsymbol{y}$. Notice that in objective (4.14), the state cost q is the rate of creation or killing of probability mass. Let us verify this postulate for known cases discussed before.

Heat kernel (4.6). Here $q \equiv 0$, and $\mathrm{d}\boldsymbol{w}(\tau) \mapsto \boldsymbol{u}(\tau)\,\mathrm{d}\tau$ in (4.7) gives the controlled ODE $\dot{\boldsymbol{z}}(\tau) = \sqrt{2}\,\boldsymbol{u}(\tau)$ where dot denotes derivative w.r.t. $\tau \in [t_0, t]$. This leads to the deterministic OCP

$$\frac{1}{2}\mathrm{dist}^2_{tt_0}(\boldsymbol{x},\boldsymbol{y}) = \min_{\boldsymbol{u}_\tau} \int_{t_0}^{t} \frac{1}{2}\|\boldsymbol{u}(\tau)\|_2^2 \mathrm{d}\tau \tag{4.15a}$$

$$\dot{\boldsymbol{z}}(\tau) = \sqrt{2}\,\boldsymbol{u}(\tau), \tag{4.15b}$$

$$\boldsymbol{z}(\tau = t_0) = \boldsymbol{x}, \quad \boldsymbol{z}(\tau = t) = \boldsymbol{y}. \tag{4.15c}$$

For solving (4.15), we apply the Pontryagin's minimum principle to get the optimal control $\boldsymbol{u}_{\text{opt}}(\tau) = -\sqrt{2}\boldsymbol{\lambda}_{\text{opt}}(\tau)$ where $\boldsymbol{\lambda}_{\text{opt}}(\tau)$ is the optimal costate. The optimal state $\ddot{\boldsymbol{z}}^{\text{opt}}(\tau) = 0$, or $\dot{\boldsymbol{z}}_{\text{opt}}(\tau) = -2\boldsymbol{\lambda}_{\text{opt}}(\tau) = \boldsymbol{\alpha}$, or equivalently $\boldsymbol{z}_{\text{opt}}(\tau) = \boldsymbol{\alpha}\tau + \boldsymbol{\beta}$ for some constant $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$.

Using (4.15c), we find

$$\boldsymbol{\alpha} = \frac{\boldsymbol{x} - \boldsymbol{y}}{t_0 - t},$$

and

$$\boldsymbol{\lambda}_{\text{opt}} = -\frac{\boldsymbol{\alpha}}{2} = \frac{\boldsymbol{x} - \boldsymbol{y}}{2(t - t_0)}.$$

Hence the optimal value in (4.15a):

$$\frac{1}{2}\text{dist}^2_{tt_0}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2} \cdot 2\|\boldsymbol{\lambda}_{\text{opt}}\|_2^2 \cdot (t - t_0) = \frac{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2}{4(t - t_0)},$$

which indeed transcribes (4.6) into the form (4.11). The normalization condition for k leads to evaluating a Gaussian integral, which determines the pre-factor

$$c(t, t_0) = (4\pi(t - t_0))^{-n/2}.$$

Linear kernel (4.8)**.** Here $q \equiv 0$, and $\mathrm{d}\boldsymbol{w} \mapsto \boldsymbol{u}\,\mathrm{d}\tau$ in (4.4) yields the deterministic OCP

$$\frac{1}{2}\text{dist}^2_{tt_0}(\boldsymbol{x}, \boldsymbol{y}) = \min_{\boldsymbol{u}} \int_{t_0}^{t} \frac{1}{2}\|\boldsymbol{u}(\tau)\|_2^2 \mathrm{d}\tau \tag{4.16a}$$

$$\dot{\boldsymbol{z}}(\tau) = \boldsymbol{A}(\tau)\boldsymbol{z}(\tau) + \sqrt{2}\,\boldsymbol{B}(\tau)\boldsymbol{u}(\tau), \tag{4.16b}$$

$$\boldsymbol{z}(\tau = t_0) = \boldsymbol{x}, \quad \boldsymbol{z}(\tau = t) = \boldsymbol{y}. \tag{4.16c}$$

Problem (4.16) is that of minimum effort state steering for a controllable linear system, and its solution is commonplace in optimal control textbooks; see e.g., [85, p.

194]. The optimal value in (4.16a):

$$\frac{1}{2}\mathrm{dist}^2_{tt_0}(\boldsymbol{x},\boldsymbol{y}) = \frac{\left(\boldsymbol{\Phi}_{tt_0}\boldsymbol{x}-\boldsymbol{y}\right)^\top \boldsymbol{\Gamma}^{-1}_{tt_0}\left(\boldsymbol{\Phi}_{tt_0}\boldsymbol{x}-\boldsymbol{y}\right)}{4\left(t-t_0\right)},$$

which indeed transcribes (4.8) into the form (4.11). As before, the normalization condition for k determines the pre-factor

$$c(t,t_0) = \left(4\pi\left(t-t_0\right)\right)^{-n/2}\det\left(\boldsymbol{\Gamma}_{tt_0}\right)^{-1/2}$$

via Gaussian integral.

The kernel in Chapter 3. The kernel derived in Chapter 3, (3.28), corresponds to the Itô diffusion (3.2) with reaction rate $\frac{1}{2}\boldsymbol{z}^\top\boldsymbol{Q}\boldsymbol{z}$, $\boldsymbol{Q}\succeq\boldsymbol{0}$. Without loss of generality [74, Sec. 4.1], we consider state coordinates where a given $\boldsymbol{Q}\succ\boldsymbol{0}$ is diagonalized to yield the diagonal matrix $2\boldsymbol{\Lambda}\succ\boldsymbol{0}$. The factor 2 scaling is unimportant but kept for consistency with Chapter 3. Importantly, the Laplacian operator is invariant under this change of coordinates. For the case $\boldsymbol{Q}\succeq\boldsymbol{0}$, see Chapter 3, Section 3.2.

In the new state coordinates, $q\equiv\frac{1}{2}\boldsymbol{z}_\tau^\top 2\boldsymbol{\Lambda}\boldsymbol{z}_\tau$, and $\mathrm{d}\boldsymbol{w}_\tau\mapsto\boldsymbol{u}\,\mathrm{d}\tau$ in (3.2) gives the following modified version of the OCP (4.15):

$$\frac{1}{2}\mathrm{dist}^2_{tt_0}(\boldsymbol{x},\boldsymbol{y}) = \min_{\boldsymbol{u}}\int_{t_0}^{t}\left(\frac{1}{2}\|\boldsymbol{u}(\tau)\|_2^2+\frac{1}{2}\boldsymbol{z}^\top(\tau)2\boldsymbol{\Lambda}\boldsymbol{z}(\tau)\right)\mathrm{d}\tau \tag{4.17a}$$

$$\dot{\boldsymbol{z}}(\tau) = \sqrt{2}\,\boldsymbol{u}(\tau), \tag{4.17b}$$

$$\boldsymbol{z}(\tau=t_0)=\boldsymbol{x},\quad \boldsymbol{z}(\tau=t)=\boldsymbol{y}. \tag{4.17c}$$

Applying Pontryagin's minimum principle, we get the optimal control

$$\boldsymbol{u}_{\mathrm{opt}}(\tau) = -\sqrt{2}\boldsymbol{\lambda}_{\mathrm{opt}}(\tau),$$

the optimal costate ODE

$$\dot{\boldsymbol{\lambda}}_{\mathrm{opt}}(\tau) = -2\boldsymbol{\Lambda}\boldsymbol{z}_{\mathrm{opt}}(\tau),$$

and the optimal state ODE

$$\ddot{\boldsymbol{z}}_{\text{opt}}(\tau) = -2\dot{\boldsymbol{\lambda}}_{\text{opt}}(\tau) = 4\boldsymbol{\Lambda}\boldsymbol{z}_{\text{opt}}(\tau). \tag{4.18}$$

Letting $\omega_i := 2\sqrt{\boldsymbol{\Lambda}_{ii}}$ for all $i = 1, \cdots, n$, we solve for the components of (4.18) as

$$\left(z_{\text{opt}}(\tau)\right)_i = a_i e^{\omega_i \tau} + b_i e^{-\omega_i \tau}, \tag{4.19}$$

and use (4.17c) to determine the constants

$$a_i = \frac{-x_i e^{-\omega_i t} + y_i e^{-\omega_i t_0}}{2\sinh\left(\omega_i\left(t - t_0\right)\right)}, b_i = \frac{x_i e^{\omega_i t} - y_i e^{\omega_i t_0}}{2\sinh\left(\omega_i\left(t - t_0\right)\right)}. \tag{4.20}$$

Hence the optimal value in (4.17a):

$$\begin{aligned}\frac{1}{2}\text{dist}^2_{tt_0}(\boldsymbol{x}, \boldsymbol{y}) &= \int_{t_0}^{t}\left(\frac{1}{2}\|\boldsymbol{u}_{\text{opt}}(\tau)\|_2^2 + \frac{1}{4}\sum_{i=1}^{n}\omega_i^2\left(\left(z_{\text{opt}}(\tau)\right)_i\right)^2\right)\mathrm{d}\tau\\ &= \int_{t_0}^{t}\left(\frac{1}{4}\|\boldsymbol{z}_{\text{opt}}(\tau)\|_2^2 + \frac{1}{4}\sum_{i=1}^{n}\omega_i^2\left(\left(z_{\text{opt}}(\tau)\right)_i\right)^2\right)\mathrm{d}\tau\\ &= \frac{1}{2}\sum_{i=1}^{n}\frac{\omega_i\left(x_i^2 + y_i^2\right)\cosh\left(\omega_i(t - t_0)\right) - 2\omega_i x_i y_i}{2\sinh\left(\omega_i(t - t_0)\right)},\end{aligned} \tag{4.21}$$

where we have used (4.19)-(4.20) followed by algebraic simplification using the identity $\sinh(2(\cdot)) = 2\sinh(\cdot)\cosh(\cdot)$.

We note that the expression (4.21) indeed transcribes the Markov kernel in (3.28) in Chapter 3 into the form (4.11). Since this Markov kernel is not a transition probability, the normalization condition no longer holds, and the pre-factor

$$c(t, t_0) = \prod_{i=1}^{n}\left(\frac{\omega_i}{4\pi\sinh\left(\omega_i(t - t_0)\right)}\right)^{1/2} \tag{4.22}$$

does not follow from there. However, having determined (4.21), the pre-factor (4.22) can be obtained by substituting (4.11) with (4.21) in (4.10). See Appendix D.1 for details of this simple but non-trivial computation.

4.4 Distance and Kernel for LQ SBP

Having seen three exemplars for the computational pipeline:

$$\text{deterministic OCP} \longrightarrow \text{dist} \longrightarrow \text{Markov kernel } k,$$

we now apply the same approach to derive k for the Itô diffusion (4.4) with reaction rate $q(\boldsymbol{z}) \equiv \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{Q}(\tau)\boldsymbol{z}$, $\tau \in [t_0, t]$. For convenience, let $\hat{\boldsymbol{B}}(\tau) := \sqrt{2}\,\boldsymbol{B}(\tau)$.

Our standing assumptions are:

A1. $(\boldsymbol{A}(\tau), \boldsymbol{B}(\tau))$, and thus $\left(\boldsymbol{A}(\tau), \hat{\boldsymbol{B}}_\tau\right)$, is controllable matrix-valued trajectory pair in $(\mathbb{R}^{n\times n}, \mathbb{R}^{n\times m})$ that is bounded and continuous in $\tau \in [t_0, t]$.

A2. the matricial trajectory $\boldsymbol{Q}(\tau) \succeq \boldsymbol{0}$ is continuous and bounded w.r.t. $\tau \in [t_0, t]$, and $\boldsymbol{Q}(s) \succ \boldsymbol{0}$ for some $s \in [t_0, t]$.

Finding the Markov kernel k in this setting enables the solution of the LQ SBP with generic non-Gaussian endpoint distributions having finite second moments, i.e., the solution of the problem (4.2). For ρ_0, ρ_1 with finite second moments, the existence-uniqueness for the solution of (4.2) is guaranteed. This follows from transcribing (4.2) to a stochastic calculus of variations problem involving the relative entropy a.k.a. the Kullback-Leibler divergence $\mathrm{D_{KL}}\left(\cdot \parallel \cdot\right)$ minimization:

$$\min_{\mathbb{P}\in\Pi_{01}} \quad \mathrm{D_{KL}}\left(\mathbb{P} \parallel \frac{\exp\left(-\frac{1}{4}\int_{t_0}^{t_1} \boldsymbol{x}^\top \boldsymbol{Q}(t)\boldsymbol{x}\,\mathrm{d}t\right)\mathbb{W}}{Z}\right) \tag{4.23}$$

where

$$\Pi_{01} := \left\{\mathbb{M} \in \mathcal{M}\left(\mathcal{C}\left([t_0, t_1]\,; \mathbb{R}^n\right)\right) \mid \mathbb{M} \text{ has marginal } \mu_k \text{ at time } t_k \;\forall k \in \{0, 1\}\right\}, \tag{4.24}$$

where $\mathcal{M}\left(\mathcal{C}\left([t_0, t_1]\,; \mathbb{R}^n\right)\right)$ is the set of probability measures on the path space $\mathcal{C}\left([t_0, t_1]\,; \mathbb{R}^n\right)$ generated by the Itô diffusion (4.4), the $\mathbb{W} \in \mathcal{M}\left(\mathcal{C}\left([t_0, t_1]\,; \mathbb{R}^n\right)\right)$ is the Wiener measure, and the Z is a normalizing constant. Problem (4.23) seeks to find the most likely measure-valued path generated by (4.4) w.r.t. a weighted Wiener (i.e., Gibbs) measure defined by the quadratic state cost, with endpoint marginal constraints given by the problem data μ_0, μ_1. For the equivalence of the formulations (4.2) and (4.23), see e.g., [10, 39, 133], [77, Sec. 4.1]. The existence-uniqueness of solution for (4.23) is a consequence of strict convexity of $\mathrm{D_{KL}}\left(\cdot \parallel \cdot\right)$ w.r.t. the first argument.

Unlike the three exemplar kernels discussed above, the k for problem (4.2) is not known. Moreover, previously mentioned approaches (Hermite polynomials, Weyl calculus) to compute k that works for simpler cases no longer generalize in any obvious way. We postulate that the form (4.11) for the Markov kernel holds here as well.

4.5 Getting dist

Inspired by the instances examined above, we define dist through the following deterministic OCP:

$$\frac{1}{2}\mathrm{dist}^2_{tt_0}(\boldsymbol{x},\boldsymbol{y}) = \min_{\boldsymbol{u}_\tau} \int_{t_0}^{t} \left(\frac{1}{2}\|\boldsymbol{u}(\tau)\|_2^2 + \frac{1}{2}(\boldsymbol{z}(\tau))^\top \boldsymbol{Q}(\tau)\boldsymbol{z}(\tau)\right)\mathrm{d}\tau \tag{4.25a}$$

$$\dot{\boldsymbol{z}}(\tau) = \boldsymbol{A}(\tau)\boldsymbol{z}(\tau) + \sqrt{2}\,\boldsymbol{B}(\tau)\boldsymbol{u}(\tau), \tag{4.25b}$$

$$\boldsymbol{z}(\tau = t_0) = \boldsymbol{x}, \quad \boldsymbol{z}(\tau = t) = \boldsymbol{y}. \tag{4.25c}$$

This is an atypical linear quadratic OCP in that when $\boldsymbol{x}, \boldsymbol{y}$ are not too close to the origin, the optimal state trajectory trades off the soft penalty (state cost-to-go) in deviating away from origin with that of meeting the hard endpoint constraints (4.25c) within the given deadline.

We need the following result from [134, p. 140-141] paraphrased below to suit our notations.

Proposition 4.1. *[134, p. 140-141] Suppose there exists symmetric matrix $\boldsymbol{K}_1$ such that the solution map*[3] $\boldsymbol{\Pi}(\tau, \boldsymbol{K}_1, t)$ *of the Riccati matrix ODE initial value problem*

$$\dot{\boldsymbol{K}}(\tau) = -\boldsymbol{A}^{\top}(\tau)\boldsymbol{K}(\tau) - \boldsymbol{K}(\tau)\boldsymbol{A}(\tau) + \boldsymbol{K}(\tau)\hat{\boldsymbol{B}}(\tau)\hat{\boldsymbol{B}}^{\top}(\tau)\boldsymbol{K}(\tau) - \boldsymbol{Q}(\tau), \tag{4.26a}$$

$$\boldsymbol{K}(\tau = t) = \boldsymbol{K}_1, \tag{4.26b}$$

exists for all $\tau \in [t_0, t]$. *Let*

$$\hat{\boldsymbol{A}}(\tau) \coloneqq \boldsymbol{A}(\tau) - \hat{\boldsymbol{B}}(\tau)\hat{\boldsymbol{B}}^{\top}(\tau)\boldsymbol{\Pi}(\tau, \boldsymbol{K}_1, t) \quad \forall \tau \in [t_0, t]. \tag{4.27}$$

Then the differentiable trajectory $\boldsymbol{z}(\tau)$ *minimizes*

$$\begin{aligned} \eta = \int_{t_0}^{t} \left(\frac{1}{2}\|\boldsymbol{u}(\tau)\|_2^2 + \frac{1}{2}\boldsymbol{z}^{\top}(\tau)\boldsymbol{Q}(\tau)\boldsymbol{z}(\tau)\right)\mathrm{d}\tau \\ \textit{subject to} \quad \dot{\boldsymbol{z}}(\tau) = \boldsymbol{A}(\tau)\boldsymbol{z}(\tau) + \hat{\boldsymbol{B}}(\tau)\boldsymbol{u}(\tau), \\ \boldsymbol{z}(\tau = t_0) = \boldsymbol{x}, \quad \boldsymbol{z}(\tau = t) = \boldsymbol{y}, \end{aligned}$$

if and only if it also minimizes

$$\begin{aligned} \hat{\eta} = \int_{t_0}^{t} \frac{1}{2}\|\boldsymbol{v}(\tau)\|_2^2 \mathrm{d}\tau \\ \textit{subject to} \quad \dot{\boldsymbol{z}}(\tau) = \hat{\boldsymbol{A}}(\tau)\boldsymbol{z}(\tau) + \hat{\boldsymbol{B}}(\tau)\boldsymbol{v}(\tau), \\ \boldsymbol{z}(\tau = t_0) = \boldsymbol{x}, \quad \boldsymbol{z}(\tau = t) = \boldsymbol{y}. \end{aligned}$$

Additionally, along any trajectory satisfying the boundary conditions, we have

$$\eta = \hat{\eta} + \frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{\Pi}(t_0, \boldsymbol{K}_1, t)\boldsymbol{x} - \frac{1}{2}\boldsymbol{y}^{\top}\boldsymbol{K}_1\boldsymbol{y}. \tag{4.28}$$

Remark 4.1 (Existence, uniqueness and positive semi-definiteness of $\boldsymbol{\Pi}$)**.** *Proposi-*

[3] The mapping $\boldsymbol{\Pi}(\tau, \boldsymbol{K}_1, t)$ is understood as the solution of (4.26a) at any $\tau \in [t_0, t]$ solved backward in time with initial condition (4.26b).

tion 4.1 does not assume the controllability of (4.25b). *However, under our standing controllability assumption* $\boldsymbol{A1}$ *and the positive semi-definiteness assumption for* $\boldsymbol{Q}(\tau)$ $\forall \tau \in [t_0, t]$ *in* $\boldsymbol{A2}$*, the existence-uniqueness of the mapping* $\boldsymbol{\Pi}$ *is guaranteed for all* $\tau \in [t_0, t]$. *Furthermore, the unique solution* $\boldsymbol{\Pi}(\tau, \boldsymbol{K}_1, t) \succeq \boldsymbol{0}$ *for any* $\boldsymbol{K}_1 \succeq \boldsymbol{0}$. *See e.g., [135, Thm. 8-10].*

Remark 4.2 (Strict positive definiteness of $\boldsymbol{\Pi}$)**.** *The assumption* $\boldsymbol{Q}(s) \succ \boldsymbol{0}$ *for some* $s \in [t_0, t]$*, stated in the latter part of* $\boldsymbol{A2}$*, further ensures that* $\boldsymbol{\Pi}(\tau, \boldsymbol{K}_1, t) \succ \boldsymbol{0}$ *for any* $\boldsymbol{K}_1 \succeq \boldsymbol{0}$. *This is a consequence of the matrix variation of constants formula [134, p. 59, exercise 1 in p. 162], [136, Sec. II].*

Since the optimal $\hat{\eta}$ in Proposition 4.1 is the cost for minimum energy state transfer from $\boldsymbol{x}$ at t_0 to $\boldsymbol{y}$ at t over the LTV system matrix pair $(\hat{\boldsymbol{A}}(\tau), \hat{\boldsymbol{B}}(\tau))$, we have

$$\hat{\eta}_{\text{opt}} = \frac{1}{2}\left(\hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x} - \boldsymbol{y}\right)^{\top} \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \left(\hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x} - \boldsymbol{y}\right), \tag{4.29}$$

where $\hat{\boldsymbol{\Phi}}_{tt_0}$ is the state transition matrix from t_0 to t for the state matrix $\hat{\boldsymbol{A}}(\tau)$ in (4.27). Likewise, $\hat{\boldsymbol{\Gamma}}_{tt_0}$ in (4.29) is the controllability Gramian for the pair $(\hat{\boldsymbol{A}}(\tau), \hat{\boldsymbol{B}}(\tau))$.

To see why (4.29) is well-defined, recall that for linear systems, controllability remains invariant under state feedback. Specifically, the following Proposition 4.2 holds. For its proof in the LTV setting, see [137]; the proof in the linear time-invariant setting appeared earlier in [138, Thm. 3] .

Proposition 4.2. *[137] Let* $\hat{\boldsymbol{A}}(\tau)$ *be given by* (4.27). *If the LTV system defined by the pair* $(\boldsymbol{A}(\tau), \hat{\boldsymbol{B}}(\tau))$ *is controllable, then so is the LTV system defined by the pair* $(\hat{\boldsymbol{A}}(\tau), \hat{\boldsymbol{B}}(\tau))$.

From Proposition 4.2, it follows that $\boldsymbol{\Gamma}_{tt_0} \succ \boldsymbol{0}$ implies $\hat{\boldsymbol{\Gamma}}_{tt_0} \succ \boldsymbol{0}$. In particular, $\hat{\boldsymbol{\Gamma}}_{tt_0}$ is non-singular, and (4.29) is well-defined.

Since the optimal η in (4.28), and thus the optimal value in (4.25a) must be independent of $\boldsymbol{K}_1$, we can set $\boldsymbol{K}_1 \equiv \boldsymbol{0}$ without loss of generality. Combining this observation with (4.29), we find the optimal value in (4.25a) as

$$\frac{1}{2}\mathrm{dist}^2_{tt_0}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^{\top} \boldsymbol{M}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}, \tag{4.30}$$

where

$$\boldsymbol{M}_{tt_0} := \begin{bmatrix}\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0} + \boldsymbol{\Pi}(t_0,\boldsymbol{0},t) & -\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\\ \\ -\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0} & \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\end{bmatrix}. \tag{4.31}$$

Using the Schur complement lemma, we can check that $\boldsymbol{M}_{tt_0} \succ \boldsymbol{0}$, as expected. Specifically, from (4.31), note that

$$\begin{aligned}&\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \succ \boldsymbol{0},\\ &\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0} + \boldsymbol{\Pi}(t_0,\boldsymbol{0},t) - \left(-\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\right)\hat{\boldsymbol{\Gamma}}_{tt_0}\left(-\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}\right) = \boldsymbol{\Pi}(t_0,\boldsymbol{0},t) \succ \boldsymbol{0},\end{aligned} \tag{4.32a}$$

thanks to Remark 4.2.

The formulae (4.30)-(4.31) generalize the previously known results in Chapter 3 for zero prior dynamics. In particular, the previously known result can be recovered from (4.30)-(4.31) as follows (proof in Appendix D.2).

Proposition 4.3. *When* $\boldsymbol{Q}(\tau) = 2\boldsymbol{\Lambda} \succ \boldsymbol{0}$ *(constant positive diagonal matrix) and* $(\boldsymbol{A}(\tau), \hat{\boldsymbol{B}}(\tau)) \equiv (\boldsymbol{0}, \sqrt{2}\boldsymbol{I})\ \forall \tau \in [t_0, t]$*, the formulas* (4.30)-(4.31) *reduce to* (4.21).

4.6 Getting k

Now that we have determined the distance functional in the generic LQ setting given by (4.30)-(4.31), to find k in the postulated form (4.11), i.e., in the form

$$k\left(t_0, \boldsymbol{x}, t, \boldsymbol{y}\right) = c(t, t_0) \exp\left(-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^{\top} \boldsymbol{M}_{tt_0} \begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\right), \tag{4.33}$$

we must compute the pre-factor $c(t, t_0)$. Unique identification of $c(t, t_0)$ serves the dual purpose of uniquely determining the kernel as well as verifying the postulated form (4.11).

As was the case in the last example in Sec. 4.3, here k is not a transition probability kernel and $\int k \neq 1$. Thus, the pre-factor $c(t, t_0)$ cannot be determined from normalization.

Instead, we find $c(t, t_0)$ by substituting (4.11) with (4.30) in (4.10) and invoking the initial condition. In this case,

$$Lk \equiv -\langle \nabla_{\boldsymbol{x}}, k\boldsymbol{A}(t)\boldsymbol{x}\rangle + \boldsymbol{B}(t)\boldsymbol{B}(t)^{\top}\Delta_{\boldsymbol{x}} k,$$

so L is an advection-diffusion operator. Recall that the reaction rate $q(\boldsymbol{x}) \equiv \frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{Q}(\tau)\boldsymbol{x}$ with $\boldsymbol{Q}(\tau)$ satisfying Assumption **A2**. Therefore, (4.10) becomes an *reaction-advection-diffusion PDE initial value problem*

$$\frac{\partial k}{\partial t} = -\langle \nabla_{\boldsymbol{x}}, k\boldsymbol{A}(t)\boldsymbol{x}\rangle + \langle \boldsymbol{B}(t)\boldsymbol{B}(t)^{\top}, \nabla^2_{\boldsymbol{x}} k\rangle - \frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{Q}(t)\boldsymbol{x}k, \tag{4.34a}$$

$$k(t_0, \boldsymbol{x}, t_0, \boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y}), \tag{4.34b}$$

where $\nabla^2_{\boldsymbol{x}}$ denotes the Hessian w.r.t. $\boldsymbol{x}$.

The main result presented in this chapter is the following theorem.

Theorem 4.1 (Kernel for LQ non-Gaussian steering). *For $\tau \in [t_0, t]$, consider as-*

sumptions ***A1-A2****. Let*

$$\theta(\tau) := \operatorname{trace} \boldsymbol{A}(\tau) + \langle \boldsymbol{B}(\tau)\boldsymbol{B}(\tau)^{\top}, \hat{\boldsymbol{\Phi}}_{\tau t_0}^{\top}\hat{\boldsymbol{\Gamma}}_{\tau t_0}^{-1}\hat{\boldsymbol{\Phi}}_{\tau t_0} + \boldsymbol{\Pi}(t_0, \boldsymbol{0}, \tau)\rangle, \tag{4.35a}$$

$$= \operatorname{trace}\left(\boldsymbol{A}(\tau) + \boldsymbol{B}(\tau)\boldsymbol{B}(\tau)^{\top}\boldsymbol{M}_{11}(\tau, t_0)\right), \tag{4.35b}$$

$$a := (2\pi)^{-n/2} \lim_{t \downarrow t_0} \det\Big(\boldsymbol{M}_{11}^{1/2}(t, t_0) \times \exp \int_{t_0}^{t} \left(\boldsymbol{A}(\tau) + \boldsymbol{B}(\tau)\boldsymbol{B}(\tau)^{\top}\boldsymbol{M}_{11}(\tau, t_0)\right) \mathrm{d}\tau\Big), \tag{4.36}$$

where $\exp$ *denotes the matrix exponential,* $\hat{\boldsymbol{\Phi}}_{\tau t_0}$ *is the state transition matrix for* $\hat{\boldsymbol{A}}(\tau)$ *given by* (4.27)*,* $\hat{\boldsymbol{\Gamma}}_{\tau t_0}$ *is the controllability Gramian for the pair* $\big(\hat{\boldsymbol{A}}(\tau), \hat{\boldsymbol{B}}(\tau)\big)$*, and* $\boldsymbol{\Pi}(t_0, \boldsymbol{0}, \tau)$ *is as in Sec. 4.5. The matrix* $\boldsymbol{M}_{11}(\tau, t_0)$ *appearing in* (4.35b) *and* (4.36) *is the* $(1,1)$ *block of* $\boldsymbol{M}_{\tau t_0}$ *given by* (4.31)*. Then the Markov kernel* k *solving the reaction-advection-diffusion PDE initial value problem* (4.34) *is*

$$k\left(t_0, \boldsymbol{x}, t, \boldsymbol{y}\right) = a \exp\left(-\int_{t_0}^{t} \theta(s)\mathrm{d}s\right) \times \exp\left(-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^{\top} \boldsymbol{M}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\right), \tag{4.37}$$

where $\boldsymbol{M}_{tt_0} \succ \boldsymbol{0}$ *is given by* (4.31)*.*

See Appendix D.3 for a proof of Theorem 4.1.

The kernel (4.37) significantly extends the results presented in Chapter 3 to the generic LQ case. The following corollary recovers that special case from Chapter 3.

Corollary 4.2. *[*(3.28) *as special case of* (4.37)*] When* $\boldsymbol{Q}(\tau) = 2\boldsymbol{D} \succ \boldsymbol{0}$ *(constant positive diagonal matrix) and* $(\boldsymbol{A}(\tau), \hat{\boldsymbol{B}}(\tau)) \equiv (\boldsymbol{0}, \sqrt{2}\boldsymbol{I})$ $\forall \tau \in [t_0, t]$*, the kernel* (4.37) *reduces to that in [75, Eq. (43)] or that in [74, Eq. (A.22)].*

See Appendix D.4 for the proof of Corollary 4.2.

Remark 4.3. *Using assumptions* ***A1-A2*** *and the standard* (ε, δ) *definition of limit, it is not too difficult to show that the upper limit in* (4.36) *exists. In special cases such* *as Corollary 4.2 above, the limit in* (4.36) *can be evaluated in closed form, as shown in Appendix D.4. Importantly, the limit in* (4.36) *cannot be distributed to the determinant and the exponential factors; see the computation in* (D.43).

The following properties are immediate from the expression of the kernel (4.37):

- (spatial symmetry) $k\left(t_0, \boldsymbol{x}, t, \boldsymbol{y}\right) = k\left(t_0, \boldsymbol{y}, t, \boldsymbol{x}\right)\ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,
- (positivity) $k\left(t_0, \boldsymbol{x}, t, \boldsymbol{y}\right) > 0$ since from (4.36), $a > 0$.

We next discuss how the derived kernel helps in solving the LQ SBP (4.2).

4.7 Using k to Solve LQ SBP with Non-Gaussian μ_0, μ_1

Using the derived kernel (4.37), we define the integral transforms

$$\widehat{\varphi}_0 \mapsto \mathcal{T}_{01}\widehat{\varphi}_0 := \int_{\mathbb{R}^n} k\left(t_0, \boldsymbol{x}, t_1, \boldsymbol{y}\right) \widehat{\varphi}_0(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}, \tag{4.38a}$$

$$\varphi_1 \mapsto \mathcal{T}_{10}\varphi_1 := \int_{\mathbb{R}^n} k\left(t_1, \boldsymbol{x}, t_0, \boldsymbol{y}\right) \varphi_1(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}, \tag{4.38b}$$

for measurable $\widehat{\varphi}_0(\cdot), \varphi_1(\cdot)$. These integral transforms help solve the LQ SBP (4.2) as follows.

Suppose that the endpoint measures are absolutely continuous with respective probability density functions (PDFs) ρ_0, ρ_1. Standard computation shows that the necessary conditions for optimality for (4.2) yield a pair of second-order coupled nonlinear PDEs:

$$\frac{\partial \rho_{\text{opt}}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\rho_{\text{opt}}\left(\boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{B}(t)\boldsymbol{B}^\top(t)\nabla_{\boldsymbol{x}} S\right)\right) = \langle \boldsymbol{B}(t)\boldsymbol{B}^\top(t), \nabla_{\boldsymbol{x}}^2 \rho^{\text{opt}} \rangle, \tag{4.39a}$$

$$\frac{\partial S}{\partial t} + \langle \nabla_{\boldsymbol{x}} S, \boldsymbol{A}(t)\boldsymbol{x} \rangle + \frac{1}{2}\langle \nabla_{\boldsymbol{x}} S, \boldsymbol{B}(t)\boldsymbol{B}^\top(t)\nabla_{\boldsymbol{x}} S \rangle + \left\langle \boldsymbol{B}(t)\boldsymbol{B}(t)^\top, \nabla_{\boldsymbol{x}}^2 S \right\rangle = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}(t)\boldsymbol{x}, \tag{4.39b}$$

$$\rho_{\text{opt}}\left(t_0, \boldsymbol{x}\right) = \rho_0(\boldsymbol{x}), \quad \rho_{\text{opt}}\left(t_1, \boldsymbol{x}\right) = \rho_1(\boldsymbol{x}), \tag{4.39c}$$

in unknown primal-dual pair $(\rho_{\text{opt}}(t, \boldsymbol{x}), S(t, \boldsymbol{x}))$, i.e., the optimally controlled joint state PDF and the value function.

Following our usual approach, we apply the Hopf-Cole transform [18,19]. In the absence of ε, we can apply the following, more simplified version of this transform:

$$(\rho_{\text{opt}}, S) \mapsto (\widehat{\varphi}, \varphi) := (\rho_{\text{opt}} \exp(-S), \exp S). \tag{4.40}$$

We thus recast (4.39) to a pair of boundary-coupled linear PDEs:

$$\frac{\partial \widehat{\varphi}}{\partial t} = -\langle \nabla_{\boldsymbol{x}}, \widehat{\varphi} \boldsymbol{A}(t)\boldsymbol{x}\rangle + \langle \boldsymbol{B}(t)\boldsymbol{B}(t)^{\top}, \nabla_{\boldsymbol{x}}^2 \widehat{\varphi}\rangle - \frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{Q}(t)\boldsymbol{x}\widehat{\varphi}, \tag{4.41a}$$

$$\frac{\partial \varphi}{\partial t} = -\langle \nabla_{\boldsymbol{x}}\varphi, \boldsymbol{A}(t)\boldsymbol{x}\rangle - \langle \boldsymbol{B}(t)\boldsymbol{B}(t)^{\top}, \nabla_{\boldsymbol{x}}^2 \varphi\rangle + \frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{Q}(t)\boldsymbol{x}\varphi, \tag{4.41b}$$

$$\widehat{\varphi}(t_0, \boldsymbol{x})\varphi(t_0, \boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \widehat{\varphi}(t_1, \boldsymbol{x})\varphi(t_1, \boldsymbol{x}) = \rho_1(\boldsymbol{x}). \tag{4.41c}$$

We note that the PDE (4.41a) is precisely the forward reaction-advection-diffusion PDE (4.34a), and the PDE (4.41b) is the associated backward reaction-advection-diffusion PDE.

The solution of (4.41) recovers the solution of (4.2). Specifically, $\forall t \in [t_0, t_1]$, the optimally controlled joint state PDF $\rho^{\text{opt}}(t, \boldsymbol{x}) = \widehat{\varphi}(t, \boldsymbol{x})\varphi(t, \boldsymbol{x})$, and the optimal control $\boldsymbol{u}_{\text{opt}}(t, \boldsymbol{x}) = \boldsymbol{B}(t)^{\top}\nabla_{\boldsymbol{x}} S(t, \boldsymbol{x}) = \boldsymbol{B}(t)^{\top}\nabla_{\boldsymbol{x}} \log \varphi(t, \boldsymbol{x})$.

For $k \in \{0, 1\}$, letting $\widehat{\varphi}_k(\cdot) := \widehat{\varphi}(t_k, \cdot)$, $\varphi_k(\cdot) := \varphi(t_k, \cdot)$, the system (4.41) can be solved by the usual dynamic Sinkhorn recursion (first described in Chapter 1 and illustrated in Figure 1.1):

$$\widehat{\varphi}_0 \xrightarrow{\mathcal{T}_{01}} \widehat{\varphi}_1 \xrightarrow{\rho_1/\widehat{\varphi}_1} \varphi_1 \xrightarrow{\mathcal{T}_{10}} \varphi_0 \xrightarrow{\rho_0/\varphi_0} (\widehat{\varphi}_0)_{\text{next}}, \tag{4.42}$$

involving the integral transforms (4.38).

From the computational point of view, having a handle for the kernel (4.37) helps us to apply $\mathcal{T}_{01}, \mathcal{T}_{10}$ in (4.38) during each pass of the recursion (4.42). This in turn helps to solve the LQ SBP (4.2) with arbitrary *non-Gaussian* ρ_0, ρ_1. The solution for (4.2) when both ρ_0, ρ_1 are *Gaussians*, appeared in [117], but did not need require the kernel. However, the techniques and results in [117] are difficult to generalize for non-Gaussian ρ_0, ρ_1.

If the kernel k in (4.38) were not available, an alternative way to implement $\mathcal{T}_{01}, \mathcal{T}_{10}$ in (4.42) is to apply the Feynman-Kac path integral [24, Ch. 8.2], [139, Ch. 3.3] resulting in randomized numerical approximations for $\widehat{\varphi}_1, \varphi_0$. For instance, the Feynman-Kac path integral representation for (4.38b) is

$$\mathcal{T}_{10}\varphi_1(\boldsymbol{x}) = \mathbb{E}\left[\varphi_1\left(\boldsymbol{x}_{t_1}\right)\exp\left(-\int_t^{t_1}\frac{1}{2}\boldsymbol{x}_\tau^\top \boldsymbol{Q}(\tau)\boldsymbol{x}_\tau \mathrm{d}\tau\right) \mid \boldsymbol{x} = \boldsymbol{x}\right], \tag{4.43}$$

and the conditional expectation above can be approximated by Monte Carlo simulation with (4.4). Chapter 5, which considers a case study in which we lack an explicit handle on k, will make use of such methods. Having an explicit handle on k removes the need for such randomized approximations of the deterministic functions $\widehat{\varphi}_1, \varphi_0$.

To further understand the action of the derived kernel, notice that combining (4.37) and (4.38a) yields

$$\begin{aligned}\mathcal{T}_{01}\widehat{\varphi}_0 = a\exp\left(-\int_{t_0}^{t}\theta(s)\mathrm{d}s\right)\exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{M}_{11}\left(t,t_0\right)\boldsymbol{x}\right)\\ \int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{M}_{22}\left(t,t_0\right)\boldsymbol{y} + \langle -\boldsymbol{M}_{12}^\top\left(t,t_0\right)\boldsymbol{x},\boldsymbol{y}\rangle\right)\widehat{\varphi}_0(\boldsymbol{y})\mathrm{d}\boldsymbol{y},\end{aligned} \tag{4.44}$$

and likewise for $\mathcal{T}_{10}\varphi_1$, where $\boldsymbol{M}_{11}, \boldsymbol{M}_{12}, \boldsymbol{M}_{22}$ refer to the respective blocks of (4.31). For suitably smooth $\widehat{\varphi}_0$, the integral in (4.44) can be evaluated using Lemma D.1 in Appendix D.1. We close with an example of such computation.

Example 4.1 (Mixture-of-Gaussian $\widehat{\varphi}_0$). *For fixed $n_c \in \mathbb{N}$, let $\widehat{\varphi}_0$ be an n_c component* *conic mixture-of-Gaussians, i.e.,*

$$\widehat{\varphi}_0(\boldsymbol{y}) = \sum_{i=1}^{n_c} w_i \exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{m}_i\right)^\top \boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{y}-\boldsymbol{m}_i\right)\right),$$

where $w_i > 0$, $\boldsymbol{m}_i \in \mathbb{R}^n$, $\boldsymbol{\Sigma}_i \succ \boldsymbol{0}$ $\forall i \in [n_c]$. *Then* (4.44) *gives*

$$\begin{aligned}
\mathcal{T}_{01}\widehat{\varphi}_0 &= a \exp\left(-\int_{t_0}^{t}\theta(s)\mathrm{d}s\right) \times \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{M}_{11}\left(t,t_0\right)\boldsymbol{x}\right)\\
&\quad \times \sum_{i=1}^{n_c} w_i \exp\left(-\frac{1}{2}\boldsymbol{m}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{m}_i\right)\\
&\quad \int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\left(\boldsymbol{M}_{22}\left(t,t_0\right)+\boldsymbol{\Sigma}_i^{-1}\right)\boldsymbol{y} + \langle -\boldsymbol{M}_{12}^\top\left(t,t_0\right)\boldsymbol{x}+\boldsymbol{\Sigma}_i^{-1}\boldsymbol{m}_i, \boldsymbol{y}\rangle\right)\mathrm{d}\boldsymbol{y}\\
&= a\left(2\pi\right)^{n/2}\exp\left(-\int_{t_0}^{t}\theta(s)\mathrm{d}s\right)\times\exp\left(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{M}_{11}\left(t,t_0\right)\boldsymbol{x}\right)\\
&\quad \times \sum_{i=1}^{n_c}\frac{w_i}{\sqrt{\det\left(\boldsymbol{M}_{22}\left(t,t_0\right)+\boldsymbol{\Sigma}_i^{-1}\right)}}\\
&\quad \times \exp\left\{-\frac{1}{2}\boldsymbol{m}_i^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{m}_i + \frac{1}{2}\left(-\boldsymbol{M}_{12}^\top\left(t,t_0\right)\boldsymbol{x}+\boldsymbol{\Sigma}_i^{-1}\boldsymbol{m}_i\right)^\top\right.\\
&\qquad\qquad \left.\times\left(\boldsymbol{M}_{22}\left(t,t_0\right)+\boldsymbol{\Sigma}_i^{-1}\right)\left(-\boldsymbol{M}_{12}^\top\left(t,t_0\right)\boldsymbol{x}+\boldsymbol{\Sigma}_i^{-1}\boldsymbol{m}_i\right)\right\}, \qquad (4.45)
\end{aligned}$$

where the last equality uses Lemma D.1.

This chapter consists of results generalizing those in Chapter 3, in which the Markov kernel was derived for the special case $(\boldsymbol{A}(t), \boldsymbol{B}(t)) = (\boldsymbol{0}, \boldsymbol{I})$ using both generalized Hermite polynomials and Weyl calculus. As those techniques become unwieldy for generic $(\boldsymbol{A}(t), \boldsymbol{B}(t))$ and $\boldsymbol{Q}(t) \succeq \boldsymbol{0}$, we pursued a new line of attack by postulating the structure of the Markov kernel in terms of a distance function induced by an underlying deterministic optimal control problem. We demonstrated that both new and existing results can be recovered in a conceptually transparent manner even when the underlying Markov kernel is not a transition probability.

The new kernel derived using this approach is the Green's function for a reaction-advection-diffusion PDE. This kernel allows us to solve the generic LQ SBP. While a solution to this problem has appeared in prior literature for the case where the endpoint distributions are Gaussians, for non-Gaussian endpoints, we have seen that the SBP can be solved via dynamic Sinkhorn recursions, which requires solving initial value problems involving this same reaction-advection-diffusion PDE within each epoch of the recursion, with updated initial conditions. By deriving the corresponding Green's function, our results facilitate this computation.

5 | Generalized SBP with State Cost

5.1 Overview

This chapter concerns a generalization of the SBP considered in Chapter 3. In Chapter 3, I set $\varepsilon = 1$ for convenience; here, I consider any (not necessarily small) $\varepsilon > 0$. More significantly, in Chapter 3, a *quadratic* state cost is added to the objective function; in this chapter, I consider a generic positive state cost $q(\boldsymbol{x})$.

Over the given time horizon $[t_0, t_1]$, the optimization problem of interest in this chapter is:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \quad \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\frac{1}{2}\|\boldsymbol{u}\|_2^2 + q(\boldsymbol{x})\right)\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{5.1a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u}) = \varepsilon\Delta_{\boldsymbol{x}}\rho, \tag{5.1b}$$

$$\boldsymbol{x}(t_0) \sim \rho_0 \text{ (given)}, \quad \boldsymbol{x}(t_1) \sim \rho_1 \text{ (given)}, \tag{5.1c}$$

hence referred to as the SBP with *additive state cost*.

Recall as in prior chapters, $\mathcal{P}_{01}$ denotes the collection of all PDF-valued trajectories $\rho(t,\boldsymbol{x})$ continuous in $t \in [t_0, t_1]$, and supported over $\mathbb{R}^n$ for each $t \in [t_0, t_1]$, such that $\rho(t_0,\boldsymbol{x}) = \rho_0$, $\rho(t_1,\boldsymbol{x}) = \rho_1$.

Note that (5.1) can be recovered from the generalized SBP (1.12) by setting $r(\boldsymbol{u}) \equiv \frac{1}{2}\|\boldsymbol{u}\|_2^2$, $\boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{u}) \equiv \boldsymbol{u}$, $\boldsymbol{\sigma}(t,\boldsymbol{x},\boldsymbol{u}) \equiv \boldsymbol{I}_{\mathrm{n}}$, and considering any nonnegative, bounded, and measurable $q(\boldsymbol{x})$.

The second half of this chapter concerns a case study in which the SBP with additive state cost is connected to Lambert's Problem, a key problem in astrodynamics. I conclude the chapter with our numerical simulation solving Lambert's Problem using our developments for the generalized SBP with additive state cost.

5.2 Conditions for Optimality of SBP with Additive State Cost

For arbitrary $\varepsilon > 0$, denote the minimizing pair for problem (5.1) as $(\rho_{\text{opt}}, \boldsymbol{u}_{\text{opt}})$. The existence-uniqueness of the solution to (5.1) is then established as follows.

Theorem 5.1. *Let $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^n)$, $\varepsilon > 0$, and the state cost $q(\boldsymbol{x})$ be positive and upper bounded. Then the minimizing tuple $(\rho_{opt}, \boldsymbol{u}_{opt})$ for problem* (5.1) *exists and is unique.*

See Appendix E.1 for a proof of Theorem 5.1.

As part of the proof of Theorem 5.1, we determine that the stochastic optimal control problem (5.1) is equivalent to the stochastic calculus of variations problem:

$$\underset{\mathbb{P}\in\Pi_{01}}{\arg\inf} \quad D_{\text{KL}}\left(\mathbb{P} \parallel \frac{\exp\left(-\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right)\mathbb{W}}{Z}\right), \tag{5.2}$$

where $\mathbb{W}$ represents the Wiener measure and Z is a normalizing constant. Recall that $D_{\text{KL}}(\mathbb{P} \parallel \mathbb{Q})$ represents the Kullback-Leibler divergence between probability measures $\mathbb{P}$ and $\mathbb{Q}$.

Thanks to the conditional Sanov's theorem [10, Sec. 4], [140], (5.2) also shows that the problem (5.1) enjoys a large deviations principle with rate function

$$D_{\text{KL}}\left(\cdot \parallel \frac{\exp\left(-\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right)\mathbb{W}}{Z}\right).$$

This can be understood intuitively as follows. For n i.i.d. random paths in an infinite-dimensional sample space Ω with distribution $\mathbb{W}$, when n is large, the more likely paths in Π_{01} correspond to those $\mathbb{P}$ which make the objective in (5.2) small. In particular, for $n \to \infty$, the path likelihood is asymptotically equivalent to

$$\exp\left(-n \inf_{\mathbb{P}\in\Pi_{01}} D_{\mathrm{KL}}\left(\mathbb{P} \parallel \frac{\exp\left(-\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right)\mathbb{W}}{Z}\right)\right).$$

The most likely path in Π_{01} corresponds to the minimizer of (5.2).

The SBP with additive state cost (5.1) can be reduced to solving a system of boundary-coupled reaction-diffusion PDEs with state-dependent nonlinear reaction rates.

To see this, let us first derive the first-order conditions for optimality for (5.1), stated next.

Theorem 5.2. *The pair* $\left(\rho_{opt}, \boldsymbol{u}_{opt}\right)$ *solving the SBP with additive state cost* (5.1)*, satisfies the system of coupled PDEs*

$$\frac{\partial S}{\partial t} + \frac{1}{2}\|\nabla_{\boldsymbol{x}} S\|_2^2 + \varepsilon\Delta_{\boldsymbol{x}} S = q(\boldsymbol{x}), \tag{5.3a}$$

$$\frac{\partial \rho_{opt}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\rho_{opt}\nabla_{\boldsymbol{x}} S\right) = \varepsilon\Delta_{\boldsymbol{x}}\rho_{opt}, \tag{5.3b}$$

with boundary conditions

$$\rho_{opt}(t_0, \boldsymbol{x}) = \rho_0(\boldsymbol{x}), \quad \rho_{opt}(t_1, \boldsymbol{x}) = \rho_1(\boldsymbol{x}), \tag{5.4}$$

where $S \in C^{1,2}\left(\mathbb{R}^n; [t_0, t_1]\right)$*, and the optimal control*

$$\boldsymbol{u}_{opt} = \nabla_{\boldsymbol{x}} S(t, \boldsymbol{x}). \tag{5.5}$$

For a proof of Theorem 5.2, see Appendix E.2.

We next show that the coupled *nonlinear* PDEs (5.3) with unconventional boundary conditions (5.4), can be exactly converted to a system of *linear* reaction-diffusion PDEs at the expense of moving the coupling to the boundary conditions.

Theorem 5.3. *Consider the SBP with additive state cost* (5.1) *with given* $[t_0, t_1]$*,* q*,*

ε, ρ_0, ρ_1 as in Theorem 5.1. Let $(\rho_{\rm opt}, S)$ be the solution of (5.3)-(5.4)*. Consider the Hopf-Cole transform $(\rho_{\rm opt}, S) \mapsto (\widehat{\varphi}_\varepsilon, \varphi_\varepsilon)$ defined as*

$$\varphi_\varepsilon = \exp\left(\frac{S}{2\varepsilon}\right), \tag{5.6a}$$

$$\widehat{\varphi}_\varepsilon = \rho_{\rm opt} \exp\left(-\frac{S}{2\varepsilon}\right). \tag{5.6b}$$

Note that this is (1.28) *with $\lambda = 2\varepsilon$; the same transform used in the classical SBP* (1.10)*. Then, the pair of Schrödinger factors $(\widehat{\varphi}_\varepsilon, \varphi_\varepsilon)$ solve the system of forward and backward linear reaction-diffusion PDEs:*

$$\frac{\partial \widehat{\varphi}_\varepsilon}{\partial t} = \left(\varepsilon \Delta_{\boldsymbol{x}} - \frac{1}{2\varepsilon} q(\boldsymbol{x})\right) \widehat{\varphi}_\varepsilon, \tag{5.7a}$$

$$\frac{\partial \varphi_\varepsilon}{\partial t} = \left(-\varepsilon \Delta_{\boldsymbol{x}} + \frac{1}{2\varepsilon} q(\boldsymbol{x})\right) \varphi_\varepsilon, \tag{5.7b}$$

with coupled boundary conditions

$$\widehat{\varphi}_\varepsilon(t_0, \boldsymbol{x}) \varphi_\varepsilon(t_0, \boldsymbol{x}) = \rho_0, \quad \widehat{\varphi}_\varepsilon(t_1, \boldsymbol{x}) \varphi_\varepsilon(t_1, \boldsymbol{x}) = \rho_1. \tag{5.8}$$

For all $t \in [t_0, t_1]$, the minimizing pair for (5.1) *can be recovered from the Schrödinger factors $(\widehat{\varphi}_\varepsilon, \varphi_\varepsilon)$ as*

$$\rho_{opt}(t, \boldsymbol{x}) = \widehat{\varphi}_\varepsilon(t, \boldsymbol{x}) \varphi_\varepsilon(t, \boldsymbol{x}), \tag{5.9a}$$

$$\boldsymbol{u}_{opt}(t, \boldsymbol{x}) = 2\varepsilon \nabla_{\boldsymbol{x}} \log \varphi_\varepsilon(t, \boldsymbol{x}). \tag{5.9b}$$

For a proof of Theorem 5.3, see Appendix E.3.

Specifically, (5.7a) (resp. (5.7b)) is a forward-in-time (resp. backward-in-time) reaction-diffusion PDE with state-dependent reaction rate $q(\boldsymbol{x})$. Once this PDE system for the Schrödinger factors is solved, the optimally controlled joint PDF and the optimal control can then be computed using (5.9). Notice that the special case

$q(\boldsymbol{x}) := 0$ corresponding to the classical SBP reduces (5.7) to a system of forward-backward heat PDEs, as described in Chapter 1.

The derived system (5.7)-(5.8) is a system of forward-backward diffusions where individual points in the state space $\mathbb{R}^n$ move forward and reverse in time according to Brownian motion with variance 2ε, together with killing or creation of probability mass at rate $q(\boldsymbol{x})$. Thus, $q(\boldsymbol{x})$ in (5.7a) is state-dependent killing rate. Likewise, $q(\boldsymbol{x})$ in (5.7b) is state-dependent creation rate. The boundary conditions (5.8) ensure that the total probability mass remains balanced.

5.3 Lambert's Problem

We now examine a case study in which the generalized SBP with state cost is used to solve a well-known problem in astrodynamics: Lambert's Problem. I first provide the relevant historical context and the formulation of Lambert's Problem, to motivate the development that follows.

Lambert's Problem in astrodynamics was originally concerned with the determination of a comet's orbit. In his 1761 book, Johann Heinrich Lambert gave derivations relating a given flight time $t_1 - t_0$ to initial and final position vectors $\boldsymbol{x}_0, \boldsymbol{x}_1 \in \mathbb{R}^3$, for the cases of parabolic, elliptic, and hyperbolic Keplerian arcs [141]. Other early pioneers contributing to the solution of the Lambert's problem include Euler, who obtained solutions for parabolic and elliptic Keplerian arcs in 1743 [142], and Lagrange, who derived several alternative proofs for such results in 1780 [143]. For a detailed chronology of related results, see [144, Sec. 9].

Lambert's Problem remains highly relevant today due to its application in solving interplanetary transfer, missile interception and rendezvous problems.

5.3.1 Deterministic Lambert's Problem

The classical Lambert's Problem in orbital mechanics is a *finite dimensional* two point boundary value problem over a fixed time horizon $[t_0, t_1]$, subject to a two-body gravitational potential $V(\boldsymbol{x})$ and endpoint constraints on the position vector. For a mathematical formulation of the problem, let $\boldsymbol{x} = (x, y, z)^\top \in \mathbb{R}^3$ denote the position, and $\boldsymbol{u} := \dot{\boldsymbol{x}} \in \mathbb{R}^3$ denote the velocity of a spacecraft with respect to the Earth Centered Inertial (ECI) frame.

The classical Lambert's Problem seeks to determine a velocity field $\boldsymbol{u} = \boldsymbol{u}(t, \boldsymbol{x})$ subject to the Keplerian equation of motion, endpoint positions, and hard flight time constraints, i.e.,

$$\ddot{\boldsymbol{x}} = -\nabla_{\boldsymbol{x}} V(\boldsymbol{x}), \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \text{ (given)}, \quad \boldsymbol{x}(t_1) = \boldsymbol{x}_1 \text{ (given)}, \tag{5.10}$$

where the nonlinear gravitational potential $V(\cdot)$ is measurable and bounded. We consider specifically the case in which

$$V(\boldsymbol{x}) := -\frac{\mu}{|\boldsymbol{x}|} - \frac{\mu J_2 R_{\text{Earth}}^2}{2|\boldsymbol{x}|^3}\left(1 - \frac{3z^2}{|\boldsymbol{x}|^2}\right), \quad \boldsymbol{x} \in [R_{\text{Earth}}, +\infty), \tag{5.11}$$

where the parameter $\mu = 398600.4415$ km^3/s^2 denotes the product of the Earth's gravitational constant and mass and the parameter $J_2 = 1.08263 \times 10^{-3}$ denotes the (unitless) second zonal harmonic coefficient, a measure of the Earth oblateness. The radius of Earth $R_{\text{Earth}} = 6378.1363$ km.

The theoretical and algorithmic results presented in this chapter apply to more general geopotential models [145], i.e., when higher order harmonics (zonal and tesseral terms) are included in $V(\boldsymbol{x})$. Indeed, the only constraint we must place on $V(\boldsymbol{x})$ for our results to remain valid is that $V(\boldsymbol{x})$ must be negative, bounded and continuously differentiable, which is indeed the case in (5.11) for all $\boldsymbol{x} \in [R_{\text{Earth}}, +\infty)$.

The deterministic Lambert's problem has a substantial literature in the guidance-control community – a non-exhaustive list of references includes [48–52], [53, Ch. 13.4]. A deterministic optimal control reformulation that inspires our development is due to [146].

5.3.2 Probabilistic Lambert's Problem

A probabilistic version of Lambert's problem softens the endpoint constraints in (5.10) to

$$\boldsymbol{x}(t_0) \sim \rho_0 \text{ (given)}, \quad \boldsymbol{x}(t_1) \sim \rho_1 \text{ (given)}. \tag{5.12}$$

Thus, (5.12) allows for stochastic uncertainties in the endpoint relative positions. Instead of steering between two given position vectors, as in the deterministic case, the probabilistic case seeks to steer between the statistics given by the respective joint PDFs ρ_0, ρ_1. Therefore, the probabilistic Lambert's problem can be expressed as:

$$\underset{\dot{\boldsymbol{x}}=\boldsymbol{u}(t,\boldsymbol{x})}{\text{find}} \quad \boldsymbol{u} \tag{5.13a}$$

$$\ddot{\boldsymbol{x}} = -\nabla_{\boldsymbol{x}} V\left(\boldsymbol{x}\right), \tag{5.13b}$$

$$\boldsymbol{x}(t_0) \sim \rho_0 \text{ (given)}, \quad \boldsymbol{x}(t_1) \sim \rho_1 \text{ (given)}. \tag{5.13c}$$

From an engineering perspective, the PDF ρ_0 captures initial condition uncertainty, which may arise from statistical estimation errors. The PDF ρ_1 on the other hand, encodes desired statistical performance specification, i.e., the allowable terminal condition uncertainty. A small terminal condition uncertainty would imply that ρ_1 is "tall and skinny," i.e., approximating Dirac delta with its Lebesgue mass supported on a small subset of $\mathbb{R}^3$. A less stringent terminal uncertainty specification would allow ρ_1 to have its mass spread over a subset of $\mathbb{R}^3$ with larger Lebesgue volume. If problem (5.13) is feasible, then intuition suggests that a more (resp. less) stringent ρ_1 specification would result in a more (resp. less) "expensive" control $\boldsymbol{u}$ but it is unclear in what sense. As such, (5.13) is posed as a *feasibility* problem and even if it admits a unique solution $\boldsymbol{u}$ (also unclear why), it is not obvious what *optimality* guarantee, if any, does that solution usher.

5.4 Lambertian OMT

In this section, we establish that the Lambert's problem with endpoint joint PDF constraints, (5.13), is a generalized variant of the OMT problem introduced in Chapter 1. Recall that generalized OMT was formulated as follows:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\frac{1}{2}\|\boldsymbol{u}\|_2^2\,\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{5.14a}$$

$$\frac{\partial\rho}{\partial t}+\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})=0, \tag{5.14b}$$

$$\rho(t_0,\boldsymbol{x})=\rho_0,\quad \rho(t_1,\boldsymbol{x})=\rho_1. \tag{5.14c}$$

Thanks to the connection between OMT and SBPs, we can specialize results from Section 5.2 to provide guarantees about this variant of the OMT.

The connection between (5.13) and (5.14) allows for the rigorous establishment of the uniqueness of the solution of (5.13). This proof of uniqueness reveals that the unique velocity field $\boldsymbol{u}$ is indeed optimal in certain minimum effort sense.

5.4.1 Formulation of Lambertian OMT

We start with a known result [146, 147] that exactly transforms Lambert's Problem (5.10) to a standard deterministic optimal control problem with state variable $\boldsymbol{x}$ and control variable $\boldsymbol{u}$, given by

$$\arg\inf_{\boldsymbol{u}} \quad \int_{t_0}^{t_1} \left(\frac{1}{2}\|\boldsymbol{u}\|_2^2 - V(\boldsymbol{x})\right) \mathrm{d}t \tag{5.15a}$$

$$\dot{\boldsymbol{x}} = \boldsymbol{u}, \tag{5.15b}$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \,\text{(given)}, \quad \boldsymbol{x}(t_1) = \boldsymbol{x}_1 \,\text{(given)}. \tag{5.15c}$$

Letting $(\boldsymbol{q}, \dot{\boldsymbol{q}}) \equiv (\boldsymbol{x}, \boldsymbol{u})$, this reformulation can be interpreted as the celebrated Hamilton's principle of least action where the objective in (5.15a) is an action integral $\int_{t_0}^{t_1} L(\boldsymbol{q}, \dot{\boldsymbol{q}})\,\mathrm{d}t$ with Lagrangian $L(\boldsymbol{q}, \dot{\boldsymbol{q}}) \coloneqq T(\dot{\boldsymbol{q}}) - V(\boldsymbol{q})$, the kinetic energy $T(\dot{\boldsymbol{q}}) \coloneqq \frac{1}{2}\langle \dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}\rangle$, and the potential energy $V(\boldsymbol{q})$ given by (5.11). The associated Hamiltonian $H(\boldsymbol{q}, \dot{\boldsymbol{q}})$ defined as the Legendre-Fenchel conjugate of the Lagrangian L, equals $H(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \langle \frac{\partial L}{\partial \dot{\boldsymbol{q}}}, \dot{\boldsymbol{q}}\rangle - L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = T(\dot{\boldsymbol{q}}) + V(\boldsymbol{q})$, the total energy. Consequently, the equations of motion in Hamilton's canonical variables $(\boldsymbol{q}, \boldsymbol{p}) \coloneqq (\boldsymbol{q}, \frac{\partial L}{\partial \dot{\boldsymbol{q}}}) = (\boldsymbol{q}, \dot{\boldsymbol{q}})$ becomes

$$\dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}}, \quad \dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}},$$

which is identical to the dynamics $\ddot{\boldsymbol{x}} = -\nabla_{\boldsymbol{q}} V(\boldsymbol{q})$ in (5.10), as expected.

Replacing the endpoint constraints in (5.10) with (5.12) is, therefore, equivalent to modifying the optimal control problem (5.15) by replacing (5.15c) with (5.12). Furthermore, the probabilistic uncertainty in the initial condition $\boldsymbol{x}(t_0) \sim \rho_0$ is advected by the controlled sample path dynamics (5.15b) resulting in the joint PDF evolution of the state $\boldsymbol{x}(t)$ governed via the Liouville PDE initial value problem (see e.g., [14, 148, 149])

$$\frac{\partial \rho}{\partial t} + \nabla_{\boldsymbol{x}} \cdot (\rho \boldsymbol{u}) = 0, \quad \rho(t = 0, \cdot) = \rho_0 \,\text{(given)}, \tag{5.16}$$

where $\rho(t, \boldsymbol{x})$ denotes the *controlled* transient joint state PDF for a given control policy $\boldsymbol{u}(t,\boldsymbol{x})$. The Liouville PDE $\dfrac{\partial\rho}{\partial t}+\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})=0$ is the continuity equation signifying the conservation of probability mass over the state space, i.e., $\int_{\mathbb{R}^3}\rho(t,\boldsymbol{x})\mathrm{d}\boldsymbol{x}=1$ for all $t\in[t_0,t_1]$. The solution of the Liouville PDE is to be understood in weak sense, i.e., for all compactly supported smooth test functions $\phi(t,\boldsymbol{x})\in C_c^\infty\left(\mathbb{R}^3\times[t_0,t_1]\right)$, the function $\rho(t,\boldsymbol{x})$ satisfies $\int_{t_0}^{t_1}\int_{\mathbb{R}^3}\left(\rho\frac{\partial\phi}{\partial t}+\rho\langle\boldsymbol{u},\nabla_{\boldsymbol{x}}\phi\rangle\right)\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_{\mathbb{R}^3}\rho_0(\boldsymbol{x})\phi(\boldsymbol{x},t_0)\mathrm{d}\boldsymbol{x}=0$.

Different feasible control policies $\boldsymbol{u}$ in (5.16) induce different PDF-valued trajectories $\rho(t,\cdot)$ connecting the prescribed endpoint joints $\rho_0,\rho_1\in\mathcal{P}_2\left(\mathbb{R}^3\right)$. Hence, the objective in (5.15a) should be averaged w.r.t. the controlled transient joint probability measure $\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$, and the resulting functional should be minimized over the pair $(\rho,\boldsymbol{u})$. We thus arrive at a *generalized OMT* formulation as follows:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{V}}{\arg\inf}\quad\int_{t_0}^{t_1}\int_{\mathbb{R}^3}\left(\frac{1}{2}\|\boldsymbol{u}\|_2^2-V(\boldsymbol{x})\right)\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{5.17a}$$

$$\frac{\partial\rho}{\partial t}+\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})=0, \tag{5.17b}$$

$$\boldsymbol{x}(t_0)\sim\rho_0\,\text{(given)},\quad \boldsymbol{x}(t_1)\sim\rho_1\,\text{(given)}, \tag{5.17c}$$

henceforth referred to as the *Lambertian optimal mass transport (L-OMT)*.

Notice that in the special case when the endpoint joint PDFs are Dirac deltas at some given positions $\boldsymbol{x}_0,\boldsymbol{x}_1\in\mathbb{R}^3$, i.e., $\rho_0(\boldsymbol{x})=\delta(\boldsymbol{x}-\boldsymbol{x}_0),\rho_1(\boldsymbol{x})=\delta(\boldsymbol{x}-\boldsymbol{x}_1)$, then problem (5.17) reduces back to (5.15), or equivalently to (5.10).

Note that if potential $V\equiv 0$, (5.17) would reduce to the classical dynamic OMT (5.14).

5.4.2 Existance/Uniqueness of L-OMT

We next establish the existence-uniqueness of the solution to (5.17) as follows:

Theorem 5.4. ***(Existence-uniqueness of solution for L-OMT)*** *Let* $\rho_0,\rho_1\in\mathcal{P}_2\left(\mathbb{R}^3\right)$, *and the gravitational potential* $V(\cdot) \in (-\infty, 0)$ *as in* (5.11). *Then the minimizing tuple* $\left(\rho_{opt}, \boldsymbol{u}_{opt}\right)$ *for problem* (5.17) *exists and is unique.*

Proof. The "cost-to-go" in (5.17a) is the average of the Lagrangian

$$L(t, \boldsymbol{x}, \dot{\boldsymbol{r}}) = \frac{1}{2}\|\dot{\boldsymbol{r}}\|_2^2 - V(\boldsymbol{x}). \tag{5.18}$$

Since $\frac{1}{2}\|\cdot\|_2^2$ is a strictly convex function, so L is strictly convex in $\dot{\boldsymbol{r}}$. We now show that L is also superlinear in $\dot{\boldsymbol{r}}$. Recall that a function is superlinear, or 1-coercive, in $\boldsymbol{x}$ if

$$\lim_{\|\boldsymbol{x}\|_2 \to \infty} \frac{f(\boldsymbol{x})}{\|\boldsymbol{x}\|_2} = +\infty.$$

Thus, to establish that L is superlinear, we notice:

$$\lim_{\|\dot{\boldsymbol{r}}\|_2 \to \infty} \frac{L}{\|\dot{\boldsymbol{r}}\|_2} = \lim_{\|\dot{\boldsymbol{r}}\|_2 \to \infty} \left(\frac{\frac{1}{2}\|\dot{\boldsymbol{r}}\|_2^2 - V(\boldsymbol{x})}{\|\dot{\boldsymbol{r}}\|_2} \right) = \left(\lim_{\|\dot{\boldsymbol{r}}\|_2 \to \infty} \frac{1}{2}\|\dot{\boldsymbol{r}}\|_2 \right) + \left(\lim_{\|\dot{\boldsymbol{r}}\|_2 \to \infty} \frac{-V(\boldsymbol{x})}{\|\dot{\boldsymbol{r}}\|_2} \right). \tag{5.19}$$

The first term $\lim_{\|\dot{\boldsymbol{r}}\|_2 \to \infty} \frac{1}{2}\|\dot{\boldsymbol{r}}\|_2 = +\infty$. For the second term in (5.19), $V(\cdot) \in (-\infty, 0)$ implies $-V(\boldsymbol{x}) \in (0, +\infty)$ and thus $\lim\limits_{\|\dot{\boldsymbol{r}}\|_2 \to \infty} \dfrac{-V(\boldsymbol{x})}{\|\dot{\boldsymbol{r}}\|_2} = +\infty$. Therefore,

$$\lim_{\|\dot{\boldsymbol{r}}\|_2 \to \infty} \frac{L}{\|\dot{\boldsymbol{r}}\|_2} = +\infty.$$

Being strictly convex and superlinear in $\dot{\boldsymbol{r}}$, the L in (5.18) is a weak Tonelli Lagrangian [150, p. 118], [151, Ch. 6.2], which in turn guarantees [151, Thm. 1.4.2] that the pair $\left(\rho_{\text{opt}}, \boldsymbol{u}_{\text{opt}}\right)$ for the generalized OMT problem (5.17) exists and is unique. $\square$

The above proof of Theorem 5.4 above only required that V has codomain $(-\infty, 0)$. Thus, the existence-uniqueness result above applies for all generic bounded and continuously differentiable gravitational potentials having codomains in the subsets of

$(-\infty, 0)$.

To explicitly demonstrate that the V in (5.11) satisfies this requirement, we rewrite (5.11) as:

$$V(\boldsymbol{x}) = -\frac{\mu}{\|\boldsymbol{x}\|_2}\left(1 + \frac{J_2 R_{\text{Earth}}^2}{2\|\boldsymbol{x}\|_2^2}\left(1 - \frac{3z^2}{\|\boldsymbol{x}\|_2^2}\right)\right). \tag{5.20}$$

Since $0 \leq z^2 \leq \|\boldsymbol{x}\|^2$, we have $-2 \leq 1 - \frac{3z^2}{\|\boldsymbol{x}\|_2^2} \leq 1$. On the other hand, $\|\boldsymbol{x}\| \geq R_{\text{Earth}}$, which yields

$$\frac{J_2 R_{\text{Earth}}^2}{2\|\boldsymbol{x}\|_2^2} \leq \frac{J_2 R_{\text{Earth}}^2}{2R_{\text{Earth}}^2} \leq \frac{J_2}{2} < 0.0006.$$

Hence,

$$(-2)(0.0006) = -0.0012 < \frac{J_2 R_{\text{Earth}}^2}{2\|\boldsymbol{x}\|_2^2}\left(1 - \frac{3z^2}{\|\boldsymbol{x}\|_2^2}\right),$$

which gives

$$0 < 1 - 0.0012 < \left(1 + \frac{J_2 R_{\text{Earth}}^2}{2\|\boldsymbol{x}\|_2^2}\left(1 - \frac{3z^2}{\|\boldsymbol{x}\|_2^2}\right)\right). \tag{5.21}$$

Combining (5.20) and (5.21) yields $V(\boldsymbol{x}) < 0$.

5.5 Lambertian SBP

Recognizing (5.13) as a generalized version of OMT naturally leads to a further generalization to the case when the velocity has additive process noise, i.e., when the controlled ODE

$$\dot{\boldsymbol{x}} = \boldsymbol{u}(t, \boldsymbol{x})$$

is replaced with the Itô stochastic differential equation (SDE) (E.1), repeated for the reader's convenience as:

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{u}(t, \boldsymbol{x})\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}\boldsymbol{w}(t),$$

where the strength of the process noise $\varepsilon > 0$ is *not necessarily* small. The process noise in (E.1) may result from noisy actuation in velocity command or from stochastic disturbance in atmospheric drag, solar radiation pressure etc, and hence of practical interest. Accounting for this process noise generalizes the L-OMT (5.17) as

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{V}}{\arg\inf} \quad \int_{t_0}^{t_1}\int_{\mathbb{R}^3}\left(\frac{1}{2}\|\boldsymbol{u}\|_2^2 - V(\boldsymbol{x})\right)\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{5.22a}$$

$$\frac{\partial\rho}{\partial t} + \nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u}) = \varepsilon\Delta_{\boldsymbol{x}}\rho, \tag{5.22b}$$

$$\boldsymbol{x}(t_0) \sim \rho_0\ \text{(given)}, \quad \boldsymbol{x}(t_1) \sim \rho_1\ \text{(given)}, \tag{5.22c}$$

which is the special case of the SBP with additive state cost, (5.1), explored in Chapter 3, in which $q(\boldsymbol{x}) := -V(\boldsymbol{x})$. This problem is henceforth referred to as *Lambertian SBP*, or L-SBP. The existence and uniqueness of the solution to (5.22) follows directly from Theorem 5.1 presented above.

Crucially, reformulating the Lambert's problem as L-SBP also allows the Lambert's problem to be solved using the fixed point recursion algorithm typically employed for SBPs.

5.6 Integral Representation for the Schrödinger Factors

In this section, we next express (5.7a) (and, by change of variables, (5.7b)) in Fredholm integral form. In the absence of a known Markov kernel k, these integral representations can be used to solve SBPs with additive state cost.

Theorem 5.5. *For $t \in [t_0, \infty)$, the reaction-diffusion PDE IVP*

$$\frac{\partial u}{\partial t} = a\Delta_{\boldsymbol{x}} u + q(\boldsymbol{x})u, \quad \boldsymbol{x} \in \mathbb{R}^n, \quad u(t_0, \boldsymbol{x}) = u_0(\boldsymbol{x}) \text{ given}, \tag{5.23}$$

where the constant $a > 0$ and the function $q(\boldsymbol{x})$ is sufficiently smooth a.e., has solution

$$\begin{aligned} u(t, \boldsymbol{x}) = &\underbrace{\frac{1}{\sqrt{(4\pi a t)^n}} \int_{\mathbb{R}^n} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2}{4at}\right) u_0(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}}_{\text{term 1}} \\ &+ \underbrace{\int_{t_0}^{t} \frac{1}{\sqrt{(4\pi a(t-\tau))^n}} \int_{\mathbb{R}^n} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2}{4a(t-\tau)}\right) q(\boldsymbol{y}) u(\tau, \boldsymbol{y})\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\tau}_{\text{term 2}}. \end{aligned} \tag{5.24}$$

The right-hand-side of (5.24) is a sum of two terms: term 1 is the solution of the heat PDE IVP accounting for the initial condition $u_0(\cdot)$; term 2 captures the effect of the reaction rate $q(\cdot)$.

Before giving the proof of Theorem 5.5, we summarize a few key results regarding Fourier transforms. To do so, let $\mathrm{i} := \sqrt{-1}$. We use the non-unitary angular frequency version of Fourier transform

$$\widehat{f}(\boldsymbol{\omega}) := \int_{\mathbb{R}^n} f(\boldsymbol{x}) \exp\left(-\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle\right) \mathrm{d}\boldsymbol{x}, \tag{5.25}$$

and the corresponding inverse Fourier transform

$$f(\boldsymbol{x}) := \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \widehat{f}(\boldsymbol{\omega}) \exp\left(+\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle\right) \mathrm{d}\boldsymbol{\omega}. \tag{5.26}$$

Symbolically write $\widehat{f}(\boldsymbol{\omega}) = \mathcal{F}[f(\boldsymbol{x})]$, $f(\boldsymbol{x}) = \mathcal{F}^{-1}\left[\widehat{f}(\boldsymbol{\omega})\right]$. For a suitably smooth function $f(\boldsymbol{x})$ on $\mathbb{R}^n$, the following holds (derivations of in the appendix).

- The Fourier transform of Laplacian on $\mathbb{R}^n$:

$$\mathcal{F}[\Delta_{\boldsymbol{x}} f] = -\|\boldsymbol{\omega}\|_2^2 \widehat{f}(\boldsymbol{\omega}). \tag{5.27}$$

- Convolution:

$$\mathcal{F}[(f * g)(\boldsymbol{x})] = \widehat{f}(\boldsymbol{\omega})\widehat{g}(\boldsymbol{\omega}), \mathcal{F}[f(\boldsymbol{x})g(\boldsymbol{x})] = \frac{1}{(2\pi)^n}\left(\widehat{f} * \widehat{g}\right)(\boldsymbol{\omega}). \tag{5.28}$$

- Inverse Fourier transform of exp-neg-square:

$$\mathcal{F}^{-1}\left[\exp\left(-a\|\boldsymbol{\omega}\|_2^2 t\right)\right] = \frac{\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{4at}\right)}{\sqrt{(4\pi a t)^n}}. \tag{5.29}$$

With these results, we proceed with the proof of Theorem 5.5.

Proof. We denote the Fourier transform of $u(\boldsymbol{x}, t)$ w.r.t. $\boldsymbol{x}$ as $\widehat{u}(t, \boldsymbol{\omega})$ for all $t \in [t_0, \infty)$, and the Laplace transform of $\widehat{u}(t, \boldsymbol{\omega})$ w.r.t. t as $\widehat{U}(s, \boldsymbol{\omega})$. Taking the Fourier transform to both sides of (5.23), we obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}[u(\boldsymbol{x}, t)] = a\mathcal{F}[\Delta u] + \mathcal{F}[q(\boldsymbol{x})u] \\
\Rightarrow \quad &\frac{\mathrm{d}}{\mathrm{d}t}\widehat{u}(t, \boldsymbol{\omega}) = -a\|\boldsymbol{\omega}\|_2^2 \widehat{u}(t, \boldsymbol{\omega}) + \frac{1}{(2\pi)^n}\left(\widehat{q}(\boldsymbol{\omega}) * \widehat{u}(t, \boldsymbol{\omega})\right),
\end{aligned} \tag{5.30}$$

where we use the above definitions of the Fourier transform of the Laplacian and convolution. Taking the Laplace transform to both sides of (5.30), yields

$$\begin{aligned}
&s\widehat{U}(s, \boldsymbol{\omega}) - \widehat{u}_0(\boldsymbol{\omega}) = -a\|\boldsymbol{\omega}\|_2^2 \widehat{U}(s, \boldsymbol{\omega}) + \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \widehat{q}(\boldsymbol{\omega} - \boldsymbol{\eta})\widehat{U}(s, \boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta} \\
\Rightarrow \quad &\widehat{U}(s, \boldsymbol{\omega}) = \frac{\widehat{u}_0(\boldsymbol{\omega})}{s + a\|\boldsymbol{\omega}\|_2^2} + \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \widehat{q}(\boldsymbol{\omega} - \boldsymbol{\eta})\frac{\widehat{U}(s, \boldsymbol{\eta})}{s + a\|\boldsymbol{\omega}\|_2^2}\,\mathrm{d}\boldsymbol{\eta}.
\end{aligned} \tag{5.31}$$

Inverse Laplace transform of (5.31) yields

$$
\begin{aligned}
\widehat{u}(t,\boldsymbol{\omega}) &= \widehat{u}_0(\boldsymbol{\omega}) \exp\left(-a\|\boldsymbol{\omega}\|_2^2 t\right) + \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \widehat{q}(\boldsymbol{\omega}-\boldsymbol{\eta})\, \mathcal{L}^{-1}\left[\frac{\widehat{U}(s,\boldsymbol{\eta})}{s+a\|\boldsymbol{\omega}\|_2^2}\right] \mathrm{d}\boldsymbol{\eta} \\
&= \widehat{u}_0(\boldsymbol{\omega}) \exp\left(-a\|\boldsymbol{\omega}\|_2^2 t\right) \\
&\quad + \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \widehat{q}(\boldsymbol{\omega}-\boldsymbol{\eta}) \left(\int_0^t \widehat{u}(\tau,\boldsymbol{\eta}) \exp\left(-a\|\boldsymbol{\omega}\|_2^2 (t-\tau)\right) \mathrm{d}\tau\right) \mathrm{d}\boldsymbol{\eta},
\end{aligned}
$$

where the last line used the convolution property for the Laplace transform, namely $\mathcal{L}^{-1}\left[F(s)G(s)\right] = (f * g)\,(t)$.

We next change the order of integration (using Fubini-Tonelli Theorem) for the space-time integral above, then apply the inverse Fourier transform to both sides, and use (5.28) to arrive at

$$
\begin{aligned}
u(\boldsymbol{x},t) &= \left(u_0 * \mathcal{F}^{-1}\left[\exp\left(-a\|\boldsymbol{\omega}\|_2^2 t\right)\right]\right)(\boldsymbol{x}) \\
&\quad + \frac{1}{(2\pi)^n}\int_{t_0}^t \left(\mathcal{F}^{-1}\left[\exp\left(-a\|\boldsymbol{\omega}\|_2^2 (t-\tau)\right)\right]\right. \\
&\qquad\qquad \left. * \mathcal{F}^{-1}\left[\left(\int_{\mathbb{R}^n} \widehat{q}(\boldsymbol{\omega}-\boldsymbol{\eta})\widehat{u}(\tau,\boldsymbol{\eta})\mathrm{d}\boldsymbol{\eta}\right)\right]\right)(\boldsymbol{x})\,\mathrm{d}\tau. \qquad (5.32)
\end{aligned}
$$

Using the definition of convolution again, we notice that $\mathcal{F}^{-1}\left[\left(\int_{\mathbb{R}^n} \widehat{q}(\boldsymbol{\omega}-\boldsymbol{\eta})\widehat{u}(\tau,\boldsymbol{\eta})\mathrm{d}\boldsymbol{\eta}\right)\right] = (2\pi)^n q(\boldsymbol{x})u(\tau,\boldsymbol{x})$, which together with (5.29), simplifies (5.32) to (5.24). □

The solution for a generalized SBP with additive (positive, bounded above) state cost can thus be determined via a numerical approach employing (5.24).

5.7 Solving Lambertian SBP with Generic Gravitational Potential

Generalizing the algorithm presented in Chapter 1, note that a SBP with additive state cost may be solved as follows:

Step 1. Make an initial guess for $\widehat{\varphi}_{\varepsilon,0}$ that is everywhere positive.

Step 2. Use the $\widehat{\varphi}_{\varepsilon,0}$ from **Step 1** to integrate (5.7a) from t_0 to t_1, to determine $\widehat{\varphi}_{\varepsilon}(t_1,\boldsymbol{x})$.

Step 3. Set $\varphi_{\varepsilon,1} = \rho_1(\boldsymbol{x})/\widehat{\varphi}_{\varepsilon}(t_1,\boldsymbol{x})$ by enforcing (5.8) at t_1.

Step 4. Use the $\varphi_{\varepsilon,1}$ from **Step 3** to integrate (5.7b) from t_1 to t_0 to determine $\varphi_{\varepsilon}(t_0,\boldsymbol{x})$.

Step 5. Redefine $\widehat{\varphi}_{\varepsilon,0} = \rho_0/\varphi_{\varepsilon}(t_0,\boldsymbol{x})$ by enforcing (5.8) at t_0.

Step 6. Repeat **Steps 2-5** until the pair $(\widehat{\varphi}_{\varepsilon,0}(\boldsymbol{x}),\varphi_{\varepsilon,1}(\boldsymbol{x}))$ has converged up to a desired numerical tolerance.

Step 7. With the converged endpoint factors $\widehat{\varphi}_{\varepsilon,0}(\boldsymbol{x}),\varphi_{\varepsilon,1}(\boldsymbol{x})$ from **Step 6**, compute the transient factors $\widehat{\varphi}_{\varepsilon}(t,\boldsymbol{x}),\varphi_{\varepsilon}(t,\boldsymbol{x})$ at desired $t\in[t_0,t_1]$ using the IVP solver for (5.7a) and (5.7b).

Step 8. Use (5.9) to return the optimal solution $\left(\rho_{\varepsilon}^{\text{opt}},\boldsymbol{u}_{\text{opt}}\right)$ for the SBP with additive state cost.

To specialize this algorithm for the L-SBP requires only the substitution $q(\boldsymbol{x}) := -V(\boldsymbol{x})$ in (5.7a) and (5.7b).

In the case of linear SBPs, the relevant Markov kernel k_{L} (4.8) in the Schrödinger system (1.20) is explicitly defined, and we thus presume the availability of an IVP solver throughout Chapter 2. In contrast, here, we present an integral formulation (5.24) to allow for forward and backward integration needed for Steps 2 and 4.

We can leverage this reformulation (5.22) to specialize (5.24) for the test case of Lambert's Problem, with $n = 3$, resulting in an IVP involving (5.7a). Specifically, we find the following integral representation formula

$$\begin{aligned}\widehat{\varphi}_{\varepsilon}(t,\boldsymbol{x}) = &\frac{1}{\sqrt{(4\pi\varepsilon t)^3}}\int_{\mathbb{R}^3}\exp\left(-\frac{\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|_2^2}{4\varepsilon t}\right)\widehat{\varphi}_{\varepsilon,0}(\tilde{\boldsymbol{x}})\,\mathrm{d}\tilde{\boldsymbol{x}} \\ &-\int_{t_0}^{t}\frac{1}{2\varepsilon\sqrt{(4\pi\varepsilon(t-\tau))^3}}\int_{\mathbb{R}^3}\exp\left(-\frac{\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|_2^2}{4\varepsilon(t-\tau)}\right)V(\tilde{\boldsymbol{x}})\widehat{\varphi}_{\varepsilon}(\tau,\tilde{\boldsymbol{x}})\,\mathrm{d}\tilde{\boldsymbol{x}}\,\mathrm{d}\tau,\end{aligned} \tag{5.33}$$

which can be employed in Step 2 of the proposed algorithm.

Additionally, consider a change of time variable $t \mapsto \bar{t} := t_0 + t_1 - t$. When the physical time $t \in [t_1, t_0]$ flows backward, then the transformed time $\bar{t} \in [t_0, t_1]$ flows forward. This allows the IVP involving (5.7b) to be rewritten for L-SBP as

$$\frac{\partial\bar{\varphi}_{\varepsilon}}{\partial\bar{t}}(\bar{t},\boldsymbol{x}) = \left(\varepsilon\Delta_{\boldsymbol{x}} - \frac{1}{2\varepsilon}V(\boldsymbol{x})\right)\bar{\varphi}_{\varepsilon}(\bar{t},\boldsymbol{x}), \qquad \bar{\varphi}_{\varepsilon}(\bar{t} = t_0, \boldsymbol{x}) = \varphi_{\varepsilon,1}. \tag{5.34}$$

From the solution of (5.34), we recover $\varphi_{\varepsilon}(t,\boldsymbol{x}) = \bar{\varphi}_{\varepsilon}(t_0+t_1-t,\boldsymbol{x})$. Since this change of variable produced a PDE (5.34) identical in form to the one that resulted in (5.24), Step 4 is also calculated via (5.24).

5.7.1 Left Riemann Sum Approximation

For the numerical implementation of (5.24), we approximate the second summand the right-hand-side of (5.33) via left Riemann sum. Recall that the first summand in the right-hand-side of (5.33) is the well-known solution of the heat equation, denoted as $\widehat{\varphi}_{\text{heat}}(t,\boldsymbol{x})$, which can be implemented by direct matrix-vector multiplication over a computational domain.

In particular, we uniformly discretize a hyper-rectangular computational domain

$[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}] \times [z_{\min}, z_{\max}] \times [t_0, t_1] \subset \mathbb{R}^3 \times [t_0, t_1]$. Along the space-time dimensions, we use constant step sizes $\Delta x, \Delta y, \Delta z$, Δt for N_x, N_y, N_z, N_t steps, respectively.

I assign $\widehat{\varphi}_\varepsilon(t, \boldsymbol{x}) \leftarrow \widehat{\varphi}_{\varepsilon,0}$ for $t \in [t_0, t_0 + \Delta t)$, and approximate the second term of (5.33), $\widehat{\varphi}_{\text{left}}(t_0 + \Delta t, \boldsymbol{x})$, as

$$\frac{1}{2\varepsilon\sqrt{(4\pi\varepsilon(\Delta t))^3}} \sum_{m=0}^{N_x} \left(\sum_{n=0}^{N_y} \left(\sum_{j=0}^{N_z} \exp\left(-\frac{\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}_{(m,n,j)}\|^2}{4\varepsilon\Delta t} \right) V(\widetilde{\boldsymbol{x}}_{(m,n,j)}) \widehat{\varphi}_{\varepsilon,0}(\widetilde{\boldsymbol{x}}_{(m,n,j)}) \right) \right) \times \Delta x \Delta y \Delta z \Delta t, \tag{5.35}$$

where $\widetilde{\boldsymbol{x}}_{(m,n,j)} := [x_{\min} + m\Delta x, y_{\min} + n\Delta y, z_{\min} + j\Delta z]^\top$ and the tuple $(m, n, j) \in \{0, 1, \cdots, N_x\} \times \{0, 1, \cdots, N_y\} \times \{0, 1, \cdots, N_z\}$. I thus get

$$\begin{aligned} \widehat{\varphi}_\varepsilon(t_0 + \Delta t, \boldsymbol{x}) &\approx \widehat{\varphi}_{\varepsilon,\text{approx}}(t_0 + \Delta t, \boldsymbol{x}) \\ &= \widehat{\varphi}_{\text{heat}}(t_0 + \Delta t, \boldsymbol{x}) + \widehat{\varphi}_{\text{left}}(t_0 + \Delta t, \boldsymbol{x}). \end{aligned}$$

For $k \in \{1, \cdots, N_t\}$, we construct a recursive formula for $\widehat{\varphi}_{\varepsilon,\text{approx}}(t_0 + k\Delta t, \boldsymbol{x})$ using $\widehat{\varphi}_{\varepsilon,\text{approx}}(t_0 + q\Delta t, \boldsymbol{x})\ \forall \{q \in \mathbb{N} \cup \{0\} \mid q < k\}$. We approximate $\widehat{\varphi}_\varepsilon(t, \boldsymbol{x}) \approx \widehat{\varphi}_{\varepsilon,\text{approx}}(t_0 + q\Delta t, \boldsymbol{x})$ for $t \in [t_0 + q\Delta t, t_0 + (q+1)\Delta t)$. To evaluate the second term of (5.33), we integrate over each interval $[t_0 + q\Delta t, t_0 + (q+1)\Delta t)$ using triple summation, as in the calculation of $\widehat{\varphi}_{\text{left}}(t_0 + \Delta t, \boldsymbol{x})$. This yields a recursive formula

$$\begin{aligned} &\widehat{\varphi}_{\text{left}}(t_0 + k\Delta t, \boldsymbol{x}) \\ &= \sum_{q=0}^{k-1} \left(\frac{1}{2\varepsilon\sqrt{(4\pi\varepsilon(k-q)\Delta t)^3}} \sum_{m=0}^{N_x} \left(\sum_{n=0}^{N_y} \left(\sum_{j=0}^{N_z} \exp\left(-\frac{\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}_{(m,n,j)}\|^2}{4\varepsilon(k-q)\Delta t} \right) \right.\right.\right. \\ &\qquad \left.\left.\left. V(\widetilde{\boldsymbol{x}}_{(m,n,j)}) \widehat{\varphi}_{\varepsilon,\text{approx}}(t_0 + q\Delta t, \widetilde{\boldsymbol{x}}_{(m,n,j)}) \right) \right) \right) \Delta x \Delta y \Delta z \Delta t. \end{aligned} \tag{5.36}$$

Notice that specializing (5.36) for $k = 1$ recovers (5.35), as expected.

As earlier, we set

$$\widehat{\varphi}_{\varepsilon,\text{approx}}(t_0 + k\Delta t, \boldsymbol{x}) = \widehat{\varphi}_{\text{heat}}(t = t_0 + k\Delta t, \boldsymbol{x}) + \widehat{\varphi}_{\text{left}}(t_0 + k\Delta t, \boldsymbol{x}).$$

Applying this result, we complete **Step 2** of our algorithm using

$$\widehat{\varphi}_{\varepsilon}(t_1, \boldsymbol{x}) \approx \widehat{\varphi}_{\varepsilon,\text{approx}}(t_1, \boldsymbol{x}) = \widehat{\varphi}_{\text{heat}}(t_1, \boldsymbol{x}) + \widehat{\varphi}_{\text{left}}(t_1, \boldsymbol{x}).$$

Likewise, **Step 4** of our algorithm is completed via this left Riemann sum approximation in conjunction with the change of variable $\bar{t} = t_0 + t_1 - t$.

Remark 5.1. *An alternative way to numerically solve the IVPs associated with* (5.7) *is to use the Feynman-Kac path integral [24, Ch. 8.2], [152], [153, Ch. 3.3], which was recently used to solve a class of SBPs [154]. Note that this also requires a function approximation oracle.*

I note that a convergent solution for the Schrödinger factor recursion determines these factors in a projectivized unique sense, i.e., the recursion finds $\left(c\widehat{\varphi}_{\varepsilon,0}(\boldsymbol{x}), \frac{1}{c}\varphi_{\varepsilon,1}(\boldsymbol{x})\right)$, and thus $\left(c\widehat{\varphi}_{\varepsilon}(t,\boldsymbol{x}), \frac{1}{c}\varphi_{\varepsilon}(t,\boldsymbol{x})\right)$ for $t \in [t_0, t_1]$, up to an arbitrary constant $c > 0$. The product of these factors is the optimally controlled joint PDF $\rho_{\text{opt}}(t, \boldsymbol{x})$, which is unique in the usual sense.

5.8 Numerical Simulation

I illustrate our proposed algorithm for solving the L-SBP in a low Earth orbit (LEO) transfer case study as in [147] and [155]. Specifically, we consider stochastic transfer of a spacecraft from initial mean position $[5000, 10000, 2100]^\top$ km to a final mean position $[-14600, 2500, 7000]^\top$ km in ECI coordinates for a fixed flight time horizon $[t_0, t_1] = [0, 1\text{ hour}]$.

For numerical conditioning, we re-scale the variables as $\boldsymbol{x}' := \boldsymbol{x}/R$, $t' := t/T$, where the normalization constants $R = 6600$ km and $T = 5399$ s. In these re-scaled coordinates, the potential (5.11) becomes

$$\overline{V}(\boldsymbol{x}') = -\frac{\mu_{\text{new}} R}{T\|\boldsymbol{x}'\|_2} - \frac{\mu_{\text{new}} J_2 R_{\text{Earth}}^2}{2RT\|\boldsymbol{x}'\|_2^3}\left(1 - \frac{3R^2(z')^2}{R^2\|\boldsymbol{x}'\|_2^2}\right) \tag{5.37}$$

$$= -\frac{\mu_{\text{new}} R}{T\|\boldsymbol{x}'\|_2}\left(1 + \frac{J_2 R_{\text{Earth}}^2}{2R\|\boldsymbol{x}'\|_2^2}\right)\left(1 - \frac{3(z')^2}{\|\boldsymbol{x}'\|_2^2}\right), \tag{5.38}$$

$$\mu_{\text{new}} := \frac{\mu T}{R^2} \text{ km/s}. \tag{5.39}$$

To ensure that all controlled state sample paths stay above the surface of Earth, we add an regularizer to the potential $\overline{V}$, and write the regularized version of (5.39) as

$$\overline{V}_{\text{reg}}(\boldsymbol{x}') = -\frac{\mu_{\text{new}} R}{T\|\boldsymbol{x}'\|_2}\left(1 + \frac{J_2 R_{\text{Earth}}^2}{2R\|\boldsymbol{x}'\|_2^2}\right)\left(1 - \frac{3(z')^2}{\|\boldsymbol{x}'\|_2^2}\right) + \gamma\left(\|\boldsymbol{x}'\|_2^2 - \frac{(R_{\text{Earth}} + c)^2}{R^2}\right), \tag{5.40}$$

where the constant buffer c is chosen in our simulation such that $R_{\text{Earth}} + c = 6560$ Km, and $\gamma > 0$ is the regularization weight.

Using the pushforward of probability measures, the initial and final PDFs in the new coordinates become

$$\rho_0'(\boldsymbol{x}') = R^3 \rho_0(R\boldsymbol{x}'), \quad \rho_1'(\boldsymbol{x}') = R^3 \rho_1(R\boldsymbol{x}'). \tag{5.41}$$

Letting $\widehat{\phi}(t', \boldsymbol{x}') := \widehat{\varphi}_\varepsilon(t, \boldsymbol{x})$, and noting that $\frac{\partial \widehat{\varphi}_\varepsilon}{\partial t}(t, \boldsymbol{x}) = \frac{1}{T}\frac{\partial \widehat{\phi}}{\partial t'}(t', \boldsymbol{x}')$, $\Delta_{\boldsymbol{x}} \widehat{\varphi}_\varepsilon(t, \boldsymbol{x}) = \frac{1}{R^2}\Delta_{\boldsymbol{x}'}\widehat{\phi}(t', \boldsymbol{x}')$, I rewrite (5.7a) as

$$\frac{\partial \widehat{\phi}}{\partial t'}(t', \boldsymbol{x}') = \left(\frac{\varepsilon T}{R^2}\Delta_{\boldsymbol{x}'} + \frac{T}{2\varepsilon}\overline{V}_{\text{reg}}(\boldsymbol{x}')\right)\widehat{\phi}(t', \boldsymbol{x}'), \tag{5.42}$$

and apply the left Riemann sum approach to solve the corresponding PDE IVP. Likewise, we apply a change of variable $\phi(t', \boldsymbol{x}') := \varphi_\varepsilon(t, \boldsymbol{x})$ to (5.7b).

I choose $\rho_0 = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\rho_1 = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ with $\boldsymbol{\mu}_0 = [5000, 10000, 2100]^\top$, $\boldsymbol{\mu}_1 = [-14600, 2500, 7000]^\top$ as in [147, 155], $\boldsymbol{\Sigma}_0 = \operatorname{diag}(\boldsymbol{\mu}_0^2)/100, \boldsymbol{\Sigma}_1 = \operatorname{diag}(\boldsymbol{\mu}_1^2)/100$, where the vector exponentiation is element-wise. Using (5.41), we implement the proposed algorithm in the re-scaled coordinates $(t', \boldsymbol{x}')$. Note that while we use ρ_0, ρ_1 as Gaussians for this specific simulation, our method applies for any non-Gaussian PDFs with finite second moments.

I consider the spatial grid $\left[-\frac{30000}{R}, \frac{10000}{R}\right] \times \left[-\frac{5000}{R}, \frac{25000}{R}\right] \times \left[-\frac{5000}{R}, \frac{25000}{R}\right]$ and set $N_x = N_y = N_z = 12$, $N_t = 50$, $\varepsilon = 15000$, $\gamma = 7.5$. We perform the Schrödinger factor recursion (**Step 6** in our algorithm) for 7 iterations which we found sufficient for convergence. The optimally controlled joint $\rho'_{\text{opt}}(t', \boldsymbol{x}')$, computed from **Step 8**, is then transformed back to

$$\rho_{\text{opt}}(t, \boldsymbol{x}) = \frac{1}{R^3}\rho'(t', \boldsymbol{x}') = \frac{1}{R^3}\rho'(t/T, \boldsymbol{x}/R).$$

Likewise, we obtain the optimal control

$$\boldsymbol{u}_{\text{opt}}(t, \boldsymbol{x}) = 2\varepsilon\nabla_{\boldsymbol{x}} \log\varphi(t, \boldsymbol{x}) = \frac{2\varepsilon}{R}\nabla_{\boldsymbol{x}'} \log\phi(t', \boldsymbol{x}').$$

We evaluate $(\rho_{\text{opt}}, \boldsymbol{u}_{\text{opt}})$ on a grid of size $100 \times 100 \times 100 \times N_t$ which is denser than the original grid of size $N_x \times N_y \times N_z \times N_t$.

To evaluate the effectiveness of our proposed solution, we perform a closed loop simulation using the computed optimal policy $\boldsymbol{u}_{\text{opt}}(t, \boldsymbol{x})$. We draw 50 samples from the known ρ_0, and use the Euler-Maruyama scheme to propagate these samples forward in time from $t_0 = 0$ s to $t_1 = 3600$ s, using the closed loop Itô SDE (E.1). As the values of $\boldsymbol{u}_{\text{opt}}(t, \boldsymbol{x})$ are known only on the $100\times100\times100\times N_t$ grid, we use the nearest neighbor approximation to query $\boldsymbol{u}_{\text{opt}}$ at a out-of-grid sample during SDE integration. The 50 optimally controlled closed-loop state sample paths thus computed, are

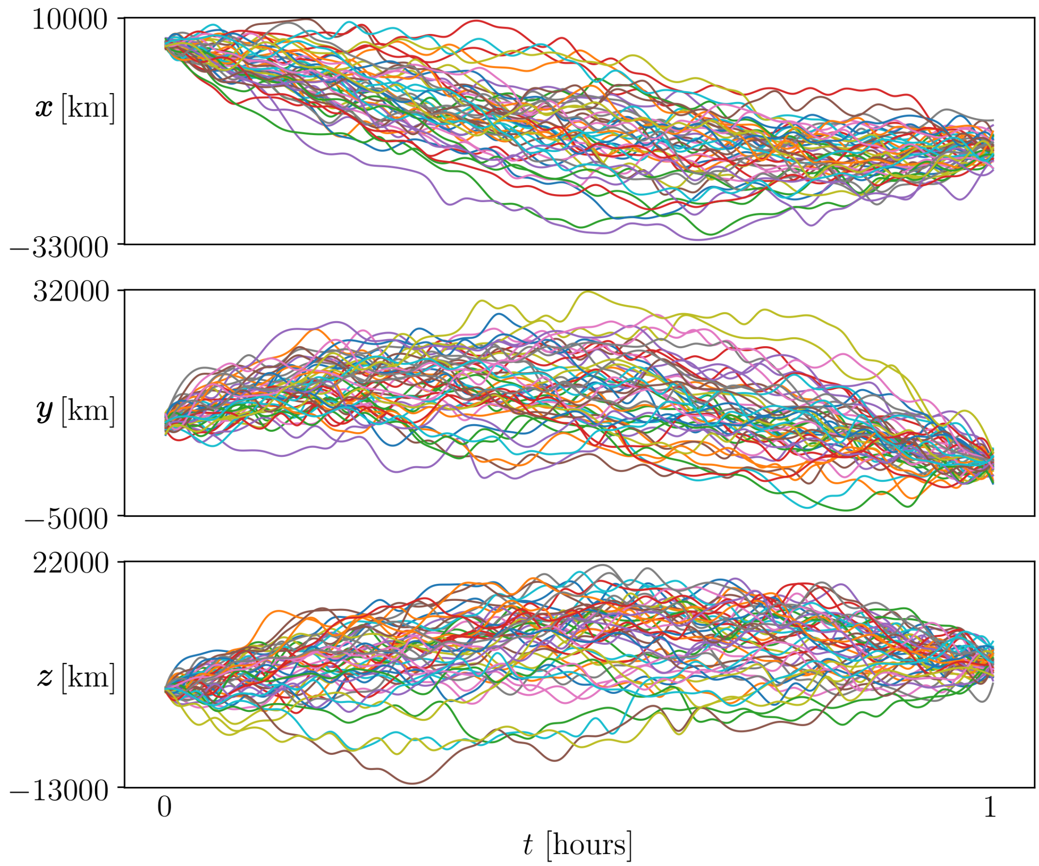

Figure 5.1: Optimally controlled closed loop state sample paths for the numerical simulation.

shown in Figure 5.1, demonstrating the transfer of the controlled stochastic state from ρ_0 to ρ_1.

Figure 5.2 shows the corresponding sample paths of the optimal controls.

In Figure 5.3, the filled circles depict the evolution (*red* initial and *blue* final) of the means of the aforesaid 50 optimally controlled sample path ensemble in $\mathbb{R}^3$. The translucent ellipsoids therein display one standard deviations around the means. We include the ECI coordinate in Figure 5.3 for reference, and to emphasize that our regularizer in (5.40) successfully keeps the optimally controlled sample paths above the Earth's surface.

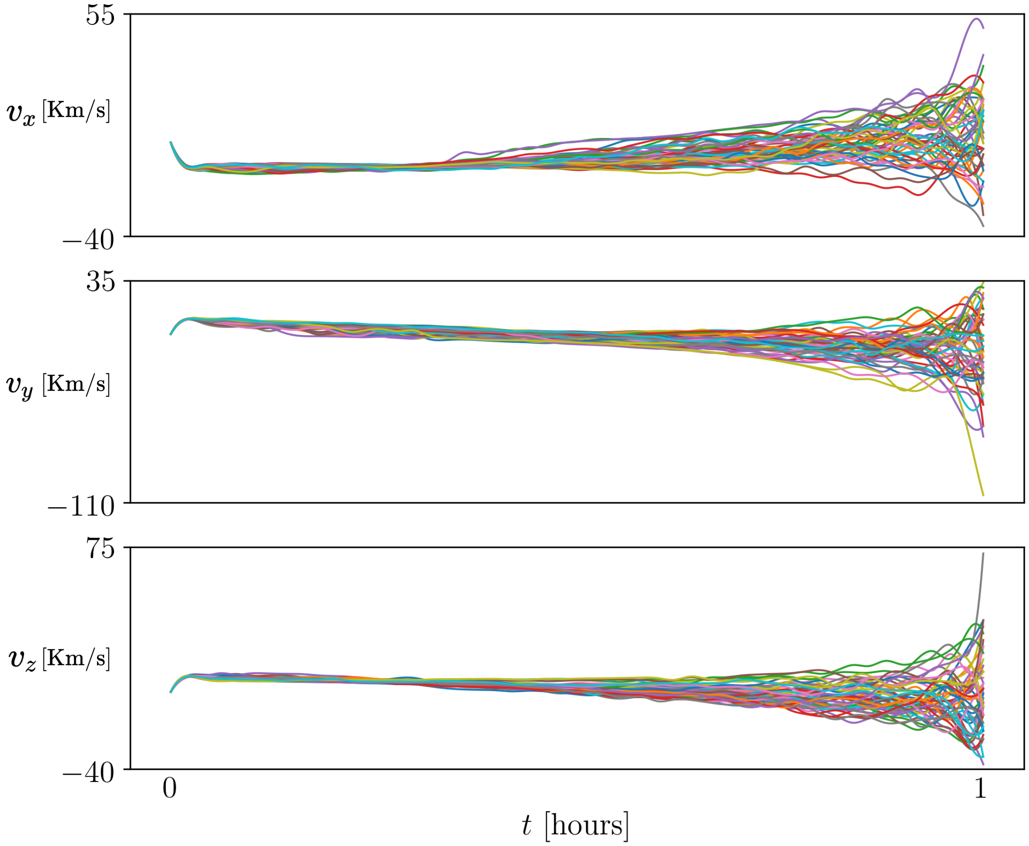

Figure 5.2: Sample paths for the components u_x, u_y, u_z of the optimal control for the numerical simulation.

Figure 5.4 plots five snapshots for the univariate x, y, z position marginals computed from the optimally controlled joint PDF $\rho_{\text{opt}}(t, \boldsymbol{x})$. This again highlights successful transfer of the stochastic state from given ρ_0 to given ρ_1 over the given flight time horizon $[t_0, t_1]$. From Figure 5.4, we also note that the optimally controlled marginals (and thus the joint PDFs) at the intermediate times are non-Gaussians even though the endpoints ρ_0, ρ_1 are Gaussians in this simulation case study.

Our numerical simulation demonstrates that the SBP with additive state cost, *not necessarily quadratic*, can be solved using the updated algorithm introduced in this chapter. Moreover, as the case study is adapted from a practical problem of interest in the aerospace community, we simultaneously demonstrate the widespread appeal of solutions to generalizations of the SBP.

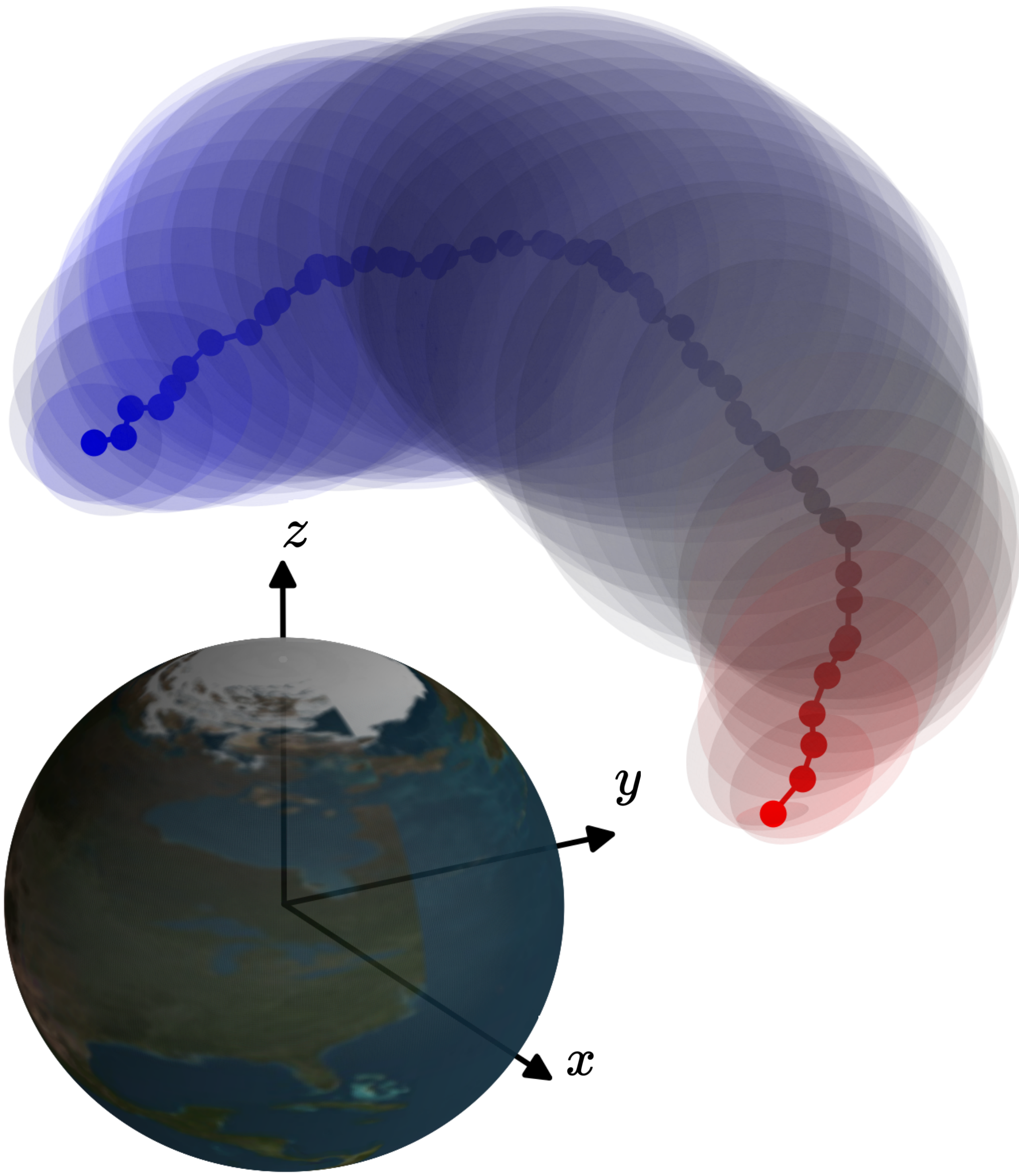

Figure 5.3: The filled circles show the mean position snapshots for the 50 optimally controlled sample paths in $\mathbb{R}^3$. The translucent ellipsoids denote one standard deviation around each mean position.

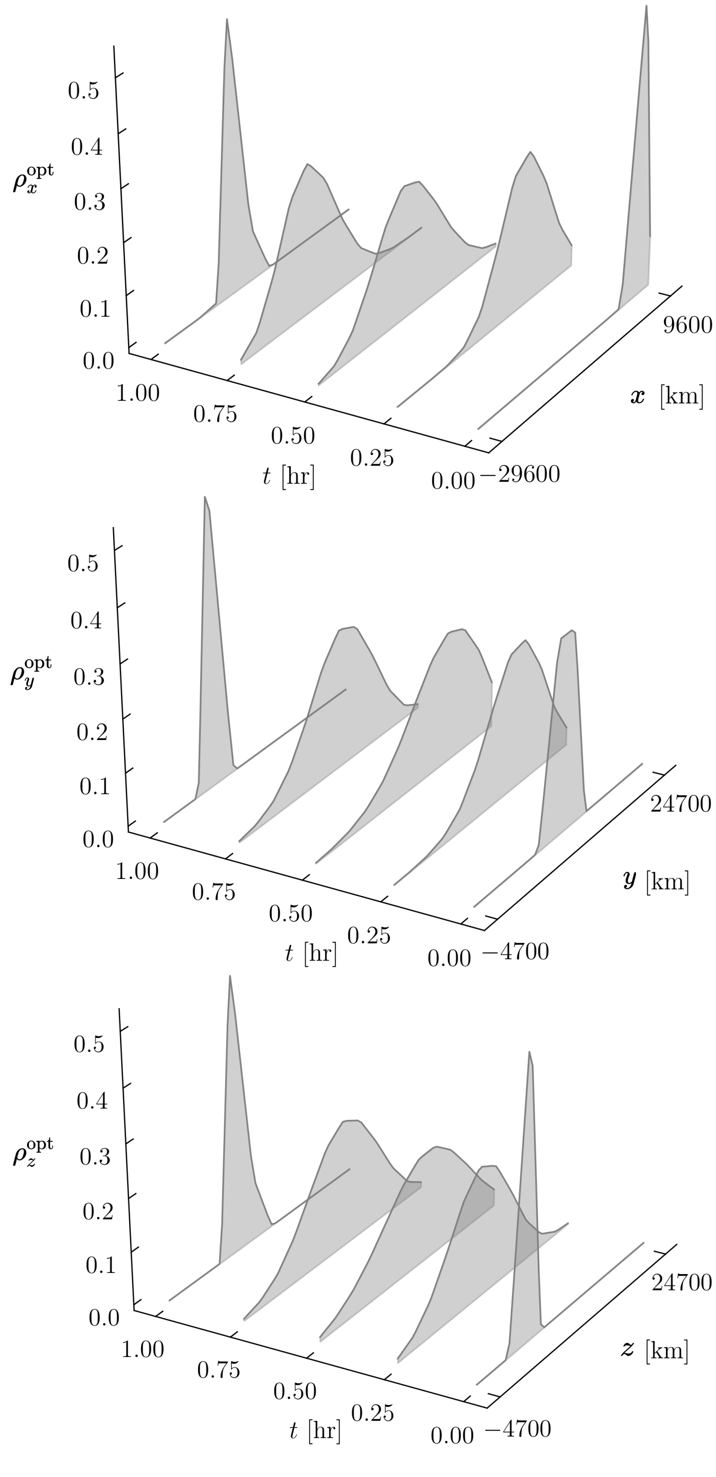

Figure 5.4: Univariate x, y, z marginals for the optimally controlled joint $\rho_{\rm opt}(t, \boldsymbol{x})$ at $t = 0, 0.25, 0.50, 0.75, 1$ hours.

6 | Future Directions and Summary

6.1 Overview

In this concluding chapter, I start by pointing out a new connection between the control-affine SBP and a two point boundary value problem for the quantum-mechanical Schrödinger PDE. These developments reveal new vistas to explore in future work following the original motivation behind Schrödinger's 1931-32 papers, namely new points of contacts between quantum mechanics and non-equilibrium statistical physics. Our original results in Chapter 6.2 give credence to the fact that much remains to be done along this direction.

In Chapter 6.3, I conclude by summarizing the technical contributions made in this dissertation.

6.2 Connection to the Schrödinger PDE

In 1931-32, Erwin Schrödinger posed [121, 122] the question: What is the most likely probability density-valued continuous curve connecting two given probability density functions when the prior dynamics is Brownian motion? The diffusion process generating this curve is now known as the *Schrödinger bridge*, and has found widespread applications in stochastic control [17, 20, 21, 30, 34, 77], generative AI [29, 70, 156–158], and throughout this dissertation. In its original incarnation, the SBP is a stochastic calculus of variations problem most naturally described through theory of large deviations [159], specifically by conditional Sanov's theorem [39, 140, 160]. That the Schrödinger bridge admits stochastic optimal control reformulation, was understood only in the late 20th century [161–164].

This section is similar in spirit to Schrödinger's original motivation: to find points of contact between quantum mechanics and non-equilibrium statistical mechanics. In particular, we focus again on the control-affine SBP discussed in Section of Chapter 1. Recall that the formulation of the control-affine SBP is as follows:

$$\underset{(\rho,\boldsymbol{u})\in\mathcal{P}_{01}\times\mathcal{U}}{\arg\inf} \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(q\left(t,\boldsymbol{x}\right)+\frac{1}{2}\|\boldsymbol{u}\|_2^2\right)\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \tag{6.1a}$$

$$\frac{\partial\rho}{\partial t}+\nabla_{\boldsymbol{x}}\cdot\left(\rho\left(\boldsymbol{A}\left(t,\boldsymbol{x}\right)+\boldsymbol{B}\left(t,\boldsymbol{x}\right)\boldsymbol{u}\right)\right)=\Delta_{\boldsymbol{\Sigma}(t,\boldsymbol{x},\boldsymbol{u})}\,\rho, \tag{6.1b}$$

$$\rho(t=t_0,\boldsymbol{x})=\rho_0,\quad \rho(t=t_1,\boldsymbol{x})=\rho_1. \tag{6.1c}$$

Recall that ρ is the probability density function for the stochastic state $\boldsymbol{x}(t)\in\mathbb{R}^n$ that follows the control-affine Itô stochastic differential equation:

$$\mathrm{d}\boldsymbol{x}_t^{\boldsymbol{u}}=\left(\boldsymbol{A}\left(t,\boldsymbol{x}\right)+\boldsymbol{B}\left(t,\boldsymbol{x}\right)\boldsymbol{u}\right)\mathrm{d}t+\boldsymbol{\sigma}(t,\boldsymbol{x})\mathrm{d}\boldsymbol{w}, \tag{6.2}$$

and the diffusion tensor $\boldsymbol{\Sigma}:=\boldsymbol{\sigma}\boldsymbol{\sigma}^\top\geq\boldsymbol{0}$. In (6.2), the $\boldsymbol{w}\in\mathbb{R}^p$ is standard Brownian motion, and the control $\boldsymbol{u}\in\mathcal{U}:=\{\boldsymbol{v}:[t_0,t_1]\times\mathbb{R}^n\mapsto\mathbb{R}^m\mid\|\boldsymbol{v}\|_2^2<\infty\}$, the collection of Markovian finite energy inputs. Note also that the composition $\nabla_{\boldsymbol{x}}\circ\nabla_{\boldsymbol{x}}=\mathrm{Hess}_{\boldsymbol{x}}$, the Hessian and the inner product $\langle\nabla_{\boldsymbol{x}},\nabla_{\boldsymbol{x}}\rangle=\Delta_{\boldsymbol{x}}$, the standard Laplacian.

The given data for problem (6.1) are: the time horizon $[t_0,t_1]$, the endpoint state PDFs ρ_0,ρ_1 having finite second moments, the bounded measurable state cost $q(\cdot)$, and the $\boldsymbol{A},\boldsymbol{B},\boldsymbol{\sigma}$ in (6.2) satisfying the standard assumptions:

A1. $\exists c_1,c_2>0$ such that $\forall\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$, $\forall t\in[t_0,t_1]$,

$$\|\boldsymbol{A}(t,\boldsymbol{x})\|_2+\|\boldsymbol{\sigma}\left(t,\boldsymbol{x}\right)\|_2\leq c_1\left(1+\|\boldsymbol{x}\|_2\right),$$

$$\|\boldsymbol{A}(t,\boldsymbol{x})-\boldsymbol{A}(t,\boldsymbol{y})\|_2\leq c_2\|\boldsymbol{x}-\boldsymbol{y}\|_2.$$

A2. $\exists c_3 > 0$ such that $\forall \boldsymbol{x} \in \mathbb{R}^n$, $\forall t \in [t_0, t_1]$,

$$\langle \boldsymbol{x}, \boldsymbol{\Sigma}(t, \boldsymbol{x})\boldsymbol{x} \rangle \geq c_3 \|\boldsymbol{x}\|_2^2.$$

From a control-theoretic perspective, the control-affine SBP is of more interest than the classical SBP. This is because for control systems of practical interest, the unforced dynamics often have nontrivial prior drift $\boldsymbol{A}$ and diffusion coefficient $\boldsymbol{\sigma}$, as opposed to standard Brownian motion. Also, practical control systems have limited control authority encoded by the input coefficient $\boldsymbol{B}$. Therefore, a better understanding of (6.1) is needed.

Again let $S \in \mathcal{C}^{1,2}([t_0, t_1]; \mathbb{R}^n)$ be the dual variable (value function) associated with the variational problem (6.1). Standard computation, as discussed throughout this dissertation, shows that the primal-dual pair (ρ_{opt}, S) for problem (6.1) solves the following system of coupled nonlinear PDEs:

$$\begin{aligned}
&\text{(primal PDE)} && \partial_t \rho_{\text{opt}} + \nabla_{\boldsymbol{x}} \cdot \left(\rho_{\text{opt}}\left(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{B}^\top \nabla_{\boldsymbol{x}} S\right)\right) \\
& && \qquad = \frac{1}{2}\Delta_{\boldsymbol{\Sigma}}\, \rho_{\text{opt}}, && \text{(6.3a)} \\
&\text{(dual PDE)} && \frac{\partial S}{\partial t} + \langle \nabla_{\boldsymbol{x}} S, \boldsymbol{A} \rangle + \frac{1}{2}\langle \nabla_{\boldsymbol{x}} S, \boldsymbol{B}\boldsymbol{B}^\top \nabla_{\boldsymbol{x}} S \rangle \\
& && \qquad + \frac{1}{2}\langle \boldsymbol{\Sigma}, \text{Hess}_{\boldsymbol{x}} S \rangle = q, && \text{(6.3b)} \\
&\text{(boundary conditions)} && \rho_{\text{opt}}(t_0, \boldsymbol{x}) = \rho_0(\boldsymbol{x}), \qquad \rho_{\text{opt}}(t_1, \boldsymbol{x}) = \rho_1(\boldsymbol{x}). && \text{(6.3c)}
\end{aligned}$$

The optimal control is

$$\boldsymbol{u}_{\text{opt}} = \boldsymbol{B}^\top \nabla_{\boldsymbol{x}} S. \tag{6.4}$$

Notice that (6.3a) is a controlled Fokker-Planck-Kolmogorov PDE in primal variable ρ_{opt} while (6.3b) is a Hamilton-Jacobi-Bellman (HJB) PDE in dual variable S.

Earlier chapters followed the standard approach of applying the Hopf-Cole transform to transform the system (6.3) into a form that is amenable for the contractive fixed point recursions, thus facilitating a numerical solution for (6.3).

Applying a different transform on (6.3) instead transcribes it to a quantum mechanical Schrödinger PDE boundary value problem (BVP) in unknown wave function. Let $\mathrm{i} := \sqrt{-1}$, the superscript † denote the complex conjugate, and

$$R := \frac{1}{2}\log \rho_{\mathrm{opt}}. \tag{6.5}$$

For fixed $\lambda > 0$, apply the Madelung transform $(\rho_{\mathrm{opt}}, S) \mapsto (\psi, \psi^{\dagger})$, or equivalently $(R, S) \mapsto (\psi, \psi^{\dagger})$ [165, Sec. 7], given by

$$\psi := \exp\left(R + \frac{\mathrm{i}}{\lambda}S\right), \tag{6.6a}$$

$$\psi^{\dagger} := \exp\left(R - \frac{\mathrm{i}}{\lambda}S\right). \tag{6.6b}$$

We refer to ψ as the *wave function*, and $\psi^{\dagger}$ as the *conjugate wave function*. Note that (6.5)-(6.6) imply *Born's relation* [166]:

$$\rho_{\mathrm{opt}}(t, \boldsymbol{x}) = \psi(t, \boldsymbol{x})\psi^{\dagger}(t, \boldsymbol{x}) \quad \forall t \in [t_0, t_1], \tag{6.7}$$

giving a factorization of ρ_{opt}. Because of complex conjugacy, it suffices to derive a single PDE BVP for the transformed variable $\psi(t, \boldsymbol{x})$, unlike the case for the Hopf-Cole transform.

At the level of generality of the control-affine SBP (6.1), it is not immediately clear whether the transformed PDE BVP can be related to the quantum mechanical Schrödinger PDE. Even if this is possible, it is unclear what the corresponding quantum potential's structure would be.

In the remainder of this chapter, I present our results, as in [78], on transcribing the optimality conditions for the control-affine SBP (6.1), namely (6.3)-(6.4), into a system of equations involving ψ. We specialize these results for the classical SBP, resulting in a Schrödinger PDE. Connections to Bohm's interpretation of quantum mechanics emerge as a consequence of these calculations.

The connections between the Schrödinger bridge and the Schrödinger PDE in quantum mechanics were explored by Nagasawa [165, Sec. 7], where a Schrödinger process was derived from the Schrödinger PDE. Similar connections were pursued by Guerra and Morato [41].

Also starting from the Schrödinger PDE, Ohsumi derived [167] a stochastic control problem in the spirit of inverse optimal control. However, the resulting formulation is not a Schrödinger bridge.

More broadly, several works [75, 168–170] have discussed connections between non-quantum stochastic control and quantum mechanics.

In contrast to the existing works, our developments start from the control-affine SBP–a concrete stochastic optimal control problem–and from there, derive suitable versions of the Schrödinger PDE.

As a point of clarity: Our results concern transforming the non-quantum Schrödinger bridge (as in classical Markov diffusion process) to the Schrödinger PDE. We do not study the quantum Schrödinger bridge [171, 172].

6.2.1 Dynamics of the Wave Function for the Control Affine SBP

We start with the following Lemma.

Lemma 6.1 (Weighted Laplacian of R)**.** *For R defined as in* (6.5)*, we have*

$$\frac{1}{4\rho_{\text{opt}}}\Delta_{\boldsymbol{\Sigma}}\rho_{\text{opt}} = \frac{1}{2}\Delta_{\boldsymbol{\Sigma}}R + \left(\nabla_{\boldsymbol{x}}R\right)^{\top}\boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}R + \left(\frac{1}{4}-\frac{1}{2}R\right)\langle Hess_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle. \tag{6.8}$$

See Appendix F.1 for a proof of Lemma 6.8.

Using Lemma 6.1, we prove that the transformed PDE BVP is in the form

$$\mathrm{i}\lambda\frac{\partial\psi}{\partial t} = -\frac{\lambda^2}{2}\Delta_{\boldsymbol{\Sigma}}\psi + V_{\text{the control-affine SBP}}\psi, \tag{6.9a}$$

$$\psi(t_0,\boldsymbol{x})\psi^{\dagger}(t_0,\boldsymbol{x}) = \rho_0,\ \psi(t_1,\boldsymbol{x})\psi^{\dagger}(t_1,\boldsymbol{x}) = \rho_1, \tag{6.9b}$$

where the quantum potential $V_{\text{the control-affine SBP}}$ is complex.

Theorem 6.1. *Let ψ be as in* (6.6a)*. Then* (6.3) *is equivalent to* (6.9) *with complex potential $V_{\textit{the control-affine SBP}}$ having real ($\Re$) and imaginary ($\Im$) parts*

$$\begin{aligned}\Re\left(V_{\textit{the control-affine SBP}}\right) =&\frac{\lambda^2}{2}\langle \mathit{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle + \frac{\lambda^2}{2}\langle\boldsymbol{\Sigma},\mathit{Hess}_{\boldsymbol{x}}R\rangle + \langle\nabla_{\boldsymbol{x}}S,\boldsymbol{A}\rangle\\ &+\frac{\lambda^2}{2}\|\nabla_{\boldsymbol{x}}R\|^2_{\boldsymbol{\Sigma}} - \frac{1}{2}\|\nabla_{\boldsymbol{x}}S\|^2_{\boldsymbol{\Sigma}} + \lambda^2\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}R\rangle\\ &+\frac{1}{2}\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}S\rangle + \frac{1}{2}\langle\boldsymbol{\Sigma},\mathit{Hess}_{\boldsymbol{x}}S\rangle - q,\end{aligned} \tag{6.10}$$

$$\begin{aligned}\Im\left(V_{\textit{the control-affine SBP}}\right) =&\lambda\Big\{\frac{1}{2}\langle\boldsymbol{\Sigma},\mathit{Hess}_{\boldsymbol{x}}S\rangle + \left(\nabla_{\boldsymbol{x}}R\right)^{\top}\boldsymbol{\Sigma}\left(\nabla_{\boldsymbol{x}}S\right) + \langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}S\rangle\\ &-\langle\nabla_{\boldsymbol{x}}R,\boldsymbol{A}+\boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}S\rangle - \frac{1}{2}\nabla_{\boldsymbol{x}}\cdot\left(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}S\right)\\ &+\frac{1}{2}\Delta_{\boldsymbol{\Sigma}}R + \left(\nabla_{\boldsymbol{x}}R\right)^{\top}\boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}R + \left(\frac{1}{4}-\frac{1}{2}R\right)\langle\mathit{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle\Big\},\end{aligned} \tag{6.11}$$

where the squared weighted norm $\|\cdot\|^2_{\boldsymbol{\Sigma}} \coloneqq (\cdot)^{\top}\boldsymbol{\Sigma}(\cdot)$.

For a proof of Theorem 6.1, see Appendix F.2.

Note that as a first step of deriving a PDE for ψ in Appendix F.2, we combine Lemma 6.1 with the primal-dual PDEs (6.3a)-(6.3b), to obtain

$$\begin{aligned}\frac{1}{\psi}\frac{\partial\psi}{\partial t} =&\frac{\partial R}{\partial t} + \frac{\mathrm{i}}{\lambda}\frac{\partial S}{\partial t}\\ =&\frac{1}{2\rho_{\mathrm{opt}}}\left(-\nabla_{\boldsymbol{x}}\cdot\left(\rho_{\mathrm{opt}}\left(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}S\right)\right) + \frac{1}{2}\Delta_{\boldsymbol{x}}\rho_{\mathrm{opt}}\right)\\ &+\frac{\mathrm{i}}{\lambda}\left(-\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{A}\rangle - \frac{1}{2}\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}S\rangle - \frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}S\rangle + q\right)\\ =&-\langle\nabla_{\boldsymbol{x}}R,\boldsymbol{A}+\boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}S\rangle - \frac{1}{2}\nabla_{\boldsymbol{x}}\cdot\left(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}S\right)\end{aligned}$$

$$+\frac{1}{2}\Delta_{\boldsymbol{\Sigma}} R + (\nabla_{\boldsymbol{x}} R)^{\top} \boldsymbol{\Sigma} \nabla_{\boldsymbol{x}} R + \left(\frac{1}{4} - \frac{1}{2}R\right) \langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma} \rangle$$
$$+\frac{\mathrm{i}}{\lambda}\left(-\langle \nabla_{\boldsymbol{x}} S, \boldsymbol{A}\rangle - \frac{1}{2}\langle \nabla_{\boldsymbol{x}} S, \boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}} S\rangle - \frac{1}{2}\langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} S\rangle + q\right). \tag{6.12}$$

This first step is worthy of note outside of the proof in Appendix F.2, because it follows from (6.12) that the dynamics of R is given by

$$\frac{\partial R}{\partial t} = -\langle \nabla_{\boldsymbol{x}} R, \boldsymbol{A} + \boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}} S\rangle - \frac{1}{2}\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}} S)$$
$$+\frac{1}{2}\Delta_{\boldsymbol{\Sigma}} R + \|\nabla_{\boldsymbol{x}} R\|^2_{\boldsymbol{\Sigma}} + \left(\frac{1}{4} - \frac{1}{2}R\right)\langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle. \tag{6.13}$$

Using the solution of the BVP (6.9), one can use (6.7) to compute the optimally controlled joint PDF at any $t \in [t_0, t_1]$. Combining (6.4) and (6.6), the optimal control

$$\boldsymbol{u}_{\mathrm{opt}} = -\frac{\mathrm{i}\lambda}{2}\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}} \log \frac{\psi}{\psi^{\dagger}}(t, \boldsymbol{x})$$

, which is a real function of $(t, \boldsymbol{x})$ since $\nabla_{\boldsymbol{x}} S$ is real.

6.2.2 The Case $\Sigma = \lambda B B^{\top}$

A special case that arises in practice is when the control and the noise coefficients are proportional, i.e., $\boldsymbol{\Sigma} \propto \boldsymbol{B}\boldsymbol{B}^{\top}$. This case occurs when the process noise models imperfect actuation, or when the noise enters through input (e.g., force, torque, current) channels.

Notice that in Sec. 6.2.1, the $\lambda > 0$ was an arbitrary constant. By interpreting λ as the proportionality constant in $\boldsymbol{\Sigma} = \lambda \boldsymbol{B}\boldsymbol{B}^{\top}$, the expressions (6.10)-(6.11) specialize to

$$\Re\left(V^{\lambda}_{\text{the control-affine SBP}}\right) = \frac{\lambda^2}{2}\langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle + \frac{\lambda^2}{2}\langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} R\rangle$$
$$+\frac{\lambda^2}{2}\|\nabla_{\boldsymbol{x}} R\|^2_{\boldsymbol{\Sigma}} + \lambda^2\langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} R\rangle + \langle \nabla_{\boldsymbol{x}} S, \boldsymbol{A}\rangle$$

$$
\begin{aligned}
&+\left(\frac{1}{2\lambda}-\frac{1}{2}\right)\|\nabla_{\boldsymbol{x}}S\|_{\boldsymbol{\Sigma}}^{2}+\frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}S\rangle-q, &(6.14)\\
\mathfrak{I}\left(V^{\lambda}_{\text{the control-affine SBP}}\right)=&\frac{\lambda-1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}S\rangle+(\lambda-1)\left\langle\nabla_{\boldsymbol{x}}R,\boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}S\right\rangle\\
&+\left(\lambda-\frac{1}{2}\right)\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}S\rangle-\lambda\langle\nabla_{\boldsymbol{x}}R,\boldsymbol{A}\rangle-\frac{\lambda}{2}\nabla_{\boldsymbol{x}}\cdot\boldsymbol{A}\\
&+\frac{\lambda}{2}\Delta_{\boldsymbol{\Sigma}}R+\lambda\|\nabla_{\boldsymbol{x}}R\|_{\boldsymbol{\Sigma}}^{2}+\left(\frac{\lambda}{4}-\frac{\lambda}{2}R\right)\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle, &(6.15)
\end{aligned}
$$

where the superscript λ in $V^{\lambda}_{\text{the control-affine SBP}}$ indicates the proportionality constant.

A further special case of interest is $\lambda=1$, i.e., when the input and noise channels are the same [20, 21]. Then, three terms drop from (6.14)-(6.15), giving

$$
\begin{aligned}
\mathfrak{R}\left(V^{1}_{\text{the control-affine SBP}}\right)=&\frac{1}{2}\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle+\frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}R\rangle+\frac{1}{2}\|\nabla_{\boldsymbol{x}}R\|_{\boldsymbol{\Sigma}}^{2}\\
&+\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}R\rangle+\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{A}\rangle+\frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}S\rangle-q, &(6.16)\\
\mathfrak{I}\left(V^{1}_{\text{the control-affine SBP}}\right)=&\frac{1}{2}\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}S\rangle-\langle\nabla_{\boldsymbol{x}}R,\boldsymbol{A}\rangle-\frac{1}{2}\nabla_{\boldsymbol{x}}\cdot\boldsymbol{A}\\
&+\frac{1}{2}\Delta_{\boldsymbol{\Sigma}}R+\|\nabla_{\boldsymbol{x}}R\|_{\boldsymbol{\Sigma}}^{2}+\left(\frac{1}{4}-\frac{1}{2}R\right)\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle, &(6.17)
\end{aligned}
$$

wherein the superscript 1 in $V^{1}_{\text{the control-affine SBP}}$ indicates $\lambda=1$.

We further specialize (6.16)-(6.17) for the classical SBP, and derive a PDE in the form of the more familiar Schrödinger PDE:

$$
\mathrm{i}\frac{\partial\psi}{\partial t}=-\frac{1}{2}\Delta_{\boldsymbol{x}}\psi+V_{\text{the classical SB}}\psi, \tag{6.18}
$$

(6.18). We then explain how the resulting $V_{\text{the classical SB}}$ generalizes the Bohm potential.

6.2.3 Reduction to Standard Schrödinger PDE

Recall that the classical SBP corresponds to the following special case of (6.1): $q \equiv 0$, $\boldsymbol{A} \equiv \boldsymbol{0}$, $\boldsymbol{B} \equiv \boldsymbol{\sigma} \equiv \boldsymbol{I}$. These choices specialize (6.16)-(6.17) to

$$\Re\left(V_{\text{the classical SB}}\right) = \frac{1}{2}\Delta_{\boldsymbol{x}} R + \frac{1}{2}\|\nabla_{\boldsymbol{x}} R\|_2^2 + \frac{1}{2}\Delta_{\boldsymbol{x}} S, \tag{6.19}$$

$$\Im\left(V_{\text{the classical SB}}\right) = \frac{1}{2}\Delta_{\boldsymbol{x}} R + \|\nabla_{\boldsymbol{x}} R\|_2^2. \tag{6.20}$$

Since $\Delta_{\boldsymbol{I}} \equiv \Delta_{\boldsymbol{x}}$ (the standard Laplacian), for the potential $V_{\text{the classical SB}}$ given by (6.19)-(6.20), the PDE (6.9a) reduces to (6.18), which is the more familiar form of the quantum mechanical Schrödinger PDE. The boundary conditions (6.9b) remain unchanged.

In the case of the classical SBP, the PDEs for R, S, given by (6.13) and (6.3b) respectively, specialize to

$$\frac{\partial R}{\partial t} = -\langle \nabla_{\boldsymbol{x}} R, \nabla_{\boldsymbol{x}} S \rangle - \frac{1}{2}\Delta_{\boldsymbol{x}} S + \Im\left(V_{\text{the classical SB}}\right), \tag{6.21a}$$

$$\frac{\partial S}{\partial t} = -\frac{1}{2}\|\nabla_{\boldsymbol{x}} S\|_2^2 - \frac{1}{2}\Delta_{\boldsymbol{x}} S. \tag{6.21b}$$

6.2.4 Connections with Bohm

We now discuss the connections between (6.21) and the Bohm's interpretation of quantum mechanics. For a particle of mass $m > 0$, Bohm's interpretation [173, 174] of the Schrödinger PDE

$$\mathrm{i}\hbar\frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta_{\boldsymbol{x}}\psi + V\psi, \tag{6.22}$$

where $\hbar > 0$ is the reduced Planck's constant and V is a suitable potential, is "analogous to (but not identical with) classical equations of motion" [173]. Similar to (6.6), Bohm considers

$$\psi = R_{\mathrm{B}} \exp(\mathrm{i} S_{\mathrm{B}}/\hbar), \tag{6.23}$$

where the subscript B indicates Bohm, and writes the PDEs [173, eq. (3)-(4)]

$$\frac{\partial R_{\mathrm{B}}}{\partial t} = -\frac{1}{2m} R_{\mathrm{B}} \Delta_{\boldsymbol{x}} S_{\mathrm{B}} - \frac{1}{m} \langle \nabla_{\boldsymbol{x}} R_{\mathrm{B}}, \nabla_{\boldsymbol{x}} S_{\mathrm{B}} \rangle, \tag{6.24a}$$

$$\frac{\partial S_{\mathrm{B}}}{\partial t} = -\frac{\|\nabla_{\boldsymbol{x}} S_{\mathrm{B}}\|_2^2}{2m} - V + \frac{\hbar^2}{2m} \frac{\Delta_{\boldsymbol{x}} R_{\mathrm{B}}}{R_{\mathrm{B}}}. \tag{6.24b}$$

Comparing (6.23) with (6.6), $R = \log R_{\mathrm{B}}$, and we can rewrite (6.24) as

$$\frac{\partial R}{\partial t} = -\frac{1}{2m} \Delta_{\boldsymbol{x}} S_{\mathrm{B}} - \frac{1}{m} \langle \nabla_{\boldsymbol{x}} R, \nabla_{\boldsymbol{x}} S_{\mathrm{B}} \rangle, \tag{6.25a}$$

$$\frac{\partial S_{\mathrm{B}}}{\partial t} = -\frac{\|\nabla_{\boldsymbol{x}} S_{\mathrm{B}}\|_2^2}{2m} - V + \frac{\hbar^2}{2m} \Delta_{\boldsymbol{x}} R. \tag{6.25b}$$

Bohm interprets the term $\frac{\hbar^2}{2m} \Delta_{\boldsymbol{x}} R$ in the HJB PDE (6.25b) as the negative of a "quantum potential"[1] that acts in addition to the "classical potential" V. This allows us to view (6.25b) as the deterministic HJB PDE for a particle whose motion is governed by the sum of these two potentials, akin to classical mechanics. Bohm's parenthetical remark that the potential is not identical to the classical case refers to the nonlocal nature of this interpretation, i.e., the potential itself depends on R, and thus on the joint state PDF.

To seek connections among our (6.21) and Bohm's (6.25), we identify the value function S_{B} with S, the potential V with V_{SB}, and consider normalizations $\hbar = 1$, $m = 1$.

We notice that if $\mathfrak{I}\left(V_{\text{the classical SB}}\right) = 0$, then (6.21a) matches with (6.25a).

To relate (6.21b) with (6.25b), notice that the aforementioned identifications allow

[1]which later came to be known as the Bohm potential [175].

us to rewrite (6.25b) as

$$\frac{\partial S}{\partial t} = -\frac{1}{2}\|\nabla_{\boldsymbol{x}} S\|_2^2 - \frac{1}{2}\Delta_{\boldsymbol{x}} S - \frac{1}{2}\|\nabla_{\boldsymbol{x}} R\|_2^2 - \mathrm{i}\Im\left(V_{\text{the classical SB}}\right).$$

In particular, if $\Im\left(V_{\text{the classical SB}}\right) = 0$, then (6.25b) becomes

$$\frac{\partial S}{\partial t} = -\frac{1}{2}\|\nabla_{\boldsymbol{x}} S\|^2 - \frac{1}{2}\Delta_{\boldsymbol{x}} S + \frac{1}{4}\Delta_{\boldsymbol{x}} R, \tag{6.26}$$

which is indeed (6.21b) except for the extra term $\frac{1}{4}\Delta_{\boldsymbol{x}} R$, that is a scaled "quantum mechanical" or Bohm potential.

We have thus shown that if the potential $V_{\text{the classical SB}}$ were real (i.e., elastic scattering), then the "complexified"[2] Schrödinger bridge admits a Bohmian interpretation. This is expected because Bohm's interpretation in [173] tacitly considers the V in (6.25b) to be real.

Our results show that transforming the conditions of optimality for the classical SBP to the Schrödinger PDE (6.18) is possible only if we allow the potential $V_{\text{the classical SB}}$ to be complex, i.e., both the real and the imaginary parts (6.19)-(6.20) are nonzero. In this sense, $V_{\text{the classical SB}}$ can be seen as a generalized Bohm potential. To the best of our knowledge, the fact that a one-to-one correspondence exists between the classical SBP and the Schrödinger PDE with a complex potential has not been described before.

Remark 6.1. *The use of a complex potential in Schrödinger PDE is common in nuclear physics [176, 177], where it is called the optical potential [178–180]. Here "optical" refers to an analogous situation of having complex refractive index, whose real and imaginary parts describe the transmission and absorption of light in an optical medium, respec-*

[2]Here "complexified" simply means that the transformed variable ψ in (6.6) is a complex function of $(t, \boldsymbol{x})$. This juxtaposes with the Hopf-Cole transform [23, Sec. IV-C], where the transformed variables, called the *Schrödinger factors*, are real functions of $(t, \boldsymbol{x})$.

tively. Likewise, the real part of the Schrödinger potential encodes elastic scattering (transmission of wave function), and the imaginary part encodes inelastic scattering (absorption of wave function).

Remark 6.2. *To explicitly see why* $\Im(V_{\mathrm{SB}}) \neq 0$*, notice from* (6.20) *that otherwise* $\frac{1}{2}\Delta_{\boldsymbol{x}} R + \|\nabla_{\boldsymbol{x}} R\|_2^2 = 0$*, which by* (6.5)*, is equivalent to* $\Delta_{\boldsymbol{x}}\rho_{\mathrm{opt}} = 0$*. Then,* (6.3a) *reduces to*

$$\frac{\partial \rho_{\mathrm{opt}}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\rho_{\mathrm{opt}} \boldsymbol{u}_{\mathrm{opt}}\right) = 0. \tag{6.27}$$

On the other hand, from (6.3a)*, the optimal control* $\boldsymbol{u}_{\mathrm{opt}}^{\varepsilon}$ *for the control-affine SBP with* $q \equiv 0$*,* $\boldsymbol{A} \equiv \boldsymbol{0}$*,* $\boldsymbol{B} = \boldsymbol{\sigma} = \sqrt{\varepsilon}\boldsymbol{I}$*,* $\varepsilon > 0$*, satisfies*

$$\frac{\partial \rho_{\mathrm{opt}}^{\varepsilon}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\rho_{\mathrm{opt}}^{\varepsilon} \boldsymbol{u}_{\mathrm{opt}}^{\varepsilon}\right) = \frac{\varepsilon}{2}\Delta_{\boldsymbol{x}}\rho_{\mathrm{opt}}^{\varepsilon}, \tag{6.28}$$

where it is known [15, 16] that

$$\boldsymbol{u}_{\mathrm{opt}}^{\varepsilon} \xrightarrow{\varepsilon \downarrow 0} \boldsymbol{u}_{\mathrm{opt}}^{\varepsilon=0} \neq \boldsymbol{u}_{\mathrm{opt}}^{\varepsilon=1} =: \boldsymbol{u}_{\mathrm{opt}},$$

thereby contradicting (6.27).

6.2.5 Remarks

Through the above calculations, we clarify the connections between the control affine Schrödinger bridge and the Schrödinger PDE. The former is a problem in (non-quantum) stochastic control about optimal steering of probability distribution that has become a topic of significant interest in control theory and generative AI. In contrast, the Schrödinger PDE is the foundational equation in quantum mechanics.

The necessary conditions of optimality for the control affine Schrödinger bridge problem can be transformed to suitable versions of the Schrödinger PDE boundary value problems. Our calculations reveal that the corresponding Schrödinger PDE must have a certain structured complex potential that is a non-trivial generalization of the Bohm potential in quantum mechanics. Like the original Bohm potential, the derived generalization is nonlocal. Our work demonstrates that the non-equilibrium statistical mechanics of the control affine Schrödinger bridge problem induces an absorbing medium for the wave function that manifests as a nonzero imaginary part of the potential.

A promising direction of future work would be to explore the possibility of inverse scattering-like computation to solve the two point boundary value problem for the Schrödinger equation derived here. This would open up the possibility to steer distributions by steering waves.

6.3 Summary of Results

This dissertation addresses two key types generalizations of the SBP. One generalization concerns the prior drift and diffusion coefficients in the dynamical constraint, and the other concerns the presence of additive state cost in the objective. We address various aspects of the Sinkhorn recursions used to solve the classical SBP under these generalizations.

We discuss guarantees on the rate of convergence of these Sinkhorn recursions in the case of classical SBP, and derive similar guarantees for the first generalization on the dynamical constraint. As a result, we uncover novel control-theoretic and geometric interpretations of the worst-case contraction coefficients quantifying the rate of convergence in solving such generalized SBPs via certain fixed point recursions.

Several variants of the second generalization are addressed. For a fixed quadratic state cost, we derive the necessary kernel k to show that this generalized SBP is exactly solveable. We provide numerical simulations of this solution. In the process, we present two pathways to this kernel k, one involving the use of Hermite polynomials and a second involving a detailed discussion of Weyl calculus.

We also investigate the linear SBP with quadratic state cost, which combines the two types of generalizations of the SBP considered in this dissertation. To solve this generalization, we present a third, most general, pathway to deriving k.

Finally, we drop the restriction that the state cost must be quadratic, and discuss existence-uniqueness results for the SBP with additive state cost, under mild assumptions using the large deviation principle over the path space with respect to suitable Gibbs path measure. Subsequently, we show that the additive state cost manifests as reaction rate in coupled reaction-diffusion PDEs, and provide a numerical algorithm for finding the solution to these PDEs. We establish that the probabilistic Lambert's problem, of interest in the astrodynamics community, precisely leads to such a formulation, and we solve the same using the tools developed as part of analyzing such SBPs.

We conclude with a recent discovery connecting the control-affine SBP with the Schrödinger PDE. This connection is very new, and worthy of future investigation.

A | Proofs for Chapter 1

A.1 Proof of Theorem 1.1

Proof. Consider the Lagrangian L for the constrained variational problem (1.24):

$$\begin{aligned}
\mathrm{L}\left(\rho, \boldsymbol{u}, S\right) := &\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(q + \frac{1}{2}\|\boldsymbol{u}\|_2^2\right)\rho\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \\
&+ \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\nabla_{\boldsymbol{x}}\cdot\left(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{u}\right)\right)S\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \\
&+ \int_{t_0}^{t_1}\int_{\mathbb{R}^n} S\frac{\partial\rho}{\partial t}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t - \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\Delta_{\boldsymbol{\Sigma}}\rho^{\boldsymbol{u}}\right)S\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t,
\end{aligned} \tag{A.1}$$

where the Lagrange multiplier $S \in \mathcal{C}^{1,2}\left(\left[t_0, t_1\right]; \mathbb{R}^n\right)$. The idea now is to perform integration-by-parts for the second, third and fourth integrands in (A.1) such that the differential operators acting on ρ gets transferred to those of S. This will then allow us to re-write (A.1) as a linear functional in ρ only.

Specifically, for the second integral in (A.1), we perform integration-by-parts w.r.t. $\boldsymbol{x}$. Assuming the limits at $\|\boldsymbol{x}\|_2 \to \infty$ are zero, we find

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\nabla_{\boldsymbol{x}}\cdot\left(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{u}\right)\right)S\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = -\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\langle\nabla_{\boldsymbol{x}}S, \boldsymbol{A}+\boldsymbol{B}\boldsymbol{u}\rangle\rho\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{A.2}$$

For the third integral in (A.1), we apply the Fubini-Torreli theorem to switch the order of integration and perform integration-by-parts w.r.t. t, to obtain

$$\begin{aligned}
&\int_{t_0}^{t_1}\int_{\mathbb{R}^n} S\frac{\partial\rho}{\partial t}\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \\
&\quad = \underbrace{\int_{\mathbb{R}^n}\left(S(t_1,\boldsymbol{x})\rho_1(\boldsymbol{x}) - S(t_0,\boldsymbol{x})\rho_0(\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x}}_{\text{constant over } \mathcal{P}_{01}\times\mathcal{U}} - \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\frac{\partial S}{\partial t}\rho\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t.
\end{aligned} \tag{A.3}$$

For the fourth integral in (A.1), two-fold integration by parts w.r.t. $\boldsymbol{x}$ yields

$$\int_{\mathbb{R}^n}\left(\Delta_{\boldsymbol{\Sigma}}\rho\right)S\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = \int_{\mathbb{R}^n}\langle\boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}}S\rangle\rho\,\mathrm{d}\boldsymbol{x}. \tag{A.4}$$

Substituting (A.3), (A.2), (A.4) in (A.1), we obtain

$$\mathrm{L}\left(\rho, \boldsymbol{u}, S\right) = \text{constant} + \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\Big(q + \frac{1}{2}\|\boldsymbol{u}\|_2^2 - \frac{\partial S}{\partial t} - \langle\nabla_{\boldsymbol{x}} S, \boldsymbol{A} + \boldsymbol{B}\boldsymbol{u}\rangle \\ - \langle\boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} S\rangle\Big)\,\rho\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \quad \text{(A.5)}$$

To minimize (A.5) w.r.t. $\boldsymbol{u}$, we set

$$\boldsymbol{u}^{\mathrm{opt}} - \boldsymbol{B}^\top\nabla_{\boldsymbol{x}} S = 0$$

which yields

$$\boldsymbol{u}_{\mathrm{opt}} = \boldsymbol{B}^\top\nabla_{\boldsymbol{x}} S,$$

the optimal control given in (1.27).

Substituting (1.27) in the primal feasibility constraint (1.24b), we arrive directly at (1.26a).

Substituting (1.27) into

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(q + \frac{1}{2}\|\boldsymbol{u}_{\mathrm{opt}}\|_2^2 - \frac{\partial S}{\partial t} - \langle\nabla_{\boldsymbol{x}} S, \boldsymbol{A} + \boldsymbol{B}\boldsymbol{u}_{\mathrm{opt}}\rangle - \langle\boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} S\rangle\right)\rho\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t,$$

and requiring the integral to be zero, yields the dynamic programming equation:

$$\begin{aligned}&\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\Big(q + \frac{1}{2}\|\boldsymbol{B}^\top\nabla_{\boldsymbol{x}} S\|_2^2 - \frac{\partial S}{\partial t} \\ &\qquad - \langle\nabla_{\boldsymbol{x}} S, \boldsymbol{A} + \boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}} S\rangle - \langle\boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} S\rangle\Big)\,\rho\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \\ &= \int_{t_0}^{t_1}\int_{\mathbb{R}^n}\Big(q - \frac{\partial S}{\partial t} - \langle\nabla_{\boldsymbol{x}} S, \boldsymbol{A}\rangle \\ &\qquad - \frac{1}{2}\langle\nabla_{\boldsymbol{x}} S, \boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}} S\rangle - \langle\boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} S\rangle\Big)\rho\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = 0. \quad \text{(A.6)}\end{aligned}$$

Since (A.6) should be satisfied for arbitrary ρ, we must have

$$q - \frac{\partial S}{\partial t} - \langle\nabla_{\boldsymbol{x}} S, \boldsymbol{A}\rangle - \frac{1}{2}\langle\nabla_{\boldsymbol{x}} S, \boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}} S\rangle - \langle\boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} S\rangle = 0,$$

which yields (1.26b).

Finally, the restriction of ρ to $\mathcal{P}_{01}$ in (1.24a) immediately yields (1.26c). □

A.2 Proof of Theorem 1.2

Proof. This proof will make use of the following lemma.

Lemma A.1 (Weighted Laplacian of a product)**.** *Given scalar fields* $\alpha, \beta : \mathbb{R}^n \mapsto \mathbb{R}$ *and symmetric matrix field* $\boldsymbol{\Sigma} : \mathbb{R}^n \mapsto \mathbb{S}^n$ *such that* $\alpha, \beta, \boldsymbol{\Sigma} \in \mathcal{C}^2(\mathbb{R}^n)$*, we have*

$$\begin{aligned}\Delta_{\boldsymbol{\Sigma}}(\alpha\beta) = {} & \alpha\langle\boldsymbol{\Sigma}, \mathit{Hess}_{\boldsymbol{x}}\beta\rangle + \beta\langle\boldsymbol{\Sigma}, \mathit{Hess}_{\boldsymbol{x}}\alpha\rangle + 2\alpha\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}}\beta\rangle \\ & + 2\beta\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}}\alpha\rangle + \alpha\beta\underbrace{\langle \mathit{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle}_{=\mathbf{1}^\top\boldsymbol{\Sigma}\mathbf{1}} + 2\underbrace{\langle\nabla_{\boldsymbol{x}}\alpha, \boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}\beta\rangle}_{=\langle\nabla_{\boldsymbol{x}}\beta, \boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}\alpha\rangle},\end{aligned} \tag{A.7}$$

wherein the divergence of a matrix field is understood as a vector with elements

$$(\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma})_i := \sum_{j=1}^{n}\frac{\partial\Sigma_{ij}}{\partial x_j} \quad \forall i \in \{1, \cdots, n\}.$$

The following special cases of (A.7) *are also of interest:*

(the case $\beta \equiv 1$*)*

$$\Delta_{\boldsymbol{\Sigma}}\,\alpha = \alpha\langle \mathit{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle + \langle\boldsymbol{\Sigma}, \mathit{Hess}_{\boldsymbol{x}}\alpha\rangle + 2\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}}\alpha\rangle, \tag{A.8}$$

(the case $\boldsymbol{\Sigma} \equiv \boldsymbol{I}$*)*

$$\Delta_{\boldsymbol{x}}(\alpha\beta) = \alpha\Delta_{\boldsymbol{x}}\beta + \beta\Delta_{\boldsymbol{x}}\alpha + 2\langle\nabla_{\boldsymbol{x}}\alpha, \nabla_{\boldsymbol{x}}\beta\rangle. \tag{A.9}$$

Proof of Lemma A.1. From the definition of the $\Delta_{\boldsymbol{\Sigma}}$, given in (1.13), we find

$$\begin{aligned}\Delta_{\boldsymbol{\Sigma}}(\alpha\beta) &= \sum_{i,j=1}^{n}\frac{\partial}{\partial x_j}\left(\frac{\partial}{\partial x_i}\left(\alpha\beta\Sigma_{ij}\right)\right)\\ &= \sum_{i,j=1}^{n}\Bigg\{\alpha\beta\frac{\partial^2\Sigma_{ij}}{\partial x_i\partial x_j} + \Sigma_{ij}\alpha\frac{\partial^2\beta}{\partial x_i\partial x_j} + \Sigma_{ij}\beta\frac{\partial^2\alpha}{\partial x_i\partial x_j}\\ &\qquad + \alpha\left(\frac{\partial\beta}{\partial x_j}\frac{\partial\Sigma_{ij}}{\partial x_i} + \frac{\partial\beta}{\partial x_i}\frac{\partial\Sigma_{ij}}{\partial x_j}\right)\\ &\qquad + \beta\left(\frac{\partial\alpha}{\partial x_j}\frac{\partial\Sigma_{ij}}{\partial x_i} + \frac{\partial\alpha}{\partial x_i}\frac{\partial\Sigma_{ij}}{\partial x_j}\right)\\ &\qquad + \Sigma_{ij}\left(\frac{\partial\alpha}{\partial x_i}\frac{\partial\beta}{\partial x_j} + \frac{\partial\alpha}{\partial x_j}\frac{\partial\beta}{\partial x_i}\right)\Bigg\},\end{aligned} \tag{A.10}$$

which yields (A.7). The identities (A.8)-(A.9) follow by substituting $\beta \equiv 1$ and $\boldsymbol{\Sigma} \equiv \boldsymbol{I}$ in (A.7). $\square$

We first note that since $S \in \mathcal{C}^{1,2}\left([t_0,t_1];\mathbb{R}^n\right)$, S is bounded over its domain, and the mapping (1.28) is bijective. In particular, $\widehat{\varphi}(t,\boldsymbol{x}), \varphi(t,\boldsymbol{x})$ are positive for all $t \in [t_0,t_1]$.

Multiplying $\widehat{\varphi}_\lambda(t,\boldsymbol{x})$ and $\varphi_\lambda(t,\boldsymbol{x})$ from (1.28) yields (1.32). Substituting $S = \lambda\log\varphi_\lambda$ in (1.27) yields (1.33). Specializing the same for $t = t_0, t_1$, and using (1.26c), we obtain (1.29c).

What remains is to deduce (1.29a)-(1.29b). Substituting

$$S = \lambda\log\varphi_\lambda$$

in (1.26b), using the Hessian identity

$$\begin{aligned}\text{Hess}_{\boldsymbol{x}}\left(\lambda\log\varphi_\lambda\right) &= \lambda\nabla_{\boldsymbol{x}}\circ\nabla_{\boldsymbol{x}}\log\varphi_\lambda\\ &= -\frac{\lambda}{\varphi_\lambda^2}\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)^\top + \frac{\lambda}{\varphi_\lambda}\text{Hess}_{\boldsymbol{x}}\varphi_\lambda,\end{aligned} \tag{A.11}$$

followed by algebraic simplification, gives

$$
\begin{aligned}
&\frac{\lambda}{\varphi_\lambda}\frac{\partial\varphi_\lambda}{\partial t}+\frac{\lambda}{\varphi_\lambda}\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{A}\rangle+\frac{\lambda^2}{2\varphi_\lambda}\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
&\qquad+\langle\boldsymbol{\Sigma},-\frac{\lambda}{\varphi_\lambda^2}\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)^\top\rangle+\langle\boldsymbol{\Sigma},\frac{\lambda}{\varphi_\lambda}\mathrm{Hess}_{\boldsymbol{x}}\varphi_\lambda\rangle=q\\
&\frac{\partial\varphi_\lambda}{\partial t}+\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{A}\rangle+\frac{\lambda}{2}\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
&\qquad-\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle+\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\varphi_\lambda\rangle=\frac{q\varphi_\lambda}{\lambda}\\
&\frac{\partial\varphi_\lambda}{\partial t}+\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{A}\rangle+\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\varphi_\lambda\rangle-\frac{q\varphi_\lambda}{\lambda}\\
&\qquad+\frac{1}{2}\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle=0.
\end{aligned}
\tag{A.12}
$$

We next add and subtract (another)

$$
\frac{1}{2}\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle
$$

to the LHS of (A.12), then group the resulting $+\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle$ with the drift term in (A.12), and the

$$
-\frac{1}{2}\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle=-\frac{\varphi_\lambda}{2}\langle\nabla_{\boldsymbol{x}}\log\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\varphi\rangle
$$

with the reaction term in (A.12). This gives (1.29b).

To derive (1.29a), we substitute (1.32) and (1.33) in the primal PDE (1.26a). These substitutions allow us to write the LHS of (1.26a) as:

$$
\begin{aligned}
\mathrm{LHS}&=\widehat{\varphi_\lambda}\frac{\partial\varphi_\lambda}{\partial t}+\varphi_\lambda\frac{\partial\widehat{\varphi_\lambda}}{\partial t}+\nabla_{\boldsymbol{x}}\cdot\left(\varphi_\lambda\widehat{\varphi_\lambda}\boldsymbol{A}\right)+\nabla_{\boldsymbol{x}}\cdot\left(\widehat{\varphi_\lambda}\,\lambda\boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}}\varphi_\lambda\right)\\
&=\widehat{\varphi_\lambda}\Big\{-\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{A}\rangle-\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle-\varepsilon\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\varphi_\lambda\rangle\\
&\qquad+\left(\frac{q}{\lambda}+\frac{1}{2}\left(\nabla_{\boldsymbol{x}}\log\varphi_\lambda\right)^\top\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\right)\varphi_\lambda\Big\}\\
&\qquad+\varphi_\lambda\frac{\partial\widehat{\varphi_\lambda}}{\partial t}+\varphi_\lambda\langle\nabla_{\boldsymbol{x}}\widehat{\varphi_\lambda},\boldsymbol{A}\rangle+\widehat{\varphi_\lambda}\langle\nabla_{\boldsymbol{x}}\varphi_\lambda,\boldsymbol{A}\rangle+\varphi\widehat{\varphi_\lambda}\nabla_{\boldsymbol{x}}\cdot\boldsymbol{A}\\
&\qquad+\langle\nabla_{\boldsymbol{x}}\widehat{\varphi_\lambda},\lambda\boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}}\varphi_\lambda\rangle+\widehat{\varphi_\lambda}\nabla_{\boldsymbol{x}}\cdot\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top\nabla_{\boldsymbol{x}}\varphi_\lambda\right)
\end{aligned}
\tag{A.13}
$$

$$
\begin{aligned}
&= \varphi_\lambda \frac{\partial \widehat{\varphi_\lambda}}{\partial t} + \varphi_\lambda \nabla_{\boldsymbol{x}} \cdot \left(\widehat{\varphi_\lambda} \boldsymbol{A}\right) - \widehat{\varphi_\lambda} \langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} \varphi_\lambda \rangle \\
&\quad - \frac{\widehat{\varphi_\lambda}}{2} \langle \nabla_{\boldsymbol{x}} \log \varphi_\lambda, \left(\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}\right) \nabla_{\boldsymbol{x}} \varphi_\lambda \rangle + \langle \nabla_{\boldsymbol{x}} \widehat{\varphi_\lambda}, \lambda \boldsymbol{B}\boldsymbol{B}^\top \nabla_{\boldsymbol{x}} \varphi_\lambda \rangle \\
&\quad + \widehat{\varphi_\lambda} \langle \nabla_{\boldsymbol{x}} \cdot \lambda \boldsymbol{B}\boldsymbol{B}^\top \nabla_{\boldsymbol{x}}, \varphi_\lambda \rangle + \widehat{\varphi_\lambda} \langle \lambda \boldsymbol{B}\boldsymbol{B}^\top, \mathrm{Hess}_{\boldsymbol{x}} \varphi_\lambda \rangle + \varphi \frac{q \widehat{\varphi_\lambda}}{\lambda},
\end{aligned} \tag{A.14}
$$

where the expression inside the curly braces is due to (1.29b). Using (A.7) in Lemma A.1, the same substitutions unpack the RHS of (1.26a) as:

$$
\begin{aligned}
\mathrm{RHS} &= \Delta_{\boldsymbol{\Sigma}} \left(\varphi_\lambda \widehat{\varphi_\lambda}\right) \\
&= \varphi_\lambda \widehat{\varphi_\lambda} \langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma} \rangle + \varphi_\lambda \langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} \widehat{\varphi_\lambda} \rangle + \widehat{\varphi_\lambda} \langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} \varphi_\lambda \rangle \\
&\quad + 2\varphi_\lambda \langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} \widehat{\varphi_\lambda} \rangle + 2\widehat{\varphi_\lambda} \langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} \varphi_\lambda \rangle + 2 \langle \nabla_{\boldsymbol{x}} \varphi_\lambda, \boldsymbol{\Sigma} \nabla_{\boldsymbol{x}} \widehat{\varphi_\lambda} \rangle.
\end{aligned} \tag{A.15}
$$

For convenience, we refer to the six summands appearing in (A.15), from left to right, as term 1 to term 6. Comparing (A.14) with (A.15), we make few observations:

- term 3 in (A.15) can be combined with the terms $-\widehat{\varphi_\lambda} \langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} \varphi_\lambda \rangle$ and $\widehat{\varphi_\lambda} \langle \lambda \boldsymbol{B}\boldsymbol{B}^\top, \mathrm{Hess}_{\boldsymbol{x}} \varphi_\lambda \rangle$ in (A.14),
- term 6 in (A.15) can be combined with the term $\langle \nabla_{\boldsymbol{x}} \widehat{\varphi_\lambda}, \lambda \boldsymbol{B}\boldsymbol{B}^\top \nabla_{\boldsymbol{x}} \varphi_\lambda \rangle$ in (A.14),
- the sum of the terms 1, 2 and 4 in (A.15) equals $\varphi_\lambda \Delta_{\boldsymbol{\Sigma}} \widehat{\varphi_\lambda}$, thanks to (A.8) in Lemma A.1.

Consequently, equating (A.14) with (A.15), we obtain

$$
\begin{aligned}
&\varphi_\lambda \left\{ \frac{\partial \widehat{\varphi_\lambda}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(\widehat{\varphi_\lambda} \boldsymbol{A}\right) + \frac{q \widehat{\varphi_\lambda}}{\lambda} - \varepsilon \Delta_{\boldsymbol{\Sigma}} \widehat{\varphi_\lambda} \right\} \\
&\quad = \frac{\widehat{\varphi_\lambda}}{2} \langle \nabla_{\boldsymbol{x}} \log \varphi_\lambda, \left(\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}\right) \nabla_{\boldsymbol{x}} \varphi_\lambda \rangle \\
&\qquad - \widehat{\varphi_\lambda} \langle \lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} \varphi_\lambda \rangle - \langle \nabla_{\boldsymbol{x}} \widehat{\varphi_\lambda}, \left(\lambda \boldsymbol{B}\boldsymbol{B}^\top - 2\boldsymbol{\Sigma}\right) \nabla_{\boldsymbol{x}} \varphi_\lambda \rangle
\end{aligned}
$$

$$-\widehat{\varphi_\lambda}\langle\nabla_{\boldsymbol{x}}\cdot\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right),\nabla_{\boldsymbol{x}}\varphi_\lambda\rangle. \tag{A.16}$$

Dviding both sides of (A.16) by φ (recall that φ is positive everywhere by definition), we find

$$\begin{aligned}
&\frac{\partial\widehat{\varphi_\lambda}}{\partial t}+\nabla_{\boldsymbol{x}}\cdot\left(\widehat{\varphi_\lambda}\boldsymbol{A}\right)+\frac{q\widehat{\varphi_\lambda}}{\lambda}-\Delta_{\boldsymbol{\Sigma}}\,\widehat{\varphi_\lambda}\\
&\quad=\frac{\widehat{\varphi_\lambda}}{2}\langle\nabla_{\boldsymbol{x}}\log\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
&\qquad-\widehat{\varphi_\lambda}\langle\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma},\frac{1}{\varphi_\lambda}\mathrm{Hess}_{\boldsymbol{x}}\varphi_\lambda\rangle-\langle\nabla_{\boldsymbol{x}}\widehat{\varphi_\lambda},\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
&\qquad-\widehat{\varphi_\lambda}\langle\nabla_{\boldsymbol{x}}\cdot\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right),\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle
\end{aligned} \tag{A.17}$$

Since

$$\mathrm{Hess}_{\boldsymbol{x}}\log\varphi_\lambda=-\frac{1}{\varphi_\lambda^2}\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)^\top+\frac{1}{\varphi_\lambda}\mathrm{Hess}\varphi_\lambda,$$

we replace the second term in the RHS of (A.17) with

$$-\widehat{\varphi_\lambda}\langle\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma},\frac{1}{\varphi_\lambda^2}\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)\left(\nabla_{\boldsymbol{x}}\varphi_\lambda\right)^\top\rangle-\widehat{\varphi_\lambda}\langle\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\log\varphi_\lambda\rangle.$$

The RHS of (A.17) then simplifies to

$$\begin{aligned}
\mathrm{RHS}=&-\frac{\widehat{\varphi_\lambda}}{2}\langle\nabla_{\boldsymbol{x}}\log\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
&-\widehat{\varphi_\lambda}\langle\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
&-\langle\nabla_{\boldsymbol{x}}\widehat{\varphi_\lambda},\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
&-\widehat{\varphi_\lambda}\langle\nabla_{\boldsymbol{x}}\cdot\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right),\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
=&-\nabla_{\boldsymbol{x}}\cdot\left(\widehat{\varphi_\lambda}\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\right)\\
&-\frac{\widehat{\varphi_\lambda}}{2}\langle\nabla_{\boldsymbol{x}}\log\varphi_\lambda,\left(\lambda\boldsymbol{B}\boldsymbol{B}^\top-2\boldsymbol{\Sigma}\right)\nabla_{\boldsymbol{x}}\log\varphi_\lambda\rangle\\
\overset{(1.30),(1.31)}{=}&-\nabla_{\boldsymbol{x}}\cdot\left(\widehat{\varphi_\lambda}\boldsymbol{A}_\varphi\right)-q_\varphi\widehat{\varphi_\lambda},
\end{aligned}$$

which is exactly (1.29a). This completes the proof. □

B | Proofs for Chapter 2

B.1 Proof of Theorem 2.1

Proof. For a given $\varepsilon > 0$, consider the uncontrolled Ornstein-Uhlenbeck (OU) process with time-varying coefficients:

$$\mathrm{d}\boldsymbol{x}(t) = \boldsymbol{A}(t)\boldsymbol{x}(t)\mathrm{d}t + \sqrt{2\varepsilon}\boldsymbol{B}(t)\mathrm{d}\boldsymbol{w}(t),$$

with the associated Markov kernel (4.8) as

$$k_{\mathrm{L}}\left(t_0, \boldsymbol{x}_0, t_1, \boldsymbol{x}_1\right) = \frac{\exp\left(-\dfrac{\left(\boldsymbol{\Phi}_{t_1 t_0}\boldsymbol{x}_0 - \boldsymbol{x}_1\right)^{\top} \boldsymbol{M}_{10}^{-1}\left(\boldsymbol{\Phi}_{t_1 t_0}\boldsymbol{x}_0 - \boldsymbol{x}_1\right)}{4\varepsilon}\right)}{\sqrt{(4\pi\varepsilon)^n \det(\boldsymbol{M}_{10})}}. \tag{B.1}$$

As $\boldsymbol{M}_{10}$ is symmetric positive definite, its inverse $\boldsymbol{M}_{10}^{-1}$ and principal square root $\boldsymbol{M}_{10}^{-1/2}$ are both symmetric positive definite, and thus (B.1) is well-defined.

As was the case for the kernel k_{B} in (2.3), the kernel k_{L} too is continuous and positive for all $(\boldsymbol{x}_0, \boldsymbol{x}_1) \in \mathcal{X}_0 \times \mathcal{X}_1$. Compactness of $\mathcal{X}_0, \mathcal{X}_1$ implies that there exist constants $\alpha_{\mathrm{L}}, \beta_{\mathrm{L}}$ given by (2.16) such that $0 < \alpha_{\mathrm{L}} \leq \beta_{\mathrm{L}} < \infty$, and

$$\alpha_{\mathrm{L}} \leq k_{\mathrm{L}}\left(t_0, \boldsymbol{x}_0, t_1, \boldsymbol{x}_1\right) \leq \beta_{\mathrm{L}} \quad \forall\ (\boldsymbol{x}_0, \boldsymbol{x}_1) \in \mathcal{X}_0 \times \mathcal{X}_1. \tag{B.2}$$

From (2.16), (B.1) and (B.2), we then obtain the following analogue of (2.8):

$$\alpha_{\mathrm{L}} = \frac{\exp\left(-\widetilde{\alpha}_{\mathrm{L}}/(4\varepsilon)\right)}{\sqrt{(4\pi\varepsilon)^n \det(\boldsymbol{M}_{10})}},\ \beta_{\mathrm{L}} = \frac{\exp\left(-\widetilde{\beta}_{\mathrm{L}}/(4\varepsilon)\right)}{\sqrt{(4\pi\varepsilon)^n \det(\boldsymbol{M}_{10})}}. \tag{B.3}$$

Consequently, the worst-case contraction coefficient

$$\gamma_{\mathrm{L}} = \tanh^2\left(\frac{1}{2}\log\left(\frac{\beta_{\mathrm{L}}}{\alpha_{\mathrm{L}}}\right)\right) = \tanh^2\left(\frac{\widetilde{\alpha}_{\mathrm{L}} - \widetilde{\beta}_{\mathrm{L}}}{8\varepsilon}\right). \tag{B.4}$$

$\square$

C | Proofs for Chapter 3

C.1 Proof of Lemma 3.1

Proof. The result follows from Taylor expansion of $e^{-(\beta x+\alpha(t-t_0))^2}$ in variable t about t_0, and then identifying the coefficients as derivatives w.r.t. x of suitable order. $\square$

C.2 Proof for Lemma 3.3

Proof. Let

$$Y(y) := \exp\left(-\frac{y^2\sqrt{d}}{2}\right) f(y)$$

be the solution of the parametric ODE (3.13) for some unknown $f(y)$. Substituting this expression for $Y(y)$ into (3.13) yields:

$$\begin{aligned}&-\sqrt{d}\exp\left(-\frac{y^2\sqrt{d}}{2}\right)f + dy^2\exp\left(-\frac{y^2\sqrt{d}}{2}\right)f - y\sqrt{d}\exp\left(-\frac{y^2\sqrt{d}}{2}\right)\frac{\mathrm{d}f}{\mathrm{d}y}\\ &\quad - y\sqrt{d}\exp\left(-\frac{y^2\sqrt{d}}{2}\right)\frac{\mathrm{d}f}{\mathrm{d}y} + \exp\left(-\frac{y^2\sqrt{d}}{2}\right)\frac{\mathrm{d}^2 f}{\mathrm{d}y^2} - (dy^2+c)\exp\left(-\frac{y^2\sqrt{d}}{2}\right)f = 0.\end{aligned}$$

Simplifying the above, we obtain

$$\frac{\mathrm{d}^2 f}{\mathrm{d}y^2} - 2y\sqrt{d}\frac{\mathrm{d}f}{\mathrm{d}y} - \left(c+\sqrt{d}\right) f = 0.$$

Introducing a change of variable $y \mapsto \xi := d^{1/4}y$ and letting $F(\xi) := f(d^{-1/4}\xi)$, we rewrite the above ODE as

$$\sqrt{d}\frac{\mathrm{d}^2 F}{\mathrm{d}\xi^2} - 2d^{-1/4}\xi d^{1/4}\sqrt{d}\frac{\mathrm{d}F}{\mathrm{d}\xi} - \left(c+\sqrt{d}\right) F = 0.$$

Dividing through by $\sqrt{d}$ gives

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\xi^2} - 2\xi\frac{\mathrm{d}F}{\mathrm{d}\xi} - \left(\frac{c}{\sqrt{d}}+1\right) F = 0, \tag{C.1}$$

which is the Hermite's ODE [181, p. 328-329].

Assuming f, and thus F, must be polynomially bounded, the solutions to (C.1) is $F(\xi) = H_n(\xi)$, the Hermite polynomials of order $n = -\frac{c}{2\sqrt{d}} - \frac{1}{2} \in \mathbb{N}_0$. Therefore, the solutions to (3.13) can be written as $Y(y) = a\exp\left(-\frac{y^2\sqrt{d}}{2}\right)H_n\left(d^{1/4}y\right)$ for some constant $a \in \mathbb{R}$ and $n = -\frac{c}{2\sqrt{d}} - \frac{1}{2} \in \mathbb{N}_0$. □

C.3 Proof of Proposition 3.1

Proof. Substituting $t = t_0$ in (3.15) gives

$$\widehat{\eta}_0(\boldsymbol{y}) = \sum_{m_n=0}^{\infty} \dots \sum_{m_1=0}^{\infty} a_{\boldsymbol{m}} \prod_{i=1}^{n} \exp\left(-t_0\sqrt{d_i}\,(2m_i+1) - \frac{y_i^2\sqrt{d_i}}{2}\right) H_{m_i}\left(d_i^{1/4}y_i\right). \tag{C.2}$$

For a fixed $j_1 \in \mathbb{N}$, multiplying both sides of (C.2) by $H_{j_1}\left(d_1^{1/4}y_1\right)\exp\left(-\frac{y_1^2\sqrt{d_1}}{2}\right)$, and integrating over y_1, we get

$$\begin{aligned}
&\int_{-\infty}^{\infty} H_{j_1}\left(d_1^{1/4}y_1\right)\exp\left(-\frac{y_1^2\sqrt{d_1}}{2}\right)\widehat{\eta}_0(\boldsymbol{y})\mathrm{d}y_1 \\
&= \int_{-\infty}^{\infty} \sum_{m_n=0}^{\infty} \dots \sum_{m_1=0}^{\infty} a_{\boldsymbol{m}} \prod_{i=2}^{n} \exp\left(-t_0\sqrt{d_i}\,(2m_i+1) - \frac{y_i^2\sqrt{d_i}}{2}\right)\left(d_i^{1/4}y_i\right) \\
&\qquad \times H_{m_1}\left(d_1^{1/4}y_1\right) H_{j_1}\left(d_1^{1/4}y_1\right) \\
&\qquad \times \exp\left(-t_0\sqrt{d_1}\,(2m_1+1) - y_1^2\sqrt{d_1}\right)\mathrm{d}y_1.
\end{aligned}$$

Using (3.7), we simplify the above as follows:

$$\begin{aligned}
&\int_{-\infty}^{\infty} H_{j_1}(d_1^{1/4}y_1)\exp\left(-\frac{y_1^2\sqrt{d_1}}{2}\right)\widehat{\eta}_0(\boldsymbol{y})\mathrm{d}y_1 \\
&= \sum_{m_n=0}^{\infty} \dots \sum_{m_2=0}^{\infty} a_{(j_1,m_2,\dots,m_n)} \prod_{i=2}^{n} \exp\left(-t_0\sqrt{d_i}\,(2m_i+1) - \frac{y_i^2\sqrt{d_i}}{2}\right) H_{m_i}\left(d_i^{1/4}y_i\right) \\
&\qquad \int_{-\infty}^{\infty} H_{j_1}(d_1^{1/4}y_1)H_{j_1}(d_1^{1/4}y_1)\exp\left(-t_0\sqrt{d_1}\,(2j_1+1) - y_1^2\sqrt{d_1}\right)dy_1 \\
&= \sum_{m_n=0}^{\infty} \dots \sum_{m_2=0}^{\infty} a_{(j_1,m_2,\dots,m_n)} \prod_{i=2}^{n} \exp\left(-t_0\sqrt{d_i}\,(2m_i+1) - \frac{y_i^2\sqrt{d_i}}{2}\right)
\end{aligned}$$

$$H_{m_i}\!\left(d_i^{1/4}y_i\right)\frac{\sqrt{\pi}\,2^{j_1}\ (j_1!)}{d_1^{1/4}}\exp\left(-t_0\sqrt{d_1}(2j_1+1)\right).$$

Repeating this process $n-1$ more times, yields

$$\begin{aligned}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} H_{j_1}\left(d_1^{1/4}y_1\right)\exp\left(-\frac{y_1^2\sqrt{d_1}}{2}\right)\\ \ldots H_{j_n}\left(d_n^{1/4}y_n\right)\exp\left(-\frac{y_n^2\sqrt{d_n}}{2}\right)\widehat{\eta}_0(\boldsymbol{y})\,\mathrm{d}y_1\ldots\mathrm{d}y_n\\ = a_{(j_1,j_2,\ldots,j_n)}\prod_{i=1}^{n}\left(d_i^{-1/4}\sqrt{\pi}2^{j_i}(j_i!)\exp\left(-t_0\sqrt{d_i}(2j_i+1)\right)\right).\end{aligned}$$

Therefore, for any index vector $\boldsymbol{j}=(j_1,j_2,\ldots,j_n)\in\mathbb{N}_0^n$, we have

$$\begin{aligned}a_{\boldsymbol{j}} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{i=1}^{n}\Bigg(\frac{d_i^{1/4}}{\sqrt{\pi}2^{j_i}\,(j_i!)}H_{j_i}\left(d_i^{1/4}y_i\right)\\ \times\exp\left(t_0\sqrt{d_i}\,(2j_i+1)-\frac{y_i^2\sqrt{d_i}}{2}\right)\Bigg)\widehat{\eta}_0\,(\boldsymbol{y})\mathrm{d}y_1\ldots\mathrm{d}y_n.\end{aligned} \tag{C.3}$$

Substituting (C.3) for the coefficients in the general solution (3.15), and using $\boldsymbol{z}$ as the dummy integration variable, we arrive at (3.16)-(3.17). □

C.4 Proof of Theorem 3.1

Proof. We start by re-writing (3.16)-(3.17) as

$$\begin{aligned}\widehat{\eta}(t,\boldsymbol{y}) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{i=1}^{n}\\ \left\{\frac{d_i^{1/4}}{\sqrt{\pi}}\exp\left(-\frac{(y_i^2+z_i^2)\sqrt{d_i}}{2}-(t-t_0)\sqrt{d_i}\right)(\heartsuit)\right\}\widehat{\eta}_0(\boldsymbol{z})\,\mathrm{d}z_1\ldots\mathrm{d}z_n,\end{aligned} \tag{C.4}$$

where

$$(\heartsuit) := \sum_{m_i=0}^{\infty}\left(\frac{1}{2^{m_i}\,(m_i!)}\exp(-2m_i\,(t-t_0)\sqrt{d_i})H_{m_i}\left(d_i^{1/4}y_i\right)H_{m_i}\left(d_i^{1/4}z_i\right)\right). \tag{C.5}$$

For fixed $c > 0$, we use the Hermite polynomial identity:

$$
\begin{aligned}
H_r(cx) &= \frac{1}{c^r} e^{(cx)^2} \left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^r \left(e^{-(cx)^2}\right) \\
&= (-2\mathrm{i})^r \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-(\theta - cx\mathrm{i})^2} \theta^r \mathrm{d}\theta, \qquad r \in \mathbb{N}_0,
\end{aligned}
$$

where the first equality is due to definition (3.6), and the second equality is due to the Fourier representation

$$
e^{-x^2} = \int e^{-\theta^2 + 2\mathrm{i}x\theta} \mathrm{d}\theta / \sqrt{\pi}. \tag{C.6}
$$

Therefore,

$$
\begin{aligned}
(\heartsuit) = &\sum_{m_i=0}^\infty \frac{1}{2^{m_i}(m_i!)} \exp\left(-2m_i(t-t_0)\sqrt{d_i}\right) \times \frac{1}{d_i^{m_i/4}} e^{\left(d_i^{1/4} y_i\right)^2} \left(\frac{\mathrm{d}}{\mathrm{d}y_i}\right)^{m_i} e^{-\left(d_i^{1/4} y_i\right)^2} \\
&\times (-2\mathrm{i})^{m_i} \frac{1}{\sqrt{\pi}} \int_{t_0}^\infty e^{-\left(\theta - t_0 - d_i^{1/4} z_i \mathrm{i}\right)^2} (\theta - t_0)^{m_i} \mathrm{d}\theta \\
= &\frac{e^{\left(d_i^{1/4} y_i\right)^2}}{\sqrt{\pi}} \int_{t_0}^\infty \sum_{m_i=0}^\infty \left(\frac{(\varrho_i \mathrm{i}(\theta - t_0))^{m_i}}{(m_i!)\left(d_i^{1/4}\right)^{m_i}} \left(\frac{\mathrm{d}}{\mathrm{d}y_i}\right)^{m_i} e^{-\left(d_i^{1/4} y_i\right)^2} \right) e^{-\left(\theta - t_0 - d_i^{1/4} z_i \mathrm{i}\right)^2} \mathrm{d}\theta,
\end{aligned}
$$

where $\varrho_i := \exp\left(-2(t-t_0)\sqrt{d_i}\right)$. Invoking Lemma 3.1, we note that

$$
\sum_{m_i=0}^\infty \left(\frac{(\varrho_i \mathrm{i}(\theta - t_0))^{m_i}}{(m_i!)\left(d_i^{1/4}\right)^{m_i}} \left(\frac{\mathrm{d}}{\mathrm{d}y_i}\right)^{m_i} \left(e^{-(d_i^{1/4} y_i)^2}\right) \right) = e^{-\left(d_i^{1/4} y_i + \mathrm{i}\varrho_i(\theta - t_0)\right)^2},
$$

and so, $(\heartsuit)$ simplifies to

$$
\frac{e^{d_i^{1/2} z_i^2}}{\sqrt{\pi}} \int_{t_0}^\infty \exp\Big(-(1-\varrho_i^2)(\theta - t_0)^2 + 2d_i^{1/4}(\theta - t_0)\mathrm{i}\,(z_i - y_i \varrho_i)\Big) \mathrm{d}\theta. \tag{C.7}
$$

Noting that $0 < \varrho_i < 1$, we evaluate (C.7) by substituting $\sigma := \sqrt{1-\varrho_i^2}\,(\theta - t_0)$ as

$$
\begin{aligned}
(\heartsuit) &= \frac{e^{d_i^{1/2} z_i^2}}{\sqrt{1-\varrho_i^2}\sqrt{\pi}} \int_0^\infty \exp\left(-\sigma^2 + 2\mathrm{i}\sigma \frac{d_i^{1/4}(z_i - y_i\varrho_i)}{\sqrt{1-\varrho_i^2}}\right) \mathrm{d}\sigma \\
&= \frac{1}{\sqrt{1-\varrho_i^2}} \exp\left(\frac{2\varrho_i d_i^{1/2} y_i z_i - \varrho_i^2 d_i^{1/2}\left(y_i^2 + z_i^2\right)}{1-\varrho_i^2}\right),
\end{aligned} \tag{C.8}
$$

where the last equality follows from (C.6).

Combining (C.8) with (C.4), we obtain

$$\widehat{\eta}(t, \boldsymbol{y}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \kappa_{++}(t_0, \boldsymbol{y}, t, \boldsymbol{z}) \widehat{\eta}_0(\boldsymbol{z}) \mathrm{d}z_1 \dots \mathrm{d}z_n,$$

where the kernel

$$\kappa_{++}(t_0, \boldsymbol{y}, t, \boldsymbol{z}) := \prod_{i=1}^{n} \Bigg\{ \frac{d_i^{1/4}}{\sqrt{\pi}} \exp\left(-\frac{(y_i^2 + z_i^2)\sqrt{d_i}}{2} - (t - t_0)\sqrt{d_i} \right) \\ \frac{1}{\sqrt{1 - \varrho_i^2}} \exp\left(\frac{2\varrho_i d_i^{1/2} y_i z_i - \varrho_i^2 d_i^{1/2} (y_i^2 + z_i^2)}{1 - \varrho_i^2} \right) \Bigg\}.$$

Next, we rearrange the above expression for κ_{++} as

$$\kappa_{++}(t_0, \boldsymbol{y}, t, \boldsymbol{z}) = \frac{1}{\pi^{n/2}} \left(\prod_{i=1}^{n} d_i^{1/4} \right) \left(\prod_{i=1}^{n} \frac{1}{\sqrt{1 - \varrho_i^2}} \right) \times \left(\prod_{i=1}^{n} \exp\left(-(t - t_0)\sqrt{d_i} \right) \right) \\ \times \left(\prod_{i=1}^{n} \exp\left(\left(-\frac{1}{2} - \frac{\varrho_i^2}{1 - \varrho_i^2} \right) (y_i^2 + z_i^2)\sqrt{d_i} \right) \right) \\ \times \left(\prod_{i=1}^{n} \exp\left(\frac{2\varrho_i}{1 - \varrho_i^2} y_i z_i \sqrt{d_i} \right) \right). \tag{C.9}$$

Recalling the definitions of hyperbolic functions in terms of exponential, and that $\varrho_i := \exp\left(-2\,(t - t_0)\sqrt{d_i}\right)$, we find:

$$\begin{aligned} &1/\sqrt{1 - \varrho_i^2} = e^{(t - t_0)\sqrt{d_i}} / \sqrt{2\sinh(2\,(t - t_0)\sqrt{d_i})}, \\ &2\varrho_i/(1 - \varrho_i^2) = 1/\sinh(2\,(t - t_0)\sqrt{d_i}), \\ &-1/2 - \varrho_i^2/(1 - \varrho_i^2) = -\coth(2\,(t - t_0)\sqrt{d_i})/2. \end{aligned}$$

Hence, (C.9) can be expressed as

$$\kappa_{++}(t_0, \boldsymbol{y}, t, \boldsymbol{z}) = \frac{\left(\prod_{i=1}^{n} d_i^{1/4} \right)}{(2\pi)^{n/2}} \left(\prod_{i=1}^{n} \frac{1}{\sqrt{\sinh(2\,(t - t_0)\sqrt{d_i})}} \right) \\ \times \left(\prod_{i=1}^{n} \exp\left(\frac{\sqrt{d_i}}{\sinh(2\,(t - t_0)\sqrt{d_i})} \left(y_i z_i + (y_i^2 + z_i^2) \right.\right.\right. \\ \times \left.\left.\left.\left(-\frac{1}{2}\cosh(2\,(t - t_0)\sqrt{d_i}) \right) \right) \right) \right). \tag{C.10}$$

Letting $\boldsymbol{M}_{tt_0}$ as in (3.20) and noting that $\det(\boldsymbol{M}_{tt_0}) = \prod_{i=1}^{n} d_i > 0$, (C.10) simplifies to (3.19). □

C.5 Proof of Proposition 3.2

Proof. (i) Follows from direct computation.

(ii) Thanks to the hyperbolic function identity $\coth^2(\cdot) - \operatorname{csch}^2(\cdot) = 1$, definition (3.21) is satisfied, i.e., $\left(\boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)}\right)^\top \boldsymbol{J}\left(\boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)}\right) = \boldsymbol{J}$. Thus $\boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)} \in \mathrm{Sp}\left(2n,\mathbb{R}\right)$.

To see positive definiteness, first note that $\boldsymbol{M}_{tt_0}^{(1)} \succ 0$ for $\boldsymbol{D} \succ 0$ since

$$
\begin{aligned}
&\begin{pmatrix}\boldsymbol{y} & \boldsymbol{z}\end{pmatrix}\boldsymbol{M}_{tt_0}^{(1)}\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix}\\
&\quad=\begin{pmatrix}\boldsymbol{y} & \boldsymbol{z}\end{pmatrix}\begin{bmatrix}\cosh\left(2\left(t-t_0\right)\sqrt{\boldsymbol{D}}\right) & -\boldsymbol{I}\\ -\boldsymbol{I} & \cosh\left(2\left(t-t_0\right)\sqrt{\boldsymbol{D}}\right)\end{bmatrix}\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix}\\
&\quad=\boldsymbol{y}^\top\underbrace{\cosh\left(2\left(t-t_0\right)\sqrt{\boldsymbol{D}}\right)}_{\succ \boldsymbol{I}}\boldsymbol{y} - 2\langle\boldsymbol{y},\boldsymbol{z}\rangle + \boldsymbol{z}^\top\underbrace{\cosh\left(2\left(t-t_0\right)\sqrt{\boldsymbol{D}}\right)}_{\succ \boldsymbol{I}}\boldsymbol{z}\\
&\quad> \|\boldsymbol{y}-\boldsymbol{z}\|_2^2 \geq 0,
\end{aligned}
$$

where the second equality follows because $\cosh(\cdot) > 1$ for nonzero argument.

Also, $\boldsymbol{D} \succ \boldsymbol{0}$ implies the positive diagonal matrix $\boldsymbol{M}_{tt_0}^{(2)} \succ \boldsymbol{0}$, and therefore,

$$
\begin{aligned}
\operatorname{eig}\left(\boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)}\right) &= \operatorname{eig}\left(\boldsymbol{M}_{tt_0}^{(1)}\left(\boldsymbol{M}_{tt_0}^{(2)}\right)^{1/2}\left(\boldsymbol{M}_{tt_0}^{(2)}\right)^{1/2}\right)\\
&= \operatorname{eig}\left(\underbrace{\left(\boldsymbol{M}_{tt_0}^{(2)}\right)^{1/2}\boldsymbol{M}_{tt_0}^{(1)}\left(\boldsymbol{M}_{tt_0}^{(2)}\right)^{1/2}}_{\text{congruence transform of } \boldsymbol{M}_{tt_0}^{(1)}}\right) > 0.
\end{aligned}
$$

Additionally, $\boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)}$ being symmetric, we have that $\boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)} \succ \boldsymbol{0}$.

(iii) That

$$\mathrm{eig}\left(\boldsymbol{M}_{tt_0}\right) = \mathrm{eig}\left(\underbrace{\begin{bmatrix}\boldsymbol{D}^{1/4} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}^{1/4}\end{bmatrix}\boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)}\begin{bmatrix}\boldsymbol{D}^{1/4} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}^{1/4}\end{bmatrix}}_{\text{congruence transform of } \boldsymbol{M}_{tt_0}^{(1)}\boldsymbol{M}_{tt_0}^{(2)}}\right) > 0,$$

and $\boldsymbol{M}_{tt_0}$ is symmetric, together imply $\boldsymbol{M}_{tt_0} \succ \boldsymbol{0}$. □

C.6 Proof of Proposition 3.3

Proof. That the minimum in (3.23) is attained for the L in (3.22), follows from the strict convexity and coercivity of $\boldsymbol{u} \mapsto L\left(\cdot, \boldsymbol{u}\right)$.

For notational convenience, let $\omega_i := 2\sqrt{d_i}\ \forall i = 1, \cdots, n$. The optimal solution $\boldsymbol{q}_{\text{opt}}(\tau)$, $\tau \in [t_0, t]$, solves the Euler-Lagrange equation $\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\boldsymbol{q}^{\text{opt}} = 4\boldsymbol{D}\boldsymbol{q}^{\text{opt}}$, or in component form: $\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}q_i{}^{\text{opt}} = \omega_i^2 q_i^{\text{opt}}$ subject to the endpoint constraints $q_i^{\text{opt}}(t_0) = y_i$, $q_i^{\text{opt}}(t) = z_i$. The general solution is $q_i^{\text{opt}}(\tau) = a_i e^{\omega_i\tau} + b_i e^{-\omega_i\tau}$, and the endpoint constraints determine the constants (a_i, b_i) as

$$\left(a_i, b_i\right) = \left(\frac{-y_i e^{-\omega_i t} + z_i e^{-\omega_i t_0}}{2\sinh\left(\omega_i\left(t - t_0\right)\right)}, \frac{y_i e^{\omega_i t} - z_i e^{\omega_i t_0}}{2\sinh\left(\omega_i\left(t - t_0\right)\right)}\right). \tag{C.11}$$

Then, (3.23) gives

$$\begin{aligned}
\mathrm{dist}_{tt_0}^2(\boldsymbol{y}, \boldsymbol{z}) &= \int_{t_0}^{t}\sum_{i=1}^{n}\left(\frac{1}{2}\left(\frac{\mathrm{d}}{\mathrm{d}\tau}q_i^{\text{opt}}(\tau)\right)^2 + \frac{1}{2}\omega_i^2\left(q_i^{\text{opt}}(\tau)\right)^2\right)\mathrm{d}\tau\\
&= \sum_{i=1}^{n}\omega_i^2\int_{t_0}^{t}\left(a_i^2 e^{2\omega_i\tau} + b_i^2 e^{-2\omega_i\tau}\right)\mathrm{d}\tau\\
&= \sum_{i=1}^{n}\frac{\omega_i}{2}\left\{a_i^2\left(e^{2\omega_i t} - e^{2\omega_i t_0}\right) - b_i^2\left(e^{-2\omega_i t} - e^{-2\omega_i t_0}\right)\right\}\\
&= \sum_{i=1}^{n}\frac{\omega_i}{2}\frac{\left(y_i^2 + z_i^2\right)\cosh\left(\omega_i\left(t - t_0\right)\right) - 2y_i z_i}{\sinh\left(\omega_i\left(t - t_0\right)\right)},
\end{aligned} \tag{C.12}$$

where in the last line, we substituted for (a_i, b_i) from (C.11), and simplified the resulting expression using the identity $\sinh(2(\cdot)) = 2\sinh(\cdot)\cosh(\cdot)$.

Since $\omega_i = 2\sqrt{d_i}\ \forall i = 1, \cdots, n$, we see that (C.12) is indeed the desired expression, c.f. the expression inside the exponential in (C.10). This completes the proof. □

C.7 Proof of Theorem 3.3

Proof. We write

$$\begin{aligned} k_{+}(t_0, \boldsymbol{y}, t, \boldsymbol{z}) &= \lim_{\left(d_{i_{n-p+1}}, \cdots, d_{i_n}\right)\downarrow \boldsymbol{0}} k_{++}(t_0, \boldsymbol{y}, t, \boldsymbol{z}) \\ &= k_{++}\left(t_0, \boldsymbol{y}_{[i_1:i_{n-p}]}, t, \boldsymbol{z}_{[i_1:i_{n-p}]}\right) \\ &\quad \times \lim_{\left(d_{i_{n-p+1}}, \cdots, d_{i_n}\right)\downarrow \boldsymbol{0}} k_{++}\left(t_0, \boldsymbol{y}_{[i_{n-p+1}:i_n]}, t, \boldsymbol{z}_{[i_{n-p+1}:i_n]}\right), \end{aligned}$$

where the last equality is due to the separable (in spatial coordinates) product structure in k_{++}. Invoking Theorem 3.2 for the remaining limit, the result follows. □

C.8 Proof of Proposition 3.4

Proof. (i) Follows from the spatial symmetries of the kernels (3.24) and (3.19).

(ii) Since (3.24) and (3.19) are both positive, so is k_{+} in (3.28). Then, $\widehat{\varphi}_0$ being everywhere positive, (3.29) is positive too.

(iii) From Proposition 3.2(iii), $\boldsymbol{M} \succ \boldsymbol{0}$. So the exponential term in (3.19) has negative definite argument. The same is true for the heat kernel (3.24). Thus the product of the exponential terms in (3.28) is bounded in $[0, 1]$. Therefore,

$k_+(t_0, \boldsymbol{V}\boldsymbol{x}, t, \boldsymbol{z})\, \widehat{\varphi}_0\left(\boldsymbol{V}^\top \boldsymbol{z}\right) \le \zeta(t, t_0)\widehat{\varphi}_0\left(\boldsymbol{V}^\top \boldsymbol{z}\right)$ where

$$\zeta(t, t_0) \coloneqq \frac{(4\pi(t-t_0))^{-\frac{p}{2}}(2\pi)^{-\frac{n-p}{2}}\left(\det(\boldsymbol{D}_{[i_1:i_{n-p}]})\right)^{\frac{1}{4}}}{\sqrt{\det\left(\sinh\left(2(t-t_0)\sqrt{\boldsymbol{D}_{[i_1:i_{n-p}]}}\right)\right)}}.$$

Because $\widehat{\varphi}_0$ is Lebesgue measurable (given), applying the dominated convergence theorem separately for $\|\boldsymbol{x}\|_2 \to \infty$ with finite t, and for $t \to \infty$, concludes the proof. □

C.9 Proof of Theorem 3.4

Proof. We proceed as before by letting $\boldsymbol{x} \coloneqq \boldsymbol{V}\boldsymbol{z}$, $\widehat{\nu}(t, \boldsymbol{x}) \coloneqq \widehat{\varphi}(t, \boldsymbol{z} = \boldsymbol{V}^\top \boldsymbol{x})$, to get

$$\frac{\partial \widehat{\nu}}{\partial t} = \left(\Delta_{\boldsymbol{x}} - \sum_{k=1}^{n}\left(\lambda_k x_k^2 + \boldsymbol{r}^\top \boldsymbol{v}_k x_k + \frac{1}{n}s\right)\right)\widehat{\nu}.$$

The formulae (3.53), (3.54) and (3.55) now become

$$Q_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(\boldsymbol{X}, \boldsymbol{D}) = s + \|\boldsymbol{D}\|_2^2 + \sum_{k=1}^{n}\left(\lambda_k X_k^2 + \boldsymbol{r}^\top \boldsymbol{v}_k X_k\right),$$

$$H_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(\boldsymbol{X}, \boldsymbol{D}) = \exp\left(-(t-t_0)Q_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(\boldsymbol{X}, \boldsymbol{D})\right),$$

$$q_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(\boldsymbol{x}, \boldsymbol{\xi}) = s + \|\boldsymbol{\xi}\|_2^2 + \sum_{k=1}^{n}\left(\lambda_k x_k^2 + \boldsymbol{r}^\top \boldsymbol{v}_k x_k\right).$$

The associated Weyl symbol $h_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}$ solves an analogue of (3.56), given by

$$\frac{\partial}{\partial t} h_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(\boldsymbol{x}, \boldsymbol{\xi}) = -\sum_{j=0}^{2}\frac{1}{j!}\{q_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}, h_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}\}_j(\boldsymbol{x}, \boldsymbol{\xi}),$$

from which, direct calculation of the Poisson brackets again yields (3.60) with the initial condition $h_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}|_{t=t_0} = 1$.

Now let

$$q_k \coloneqq \xi_k^2 + \lambda_k x_k^2 + \boldsymbol{r}^\top \boldsymbol{v}_k x_k + \frac{1}{n}s, \quad \forall k \in [n], \tag{C.13}$$

so that $\sum_{k=1}^{n} q_k = q_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}$. Following (3.61), we define the function g via

$$h_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(\boldsymbol{x},\boldsymbol{\xi}) = g(t, q_1, \cdots, q_n), \tag{C.14}$$

and using (3.60), arrive at the PDE (3.62).

Using the definition (C.13), we next simplify (3.62) as

$$\frac{\partial g}{\partial t} = \sum_{k=1}^{n} \left(-q_k g + \lambda_k \frac{\partial g}{\partial q_k} + \frac{\partial^2 g}{\partial q_k^2} \left(\lambda_k \left(q_k + c_k \right) \right) \right). \tag{C.15}$$

The PDE (C.15) generalizes (3.63).

To solve (C.15), we use the anstaz (3.64) with initial conditions (3.65). Following the same steps as before, we again find that $\beta_k(t)$ are given by (3.69). However, $\alpha(t)$ now solves a modified version of (3.70), given by

$$\begin{aligned} \frac{\dot{\alpha}}{\alpha} &= -\sum_{k=1}^{n} \left\{ \sqrt{\lambda_k} \tanh\left(\sqrt{\lambda_k}(t-t_0)\right) + c_k \tanh^2\left(\sqrt{\lambda_k}(t-t_0)\right) \right\} \\ \alpha\left(t_0\right) &= 1. \end{aligned} \tag{C.16}$$

By direct integration, the solution for (C.16) is

$$\begin{aligned} \alpha(t) = &\left(\prod_{k=1}^{n} \frac{1}{\cosh\left(\sqrt{\lambda_k}(t-t_0)\right)} \right) \\ &\times \exp\left(\sum_{k=1}^{n} \left(-c_k(t-t_0) + \frac{c_k}{\sqrt{\lambda_k}} \tanh \sqrt{\lambda_k}(t-t_0) \right) \right), \end{aligned} \tag{C.17}$$

which generalizes the earlier formula (3.71).

Combining these $\alpha(t), \beta_k(t)$ together with (3.64) and (C.14), we obtain the Weyl symbol

$$\begin{aligned} &h_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}(\boldsymbol{x},\boldsymbol{\xi}) \\ &= \left(\prod_{k=1}^{n} \frac{1}{\cosh\left(\sqrt{\lambda_k}(t-t_0)\right)} \right) \times \exp\left(-\sum_{k=1}^{n} c_k(t-t_0) \right) \end{aligned}$$

$$\times \exp\left(-\sum_{k=1}^{n} \frac{\lambda_k x_k^2 + \xi_k^2 + \boldsymbol{r}^\top \boldsymbol{v}_k x_k + (\boldsymbol{r}^\top \boldsymbol{v}_k)^2/(4\lambda_k)}{\sqrt{\lambda_k}} \times \tanh\left(\sqrt{\lambda_k}(t-t_0)\right)\right), \tag{C.18}$$

which generalizes the earlier derived (3.73).

To derive the kernel $k_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}$ from the Weyl symbol $h_{(\boldsymbol{\Lambda},\boldsymbol{r},s)}$, we next apply (3.44) to (C.18). Following the computation similar the derivation of (3.77) in Chapter 3, we arrive at (3.82). Notice that for $\boldsymbol{r} = \boldsymbol{0}, s = 0$, the kernel (3.82) reduces to (3.77), as expected. □

D | Proofs for Chapter 4

D.1 Derivation of (4.22)

Consider k as in (4.11) with $\frac{1}{2}\mathrm{dist}^2_{tt_0}$ given by (4.21). For finding the pre-factor $c(t,t_0)$, we substitute (4.11) with (4.21) in the reaction-diffusion PDE in (4.10), i.e., in

$$\begin{aligned}\frac{\partial k}{\partial t} &= \left(\Delta_{\boldsymbol{x}} - \frac{1}{2}\boldsymbol{x}^\top 2\boldsymbol{D}\boldsymbol{x}\right)k\\ &= \left(\sum_{i=1}^{n}\partial^2_{x_{ii}} - \frac{\omega_i^2}{4}x_{ii}^2\right)k,\end{aligned} \tag{D.1}$$

since $\omega_i^2 = 4D_{ii}$ for all $i = 1,\cdots,n$. This substitution in (D.1), after algebraic simplification, results in the ODE

$$\dot{c} = c\left(-\frac{1}{2}\sum_{i=1}^{n}\omega_i\coth\left(\omega_i(t-t_0)\right)\right), \tag{D.2}$$

where $\dot{c}$ denotes the derivative of $c(t,t_0)$ w.r.t. t.

Integrating (D.2) yields

$$\log c = \log a - \frac{1}{2}\sum_{i=1}^{n}\log\sinh(\omega_i(t-t_0)),$$

where $\log a$ is the numerical constant of integration. Thus,

$$c(t,t_0) = a\prod_{i=1}^{n}\frac{1}{\left(\sinh(\omega_i(t-t_0))\right)^{1/2}}. \tag{D.3}$$

Combining (D.3) with (4.11) and (4.21), we obtain

$$\begin{aligned}k(t_0,\boldsymbol{x},t,\boldsymbol{y}) =& a\prod_{i=1}^{n}\frac{1}{\sqrt{\sinh(\omega_i(t-t_0))}}\\ &\times\exp\left(-\frac{1}{2}\frac{\omega_i\left(x_i^2+y_i^2\right)\cosh\left(\omega_i(t-t_0)\right) - 2\omega_i x_i y_i}{2\sinh\left(\omega_i(t-t_0)\right)}\right),\end{aligned} \tag{D.4}$$

for all $0 \le t_0 \le t < \infty$. All that remains is to find the constant a.

Letting $\tau := t - t_0$, and invoking the initial condition in (4.10) for (D.4), we have

$$\delta(\boldsymbol{x}-\boldsymbol{y}) = a \lim_{\tau\downarrow 0} \prod_{i=1}^{n} \frac{1}{\sqrt{\sinh(\omega_i\tau)}} \times \exp\left(-\frac{1}{2}\frac{\omega_i\left(x_i^2+y_i^2\right)\cosh\left(\omega_i\tau\right) - 2\omega_i x_i y_i}{2\sinh\left(\omega_i\tau\right)}\right). \tag{D.5}$$

Integrating both sides of (D.5) w.r.t. $\boldsymbol{x}$ over $\mathbb{R}^n$ gives

$$\begin{aligned} 1 = a &\left(\lim_{\tau\downarrow 0} \prod_{i=1}^{n} \frac{1}{\sqrt{\sinh(\omega_i\tau)}}\right) \\ &\times \int_{\mathbb{R}^n} \lim_{\tau\downarrow 0} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\frac{\omega_i\left(x_i^2+y_i^2\right)\cosh\left(\omega_i\tau\right) - 2\omega_i x_i y_i}{2\sinh\left(\omega_i\tau\right)}\right)\mathrm{d}\boldsymbol{x}, \end{aligned} \tag{D.6}$$

wherein the LHS used the shift property of the Dirac delta: $\int_{\mathbb{R}^n} \delta(\boldsymbol{x}-\boldsymbol{y}) \cdot 1 \cdot \mathrm{d}\boldsymbol{x} = 1$.

Since

$$\frac{\omega_i\left(x_i^2+y_i^2\right)\cosh\left(\omega_i\tau\right) - 2\omega_i x_i y_i}{2\sinh\left(\omega_i\tau\right)} = \begin{pmatrix} \sqrt{\omega}_i x_i \\ \sqrt{\omega}_i y_i \end{pmatrix}^{\top} \begin{pmatrix} \coth(\omega_i\tau) & -\operatorname{csch}(\omega_i\tau) \\ -\operatorname{csch}(\omega_i\tau) & \coth(\omega_i\tau) \end{pmatrix} \begin{pmatrix} \sqrt{\omega}_i x_i \\ \sqrt{\omega}_i y_i \end{pmatrix}$$

is a positive definite quadratic form, we have

$$0 \leq \exp\left(-\frac{1}{2}\frac{\omega_i\left(x_i^2+y_i^2\right)\cosh\left(\omega_i\tau\right) - 2\omega_i x_i y_i}{2\sinh\left(\omega_i\tau\right)}\right) \leq 1.$$

Hence, using the dominated convergence theorem [182, Thm. 1.13], we exchange the limit and integral in (D.6) to yield

$$\begin{aligned} 1 = a \lim_{\tau\downarrow 0} \prod_{i=1}^{n} &\frac{\exp\left(-\frac{1}{2}\frac{\omega_i y_i^2 \cosh(\omega_i\tau)}{2\sinh(\omega_i\tau)}\right)}{\sqrt{\sinh(\omega_i\tau)}} \\ &\times \left(\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{\omega_i x_i^2\cosh(\omega_i\tau) - 2\omega_i x_i y_i}{2\sinh(\omega_i\tau)}\right)\mathrm{d}x_i\right). \end{aligned} \tag{D.7}$$

We need the following auxiliary lemma.

Lemma D.1 (Central identity of quantum field theory). *[100, p. 2], [99, p. 15] For* *fixed $n \times n$ matrix $\boldsymbol{A} \succ \boldsymbol{0}$ and suitably smooth $f : \mathbb{R}^n \mapsto \mathbb{R}$,*

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}\right) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \sqrt{\frac{(2\pi)^n}{\det(\boldsymbol{A})}} \exp\left(\frac{1}{2}\nabla_{\boldsymbol{x}}^\top \boldsymbol{A}^{-1}\nabla_{\boldsymbol{x}}\right) f(\boldsymbol{x})\Bigg|_{\boldsymbol{x}=\boldsymbol{0}}, \tag{D.8}$$

where the exponential of differential operator is understood as a power series. As a special case, for $\boldsymbol{b} \in \mathbb{R}^n$, we have $\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \langle \boldsymbol{b}, \boldsymbol{x}\rangle\right) \mathrm{d}\boldsymbol{x} = \sqrt{\frac{(2\pi)^n}{\det \boldsymbol{A}}} \exp\left(\frac{1}{2}\boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}\right)$. In particular, for fixed scalars $a > 0, b \in \mathbb{R}$, we have

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}ax^2 + bx\right) \mathrm{d}x = \sqrt{\frac{2\pi}{a}} \exp\left(\frac{1}{2}\frac{b^2}{a}\right).$$

Using Lemma D.1, the integral in (D.7) evaluates to

$$\sqrt{\frac{4\pi \sinh(\omega_i \tau)}{\omega_i \cosh(\omega_i \tau)}} \exp\left(\frac{\omega_i y_i^2}{4\sinh(\omega_i \tau)\cosh(\omega_i \tau)}\right).$$

Substituting back this value in (D.7), followed by simplification via identity: $1 - \cosh^2(\cdot) = -\sinh^2(\cdot)$, gives

$$\begin{aligned} 1 &= a \lim_{\tau \downarrow 0} \prod_{i=1}^{n} \sqrt{\frac{4\pi}{\omega_i \cosh(\omega_i \tau)}} \exp\left(-\frac{\omega_i y_i^2 \sinh(\omega_i \tau)}{4}\right) \\ &= a \prod_{i=1}^{n} \sqrt{\frac{4\pi}{\omega_i}}. \end{aligned} \tag{D.9}$$

Therefore, $a = \prod_{i=1}^{n} \sqrt{\frac{\omega_i}{4\pi}}$, which upon substitution in (D.3), yields (4.22).

D.2 Proof of Proposition 4.3

As is well-known [134, p. 156], [183], for any $\tau \in [t_0, t]$, the solution of the Riccati ODE initial value problem (4.26) admits linear fractional representation

$$\boldsymbol{\Pi}(\tau, \boldsymbol{K}_1, t) = \left(\boldsymbol{\Psi}_{21}(t,\tau) + \boldsymbol{\Psi}_{22}(t,\tau)\boldsymbol{K}_1\right)\left(\boldsymbol{\Psi}_{11}(t,\tau) + \boldsymbol{\Psi}_{12}(t,\tau)\boldsymbol{K}_1\right)^{-1}, \tag{D.10}$$

where

$$\boldsymbol{\Psi}_{t\tau} := \begin{bmatrix} \boldsymbol{\Psi}_{11}(t,\tau) & \boldsymbol{\Psi}_{12}(t,\tau) \\ \boldsymbol{\Psi}_{21}(t,\tau) & \boldsymbol{\Psi}_{22}(t,\tau) \end{bmatrix} \in \mathbb{R}^{2n\times 2n} \tag{D.11}$$

is the state transition matrix of the linear Hamiltonian matrix ODE

$$\begin{pmatrix} \dot{\boldsymbol{X}}_\tau \\ \dot{\boldsymbol{\Lambda}}_\tau \end{pmatrix} = \begin{bmatrix} \boldsymbol{A}_\tau & -\hat{\boldsymbol{B}}_\tau \hat{\boldsymbol{B}}_\tau^\top \\ -\boldsymbol{Q}_\tau & -\boldsymbol{A}_\tau^\top \end{bmatrix} \begin{pmatrix} \boldsymbol{X}_\tau \\ \boldsymbol{\Lambda}_\tau \end{pmatrix}, \quad \boldsymbol{X}_\tau, \boldsymbol{\Lambda}_\tau \in \mathbb{R}^{n\times n}. \tag{D.12}$$

For the special case in hand, the coefficient matrix in (D.12) equals

$$\begin{bmatrix} \boldsymbol{0} & -2\boldsymbol{I} \\ -2\boldsymbol{D} & \boldsymbol{0} \end{bmatrix},$$

and its (backward in time) state transition matrix becomes

$$\begin{aligned} \boldsymbol{\Psi}_{t\tau} &= \exp\left(\begin{bmatrix} \boldsymbol{0} & 2\boldsymbol{I} \\ 2\boldsymbol{D} & \boldsymbol{0} \end{bmatrix}(t-\tau)\right) \\ &= \begin{bmatrix} \cosh\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right) & \left(\sqrt{\boldsymbol{D}}\right)^{-1}\sinh\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right) \\ \sqrt{\boldsymbol{D}}\sinh\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right) & \cosh\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right) \end{bmatrix}, \end{aligned} \tag{D.13}$$

where all hyperbolic functions act element-wise. To see the last equality, it suffices to note that for 2×2 matrix $\begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix}$ with $b, c > 0$, we have

$$\begin{aligned} \exp\left(\begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix}(t-\tau)\right) &= \boldsymbol{I}\sum_{k=0}^{\infty} \frac{\left(\sqrt{bc}(t-\tau)\right)^{2k}}{(2k)!} + \begin{bmatrix} 0 & \sqrt{b/c} \\ \sqrt{c/b} & 0 \end{bmatrix} \sum_{k=0}^{\infty} \frac{\left(\sqrt{bc}(t-\tau)\right)^{2k+1}}{(2k+1)!} \\ &= \begin{bmatrix} \cosh(\sqrt{bc}(t-\tau)) & \sqrt{b/c}\,\sinh(\sqrt{bc}(t-\tau)) \\ \sqrt{c/b}\,\sinh(\sqrt{bc}(t-\tau)) & \cosh(\sqrt{bc}(t-\tau)) \end{bmatrix}. \end{aligned}$$

From (D.10) and (D.13), we then get

$$\boldsymbol{\Pi}(\tau, \boldsymbol{0}, t) = \boldsymbol{\Psi}_{21}(t,\tau)\left(\boldsymbol{\Psi}_{11}(t,\tau)\right)^{-1}$$

$$= \sqrt{\boldsymbol{D}} \tanh\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right). \tag{D.14}$$

Following (4.27), in this case, we also have

$$\hat{\boldsymbol{A}}_\tau = -2\boldsymbol{\Pi}(\tau, \boldsymbol{0}, t) = -2\sqrt{\boldsymbol{D}} \tanh\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right) \tag{D.15}$$

for all $\tau \in [t_0, t]$. From (D.15), $\hat{\boldsymbol{A}}_\tau$ and $\int_{t_0}^{\tau} \hat{\boldsymbol{A}}_\sigma \mathrm{d}\sigma$ commute. Hence the state transition matrix for the coefficient matrix $\hat{\boldsymbol{A}}_\tau$ equals

$$\begin{aligned}
\hat{\boldsymbol{\Phi}}_{t\tau} &= \exp\left(-2\sqrt{\boldsymbol{D}} \int_\tau^t \tanh\left(2\sqrt{\boldsymbol{D}}(t-\sigma)\right) \mathrm{d}\sigma\right) \\
&= \exp\left(-2\sqrt{\boldsymbol{D}} \int_0^{t-\tau} \tanh\left(2\sqrt{\boldsymbol{D}}s\right) \mathrm{d}s\right) \\
&= \exp\left(-\log\cosh\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right)\right) \\
&= \operatorname{sech}\left(2\sqrt{\boldsymbol{D}}(t-\tau)\right).
\end{aligned} \tag{D.16}$$

Thus

$$\begin{aligned}
\hat{\boldsymbol{\Gamma}}_{t\tau} &= \int_\tau^t \hat{\boldsymbol{\Phi}}_{t\sigma} \hat{\boldsymbol{B}}_\sigma \hat{\boldsymbol{B}}_\sigma^\top \hat{\boldsymbol{\Phi}}_{t\sigma}^\top \mathrm{d}\sigma \\
&= 2\int_\tau^t \operatorname{sech}^2(2\sqrt{\boldsymbol{D}}(t-\sigma))\mathrm{d}\sigma \\
&= \boldsymbol{D}^{-\frac{1}{2}} \tanh(2\sqrt{\boldsymbol{D}}(t-\tau)),
\end{aligned}$$

so,

$$\hat{\boldsymbol{\Gamma}}_{t\tau}^{-1} = \sqrt{\boldsymbol{D}} \coth(2\sqrt{\boldsymbol{D}}(t-\tau)). \tag{D.17}$$

Therefore, (4.31) simplifies to

$$\boldsymbol{M}_{tt_0} = \begin{bmatrix} \sqrt{\boldsymbol{D}} \coth\left(2\sqrt{\boldsymbol{D}}(t-t_0)\right) & -\sqrt{\boldsymbol{D}}\operatorname{csch}\left(2\sqrt{\boldsymbol{D}}(t-t_0)\right) \\ -\sqrt{\boldsymbol{D}}\operatorname{csch}\left(2\sqrt{\boldsymbol{D}}(t-t_0)\right) & \sqrt{\boldsymbol{D}} \coth\left(2\sqrt{\boldsymbol{D}}(t-t_0)\right) \end{bmatrix} \tag{D.18}$$

which is exactly formula (4.10) in [74] derived via completely different computation.

Substituting (D.18) in (4.30), and recalling that $\omega_i := 2\sqrt{D_{ii}}$ for all $i = 1, \cdots, n$, we recover (4.21). □

D.3 Proof of Theorem 4.1

The proof is divided into three steps.

Step 1: Determining $c(t, t_0)$ up to a constant a.

Combining (4.11) with (4.30), we get

$$\log k = \log c(t, t_0) - \frac{1}{2} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}^\top \boldsymbol{M}_{tt_0} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}. \tag{D.19}$$

Applying ∂_t to both sides of (D.19) and rearranging, gives

$$\partial_t k = k \left(\frac{\dot{c}}{c} - \frac{1}{2} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}^\top \dot{\boldsymbol{M}}_{tt_0} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \right), \tag{D.20}$$

where dot denotes derivative w.r.t. t.

Applying $\nabla_{\boldsymbol{x}}$ to both sides of (D.19), and using (4.31), we find

$$\nabla_{\boldsymbol{x}} \log k = \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \left(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0} \boldsymbol{x} \right) - \boldsymbol{\Pi}(t_0, \boldsymbol{0}, t) \boldsymbol{x}, \tag{D.21}$$

and since $\nabla_{\boldsymbol{x}}^2 = \nabla_{\boldsymbol{x}} \circ \nabla_{\boldsymbol{x}}$,

$$\nabla_{\boldsymbol{x}}^2 \log k = -\hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} - \boldsymbol{\Pi}(t_0, \boldsymbol{0}, t). \tag{D.22}$$

We next use the identity

$$\begin{aligned} \nabla_{\boldsymbol{x}}^2 \log k &= \frac{\nabla_{\boldsymbol{x}}^2 k}{k} - \left(\nabla_{\boldsymbol{x}} \log k \right) \left(\nabla_{\boldsymbol{x}} \log k \right)^\top \\ \Rightarrow \nabla_{\boldsymbol{x}}^2 k &= k \left\{ \left(\nabla_{\boldsymbol{x}} \log k \right) \left(\nabla_{\boldsymbol{x}} \log k \right)^\top + \nabla_{\boldsymbol{x}}^2 \log k \right\}. \end{aligned} \tag{D.23}$$

Substituting (D.21) and (D.22) in (D.23), we obtain

$$\nabla_{\boldsymbol{x}}^2 k = k\Big\{ \left(\hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\left(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x}\right) - \boldsymbol{\Pi}(t_0, \boldsymbol{0}, t)\boldsymbol{x}\right) \times \left(\hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\left(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x}\right) - \boldsymbol{\Pi}(t_0, \boldsymbol{0}, t)\boldsymbol{x}\right)^\top \tag{D.24}$$

$$- \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} - \boldsymbol{\Pi}(t_0, \boldsymbol{0}, t)\Big\}, \tag{D.25}$$

and thus, the diffusion term in the RHS of (4.34a) is

$$\begin{aligned}
&\langle \boldsymbol{B}(t)\boldsymbol{B}(t)^\top, \nabla_{\boldsymbol{x}}^2 k\rangle \\
=&k\langle \boldsymbol{B}(t)\boldsymbol{B}(t)^\top, \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} \\
&- \boldsymbol{\Pi}\boldsymbol{x}(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} - \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})\boldsymbol{x}^\top \boldsymbol{\Pi} \\
&+ \boldsymbol{\Pi}\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{\Pi} - \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} - \boldsymbol{\Pi}(t_0, \boldsymbol{0}, t)\rangle.
\end{aligned} \tag{D.26}$$

We note that the advection term in the RHS of (4.34a) is

$$-\langle \nabla_{\boldsymbol{x}}, k\, \boldsymbol{A}(t)\boldsymbol{x}\rangle = -k\langle \nabla_{\boldsymbol{x}} \log k, \boldsymbol{A}(t)\boldsymbol{x}\rangle - k\,\mathrm{trace}\boldsymbol{A}(t). \tag{D.27}$$

Substituting (D.20), (D.21), (D.26) and (D.27) in (4.34a), we arrive at

$$\begin{aligned}
\frac{\dot{c}}{c} - \frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^\top \dot{\boldsymbol{M}}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} =& -\langle \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x}), \boldsymbol{A}(t)\boldsymbol{x}\rangle - \mathrm{trace}\boldsymbol{A}(t) \\
&+ \langle \boldsymbol{B}(t)\boldsymbol{B}(t)^\top, \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} \\
&- \boldsymbol{\Pi}\boldsymbol{x}(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} - \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\Phi}}_{tt_0}\boldsymbol{x})\boldsymbol{x}^\top \boldsymbol{\Pi} \\
&+ \boldsymbol{\Pi}\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{\Pi} - \hat{\boldsymbol{\Phi}}_{tt_0}^\top \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} \hat{\boldsymbol{\Phi}}_{tt_0} - \boldsymbol{\Pi}(t_0, \boldsymbol{0}, t)\rangle - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}(t)\boldsymbol{x} \\
=& -\theta(t) - \frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^\top \boldsymbol{S}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix},
\end{aligned} \tag{D.28}$$

where the last equality grouped the spatially independent terms and spatially dependent (quadratic in $\boldsymbol{x}, \boldsymbol{y}$) terms, then used the definition of $\theta(t)$ from (4.35a). In particular, the matrix

$$\boldsymbol{S}_{tt_0} = \begin{bmatrix} \boldsymbol{S}_{11}(t,t_0) & \boldsymbol{S}_{12}(t,t_0) \\ \boldsymbol{S}_{12}^{\top}(t,t_0) & \boldsymbol{S}_{22}(t,t_0) \end{bmatrix} \in \mathbb{R}^{2n\times 2n}$$

comprises of $n \times n$ blocks

$$\begin{aligned} \boldsymbol{S}_{11}(t,t_0) :=& -2\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}\hat{\boldsymbol{A}}_t - \hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}\hat{\boldsymbol{B}}_t\hat{\boldsymbol{B}}_t^{\top}\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0} \\ & - \boldsymbol{\Pi}(t_0,\boldsymbol{0},t)\hat{\boldsymbol{B}}_t\hat{\boldsymbol{B}}_t^{\top}\boldsymbol{\Pi}(t_0,\boldsymbol{0},t) + \boldsymbol{Q}(t), \end{aligned} \tag{D.29a}$$

$$\boldsymbol{S}_{12}(t,t_0) := \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}\left(\hat{\boldsymbol{A}}_t + \hat{\boldsymbol{B}}_t\hat{\boldsymbol{B}}_t^{\top}\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}\right), \tag{D.29b}$$

$$\boldsymbol{S}_{22}(t,t_0) := -\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}\hat{\boldsymbol{B}}_t\hat{\boldsymbol{B}}_t^{\top}\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}. \tag{D.29c}$$

Since (D.28) holds for arbitrary $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$, equating the $\boldsymbol{x},\boldsymbol{y}$ independent terms from both sides of (D.28), we conclude that c solves the linear ODE

$$\dot{c} = -\theta(t)c. \tag{D.30}$$

The solution of (D.30) is

$$c(t,t_0) = a\exp\left(-\int_{t_0}^{t}\theta(s)\mathrm{d}s\right), \tag{D.31}$$

for to-be-determined constant a. In Step 2 detailed next, we determine a.

As a by-product of the above calculation, we get $\dot{\boldsymbol{M}}_{tt_0} = \boldsymbol{S}_{tt_0}$, but this will not be used hereafter.

Step 2: Determining the constant a.

Substituting (D.31) and (4.30)-(4.31) in (4.11), letting $t \downarrow t_0$, and invoking the initial condition (4.34b), we get

$$\delta(\boldsymbol{x}-\boldsymbol{y}) = a\lim_{t\downarrow t_0}\exp\left(-\int_{t_0}^{t}\theta(\tau)\mathrm{d}\tau - \frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^{\top}\boldsymbol{M}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\right). \tag{D.32}$$

Integrating both sides of (D.32) w.r.t. $\boldsymbol{x}$, and using the shift property of the Dirac delta: $\int_{\mathbb{R}^n}\delta(\boldsymbol{x}-\boldsymbol{y})\cdot 1\cdot \mathrm{d}\boldsymbol{x}=1$, we obtain

$$1 = a\int_{\mathbb{R}^n}\lim_{t\downarrow t_0}\exp\left(-\int_{t_0}^{t}\theta(\tau)\mathrm{d}\tau\right)\exp\left(\frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^{\top}\boldsymbol{M}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\right)\mathrm{d}\boldsymbol{x}. \tag{D.33}$$

As noted in Sec. 4.5, $\boldsymbol{M}_{tt_0}\succ\boldsymbol{0}$, so for any $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$, the second exponential term in (D.33) takes values in $[0,1]$.

Also, under assumption $\mathbf{A1}$, $\theta(\tau)$ in (4.35a) is bounded for $0\leq t_0<t<\infty$, and so is $\exp(-\int_{t_0}^{t}\theta(\tau)\mathrm{d}\tau)$. Thus, the product of the two exponential terms in (D.33) is bounded w.r.t. $\boldsymbol{x}$. Therefore, by dominated convergence theorem, we exchange the order of integral and limit in (D.33), to find

$$\begin{aligned}
1 &= a\lim_{t\downarrow t_0}\exp\left(-\int_{t_0}^{t}\theta(\tau)\mathrm{d}\tau\right)\int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^{\top}\boldsymbol{M}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\right)\mathrm{d}\boldsymbol{x}\\
&= a\lim_{t\downarrow t_0}\exp\left(-\int_{t_0}^{t}\theta(\tau)\mathrm{d}\tau\right)\exp\left(-\frac{1}{2}\boldsymbol{y}^{\top}\boldsymbol{M}_{22}(t,t_0)\boldsymbol{y}\right)\\
&\quad\times\int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{M}_{11}(t,t_0)\boldsymbol{x}+\langle-\boldsymbol{M}_{12}(t,t_0)\boldsymbol{y},\boldsymbol{x}\rangle\right)\mathrm{d}\boldsymbol{x}\\
&= a\,(2\pi)^{\frac{n}{2}}\lim_{t\downarrow t_0}\exp\left(-\int_{t_0}^{t}\theta(\tau)\mathrm{d}\tau\right)\left(\det\boldsymbol{M}_{11}(t,t_0)\right)^{-\frac{1}{2}}\\
&\quad\exp\left(-\frac{1}{2}\boldsymbol{y}^{\top}\{\boldsymbol{M}_{22}(t,t_0)-\boldsymbol{M}_{12}^{\top}(t,t_0)\boldsymbol{M}_{11}^{-1}(t,t_0)\boldsymbol{M}_{12}(t,t_0)\}\boldsymbol{y}\right),
\end{aligned} \tag{D.34}$$

wherein $\boldsymbol{M}_{11},\boldsymbol{M}_{12},\boldsymbol{M}_{22}$ are the respective $n\times n$ blocks of (4.31), and the last step follows from Lemma D.1. Invoking Lemma D.1 here is in turn made possible by the fact that $\boldsymbol{M}_{11}$ being a sum of two positive definite matrices (see (4.31)), is positive definite.

Now the idea is to evaluate the limits in (D.34). For the exponential of negative quadratic term, using (4.31), we find

$$\boldsymbol{M}_{22}(t,t_0)-\boldsymbol{M}_{12}^{\top}(t,t_0)\boldsymbol{M}_{11}^{-1}(t,t_0)\boldsymbol{M}_{12}(t,t_0) \tag{D.35}$$

$$
\begin{aligned}
&= \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} - \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}\left(\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0} + \boldsymbol{\Pi}(t_0,\mathbf{0},t)\right)^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}^{-1} \\
&= \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} - \left(\hat{\boldsymbol{\Phi}}_{tt_0}^{-1}\hat{\boldsymbol{\Gamma}}_{tt_0}\right)^{-1}\left(\hat{\boldsymbol{\Phi}}_{tt_0}^{\top}\hat{\boldsymbol{\Gamma}}_{tt_0}^{-1}\hat{\boldsymbol{\Phi}}_{tt_0} + \boldsymbol{\Pi}(t_0,\mathbf{0},t)\right)^{-1}\left(\hat{\boldsymbol{\Gamma}}_{tt_0}\hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\right)^{-1} \\
&= \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} - \left(\hat{\boldsymbol{\Gamma}}_{tt_0} + \hat{\boldsymbol{\Gamma}}_{tt_0}\hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\boldsymbol{\Pi}(t_0,\mathbf{0},t)\hat{\boldsymbol{\Phi}}_{tt_0}^{-1}\hat{\boldsymbol{\Gamma}}_{tt_0}\right)^{-1}, \qquad \text{(D.36)}
\end{aligned}
$$

thanks to the invertibility of $\hat{\boldsymbol{\Phi}}_{tt_0}, \hat{\boldsymbol{\Gamma}}_{tt_0}$. Using the Woodbury identity [184], [185, p. 19],

$$
\begin{aligned}
&\left(\hat{\boldsymbol{\Gamma}}_{tt_0} + \hat{\boldsymbol{\Gamma}}_{tt_0}\hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\boldsymbol{\Pi}(t_0,\mathbf{0},t)\hat{\boldsymbol{\Phi}}_{tt_0}^{-1}\hat{\boldsymbol{\Gamma}}_{tt_0}\right)^{-1} \\
&\qquad = \hat{\boldsymbol{\Gamma}}_{tt_0}^{-1} - \hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\left(\boldsymbol{\Pi}^{-1}(t_0,\mathbf{0},t) + \hat{\boldsymbol{\Phi}}_{tt_0}^{-1}\hat{\boldsymbol{\Gamma}}_{tt_0}\hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\right)^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}^{-1}, \qquad \text{(D.37)}
\end{aligned}
$$

which upon substituting in (D.36), gives

$$
\begin{aligned}
&\boldsymbol{M}_{22}(t,t_0) - \boldsymbol{M}_{12}^{\top}(t,t_0)\boldsymbol{M}_{11}^{-1}(t,t_0)\boldsymbol{M}_{12}(t,t_0) \\
&\qquad = \hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\left(\boldsymbol{\Pi}^{-1}(t_0,\mathbf{0},t) + \hat{\boldsymbol{\Phi}}_{tt_0}^{-1}\hat{\boldsymbol{\Gamma}}_{tt_0}\hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\right)^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}^{-1}. \qquad \text{(D.38)}
\end{aligned}
$$

Since $\hat{\boldsymbol{\Phi}}_{t_0t_0} = \boldsymbol{I}$, $\hat{\boldsymbol{\Gamma}}_{t_0t_0} = \mathbf{0}$, $\boldsymbol{\Pi}(t_0,\mathbf{0},t_0) = \mathbf{0}$, we then have

$$
\begin{aligned}
&\lim_{t\downarrow t_0}\exp\left(-\frac{1}{2}\boldsymbol{y}^{\top}\{\boldsymbol{M}_{22}(t,t_0) - \boldsymbol{M}_{12}^{\top}(t,t_0)\boldsymbol{M}_{11}^{-1}(t,t_0)\boldsymbol{M}_{12}(t,t_0)\}\boldsymbol{y}\right) \\
&= \exp\left(-\frac{1}{2}\boldsymbol{y}^{\top}\underbrace{\lim_{t\downarrow t_0}\{\hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\left(\boldsymbol{\Pi}^{-1}(t_0,\mathbf{0},t) + \hat{\boldsymbol{\Phi}}_{tt_0}^{-1}\hat{\boldsymbol{\Gamma}}_{tt_0}\hat{\boldsymbol{\Phi}}_{tt_0}^{-\top}\right)^{-1}\hat{\boldsymbol{\Phi}}_{tt_0}^{-1}\}}_{=\mathbf{0}}\boldsymbol{y}\right) \\
&= 1. \qquad \text{(D.39)}
\end{aligned}
$$

Therefore, (D.34) simplifies to

$$
1 = a\,(2\pi)^{\frac{n}{2}}\lim_{t\downarrow t_0}\exp\left(-\int_{t_0}^{t}\theta(\tau)\mathrm{d}\tau\right)\left(\det\boldsymbol{M}_{11}(t,t_0)\right)^{-\frac{1}{2}}. \qquad \text{(D.40)}
$$

Notice that for the limit in (D.40) to be defined, the limit cannot be further distributed to the exponential and determinant factors.

Since the limit of reciprocal equals to the reciprocal of the limit, and by Jacobi identity: $\exp\operatorname{trace}(\cdot) = \det\exp(\cdot)$, we re-write (D.40) as (4.36).

Step 3: Putting everything together.

Substituting (4.36) in (D.31), and then substituting the resulting expression for $c(t,t_0)$ in (4.33), the statement follows. □

D.4 Proof of Corollary 4.2

Under the stated conditions, we proved in Proposition 4.3 that the term

$$\frac{1}{2}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^{\top}\boldsymbol{M}_{tt_0}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}$$

in (4.37) reduces to (4.21). What remains is to show that the term $\exp\left(-\int_{t_0}^{t}\theta(s)\mathrm{d}s\right)$ in (4.37) reduces to $\prod_{i=1}^{n}\frac{1}{(\sinh(\omega_i(t-t_0)))^{1/2}}$ as found in (D.3), and to compute the pre-factor a.

To this end, notice that (4.35a) in this case specializes to

$$\begin{aligned}\theta(s) &= \operatorname{trace}\left(\hat{\boldsymbol{\Phi}}_{st_0}^{\top}\hat{\boldsymbol{\Gamma}}_{st_0}^{-1}\hat{\boldsymbol{\Phi}}_{st_0}+\boldsymbol{\Pi}(t_0,\boldsymbol{0},s)\right)\\ &= \operatorname{trace}\left(\sqrt{\boldsymbol{D}}\coth\left(2\sqrt{\boldsymbol{D}}(s-t_0)\right)\right)\\ &= \sum_{i=1}^{n}\sqrt{D_{ii}}\coth\left(\sqrt{D_{ii}}(s-t_0)\right),\end{aligned}\tag{D.41}$$

where the second equality follows from (D.14), (D.16) and (D.17). Then

$$\begin{aligned}\exp\left(-\int_{t_0}^{t}\theta(s)\mathrm{d}s\right) &= \lim_{\varepsilon\downarrow 0}\exp\left(-\sum_{i=1}^{n}\sqrt{D}_{ii}\int_{\varepsilon}^{t-t_0}\coth\left(2\sqrt{D}_{ii}\tau\right)\mathrm{d}\tau\right)\\ &= \exp\left(-\sum_{i=1}^{n}\frac{1}{2}\log\sinh\left(2\sqrt{D}_{ii}(t-t_0)\right)\right)\\ &= \prod_{i=1}^{n}\frac{1}{(\sinh(\omega_i(t-t_0)))^{1/2}}.\end{aligned}\tag{D.42}$$

In above, the first equality used (D.41) and a change-of-variable $s-t_0\mapsto\tau$. The second equality follows from integration and the limit. The last equality used $\omega_i := 2\sqrt{D}_{ii}$.

To compute the pre-factor a for this case, we note from (D.18) that $\det \boldsymbol{M}_{11}^{1/2}(t, t_0) = \boldsymbol{D}^{1/4} \left(\coth\left(2\sqrt{\boldsymbol{D}}(t - t_0)\right)\right)^{1/2}$. Then using the Jacobi identity $\det\exp(\cdot) = \exp\operatorname{trace}(\cdot)$, we have

$$
\begin{aligned}
a &= (2\pi)^{-n/2} \lim_{t \downarrow t_0} \left\{ \det \boldsymbol{M}_{11}^{1/2}(t, t_0) \left(\exp \int_{t_0}^{t} \theta(\tau) \mathrm{d}\tau \right) \right\} \\
&= (2\pi)^{-n/2} \boldsymbol{D}^{1/4} \lim_{t \downarrow t_0} \left\{ \left(\coth\left(2\sqrt{\boldsymbol{D}}(t - t_0)\right)\right)^{1/2} \times \left(\sinh\left(2\sqrt{\boldsymbol{D}}(t - t_0)\right)\right)^{1/2} \right\} \\
&= (2\pi)^{-n/2} \boldsymbol{D}^{1/4} \lim_{t \downarrow t_0} \left(\cosh\left(2\sqrt{\boldsymbol{D}}(t - t_0)\right)\right)^{1/2} \\
&= (2\pi)^{-n/2} \boldsymbol{D}^{1/4}, \qquad \text{(D.43)}
\end{aligned}
$$

where the second equality is due to (D.42).

With the a given by (D.43), $\exp(-\int_{t_0}^{t} \theta(s) \mathrm{d}s)$ given by (D.42), and the specialization of $\boldsymbol{M}_{tt_0}$ as in Proposition 4.3, the kernel (4.37) specializes to (3.77).

E | Proofs for Chapter 5

E.1 Proof of Theorem 5.1

Proof. The main idea behind this proof is to recast (5.1) as a relative entropy minimization problem w.r.t. a reference Gibbs measure arising from the state cost $q(\boldsymbol{x})$. Here, the role of the reference measure will be played by the Wiener measure on the path space $\Omega := C([t_0, t_1]; \mathbb{R}^n)$, i.e., the space of continuous curves in $\mathbb{R}^n$ parameterized by $t \in [t_0, t_1]$. To this end, we follow the developments as in [10] and [39]; see also [133].

Let $\mathcal{M}(\Omega)$ be the collection of all probability measures on the path space Ω. Recall that (5.1b) corresponds to the Itô diffusion process

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{u}(t, \boldsymbol{x})\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}\boldsymbol{w}(t), \tag{E.1}$$

(E.1) where $\boldsymbol{v}$ is non-anticipating, i.e., adapted to the filtration generated up until time t for all $t \in [t_0, t_1]$. Let $\mathbb{W} \in \mathcal{M}(\Omega)$ be a Wiener measure, and consider a measure $\mathbb{P} \in \mathcal{M}(\Omega)$ generated by the Itô diffusion (E.1). Using Girsanov's theorem [24, Thm. 8.6.6], [186, Ch. 3.5], we find

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}} = \frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{W}_0} \exp\left(\int_{t_0}^{t_1} \frac{\boldsymbol{u}}{\sqrt{2\varepsilon}}\mathrm{d}\boldsymbol{w}(t) + \int_{t_0}^{t_1} \frac{1}{2}\frac{\|\boldsymbol{u}\|_2^2}{2\varepsilon}\mathrm{d}t\right), \qquad \mathbb{P} \text{ a.s.,} \tag{E.2a}$$

$$\Leftrightarrow \quad \log\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}} = \log\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{W}_0} + \frac{1}{\sqrt{2\varepsilon}}\int_{t_0}^{t_1} \boldsymbol{u}\mathrm{d}\boldsymbol{w}(t) + \frac{1}{4\varepsilon}\int_{t_0}^{t_1} \|\boldsymbol{u}\|_2^2\mathrm{d}t, \qquad \mathbb{P} \text{ a.s.,} \tag{E.2b}$$

where $\mathbb{P}_0, \mathbb{W}_0$ are the distributions of $\boldsymbol{x}(t = t_0)$ under $\mathbb{P}$ and $\mathbb{W}$, respectively.

Since $\boldsymbol{u} \in \mathcal{U}$ as defined in (1.3),

$$\mathbb{E}_{\mathbb{P}}\left[\int_{t_0}^{t_1} \|\boldsymbol{u}\|_2^2\mathrm{d}t\right] < \infty. \tag{E.3}$$

So for any $t \geq t_0$, the stochastic process

$$\overline{\boldsymbol{u}}(t) := \int_{t_0}^{t} \boldsymbol{u} \mathrm{d}\boldsymbol{w}(\tau)$$

is a martingale, and consequently has constant expected value $\mathbb{E}_{\mathbb{P}}\left[\overline{\boldsymbol{u}}(t)\right] = \mathbb{E}_{\mathbb{P}}\left[\overline{\boldsymbol{u}}(t_0)\right] = 0$. Therefore, applying the expectation operator w.r.t. $\mathbb{P}$ to both sides of (E.2) obtains

$$\mathbb{E}_{\mathbb{P}}\left[\log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}}\right] = \mathbb{E}_{\mathbb{P}}\left[\log \frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{W}_0}\right] + \frac{1}{4\varepsilon} \int_{t_0}^{t_1} \mathbb{E}_{\mathbb{P}}\left[\|\boldsymbol{u}\|_2^2\right] \mathrm{d}t. \tag{E.4}$$

Since we assume the state cost $q(\boldsymbol{x})$ is positive and upper bounded, it follows that $-q(\boldsymbol{x})$ is negative and lower bounded. We next consider the Gibbs measure

$$\frac{\exp\left(-\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right)\mathbb{W}}{Z} \in \mathcal{M}\left(\Omega\right),$$

where the normalization constant

$$Z := \int_{\Omega} \exp\left(-\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right)\mathrm{d}\mathbb{W}.$$

Recall that the *relative entropy* a.k.a. *Kullback-Leibler divergence* is defined as follows: Given two probability measures $\mathbb{P}, \mathbb{Q}$ on some measure space Ω, the relative entropy or Kullback-Leibler divergence

$$D_{\mathrm{KL}}\left(\mathbb{P} \parallel \mathbb{Q}\right) := \mathbb{E}_{\mathbb{P}}\left[\log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right] = \begin{cases} \int_{\Omega} \log\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right)\mathrm{d}\mathbb{P} & \text{if} \quad \mathbb{P} \ll \mathbb{Q}, \\ +\infty & \text{otherwise,} \end{cases} \tag{E.5}$$

where $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$ denotes the Radon-Nikodym derivative, and $\mathbb{P} \ll \mathbb{Q}$ means "$\mathbb{P}$ is absolutely continuous w.r.t. $\mathbb{Q}$".

Applying this definition yields:

$$D_{\mathrm{KL}}\left(\mathbb{P} \parallel \frac{\exp\left(-\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right)\mathbb{W}}{Z}\right) \tag{E.6}$$

$$= \log Z + \mathbb{E}_{\mathbb{P}}\left[\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right] + \mathbb{E}_{\mathbb{P}}\left[\log\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}}\right]$$

$$= \log Z + \mathbb{E}_{\mathbb{P}}\left[\log\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{W}_0}\right] + \frac{1}{2\varepsilon}\int_{t_0}^{t_1}\mathbb{E}_{\mathbb{P}}\left[\frac{1}{2}\|\boldsymbol{u}\|_2^2 + q(\boldsymbol{x})\right]\mathrm{d}t, \tag{E.7}$$

where the last equality uses (E.4).

Let

$$\Pi_{01} := \Big\{\mathbb{M} \in \mathcal{M}(\Omega) \mid \mathbb{M} \text{ has marginal } \rho_i \,\mathrm{d}\boldsymbol{x} \text{ at time } t_i \;\forall i \in \{0,1\},\; \rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^n)\Big\}. \tag{E.8}$$

Since $\log Z + \mathbb{E}_{\mathbb{P}}\left[\log\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{W}_0}\right]$ is constant over Π_{01}, we conclude from (E.7) that the stochastic optimal control problem (5.1) is equivalent to the stochastic calculus of variations problem:

$$\underset{\mathbb{P}\in\Pi_{01}}{\arg\inf}\quad D_{\mathrm{KL}}\left(\mathbb{P} \parallel \frac{\exp\left(-\frac{1}{2\varepsilon}\int_{t_0}^{t_1} q(\boldsymbol{x})\mathrm{d}t\right)\mathbb{W}}{Z}\right). \tag{E.9}$$

From (E.8), the set Π_{01} is closed and convex with non-empty interior. Furthermore, the mapping $\mathbb{P} \mapsto D_{\mathrm{KL}}(\mathbb{P} \parallel \mathbb{Q})$ is strictly convex over Π_{01} for a fixed measure $\mathbb{Q}$. Hence there exists unique solution for (E.9), or equivalently for (5.1). □

E.2 Proof of Theorem 5.2

Proof. The Lagrangian for the SBP with additive state cost (5.1) is the functional

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\left(\frac{1}{2}\|\boldsymbol{u}\|_2^2+q(\boldsymbol{x})\right)\rho_q(t,\boldsymbol{x})+\psi_q(t,\boldsymbol{x})\right.$$
$$\left.\times\left(\frac{\partial\rho_q}{\partial t}+\nabla_{\boldsymbol{x}}\cdot(\rho_q\boldsymbol{u})-\varepsilon\Delta_{\boldsymbol{x}}\rho_q\right)\right)\mathrm{d}\boldsymbol{x}\,\mathrm{d}t, \tag{E.10}$$

where the Lagrange multiplier $\psi_q\in C^{1,2}(\mathbb{R}^n;[t_0,t_1])$. The functional (E.10) depends on the primal-dual variable trio $(\rho,\boldsymbol{u},\psi_q)$. The idea now is to perform unconstrained minimization of (E.10) over the feasible space $\mathcal{P}_{01}\times\mathcal{U}$.

For the term $\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\psi_q(t,\boldsymbol{x})\frac{\partial\rho}{\partial t}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t$ in (E.10), we apply Fubini-Tonelli Theorem to switch the order of the integrals, and perform integration by parts w.r.t. t, to obtain

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\psi_q(t,\boldsymbol{x})\frac{\partial\rho}{\partial t}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t$$
$$=\underbrace{\int_{\mathbb{R}^n}\left(\psi_q(t_1,\boldsymbol{x})\rho_1(\boldsymbol{x})-\psi_\varepsilon(t_0,\boldsymbol{x})\rho_0(\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x}}_{\text{constant over }\mathcal{P}_{01}\times\mathcal{V}}-\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\rho(t,\boldsymbol{x})\frac{\partial\psi_q}{\partial t}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{E.11}$$

Next, for the term $\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\psi_q(t,\boldsymbol{x})\left(\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})-\varepsilon\Delta_{\boldsymbol{x}}\rho\right)\mathrm{d}\boldsymbol{x}\,\mathrm{d}t$ in (E.10), integration by parts w.r.t. $\boldsymbol{x}$ while assuming the limits at $|\boldsymbol{x}|\to\infty$ are zero, gives

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\psi_\varepsilon(t,\boldsymbol{x})\left(\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})-\varepsilon\Delta_{\boldsymbol{x}}\rho(t,\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x}\,\mathrm{d}t$$
$$=-\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\langle\nabla_{\boldsymbol{x}}\psi_q,\boldsymbol{u}\rangle\rho(t,\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t-\varepsilon\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\langle\nabla_{\boldsymbol{x}}\rho,\nabla_{\boldsymbol{x}}\psi_q\rangle\mathrm{d}\boldsymbol{x}\mathrm{d}t. \tag{E.12}$$

For the second integral above, performing integration by parts w.r.t. $\boldsymbol{x}$ once more, (E.12) becomes

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\psi_q(t,\boldsymbol{x})\left(\nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u})-\varepsilon\Delta_{\boldsymbol{x}}\rho(t,\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x}\,\mathrm{d}t$$
$$=-\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\langle\nabla_{\boldsymbol{x}}\psi_q,\boldsymbol{u}\rangle\rho(t,\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t+\varepsilon\int_{t_0}^{t_1}\rho(t,\boldsymbol{x})\Delta_{\boldsymbol{x}}\psi_q\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{E.13}$$

Substituting (E.11) and (E.13) back in (E.10), and dropping the constant term, we arrive at the expression

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(\frac{1}{2}\|\boldsymbol{u}\|_2^2+q(\boldsymbol{x})-\frac{\partial\psi_\varepsilon}{\partial t}-\langle\nabla_{\boldsymbol{x}}\psi_q,\boldsymbol{u}\rangle-\varepsilon\Delta_{\boldsymbol{x}}\psi_q\right)\rho(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{E.14}$$

Pointwise minimization of (E.14) with respect to $\boldsymbol{u}$ for a fixed $\rho\in\mathcal{P}_{01}$, yields the optimal control

$$\boldsymbol{u}_q^{\mathrm{opt}}=\nabla_{\boldsymbol{x}}\psi_q(t,\boldsymbol{x}),$$

which is precisely (5.5). Evaluating (E.14) at this optimal solution, and equating the resulting expression to zero yields the dynamic programming equation

$$0=\int_{t_0}^{t_1}\int_{\mathbb{R}^n}\left(-\frac{1}{2}\|\nabla_{\boldsymbol{x}}\psi_q\|_2^2+q(\boldsymbol{x})-\frac{\partial\psi_q}{\partial t}-\varepsilon\Delta_{\boldsymbol{x}}\psi_q\right)\rho_q^{\mathrm{opt}}(t,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{E.15}$$

For (E.15) to hold for arbitrary $\rho\in\mathcal{P}_{01}$, we must have

$$\frac{1}{2}\|\nabla_{\boldsymbol{x}}\psi_q\|_2^2+\frac{\partial\psi_q}{\partial t}+\varepsilon\Delta_{\boldsymbol{x}}\psi_q=q(\boldsymbol{x}), \tag{E.16}$$

which is the PDE (5.3a).

Substituting (5.5) into the primal feasibility constraint (5.1b) yields (5.3b). Finally, the boundary conditions (5.4) follow from (5.1c). □

E.3 Proof of Theorem 5.3

Proof. By definition, $\psi_q\in C^{1,2}([t_0,t_1],\mathbb{R}^n)$. Being continuous over its entire domain, ψ_q is bounded. So the mapping (5.6) is bijective. In particular, both $\widehat{\varphi}_q,\varphi_q$ are positive over the support of ρ_q^{opt} for all $t\in[t_0,t_1]$.

From (5.6a),

$$\psi_q=2\varepsilon\log\varphi_{2\varepsilon}. \tag{E.17}$$

From (5.6b), $\widehat{\varphi}_\varepsilon = \rho_q^{\text{opt}} \exp(-\log(\varphi_{2\varepsilon}))$, which immediately gives (5.9a). Substituting (E.17) in (5.5) gives (5.9b). Since (5.9a) holds for all $t \in [t_0, t_1]$, so combining (5.4) and (5.9a) yields (5.8). All that remains is to derive (5.7).

Substituting (E.17) in (5.3a), we get

$$2\varepsilon^2 \left\| \frac{1}{\varphi_{2\varepsilon}} \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon} \right\|_2^2 + 2\varepsilon \frac{1}{\varphi_{2\varepsilon}} \frac{\partial \varphi_{2\varepsilon}}{\partial t} + \varepsilon \Delta_{\boldsymbol{x}} (2\varepsilon \log \varphi_{2\varepsilon}) = q(\boldsymbol{x}). \tag{E.18}$$

Since

$$\begin{aligned}
&\Delta_{\boldsymbol{x}}(2\varepsilon \log \varphi_{2\varepsilon}) \\
&\quad = 2\varepsilon \nabla_{\boldsymbol{x}} \nabla_{\boldsymbol{x}} \log \varphi_{2\varepsilon} \\
&\quad = 2\varepsilon \nabla_{\boldsymbol{x}} \left(\frac{1}{\varphi_{2\varepsilon}} \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon} \right) \\
&\quad = 2\varepsilon \left(-\frac{1}{\varphi_{2\varepsilon}^2} \langle \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon}, \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon} \rangle + \frac{1}{\varphi_{2\varepsilon}} \Delta_{\boldsymbol{x}} \varphi_{2\varepsilon} \right),
\end{aligned}$$

(E.18) simplifies to

$$\begin{aligned}
& 2\varepsilon^2 \left\| \frac{1}{\varphi_{2\varepsilon}} \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon} \right\|_2^2 + 2\varepsilon \frac{1}{\varphi_{2\varepsilon}} \frac{\partial \varphi_{2\varepsilon}}{\partial t} \\
& + 2\varepsilon^2 \left(-\frac{1}{\varphi_{2\varepsilon}^2} \langle \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon}, \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon} \rangle + \frac{1}{\varphi_{2\varepsilon}} \Delta_{\boldsymbol{x}} \varphi_{2\varepsilon} \right) = q(\boldsymbol{x}), \\
\Leftrightarrow \quad & 2\varepsilon \frac{1}{\varphi_{2\varepsilon}} \frac{\partial \varphi_{2\varepsilon}}{\partial t} + 2\varepsilon^2 \left(\frac{1}{\varphi_{2\varepsilon}} \Delta_{\boldsymbol{x}} \varphi_{2\varepsilon} \right) = q(\boldsymbol{x}).
\end{aligned} \tag{E.19}$$

Rearranging (E.19), we obtain (5.7b).

Next, substituting (5.9a) and (E.17) in (5.3b), we find

$$\begin{aligned}
& \frac{\partial}{\partial t} (\widehat{\varphi}_\varepsilon \varphi_{2\varepsilon}) + \nabla_{\boldsymbol{x}} \cdot (\widehat{\varphi}_\varepsilon \varphi_{2\varepsilon} \nabla_{\boldsymbol{x}} (2\varepsilon \log \varphi_{2\varepsilon})) = \varepsilon \Delta_{\boldsymbol{x}} (\widehat{\varphi}_\varepsilon \varphi_{2\varepsilon}), \\
\Leftrightarrow \quad & \varphi_{2\varepsilon} \frac{\partial \widehat{\varphi}_\varepsilon}{\partial t} + \widehat{\varphi}_\varepsilon \frac{\partial \varphi_{2\varepsilon}}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left(2\varepsilon \widehat{\varphi}_\varepsilon \varphi_{2\varepsilon} \frac{1}{\varphi_{2\varepsilon}} \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon} \right) \\
& = \varepsilon \left(\varphi_{2\varepsilon} \Delta_{\boldsymbol{x}} \widehat{\varphi}_\varepsilon + 2 \langle \nabla_{\boldsymbol{x}} \widehat{\varphi}_\varepsilon, \nabla_{\boldsymbol{x}} \varphi_{2\varepsilon} \rangle + \widehat{\varphi}_\varepsilon \Delta_{\boldsymbol{x}} \varphi_{2\varepsilon} \right),
\end{aligned} \tag{E.20}$$

where the right-hand-side of (E.20) used the identity $\Delta(fg) = f\Delta g + 2\langle \nabla f, \nabla g \rangle +$

$g\Delta f$ for twice differentiable f, g.

We expand the term $\nabla_{\boldsymbol{x}} \cdot \left(2\varepsilon\widehat{\varphi}_\varepsilon\varphi_{2\varepsilon}\frac{1}{\varphi_{2\varepsilon}}\nabla_{\boldsymbol{x}}\varphi_\varepsilon\right)$ appearing in the left-hand-side of (E.20) as

$$\nabla_{\boldsymbol{x}} \cdot \left(2\varepsilon\widehat{\varphi}_\varepsilon\varphi_{2\varepsilon}\frac{1}{\varphi_{2\varepsilon}}\nabla_{\boldsymbol{x}}\varphi_{2\varepsilon}\right) = 2\varepsilon\langle\nabla_{\boldsymbol{x}}\widehat{\varphi}_\varepsilon, \nabla_{\boldsymbol{x}}\varphi_{2\varepsilon}\rangle + 2\varepsilon\widehat{\varphi}_\varepsilon\Delta_{\boldsymbol{x}}\varphi_{2\varepsilon},$$

and therefore, (E.20) simplifies to

$$\begin{aligned}
&\varphi_{2\varepsilon}\frac{\partial\widehat{\varphi}_\varepsilon}{\partial t} + \widehat{\varphi}_\varepsilon\frac{\partial\varphi_{2\varepsilon}}{\partial t} + \varepsilon\widehat{\varphi}_\varepsilon\Delta_{\boldsymbol{x}}\varphi_{2\varepsilon} = \varepsilon\varphi_{2\varepsilon}\Delta_{\boldsymbol{x}}\widehat{\varphi}_\varepsilon,\\
\Leftrightarrow\quad &\varphi_{2\varepsilon}\frac{\partial\widehat{\varphi}_\varepsilon}{\partial t} + \widehat{\varphi}_\varepsilon\left(\frac{\partial\varphi_{2\varepsilon}}{\partial t} + \varepsilon\Delta_{\boldsymbol{x}}\varphi_{2\varepsilon}\right) = \varepsilon\varphi_{2\varepsilon}\Delta_{\boldsymbol{x}}\widehat{\varphi}_\varepsilon,\\
\Leftrightarrow\quad &\varphi_{2\varepsilon}\frac{\partial\widehat{\varphi}_\varepsilon}{\partial t} + \widehat{\varphi}_\varepsilon\left(\frac{1}{2\varepsilon}q(\boldsymbol{x})\varphi_{2\varepsilon}\right) = \varepsilon\varphi_{2\varepsilon}\Delta_{\boldsymbol{x}}\widehat{\varphi}_\varepsilon, \qquad \text{(E.21)}
\end{aligned}$$

where the last line's left-hand-side used the already derived PDE (5.7b). Rearranging (E.21), we arrive at (5.7a). This completes the proof. □

F | Proofs for Chapter 6

F.1 Proof of Lemma 6.1

Proof. Using [23, Lemma 1, eq. (16)], we have[a]

$$\begin{aligned}
&\Delta_{\boldsymbol{\Sigma}} R = R\langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle + \langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} R\rangle + 2\langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} R\rangle \\
\Rightarrow \quad & \frac{1}{2}\langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} R\rangle + \langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} R\rangle \\
&\qquad = \frac{1}{2}\Delta_{\boldsymbol{\Sigma}} R - \frac{1}{2} R\langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle. \qquad \text{(F.1)}
\end{aligned}$$

By the same [23, Lemma 1, eq. (16)], we also have

$$\begin{aligned}
\frac{1}{4\rho_{\mathrm{opt}}}\Delta_{\boldsymbol{\Sigma}}\rho_{\mathrm{opt}} &= \frac{1}{4}\langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle + \frac{1}{4\rho_{\mathrm{opt}}}\langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}}\rho_{\mathrm{opt}}\rangle \\
&\quad + \frac{1}{2\rho_{\mathrm{opt}}}\langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}}\rho_{\mathrm{opt}}\rangle. \qquad \text{(F.2)}
\end{aligned}$$

To express the RHS of (F.2) in terms of R, notice that the first summand in this RHS is independent of R, and thanks to (6.5), the last summand is $\langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} R\rangle$. For the middle summand in the RHS of (F.2), we find

$$\begin{aligned}
\mathrm{Hess}_{\boldsymbol{x}} R &= \mathrm{Hess}_{\boldsymbol{x}}\left(\frac{1}{2}\log\rho_{\mathrm{opt}}\right) \\
&= -\frac{1}{2}\left(\frac{\nabla_{\boldsymbol{x}}\rho_{\mathrm{opt}}}{\rho_{\mathrm{opt}}}\right)\left(\frac{\nabla_{\boldsymbol{x}}\rho_{\mathrm{opt}}}{\rho_{\mathrm{opt}}}\right)^{\top} + \frac{1}{2\rho_{\mathrm{opt}}}\mathrm{Hess}_{\boldsymbol{x}}\rho_{\mathrm{opt}} \\
&= -\frac{1}{2}\left(2\nabla_{\boldsymbol{x}} R\right)\left(2\nabla_{\boldsymbol{x}} R\right)^{\top} + \frac{1}{2\rho_{\mathrm{opt}}}\mathrm{Hess}_{\boldsymbol{x}}\rho_{\mathrm{opt}},
\end{aligned}$$

which yields

$$\frac{1}{4\rho_{\mathrm{opt}}}\mathrm{Hess}_{\boldsymbol{x}}\rho_{\mathrm{opt}} = \left(\nabla_{\boldsymbol{x}} R\right)\left(\nabla_{\boldsymbol{x}} R\right)^{\top} + \frac{1}{2}\mathrm{Hess}_{\boldsymbol{x}} R. \qquad \text{(F.3)}$$

Hence, we can rewrite (F.2) as

$$\frac{1}{4\rho_{\mathrm{opt}}}\Delta_{\boldsymbol{\Sigma}}\rho_{\mathrm{opt}} = \frac{1}{4}\langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma}\rangle + \left(\nabla_{\boldsymbol{x}} R\right)^{\top}\boldsymbol{\Sigma}\left(\nabla_{\boldsymbol{x}} R\right)$$

$$
\begin{aligned}
&+\frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}R\rangle+\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}R\rangle\\
=&\frac{1}{4}\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle+(\nabla_{\boldsymbol{x}}R)^{\top}\boldsymbol{\Sigma}\,(\nabla_{\boldsymbol{x}}R)\\
&+\frac{1}{2}\Delta_{\boldsymbol{\Sigma}}R-\frac{1}{2}R\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle,
\end{aligned}\tag{F.4}
$$

using (F.1). Grouping the first and the last summand in the RHS of (F.4), we obtain (6.8). □

[a] The divergence of a matrix field $\boldsymbol{\Sigma}$, denoted as $\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma}$, is understood as a vector with elements $(\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma})_i := \sum_{j=1}^{n}\frac{\partial\Sigma_{ij}}{\partial x_j} \quad \forall i\in\{1,\cdots,n\}$.

F.2 Proof of Theorem 6.1

Proof. To derive a PDE for ψ, we combine Lemma 6.1 with the primal-dual PDEs (1.26a)-(1.26b), to obtain

$$
\begin{aligned}
&\frac{1}{\psi}\partial_t\psi=\partial_t R+\frac{\mathrm{i}}{\lambda}\partial_t S\\
&=\frac{1}{2\rho_{\mathrm{opt}}}\left(-\nabla_{\boldsymbol{x}}\cdot\left(\rho_{\mathrm{opt}}\left(\boldsymbol{f}+\boldsymbol{g}\boldsymbol{g}^{\top}\nabla_{\boldsymbol{x}}S\right)\right)+\frac{1}{2}\Delta_{\boldsymbol{x}}\rho_{\mathrm{opt}}\right)\\
&+\frac{\mathrm{i}}{\lambda}\left(-\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{f}\rangle-\frac{1}{2}\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{g}\boldsymbol{g}^{\top}\nabla_{\boldsymbol{x}}S\rangle\right.\\
&\qquad\qquad\left.-\frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}S\rangle+q\right)\\
&=-\langle\nabla_{\boldsymbol{x}}R,\boldsymbol{f}+\boldsymbol{g}\boldsymbol{g}^{\top}\nabla_{\boldsymbol{x}}S\rangle-\frac{1}{2}\nabla_{\boldsymbol{x}}\cdot\left(\boldsymbol{f}+\boldsymbol{g}\boldsymbol{g}^{\top}\nabla_{\boldsymbol{x}}S\right)\\
&+\frac{1}{2}\Delta_{\boldsymbol{\Sigma}}R+(\nabla_{\boldsymbol{x}}R)^{\top}\boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}R+\left(\frac{1}{4}-\frac{1}{2}R\right)\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle\\
&+\frac{\mathrm{i}}{\lambda}\left(-\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{f}\rangle-\frac{1}{2}\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{g}\boldsymbol{g}^{\top}\nabla_{\boldsymbol{x}}S\rangle\right.\\
&\qquad\qquad\left.-\frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}S\rangle+q\right).
\end{aligned}\tag{F.5}
$$

We re-write (F.5) as

$$\begin{aligned}\mathrm{i}\lambda\partial_t\psi =& \mathrm{i}\lambda\psi\Big(\langle\nabla_{\boldsymbol{x}}R, \boldsymbol{f}+\boldsymbol{g}\boldsymbol{g}^\top\nabla_{\boldsymbol{x}}S\rangle + (\nabla_{\boldsymbol{x}}R)^\top\boldsymbol{\Sigma}\nabla_{\boldsymbol{x}}R\\&-\frac{1}{2}\nabla_{\boldsymbol{x}}\cdot(\boldsymbol{f}+\boldsymbol{g}\boldsymbol{g}^\top\nabla_{\boldsymbol{x}}S)+\frac{1}{2}\Delta_{\boldsymbol{\Sigma}}R\\&+\left(\frac{1}{4}-\frac{1}{2}R\right)\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle\Big)\\&+\frac{\mathrm{i}}{\lambda}\Bigg(-\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{f}\rangle-\frac{1}{2}\langle\nabla_{\boldsymbol{x}}S,\boldsymbol{g}\boldsymbol{g}^\top\nabla_{\boldsymbol{x}}S\rangle\\&-\frac{1}{2}\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}S\rangle+q\Bigg).\end{aligned}\tag{F.6}$$

We will come back to (F.6) in a bit.

From [23, Lemma 1, eq. (16)],

$$\Delta_{\boldsymbol{\Sigma}}\psi=\psi\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle+\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\psi\rangle+2\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}\psi\rangle.\tag{F.7}$$

Taking the gradient of

$$\log\psi=R+\frac{\mathrm{i}}{\lambda}S,\tag{F.8}$$

we get $\nabla_{\boldsymbol{x}}\psi=\psi\left(\nabla_{\boldsymbol{x}}R+\frac{\mathrm{i}}{\lambda}\nabla_{\boldsymbol{x}}S\right)$, and re-write (F.7) as

$$\begin{aligned}\Delta_{\boldsymbol{\Sigma}}\psi =& \psi\langle\mathrm{Hess}_{\boldsymbol{x}},\boldsymbol{\Sigma}\rangle+\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\psi\rangle\\&+2\psi\left\langle\nabla_{\boldsymbol{x}}\cdot\boldsymbol{\Sigma},\nabla_{\boldsymbol{x}}R+\frac{\mathrm{i}}{\lambda}\nabla_{\boldsymbol{x}}S\right\rangle.\end{aligned}\tag{F.9}$$

Likewise, we use (F.8) to express $\langle\boldsymbol{\Sigma},\mathrm{Hess}_{\boldsymbol{x}}\psi\rangle$ in terms of R,S by first writing

$$\mathrm{Hess}_{\boldsymbol{x}}\log\psi=-\frac{\nabla_{\boldsymbol{x}}\psi}{\psi}\left(\frac{\nabla_{\boldsymbol{x}}\psi}{\psi}\right)^\top+\frac{1}{\psi}\mathrm{Hess}_{\boldsymbol{x}}\psi.$$

Substituting (F.8) in above and rearranging yields

$$\begin{aligned}\mathrm{Hess}_{\boldsymbol{x}}\psi=\psi\Big\{&\mathrm{Hess}_{\boldsymbol{x}}R+\frac{\mathrm{i}}{\lambda}\mathrm{Hess}_{\boldsymbol{x}}S+(\nabla_{\boldsymbol{x}}R)(\nabla_{\boldsymbol{x}}R)^\top\\&-\frac{1}{\lambda^2}(\nabla_{\boldsymbol{x}}S)(\nabla_{\boldsymbol{x}}S)^\top+\frac{\mathrm{i}}{\lambda}(\nabla_{\boldsymbol{x}}R)(\nabla_{\boldsymbol{x}}S)^\top\end{aligned}$$

$$+ \frac{\mathrm{i}}{\lambda} (\nabla_{\boldsymbol{x}} S) (\nabla_{\boldsymbol{x}} R)^{\top} \Bigg\},$$

and so

$$\langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} \psi \rangle = \psi \Bigg\{ \Bigg\langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} R + \frac{\mathrm{i}}{\lambda} \mathrm{Hess}_{\boldsymbol{x}} S \Bigg\rangle + \|\nabla_{\boldsymbol{x}} R\|_{\boldsymbol{\Sigma}}^2 - \frac{1}{\lambda^2} \|\nabla_{\boldsymbol{x}} S\|_{\boldsymbol{\Sigma}}^2 + \frac{2\mathrm{i}}{\lambda} (\nabla_{\boldsymbol{x}} R)^{\top} \boldsymbol{\Sigma} (\nabla_{\boldsymbol{x}} S) \Bigg\}.$$

Substituting this result back into (F.9) and simplifying algebraically, we obtain

$$\begin{aligned} 0 = &- \frac{\lambda^2}{2} \Delta_{\boldsymbol{\Sigma}} \psi + \psi \Bigg\{ \frac{\lambda^2}{2} \langle \mathrm{Hess}_{\boldsymbol{x}}, \boldsymbol{\Sigma} \rangle + \frac{\lambda^2}{2} \langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} R \rangle \\ &- \frac{1}{2} \|\nabla_{\boldsymbol{x}} S\|_{\boldsymbol{\Sigma}}^2 + \frac{\lambda^2}{2} \|\nabla_{\boldsymbol{x}} R\|_{\boldsymbol{\Sigma}}^2 + \lambda^2 \langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} R \rangle \Bigg\} \\ &+ \mathrm{i} \lambda \psi \Bigg\{ \frac{1}{2} \langle \boldsymbol{\Sigma}, \mathrm{Hess}_{\boldsymbol{x}} S \rangle + (\nabla_{\boldsymbol{x}} R)^{\top} \boldsymbol{\Sigma} (\nabla_{\boldsymbol{x}} S) \\ &+ \langle \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma}, \nabla_{\boldsymbol{x}} S \rangle \Bigg\}. \end{aligned} \tag{F.10}$$

Summing (F.10) and (F.6), we arrive at (6.9a), where $\mathfrak{R}(V_{\mathsf{caSB}})$ and $\mathfrak{I}(V_{\mathsf{caSB}})$ are as in (6.10) and (6.11), respectively. By specializing (6.7) at the initial and terminal times t_0, t_1, we obtain the boundary conditions (6.9b). $\square$

Bibliography

[1] G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Mem. Math. Phys. Acad. Royale Sci.*, pp. 666–704, 1781.

[2] L. V. Kantorovich, "On the translocation of masses," in *Dokl. Akad. Nauk. USSR (NS)*, vol. 37, pp. 199–201, 1942.

[3] G. Peyré, M. Cuturi, *et al.*, "Computational optimal transport: with applications to data science," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[4] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem," *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393, 2000.

[5] C. Villani, *Topics in optimal transportation.* No. 58, American Mathematical Soc., 2003.

[6] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, 2005.

[7] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on pure and applied mathematics*, vol. 44, no. 4, pp. 375–417, 1991.

[8] E. Schrödinger, *Über die umkehrung der naturgesetze.* Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u. Company., 1931.

[9] E. Schrödinger, "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique," in *Annales de l'institut Henri Poincaré*, vol. 2, pp. 269–310, 1932.

[10] A. Wakolbinger, "Schrödinger bridges from 1931 to 1991," in *Proc. of the 4th Latin American Congress in Probability and Mathematical Statistics, Mexico City*, pp. 61–79, 1990.

[11] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," *The annals of mathematical statistics*, vol. 35, no. 2, pp. 876–879, 1964.

[12] R. Sinkhorn, "Diagonal equivalence to matrices with prescribed row and column sums," *The American Mathematical Monthly*, vol. 74, no. 4, pp. 402–405, 1967.

[13] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," *Advances in neural information processing systems*, vol. 26, 2013.

[14] A. Halder and R. Bhattacharya, "Dispersion analysis in hypersonic flight during planetary entry using stochastic liouville equation," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 2, pp. 459–474, 2011.

[15] T. Mikami, "Monge's problem with a quadratic cost by the zero-noise limit of h-path processes," *Probability theory and related fields*, vol. 129, no. 2, pp. 245–260, 2004.

[16] C. Léonard, "From the Schrödinger problem to the Monge–Kantorovich problem," *Journal of Functional Analysis*, vol. 262, no. 4, pp. 1879–1920, 2012.

[17] C. Léonard, "A survey of the Schrödinger problem and some of its connections with optimal transport," *Discrete and Continuous Dynamical Systems-Series A*, vol. 34, no. 4, pp. 1533–1574, 2014.

[18] E. Hopf, "The partial differential equation $u_t + uu_x = \mu_{xx}$," *Communications on Pure and Applied mathematics*, vol. 3, no. 3, pp. 201–230, 1950.

[19] J. D. Cole, "On a quasi-linear parabolic equation occurring in aerodynamics," *Quarterly of applied mathematics*, vol. 9, no. 3, pp. 225–236, 1951.

[20] Y. Chen, T. T. Georgiou, and M. Pavon, "Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a Schrödinger bridge," *Siam Review*, vol. 63, no. 2, pp. 249–313, 2021.

[21] K. F. Caluya and A. Halder, "Wasserstein proximal algorithms for the Schrödinger bridge problem: density control with nonlinear drift," *IEEE Transactions on Automatic Control*, vol. 67, no. 3, pp. 1163–1178, 2021.

[22] A. M. Teter, I. Nodozi, and A. Halder, "Solution of the probabilistic Lambert problem: Connections with optimal mass transport, Schrödinger bridge and reaction-diffusion PDEs," *arXiv preprint arXiv:2401.07961*, 2024.

[23] A. Teter and A. Halder, "On the Hopf-Cole transform for control-affine Schrödinger bridge," *arXiv preprint arXiv:2503.17640*, 2025.

[24] B. Oksendal, *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[25] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, vol. 113. springer, 2014.

[26] H. J. Kappen, "Path integrals and symmetry breaking for optimal control theory," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, p. P11011, 2005.

[27] K. F. Caluya and A. Halder, "Finite horizon density control for static state feedback linearizable systems," *arXiv preprint arXiv:1904.02272*, 2019.

[28] I. Nodozi and A. Halder, "Schrödinger meets Kuramoto via Feynman-Kac: Minimum effort distribution steering for noisy nonuniform Kuramoto oscillators," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 2953–2960, IEEE, 2022.

[29] G.-H. Liu, A. Vahdat, D.-A. Huang, E. A. Theodorou, W. Nie, and A. Anandkumar, "I^{2}SB: Image-to-Image Schrödinger Bridge," in *Proceedings of the 40th International Conference on Machine Learning*, pp. 22042–22062, 2023.

[30] I. Nodozi, C. Yan, M. Khare, A. Halder, and A. Mesbah, "Neural Schrödinger bridge with Sinkhorn losses: Application to data-driven minimum effort control of colloidal self-assembly," *IEEE Transactions on Control Systems Technology*, vol. 32, no. 3, pp. 960–973, 2023.

[31] I. Nodozi, J. O'Leary, A. Mesbah, and A. Halder, "A physics-informed deep learning approach for minimum effort stochastic control of colloidal self-assembly," in *2023 American Control Conference (ACC)*, pp. 609–615, IEEE, 2023.

[32] F. Vargas, S. Padhy, D. Blessing, and N. Nüsken, "Transport meets variational inference: Controlled Monte Carlo diffusions," in *The Twelfth International Conference on Learning Representations*, 2024.

[33] K. F. Caluya and A. Halder, "Finite horizon density steering for multi-input state feedback linearizable systems," in *2020 American Control Conference (ACC)*, pp. 3577–3582, IEEE, 2020.

[34] S. Haddad, K. F. Caluya, A. Halder, and B. Singh, "Prediction and optimal feedback steering of probability density functions for safe automated driving," *IEEE Control Systems Letters*, vol. 5, no. 6, pp. 2168–2173, 2020.

[35] C. Yan, I. Nodozi, and A. Halder, “Optimal mass transport over the Euler equation,” in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 6819–6826, IEEE, 2023.

[36] K. Elamvazhuthi and A. Halder, “The ground cost for optimal transport of angular velocity,” *arXiv preprint arXiv:2504.03190*, 2025.

[37] K. F. Caluya and A. Halder, “Reflected Schrödinger bridge: density control with path constraints,” in *2021 American Control Conference (ACC)*, pp. 1137–1142, IEEE, 2021.

[38] W. Deng, Y. Chen, N. T. Yang, H. Du, Q. Feng, and R. T. Chen, “Reflected Schrödinger bridge for constrained generative modeling,” *arXiv preprint arXiv:2401.03228*, 2024.

[39] D. Dawson, L. Gorostiza, and A. Wakolbinger, “Schrödinger processes and large deviations,” *Journal of mathematical physics*, vol. 31, no. 10, pp. 2385–2388, 1990.

[40] M. Nagasawa, *Schrödinger equations and diffusion theory*, vol. 86. Birkhäuser, 2012.

[41] F. Guerra and L. M. Morato, “Quantization of dynamical systems and stochastic control theory,” *Physical review D*, vol. 27, no. 8, p. 1774, 1983.

[42] Y. Chen, T. Georgiou, and M. Pavon, “Optimal steering of inertial particles diffusing anisotropically with losses,” in *2015 American Control Conference (ACC)*, pp. 1252–1257, IEEE, 2015.

[43] L. Cohen, *The Weyl operator and its generalization.* Springer Science & Business Media, 2012.

[44] D. Robert and M. Combescure, *Coherent states and applications in mathematical physics.* Springer, 2021.

[45] J. Derezinski, “Some remarks on Weyl pseudodifferential operators,” *Journées Équations aux dérivées partielles*, pp. 1–14, 1993.

[46] J. Derezinski, “Introduction to quantization.” `https://www.fuw.edu.pl/~derezins/quantize.pdf`, 2021. Accessed: 2024-06-09.

[47] B. V. Fedosov, “Deformation quantization and index theory,” *Mathematical topics*, vol. 9, 1996.

[48] R. H. Battin, “Lambert’s problem revisited,” *AIAA Journal*, vol. 15, no. 5, pp. 707–713, 1977.

[49] R. H. Battin and R. M. Vaughan, "An elegant Lambert algorithm," *Journal of Guidance, Control, and Dynamics*, vol. 7, no. 6, pp. 662–670, 1984.

[50] R. Engels and J. Junkins, "The gravity-perturbed Lambert problem: a ks variation of parameters approach," *Celestial mechanics*, vol. 24, no. 1, pp. 3–21, 1981.

[51] H. Shen and P. Tsiotras, "Optimal two-impulse rendezvous using multiple-revolution Lambert solutions," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 1, pp. 50–61, 2003.

[52] J. W. McMahon and D. J. Scheeres, "Linearized Lambert's problem solution," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 10, pp. 2205–2218, 2016.

[53] J. L. Junkins and H. Schaub, *Analytical mechanics of space systems.* American Institute of Aeronautics and Astronautics, 2009.

[54] R. Armellin, P. Di Lizia, and M. Lavagna, "High-order expansion of the solution of preliminary orbit determination problem," *Celestial Mechanics and Dynamical Astronomy*, vol. 112, pp. 331–352, 2012.

[55] P. W. Schumacher Jr, C. Sabol, C. C. Higginson, and K. T. Alfriend, "Uncertain Lambert problem," *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 9, pp. 1573–1584, 2015.

[56] G. Zhang, D. Zhou, D. Mortari, and M. R. Akella, "Covariance analysis of Lambert's problem via Lagrange's transfer-time formulation," *Aerospace Science and Technology*, vol. 77, pp. 765–773, 2018.

[57] N. Adurthi and M. Majji, "Uncertain Lambert problem: a probabilistic approach," *The Journal of the Astronautical Sciences*, vol. 67, pp. 361–386, 2020.

[58] Z. Hall and P. Singla, "Higher order polynomial series expansion for uncertain Lambert problem," in *AAS Astrodynamics Specialist Conference*, 2018.

[59] D. Guého, P. Singla, R. G. Melton, and D. Schwab, "A comparison of parametric and non-parametric machine learning approaches for the uncertain Lambert problem," in *AIAA Scitech 2020 Forum*, p. 1911, 2020.

[60] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal transport in systems and control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, 2021.

[61] F. Santambrogio, "Optimal transport for applied mathematicians," *Birkäuser, NY*, vol. 55, no. 58-63, p. 94, 2015.

[62] A. Tong, J. Huang, G. Wolf, D. Van Dijk, and S. Krishnaswamy, "Trajectorynet: a dynamic optimal transport network for modeling cellular dynamics," in *International conference on machine learning*, pp. 9526–9536, PMLR, 2020.

[63] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi, "Optimal-transport formulation of electronic density-functional theory," *Physical Review A*, vol. 85, no. 6, p. 062502, 2012.

[64] S. Mei, A. Montanari, and P.-M. Nguyen, "A mean field view of the landscape of two-layer neural networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 33, pp. E7665–E7671, 2018.

[65] A. Halder and R. Bhattacharya, "Probabilistic model validation for uncertain nonlinear systems," *Automatica*, vol. 50, no. 8, pp. 2038–2050, 2014.

[66] K. Lee, A. Halder, and R. Bhattacharya, "Performance and robustness analysis of stochastic jump linear systems using Wasserstein metric," *Automatica*, vol. 51, pp. 341–347, 2015.

[67] K. F. Caluya and A. Halder, "Gradient flow algorithms for density propagation in stochastic systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 3991–4004, 2019.

[68] Y. Chen, T. T. Georgiou, and M. Pavon, "The most likely evolution of diffusing and vanishing particles: Schrödinger bridges with unbalanced marginals," *SIAM Journal on Control and Optimization*, vol. 60, no. 4, pp. 2016–2039, 2022.

[69] K. F. Caluya and A. Halder, "Reflected Schrödinger bridge: Density control with path constraints," *arXiv preprint arXiv:2003.13895*, 2020.

[70] V. De Bortoli, J. Thornton, J. Heng, and A. Doucet, "Diffusion Schrödinger bridge with applications to score-based generative modeling," *Advances in Neural Information Processing Systems*, vol. 34, pp. 17695–17709, 2021.

[71] A. Stromme, "Sampling from a Schrödinger bridge," in *International Conference on Artificial Intelligence and Statistics*, pp. 4058–4067, PMLR, 2023.

[72] Y. Chen, T. Georgiou, and M. Pavon, "Entropic and displacement interpolation: a computational approach using the Hilbert metric," *SIAM Journal on Applied Mathematics*, vol. 76, no. 6, pp. 2375–2396, 2016.

[73] A. M. Teter, Y. Chen, and A. Halder, "On the contraction coefficient of the Schrödinger bridge for stochastic linear systems," *IEEE Control Systems Letters*, vol. 7, pp. 3325–3330, 2023.

[74] A. M. Teter, W. Wang, and A. Halder, "Schrödinger bridge with quadratic state cost is exactly solvable," *arXiv preprint arXiv:2406.00503*, 2024.

[75] A. M. Teter, W. Wang, and A. Halder, "Weyl calculus and exactly solvable Schrödinger bridges with quadratic state cost," in *2024 60th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1–8, IEEE, 2024.

[76] A. M. Teter, W. Wang, S. Shivakumar, and A. Halder, "Markov kernels, distances and optimal control: A parable of linear quadratic non-Gaussian distribution steering," *arXiv preprint arXiv:2504.15753*, 2025.

[77] A. M. Teter, I. Nodozi, and A. Halder, "Probabilistic Lambert problem: Connections with optimal mass transport, Schrödinger bridge, and reaction-diffusion PDEs," *SIAM Journal on Applied Dynamical Systems*, vol. 24, no. 1, pp. 16–43, 2025.

[78] A. M. Teter, A. Halder, M. D. Schneider, A. S. Perloff, J. Pratt, C. M. Artman, and M. Demireva, "Control-affine Schrödinger bridge and generalized Bohm potential," *arXiv preprint arXiv:2508.08511*, 2025.

[79] A. Teter, I. Nodozi, and A. Halder, "Proximal mean field learning in shallow neural networks," *arXiv preprint arXiv:2210.13879*, 2022.

[80] W. Wang, A. M. Teter, M. Arcak, and A. Halder, "Set invariance with probability one for controlled diffusion: Score-based approach," *arXiv preprint arXiv:2507.22385*, 2025.

[81] D. Hilbert, "über die gerade linie als kürzeste verbindung zweier punkte: Aus einem an herrn f. klein gerichteten briefe," *Mathematische Annalen*, vol. 46, no. 1, pp. 91–96, 1895.

[82] P. J. Bushell, "Hilbert's metric and positive contraction mappings in a Banach space," *Archive for Rational Mechanics and Analysis*, vol. 52, pp. 330–338, 1973.

[83] J. Franklin and J. Lorenz, "On the scaling of multidimensional matrices," *Linear Algebra and its applications*, vol. 114, pp. 717–735, 1989.

[84] Y. Chen, T. T. Georgiou, and M. Pavon, "On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint," *Journal of Optimization Theory and Applications*, vol. 169, pp. 671–691, 2016.

[85] E. B. Lee and L. Markus, *Foundations of optimal control theory*, vol. 87. Wiley New York, 1967.

[86] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Finding best approximation pairs relative to two closed convex sets in Hilbert spaces," *Journal of Approximation theory*, vol. 127, no. 2, pp. 178–192, 2004.

[87] E. G. Gilbert, D. W. Johnson, and S. S. Keerthi, "A fast procedure for computing the distance between complex objects in three-dimensional space," *IEEE Journal on Robotics and Automation*, vol. 4, no. 2, pp. 193–203, 1988.

[88] S. Cameron, "Enhancing GJK: Computing minimum and penetration distances between convex polyhedra," in *Proceedings of International Conference on Robotics and Automation*, vol. 4, pp. 3112–3117, IEEE, 1997.

[89] M. Montanari, N. Petrinic, and E. Barbieri, "Improving the GJK algorithm for faster and more reliable distance queries between convex objects," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 3, pp. 1–17, 2017.

[90] J. V. Neumann, "Zur theorie der gesellschaftsspiele," *Mathematische annalen*, vol. 100, no. 1, pp. 295–320, 1928.

[91] D.-Z. Du and P. M. Pardalos, eds., *Minimax and Applications*, vol. 4 of *Nonconvex Optimization and Its Applications.* Springer Science & Business Media, 1995.

[92] M. Kuang and E. G. Tabak, "Preconditioning of optimal transport," *SIAM Journal on Scientific Computing*, vol. 39, no. 4, pp. A1793–A1810, 2017.

[93] S. D. Marino and A. Gerolin, "An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm," *Journal of Scientific Computing*, vol. 85, no. 2, p. 27, 2020.

[94] B. D. Anderson and J. B. Moore, *Optimal control: linear quadratic methods.* Courier Corporation, 2007.

[95] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution—part III," *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3112–3118, 2018.

[96] C. Bunne, Y.-P. Hsieh, M. Cuturi, and A. Krause, "The Schrödinger bridge between Gaussian measures has a closed form," in *International Conference on Artificial Intelligence and Statistics*, pp. 5802–5833, PMLR, 2023.

[97] H. Behjoo and M. Chertkov, "Harmonic path integral diffusion," *arXiv preprint arXiv:2409.15166*, 2024.

[98] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Washington, DC, USA: U.S. Government Printing Office, tenth printing ed., 1972.

[99] A. Zee, *Quantum field theory in a nutshell*, vol. 7. Princeton university press, 2010.

[100] J. Zinn-Justin, *Quantum field theory and critical phenomena*, vol. 171. Oxford university press, 2021.

[101] F. R. Gantmakher, *The theory of matrices, Vol. 1.* American Mathematical Soc., 2000.

[102] N. J. Higham, *Functions of matrices: theory and computation.* SIAM, 2008.

[103] F. G. Mehler, "Ueber die entwicklung einer function von beliebig vielen variablen nach laplaceschen functionen höherer ordnung.," 1866.

[104] W. Kibble, "An extension of a theorem of Mehler's on Hermite polynomials," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 41, pp. 12–15, Cambridge University Press, 1945.

[105] D. Slepian, "On the symmetrized kronecker power of a matrix and extensions of Mehler's formula for hermite polynomials," *SIAM Journal on Mathematical Analysis*, vol. 3, no. 4, pp. 606–616, 1972.

[106] J. Louck, "Extension of the Kibble-Slepian formula for Hermite polynomials using boson operator methods," *Advances in Applied Mathematics*, vol. 2, no. 3, pp. 239–249, 1981.

[107] L. Hörmander, "Symplectic classification of quadratic forms, and general Mehler formulas," *Mathematische Zeitschrift*, vol. 219, pp. 413–449, 1995.

[108] K. Pravda-Starov, "Generalized Mehler formula for time-dependent non-selfadjoint quadratic operators and propagation of singularities," *Mathematische Annalen*, vol. 372, no. 3, pp. 1335–1382, 2018.

[109] A. Halder, K. F. Caluya, B. Travacca, and S. J. Moura, "Hopfield neural network flow: A geometric viewpoint," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4869–4880, 2020.

[110] D. M. Himmelblau *et al.*, *Applied nonlinear programming.* McGraw-Hill, 2018.

[111] B. C. Hall, *Quantum theory for mathematicians.* Springer, 2013.

[112] M. E. Taylor, *Noncommutative microlocal analysis.* No. 313, American Mathematical Soc., 1984.

[113] D. W. Stroock, *Partial differential equations for probabalists.* No. 112, Cambridge University Press, 2008.

[114] L. C. Evans, *Partial differential equations*, vol. 19. American Mathematical Society, 2022.

[115] L. Hörmander, *General Mehler Formulas and the Weyl Calculus, in Unpublished Manuscripts: from 1951 to 2007*, pp. 160–171. Springer International Publishing, 2018.

[116] Y. Chen, T. Georgiou, and M. Pavon, "Optimal steering of inertial particles diffusing anisotropically with losses," in *2015 American Control Conference (ACC)*, pp. 1252–1257, IEEE, 2015.

[117] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution—part iii," *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3112–3118, 2018.

[118] H. Behjoo and M. Chertkov, "Harmonic path integral diffusion," *IEEE Access*, 2025.

[119] M. Chertkov, S. Ahn, and H. Behjoo, "Sampling decisions," *arXiv preprint arXiv:2503.14549*, 2025.

[120] Y. Mei, M. Al-Jarrah, A. Taghvaei, and Y. Chen, "Flow matching for stochastic linear control systems," *arXiv preprint arXiv:2412.00617*, 2024.

[121] E. Schrödinger, "Über die Umkehrung der Naturgesetze," *Sitzungsberichte der Preuss Akad. Wissen. Phys. Math. Klasse, Sonderausgabe*, vol. IX, pp. 144–153, 1931.

[122] E. Schrödinger, "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique," in *Annales de L'Institut Henri Poincaré*, vol. 2, pp. 269–310, Presses universitaires de France, 1932.

[123] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, part i," *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1158–1169, 2015.

[124] S. R. S. Varadhan, "On the behavior of the fundamental solution of the heat equation with variable coefficients," *Communications on Pure and Applied Mathematics*, vol. 20, no. 2, pp. 431–455, 1967.

[125] S. A. Molchanov, "Diffusion processes and Riemannian geometry," *Russian Mathematical Surveys*, vol. 30, no. 1, p. 1, 1975.

[126] K. Crane, C. Weischedel, and M. Wardetzky, "The heat method for distance computation," *Communications of the ACM*, vol. 60, no. 11, pp. 90–99, 2017.

[127] I. Chavel, *Eigenvalues in Riemannian geometry.* Academic press, 1984.

[128] A. Grigor'yan and M. Noguchi, "The heat kernel on hyperbolic space," *Bulletin of the London Mathematical Society*, vol. 30, no. 6, pp. 643–650, 1998.

[129] E. P. Hsu, *Stochastic analysis on manifolds.* No. 38, American Mathematical Soc., 2002.

[130] R. W. Brockett, "Control theory and singular Riemannian geometry," in *New Directions in Applied Mathematics: Papers Presented April 25/26, 1980, on the Occasion of the Case Centennial Celebration*, pp. 11–27, Springer, 1982.

[131] R. S. Strichartz, "Sub-Riemannian geometry," *Journal of Differential Geometry*, vol. 24, no. 2, pp. 221–263, 1986.

[132] M. Gromov, "Carnot-Carathéodory spaces seen from within," in *Sub-Riemannian geometry*, pp. 79–323, Springer, 1996.

[133] C. Léonard, "Stochastic derivatives and generalized h-transforms of Markov processes," *arXiv preprint arXiv:1102.3172*, 2011.

[134] R. W. Brockett, *Finite dimensional linear systems.* John Wiley & Sons Inc., 2015.

[135] V. Kučera, "A review of the matrix Riccati equation," *Kybernetika*, vol. 9, no. 1, pp. 42–61, 1973.

[136] W. Porter, "On the matrix Riccati equation," *IEEE Transactions on Automatic Control*, vol. 12, no. 6, pp. 746–749, 1967.

[137] N. Viswanadham and D. Atherton, "On invariance of degree of controllability under state feedback," *IEEE Transactions on Automatic Control*, vol. 20, no. 2, pp. 271–273, 1975.

[138] R. Brockett, "Poles, zeros, and feedback: State space interpretation," *IEEE Transactions on Automatic Control*, vol. 10, no. 2, pp. 129–135, 1965.

[139] J. Yong and X. Y. Zhou, *Stochastic controls: Hamiltonian systems and HJB equations*, vol. 43. Springer Science & Business Media, 2012.

[140] I. Csiszár, "Sanov property, generalized I-projection and a conditional limit theorem," *The Annals of Probability*, pp. 768–793, 1984.

[141] J. H. Lambert, *"IH Lambert Academiae Scientiarvm Electoralis Boicae Membri Et Professoris Honorarii..." Insigniores Orbitae Cometarvm Proprietates.* Svmtibvs Eberhardi Klett Vidvae, 1761.

[142] L. Euler, "Determinatio orbitae cometae qui mense martio huius anni 1742 potissimum fuit observatus," *Miscellanea Berolinensia*, pp. 1–90, 1743.

[143] J. L. Lagrange, *Sur une maniere particuliere d'exprimer le tems dans les sections coniques: décrites par des forces tendantes au foyer et réciproquement proportionnelles aux carrés des distances.* imprimé chez George Jacques Decker, 1780.

[144] A. Albouy, "Lambert's theorem: geometry or dynamics?," *Celestial Mechanics and Dynamical Astronomy*, vol. 131, no. 9, p. 40, 2019.

[145] "Earth gravitational model 2008 (EGM2008) data and apps." `https://earth-info.nga.mil/index.php?dir=wgs84&action=wgs84#egm2008`, 2008 (accessed May 24, 2023).

[146] M. Bando and H. Yamakawa, "New Lambert algorithm using the Hamilton-Jacobi-Bellman equation," *Journal of Guidance, Control, and Dynamics*, vol. 33, no. 3, pp. 1000–1008, 2010.

[147] M. Kim and S. Park, "Optimal control approach to Lambert's problem and Gibbs' method," *Applied Sciences*, vol. 10, no. 7, p. 2419, 2020.

[148] A. Halder and R. Bhattacharya, "Beyond Monte Carlo: a computational framework for uncertainty propagation in planetary entry, descent and landing," in *AIAA Guidance, Navigation, and Control Conference*, p. 8029, 2010.

[149] S. Haddad, A. Halder, and B. Singh, "Density-based stochastic reachability computation for occupancy prediction in automated driving," *IEEE Transactions on Control Systems Technology*, 2022.

[150] C. Villani, *Optimal transport: old and new*, vol. 338. Springer, 2009.

[151] A. Figalli, *Optimal transportation and action-minimizing measures.* PhD thesis, Lyon, École normale supérieure (sciences), 2007.

[152] J. Yong, "Relations among ODEs, PDEs, FSDEs, BSDEs, and FBSDEs," in *Proceedings of the 36th IEEE Conference on Decision and Control*, vol. 3, pp. 2779–2784, IEEE, 1997.

[153] J. Yong and X. Y. Zhou, *Stochastic controls: Hamiltonian systems and HJB equations*, vol. 43. Springer Science & Business Media, 1999.

[154] I. Nodozi and A. Halder, "Schrödinger meets Kuramoto via Feynman-Kac: minimum effort distribution steering for noisy nonuniform Kuramoto oscillators," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 2953–2960, IEEE, 2022.

[155] H. D. Curtis, *Orbital mechanics for engineering students.* Butterworth-Heinemann, 2013.

[156] Y. Shi, V. De Bortoli, G. Deligiannidis, and A. Doucet, "Conditional simulation using diffusion Schrödinger bridges," in *Uncertainty in Artificial Intelligence*, pp. 1792–1802, PMLR, 2022.

[157] G.-H. Liu, T. Chen, O. So, and E. Theodorou, "Deep generalized Schrödinger bridge," *Advances in Neural Information Processing Systems*, vol. 35, pp. 9374–9388, 2022.

[158] W. Xie, R. Zhou, H. Wang, T. Shen, and E. Chen, "Bridging user dynamics: Transforming sequential recommendations with Schrödinger bridge and diffusion models," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 2618–2628, 2024.

[159] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, vol. 38. Springer Science & Business Media, 2009.

[160] R. Aebi and M. Nagasawa, "Large deviations and the propagation of chaos for schrödinger processes," *Probability Theory and Related Fields*, vol. 94, no. 1, pp. 53–68, 1992.

[161] H. Föllmer, "Random fields and diffusion processes," *Lect. Notes Math*, vol. 1362, pp. 101–204, 1988.

[162] P. Dai Pra, "A stochastic control approach to reciprocal diffusion processes," *Applied Mathematics and Optimization*, vol. 23, no. 1, pp. 313–329, 1991.

[163] M. Pavon and A. Wakolbinger, "On free energy, stochastic control, and Schrödinger processes," in *Modeling, Estimation and Control of Systems with Uncertainty: Proceedings of a Conference held in Sopron, Hungary, September 1990*, pp. 334–348, Springer, 1991.

[164] A. Blaquiere, "Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version)," *Dynamics and Control*, vol. 2, no. 3, pp. 235–253, 1992.

[165] M. Nagasawa, "Transformations of diffusion and Schrödinger processes," *Probability Theory and Related Fields*, vol. 82, no. 1, pp. 109–136, 1989.

[166] M. Born, "Quantenmechanik der stoßvorgänge," *Zeitschrift für physik*, vol. 38, no. 11, pp. 803–827, 1926.

[167] A. Ohsumi, "An interpretation of the Schrödinger equation in quantum mechanics from the control-theoretic point of view," *Automatica*, vol. 99, pp. 181–187, 2019.

[168] E. Nelson, "Derivation of the Schrödinger equation from Newtonian mechanics," *Physical review*, vol. 150, no. 4, p. 1079, 1966.

[169] K. Yasue, "Quantum mechanics and stochastic control theory," *Journal of Mathematical Physics*, vol. 22, no. 5, pp. 1010–1020, 1981.

[170] H. H. Rosenbrock, "Doing quantum mechanics with control theory," *IEEE Transactions on Automatic Control*, vol. 45, no. 1, pp. 73–77, 2002.

[171] M. Pavon, "Quantum Schrödinger bridges," in *Directions in mathematical systems theory and optimization*, pp. 227–238, Springer, 2002.

[172] M. Pavon and F. Ticozzi, "Schrödinger bridges for discrete-time, classical and quantum Markovian evolutions," in *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems–MTNS*, vol. 5, 2010.

[173] D. Bohm, "A suggested interpretation of the quantum theory in terms of "hidden" variables. I," *Physical review*, vol. 85, no. 2, p. 166, 1952.

[174] D. Bohm and J.-P. Vigier, "Model of the causal interpretation of quantum theory in terms of a fluid with irregular fluctuations," *Physical Review*, vol. 96, no. 1, p. 208, 1954.

[175] F. Guerra, "Structural aspects of stochastic mechanics and stochastic field theory," *Physics Reports*, vol. 77, no. 3, pp. 263–312, 1981.

[176] H. A. Bethe, "Theory of disintegration of nuclei by neutrons," *Physical Review*, vol. 47, no. 10, p. 747, 1935.

[177] H. Feshbach, "Unified theory of nuclear reactions," *Annals of Physics*, vol. 5, no. 4, pp. 357–390, 1958.

[178] R. C. Arnold, "Optical potential for high-energy physics: theory and applications," *Physical Review*, vol. 153, no. 5, p. 1523, 1967.

[179] L. Foldy and J. D. Walecka, "On the theory of the optical potential," *Annals of Physics*, vol. 54, no. 3, pp. 447–504, 1969.

[180] B. Sinha, "The optical potential and nuclear structure," *Physics Reports*, vol. 20, no. 1, pp. 1–57, 1975.

[181] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Volume 1*, vol. 1. John Wiley & Sons, 2008.

[182] E. M. Stein and R. Shakarchi, *Real analysis: measure theory, integration, and Hilbert spaces.* Princeton University Press, 2009.

[183] M. A. Shayman, "Phase portrait of the matrix Riccati equation," *SIAM Journal on Control and Optimization*, vol. 24, no. 1, pp. 1–65, 1986.

[184] W. W. Hager, "Updating the inverse of a matrix," *SIAM review*, vol. 31, no. 2, pp. 221–239, 1989.

[185] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge university press, 2nd ed., 2012.

[186] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, vol. 113. Springer Science & Business Media, 2012.