SIAM REVIEW
Vol. 63, No. 2, pp. 249–313

Stochastic Control Liaisons: Richard Sinkhorn Meets Gaspard Monge on a Schrödinger Bridge*

Yongxin Chen†
Tryphon T. Georgiou‡
Michele Pavon§

Abstract. In 1931–1932, Erwin Schrödinger studied a *hot gas Gedankenexperiment* (an instance of large deviations of the empirical distribution). Schrödinger's problem represents an early example of a fundamental *inference method*, the so-called maximum entropy method, having roots in Boltzmann's work and being developed in subsequent years by Jaynes, Burg, Dempster, and Csiszár. The problem, known as the *Schrödinger bridge problem* (SBP) with "uniform" prior, was more recently recognized as a regularization of the Monge–Kantorovich *optimal mass transport* (OMT) problem, leading to effective computational schemes for the latter. Specifically, OMT with quadratic cost may be viewed as a zero-temperature limit of the problem posed by Schrödinger in the early 1930s. The latter amounts to minimization of *Helmholtz's free energy* over probability distributions that are constrained to possess two given marginals. The problem features a delicate compromise, mediated by a temperature parameter, between minimizing the internal energy and maximizing the entropy. These concepts are central to a rapidly expanding area of modern science dealing with the so-called *Sinkhorn algorithm*, which appears as a special case of an algorithm first studied in the more challenging continuous space setting by the French analyst Robert Fortet in 1938–1940 specifically for Schrödinger bridges. Due to the constraint on end-point distributions, dynamic programming is not a suitable tool to attack these problems. Instead, Fortet's iterative algorithm and its discrete counterpart, the Sinkhorn iteration, permit computation of the optimal solution by iteratively solving the so-called *Schrödinger system*. Convergence of the iteration is guaranteed by contraction along the steps in suitable metrics, such as Hilbert's projective metric.

In both the continuous as well as the discrete time and space settings, *stochastic control* provides a reformulation of and a context for the dynamic versions of general Schrödinger bridge problems and of their zero-temperature limit, the OMT problem. These problems, in turn, naturally lead to *steering problems for flows of one-time marginals* which represent a new paradigm for *controlling uncertainty*. The zero-temperature problem in the continuous-time and space setting turns out to be the celebrated Benamou–Brenier characterization of the *McCann displacement interpolation* flow in OMT. The formalism and techniques behind these control problems on flows of probability distributions have attracted significant attention in recent years as they lead to a variety of new applications in spacecraft guidance, control of robot or biological swarms, sensing, active cooling, and network routing as well as in computer and data science. This multifaceted and versatile framework, intertwining SBP and OMT, provides the substrate for the historical and technical overview

*Received by the editors May 22, 2020; accepted for publication (in revised form) October 29, 2020; published electronically May 6, 2021.
https://doi.org/10.1137/20M1339982
Funding: This work was partially supported by the NSF under grants 1807664, 1839441, 1901599, and 1942523, by the AFOSR under grant FA9550-17-1-0435, and by University of Padova Research Project CPDA 140897.
†School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (yongchen@gatech.edu).
‡Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697 USA (tryphon@uci.edu).
§Dipartimento di Matematica "Tullio Levi-Civita," Università di Padova, 35121 Padova, Italy (pavon@math.unipd.it).

of the field given in this paper. A key motivation has been to highlight links between the classical early work in both topics and the more recent stochastic control viewpoint, which naturally lends itself to efficient computational schemes and interesting generalizations.

Key words. stochastic control, Schrödinger bridge problem, optimal mass transport, Sinkhorn algorithm, Hilbert metric

AMS subject classifications. 49L, 60J60, 49J20, 35Q35, 28A50

DOI. 10.1137/20M1339982

Contents

1 Prelude: Sinkhorn's Algorithm 251

2 Overture: Science Dramas 252
2.1 Schrödinger's Drama . . . 252
2.2 Fourier's Drama . . . 252
2.3 Kantorovich . . . 254

3 Elements of Optimal Mass Transport Theory 256
3.1 The Monge–Kantorovich Static Problem . . . 256
3.2 The Dynamic Problem . . . 257
3.3 Optimal Mass Transport as a Stochastic Control Problem . . . 258
3.4 Optimal Mass Transport with a "Prior" . . . 261

4 Schrödinger's Bridges 263
4.1 The Hot Gas Gedankenexperiment . . . 263
4.2 Large Deviations and Maximum Entropy . . . 263
4.3 Derivation of the Schrödinger System . . . 267
4.4 Stochastic Control Formulation . . . 268
4.5 Fluid-Dynamic Formulation . . . 270
4.6 General Prior . . . 271

5 Stochastic Control and General Bridge Problems 272

6 The Atomic Hypothesis, Microscopes, and Stochastic Oscillators 277

7 Minimizing the Free Energy 280
7.1 Regularized Transport Problems . . . 280
7.2 Thermodynamic Systems: Statics . . . 281
7.3 Schrödinger and Sinkhorn, Redux . . . 283

8 The Fortet-IPF-Sinkhorn Algorithm 285
8.1 Continuous Case . . . 285
8.2 Discrete Case . . . 286
8.3 Hilbert's Projective Metric . . . 287
8.4 Proof of Theorem 8.1 . . . 289

9 Efficient vs. Robust Routing for Network Flows 291
9.1 Generalized Bridge Problems . . . 291
9.2 Invariance of Most Probable Paths . . . 293
9.3 Robust Network Routing . . . 294

9.4 Optimal Flows on Weighted Graphs 298

10 Closing Comments **302**

References **304**

1. Prelude: Sinkhorn's Algorithm.

In 1962, Richard Sinkhorn completed his doctoral dissertation entitled *On Two Problems Concerning Doubly Stochastic Matrices* and submitted [223], which would appear in *The Annals of Mathematical Statistics* in 1964. He showed there that the iterative process of alternatively normalizing the rows and columns of a matrix A with strictly positive elements converges to a doubly stochastic matrix D_1AD_2. Here, D_1 and D_2 are diagonal matrices with positive diagonal elements which are unique up to a scalar factor. By 1970, a survey by Fienberg had appeared [94], in which many other contributions to this topic, including [224, 225, 127], were cited. In contrast to Sinkhorn, Ireland, and Kullback [127], who studied a more general problem than Sinkhorn, credited Deming and Stephan (1940) [83] for first introducing the *iterative proportional fitting* (IPF) procedure.[1] Several other significant contributions on this problem followed, including generalizations to multidimensional matrices such as [18, 108], and this line of research continues to this day; see, e.g., [76, 4, 234, 198, 25, 85]. This large body of literature often ignores two papers by Erwin Schrödinger from the early 1930s [215, 216] as well as important contributions by Robert Fortet (1938–1940) [105, 106] and Arne Beurling (1960) [27]. The same goes for some later contributions on the *Schrödinger system* [130, 103]. But why should this body of work on Sinkhorn algorithms be related to Schrödinger's quest for the most likely random evolution between two given marginals for a cloud of diffusive particles?

One of the main goals of this paper is to provide an exhaustive answer to this question. To do that, we shall have to address several other questions such as, What is the relationship between Schrödinger's problem and Gaspar Monge's 18th century "Mémoire sur la théorie des déblais et des remblais" (1781) [173]? Was Schrödinger interested in *regularizing* the optimal mass transport (OMT) problem, as some recent papers seem to hint [23, 66, 161]? And how does the latter regularization relate to von Helmholtz's *free energy* of thermodynamics [119]? Is there a relation to Boltzmann's maximum entropy problem (1877) [34]? Is there a connection to multiplicative functional transformations of Markov processes [128], or to Bernstein's reciprocal processes [24, 129]? And what about connections to the Fisher information functional [242] and to positive maps on cones contracting Hilbert's projective metric [28, 42]? Last and surprisingly, but definitely not least, how is this scientific exploration intertwined with the Democritus atomic hypothesis?

As it turns out, all of these topics are indeed tightly connected. Therefore, what we would like to discuss in this paper lies right at the crossroads of major areas of science, some still in rapid development. Clearly, given the overwhelming spectrum of ideas and concepts, attempting to sort this all out may result in a fuzzy picture. Given our limited competence, how can we ever hope to give at least a reasonable/interesting account of all these intersecting areas (*sutor, ne ultra crepidam*!)?[2] Our choice is to

[1] Even earlier contributions are [252, 148].

[2] "Cobbler, not beyond the sandal!" In other words, don't pass judgment beyond your expertise. Attributed to the painter Apelles of Kos in Pliny the Elder's *Naturalis Historia*.

discuss this junction in science from an angle which is not the most common in the literature, namely, *stochastic control*. We shall try to provide evidence that once Schrödinger's problem has been converted to a stochastic control problem, it lends itself naturally to computational schemes as well as to interesting generalizations. It leads naturally to a *steering problem for probability distributions*, namely, a relaxed version of a most central problem in deterministic optimal control [98].

Our tale is apparently going to be long and complex and in permanent danger of excessive branching out and continuous flashbacks. The only way to avoid total chaos, it seems to us, is to start with Erwin Schrödinger's drama in the early 1930s.

2. Overture: Science Dramas.

We briefly recall two famous dramas from the history of science.

2.1. Schrödinger's Drama.

Edward Nelson concludes his 1967 jewel of a book [176] with a 1926 quotation from Erwin Schrödinger [214, Paragraph 14]: "...It has even been doubted whether what goes on in the atom could ever be described within the scheme of space and time. From the philosophical standpoint, I would consider a conclusive decision in this sense as equivalent to a complete surrender. For we cannot really alter our manner of thinking in space and time, and what we cannot comprehend within it we cannot understand at all."

Several years later, in 1953, Schrödinger writes [217]: "For it must have given to de Broglie the same shock and disappointment as it gave to me, when we learnt that a sort of transcendental, almost psychical interpretation of the wave phenomenon had been put forward, which was very soon hailed by the majority of leading theorists as the only one reconcilable with experiments, and which has now become the orthodox creed, accepted by almost everybody, with a few notable exceptions."[3]

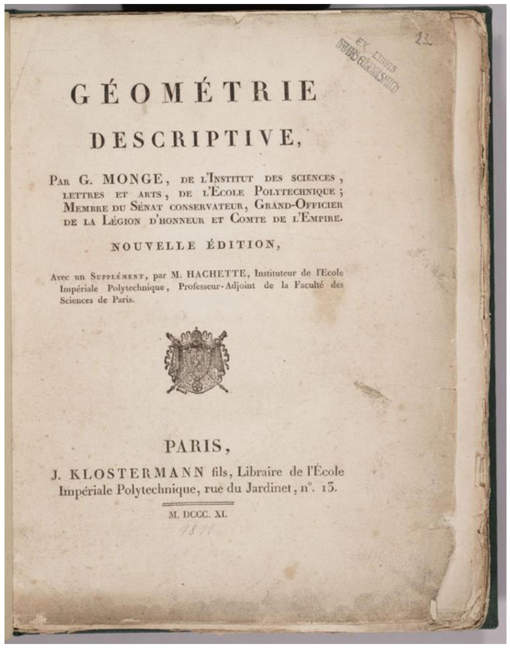
Between these two dramatic statements, however, there is a time in the early 1930s when Schrödinger has hope again and when he introduces what we now call the *Schrödinger bridges* in two remarkable papers [215, 216]. He states: "Merkwürdige Analogien zur Quantenmechanik, die mir sehr des Hindenkens wert erscheinen."[4] In this respect, the title of [215] is revealing: "Über die Umkehrung der Naturgesetze," namely, "On the reversal of natural laws." A few years later in 1937, another scientific giant, Andrey Kolmogorov, publishes a paper [147] with a very similar title, "Zur Umkehrbarkeit der statistischen Naturgesetze," namely, "On the reversibility of statistical natural laws." What had Schrödinger glimpsed? Before we discuss this, let's go back some 120 years to another dramatic moment in the history of science.

2.2. Fourier's Drama.

On December 21, 1807, Joseph B. Fourier submits a manuscript to the Institute of France in Paris entitled "Sur la propagation de la chaleur." He had been working on this memoir during his stay in Grenoble as Prefect of the Department of Isère, a post to which he had been appointed by Napoleon himself. The surprising, but not fully justified, results provoke an animated discussion among the examiners. The committee consists of Lagrange, Laplace, Lacroix, and Monge. Lagrange and Laplace, who criticize Fourier's expansion of functions as trigonometrical series, and Monge had all been teachers of Fourier at the École Normale. Fourier describes Monge [180] as "having a loud voice and is active, ingenious and very learned." In 1785, while teaching at the École Militaire in Paris, Monge, the father of *descriptive geometry* (see Figure 1), has among his students a sixteen-year-

[3] According to Nelson, "a realistic interpretation of quantum mechanics is, in my view, as unresolved as it was in the 1920s" [178, p. 230].

[4] "Remarkable analogies to quantum mechanics which appear to me very worthy of reflection."



GÉOMÉTRIE

DESCRIPTIVE,

Par G. MONGE, de l'Institut des sciences, lettres et arts, de l'École Polytechnique; Membre du Sénat conservateur, Grand-Officier de la Légion d'honneur et Comte de l'Empire.

NOUVELLE ÉDITION,

Avec un Supplément, par M. HACHETTE, Instituteur de l'École Impériale Polytechnique, Professeur-Adjoint de la Faculté des Sciences de Paris.

PARIS,

J. KLOSTERMANN fils, Libraire de l'École Impériale Polytechnique, rue du Jardinet, n°. 13.

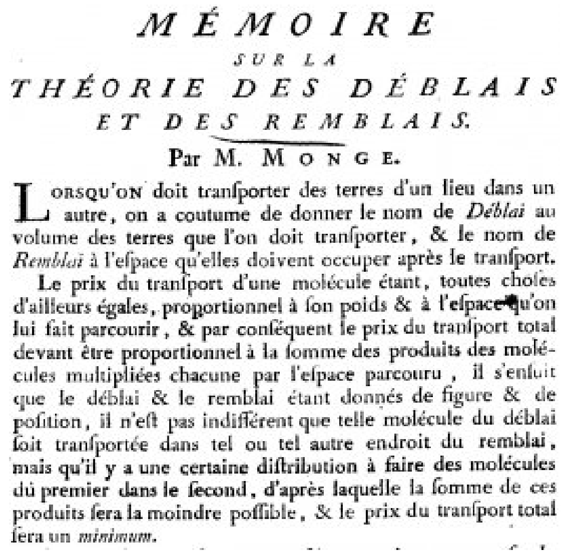
M. DCCC. XI.

Fig. 1 *G. Monge,* Géométrie Descriptive.

old Corsican with a gift for mathematics by the name of Napoleon Bonaparte. The latter will be examined upon graduation by Pierre-Simon de Laplace. Of the group of outstanding mathematicians active in France in that time, Fourier and Monge are those who take on important political positions. Monge is Minister of the Marine during the French revolution, and in between two assignments in Italy, he directs the École Polytechnique which he had coestablished in 1794 with Lazare Carnot and Napoleon. Fourier and Monge had also been together in the expedition to Egypt in 1798 and were members of the mathematics division of the Cairo Institute together with Malus and Napoleon himself. In spite of all of this, the manuscript is eventually rejected (Fourier's *Théorie analytique de la chaleur* appears only in 1822).[5]

Let us zoom in on Gaspard Monge, Comte de Péluse. A first version (unfortunately lost) of his *Mémoire sur la théorie des déblais et des remblais* [*Dissertation on the theory of earth-moving and embankments*] is read at the Académie des Sciences on January 27 and February 7, 1776. Although the memoir is proposed for publication by the secretary Condorcet, only on March 28, 1781, does Monge read a second version of his memoir; cf. Figure 2. A 40-page publication follows in 1784. Monge, who is of rather humble origins, had worked for military institutions for a while, taking advantage of his extraordinary drawing skills and geometric intuition. He had applied descriptive geometry to such problems as designing cannons (Figure 3), cutting stones, planning city walls, and drawing shadows. Now, however, he is interested in the following problem, which has both civil and military applications: Suppose you need to build some embankments carrying debris from another location; how should the transport occur so that the average distance is minimized? Monge

[5]In 1826, Fourier announces a method for the solution of systems of linear inequalities [107] that has elements in common with the simplex method of linear programming.



MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.

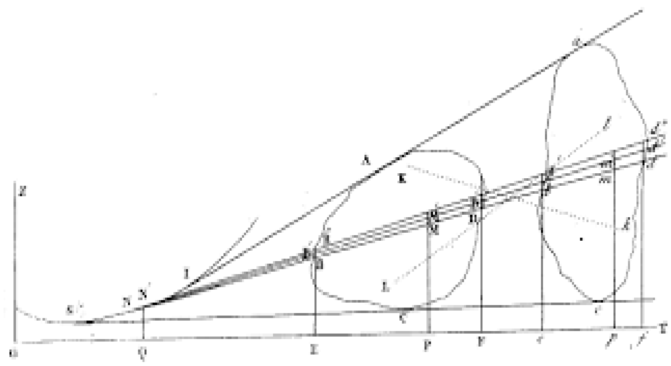
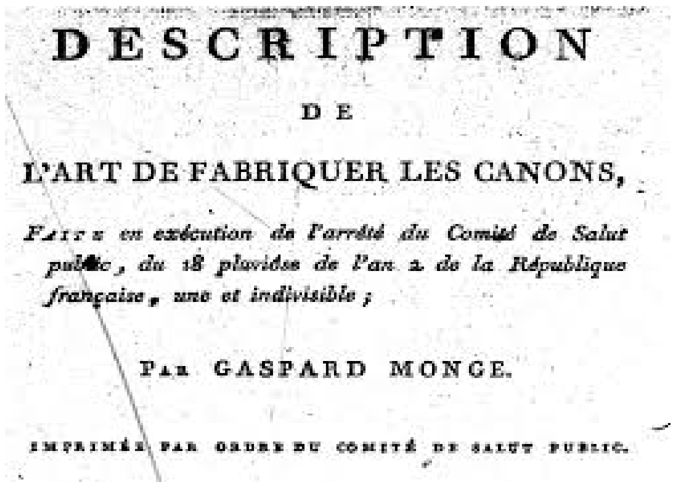


Fig. 2 *Front page of Monge's mémoire and Monge's drawing.*

discusses two- and three-dimensional problems showing a profound understanding of the problem and its challenging aspects. In particular, his intuition of the normality of optimal transport paths to a certain one-parameter family of surfaces was proven to be correct more than one hundred years later in a 200-page memoire by Appel [8].

2.3. Kantorovich. Not much happens until the 1920s and 1930s, when the transportation problem is first studied mathematically by A. N. Tolstoi [236, 237]. Then, in 1939, the transport problem is briefly mentioned by the Soviet mathematician Leonid Kantorovich in the booklet [137], where he lays the foundations of linear programming, including duality theory and a variant of the simplex method. In 1942, Kantorovich provides in [138] the generalization of linear programming to an abstract setting. He considers the following problem: Let X be compact metric space with distance function d. Let ν_0 and ν_1 be probability measures on X, and let $\Pi(\nu_0,\nu_1)$ denote all the probability distributions on $X\times X$ having ν_0 and ν_1 as marginal distributions. Consider the problem

$$K(\nu_0,\nu_1)=\inf_{\pi\in\Pi(\nu_0,\nu_1)}\int_{X\times X} d(x,y)d\pi(x,y). \tag{2.1}$$



DESCRIPTION

DE

L'ART DE FABRIQUER LES CANONS,

Faite en exécution de l'arrêté du Comité de Salut public, du 18 pluviôse de l'an 2 de la République française, une et indivisible ;

PAR GASPARD MONGE.

IMPRIMÉE PAR ORDRE DU COMITÉ DE SALUT PUBLIC.

Fig. 3 *G. Monge,* L'Art de fabriquer les Canons.

Kantorovich establishes the following fundamental duality theorem (see also Theorem 3.1 for the case of a nondistance cost function).

THEOREM 2.1.

$$K(\nu_0,\nu_1)=\sup_{\varphi}\left\{\int\varphi(x)d\,(\nu_0-\nu_1)\,;\|\varphi\|_{\mathrm{Lip}}\le 1,\ \varphi\in L^1(|\nu_0-\nu_1|)\right\}. \tag{2.2}$$

Lipschitz functions with constant 1 relative to the metric d are called *potentials* by Kantorovich. The transport plan π is called a *potential plan* if

$$\varphi(x)-\varphi(y)=d(x,y),\quad \pi \text{ a.s.},$$

for an (optimal) potential function φ (this theorem will be complemented in 1958 by the paper [140] written jointly with his student Rubinstein). In 1947 Kantorovich [139], seeing the proceedings of a conference held in Leningrad (St. Petersburg) on the bicentennial of Monge's birth, realizes that the surfaces of Monge are just the level surfaces of the optimal potentials (dual functions) he had defined in the Doklady note [138]. Thus, Kantorovich, a functional analyst motivated by economics applications, provides a more manageable (relaxed) formulation of the transport problem and major advances, opening up the avenue for the impressive developments of the past twenty years [202, 203, 92, 242, 5, 243, 185]; see [241] for a full historical account of Kantorovich's contributions.[6]

[6]It would have been possible to call this subsection "Kantorovich's Drama." Indeed, to this day it is little known that he should be credited for linear programming, including the simplex method, and for duality theory even in an abstract setting. His metric is curiously called Wasserstein (Vasershtein) due to Dobrushin being aware of the Vasershtein paper [240] (Vasershtein worked in the laboratory he headed) but not of Kantorovich's publications. In addition, Kantorovich's ideas on mathematical economics were long considered in official Soviet circles as anti-Marxist. Consequently, they suffered for many years a sort of ostracism. For more detail on all of this, see [241]. Fortunately, as partial compensation, in 1975 he was awarded the Nobel prize for economics.

3. Elements of Optimal Mass Transport Theory. The literature on this problem is by now so vast that given our degree of competence we shall not even attempt here to give a reasonable and/or balanced introduction to the various fascinating aspects of this theory. Fortunately, there exist excellent monographs and survey papers on this topic; see [202, 203, 92, 242, 5, 243, 185]. The range of applications has also increased exponentially; we mention quality control, industrial manufacturing, vehicle path planning [205, 206], image processing [23], computer graphics [226, 227, 229, 230, 231, 150], machine learning [228, 9, 174], and econometrics [109], among others. We shall only briefly review some concepts and results that are relevant to the topics of this paper.

3.1. The Monge–Kantorovich Static Problem. Let ν_0 and ν_1 be probability measures on the separable, complete metric spaces X and Y, respectively. Let $c : X \times Y \to [0, +\infty)$ be a lower semicontinuous map with $c(x, y)$ representing the cost of transporting a unit of mass from location x to location y. Let $\mathcal{T}_{\nu_0\nu_1}$ be the family of measurable maps $T : X \to Y$ such that $T\#\nu_0 = \nu_1$, namely, such that ν_1 is the *push-forward* of ν_0 under T. Any $T \in \mathcal{T}_{\nu_0\nu_1}$ is called a *transport map*. Then Monge's OMT problem is

$$\inf_{T\in\mathcal{T}_{\nu_0\nu_1}} \int_X c(x, T(x))d\nu_0(x) \tag{3.1}$$

for the particular case where $c(x, y) = d(x, y)$, i.e., it is a metric on X and $X = Y$. This problem may be unfeasible and the family $\mathcal{T}_{\nu_0\nu_1}$ may be empty.[7] This is never the case for the "relaxed" version of the problem studied by Kantorovich in the 1940s,

$$\inf_{\pi\in\Pi(\nu_0,\nu_1)} \int_{X\times Y} c(x, y)d\pi(x, y). \tag{3.2}$$

Here $\Pi(\nu_0, \nu_1)$ are "couplings" of ν_0 and ν_1, namely, probability distributions on $X\times Y$ with marginals ν_0 and ν_1 called *transport plans*. Indeed, $\Pi(\nu_0, \nu_1)$ always contain the product measure $\nu_0 \otimes \nu_1$. We have Kantorovich's duality theorem given as follows.

THEOREM 3.1. *Suppose c is lower semicontinuous; then there exists a solution to problem* (3.2). *Moreover,*

$$\min_{\pi\in\Pi(\nu_0,\nu_1)} \int_{X\times Y} c(x, y)d\pi(x, y) = \sup_{(\varphi,\psi)\in\Phi_c} \left[\int_X \varphi d\nu_0 + \int_Y \psi d\nu_1\right],$$
$$\Phi_c = \{(\varphi, \psi)|\varphi \in L^1(\nu_0), \psi \in L^1(\nu_1), \varphi(x) + \psi(y) \le c(x, y)\}.$$

Let us specialize the Monge–Kantorovich problem (3.2) to the case $X = Y = \mathbb{R}^n$ and $c(x, y) = \|x - y\|^2$. Then if ν_1 does not give mass to sets of dimension $\le n - 1$, by Brenier's theorem [242, p. 66] there exists a unique optimal transport plan π (Kantorovich) induced by a $d\nu_0$ a.e. unique (Monge) map T, of the form $T = \nabla\varphi$ with φ convex, such that

$$\pi = (I \times \nabla\varphi)\#\nu_0, \quad \nabla\varphi\#\nu_0 = \nu_1. \tag{3.3}$$

Here I denotes the identity map. Among the extensions of this result, we mention that to strictly convex, superlinear costs c by Gangbo and McCann [110]. The optimal

[7] This is the case when, e.g., ν_0 is a Dirac distribution and ν_1 is the sum of two Dirac distributions of half the magnitude. Since ν_0 needs to be "split" so as to be transferred at two separate locations, a transference plan $\mathcal{T}$ does not exist as a map.

transport problem may be used to introduce a useful distance between probability measures. Indeed, let $\mathcal{P}_2(\mathbb{R}^n)$ be the set of probability measures μ on $\mathbb{R}^n$ with finite second moment. For $\nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, the Kantorovich–Wasserstein quadratic distance is defined by

$$W_2(\nu_0, \nu_1) = \left(\inf_{\pi \in \Pi(\nu_0, \nu_1)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2 d\pi(x, y) \right)^{1/2}. \tag{3.4}$$

As is well known [242, Theorem 7.3], W_2 is a *bona fide* distance. Moreover, it provides a most natural way to "metrize" weak convergence[8] in $\mathcal{P}_2(\mathbb{R}^n)$ [242, Theorem 7.12], [5, Proposition 7.1.5] (the same applies to the case $p \geq 1$ replacing 2 with p everywhere). The *Kantorovich–Wasserstein space* $\mathcal{W}_2$ is defined as the metric space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$. It is a *Polish space*, namely, a separable, complete metric space.

3.2. The Dynamic Problem. So far, we have dealt with the *static* optimal transport problem. Nevertheless, in [22, p. 378] it is observed that "... a continuum mechanics formulation was already implicitly contained in the original problem addressed by Monge... Eliminating the time variable was just a clever way of reducing the dimension of the problem." Thus, a *dynamic* (*Eulerian*) version of the OMT problem was already *in fieri* in Gaspar Monge's 1781 *Mémoire sur la théorie des déblais et des remblais*! It was elegantly accomplished by Benamou and Brenier in [22] by showing that

$$W_2^2(\nu_0, \nu_1) = \inf_{(\mu, v)} \int_0^1 \int_{\mathbb{R}^n} \|v(t, x)\|^2 \mu_t(dx) dt, \tag{3.5a}$$

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (v\mu) = 0, \tag{3.5b}$$

$$\mu_0 = \nu_0, \quad \mu_1 = \nu_1. \tag{3.5c}$$

Here the flow $\{\mu_t; 0 \leq t \leq 1\}$ varies over continuous maps from $[0, 1]$ to $\mathcal{P}_2(\mathbb{R}^n)$ and v over smooth fields. Benamou and Brenier were motivated by computational considerations, a topic which had not received much attention in OMT; see [7]. In [243], Villani states at the beginning of Chapter 7 that two main motivations for the time-dependent version of OMT are that

- a time-dependent model gives a more complete description of the transport;
- the richer mathematical structure will be useful later on.

We can add three further reasons:

1. It allows us to view the optimal transport problem as an (atypical) *optimal control* problem; see section 3.3 below and [50, 51, 52, 53, 54, 55, 56, 57, 58, 59].
2. It opens the way to establish a connection with the *Schrödinger bridge* problem, where the latter appears as a regularization of the former [168, 169, 170, 157, 156, 52, 54, 46].
3. In some applications such as computer graphics [229, 150], interpolation of images [55], spectral morphing [134], machine learning [174], and network routing [64, 64, 65], the interpolating flow is essential!

Let $\{\mu_t^*; 0 \leq t \leq 1\}$ and $\{v^*(t, x); (t, x) \in [0, 1] \times \mathbb{R}^n\}$ be optimal for (3.5). Then

$$\mu_t^* = [(1 - t)I + t\nabla\varphi] \# \nu_0,$$

[8] Here, we say that as $k \to \infty$, μ_k converges weakly to μ in $\mathcal{P}_2(\mathbb{R}^n)$ if $\int_{\mathbb{R}^n} f d\mu_k \to \int_{\mathbb{R}^n} f d\mu$ for any continuous function f satisfying $f(x) \leq c\left(1 + d(x, x_0)^2\right)$ for every $x_0 \in \mathbb{R}^n$, the latter condition guaranteeing tightness of the sequence $\{\mu_k\}$.

with $T = \nabla\varphi$ solving Monge's problem, provides McCann's *displacement interpolation* between ν_0 and ν_1 [167]. In fact, $\{\mu_t^*; 0 \leq t \leq 1\}$ may be seen as a constant-speed geodesic joining ν_0 and ν_1 in $\mathcal{W}_2$ and, moreover, as realized by Otto in [189], $\mathcal{W}_2$ can be endowed with a Riemannian-like structure[9] which is consistent with W_2.

McCann discovered [167] that certain functionals are *displacement convex*, namely, convex along Wasserstein geodesics. This has led to a variety of applications. Following one of Otto's main discoveries [136, 189], it turns out that a large class of PDEs may be viewed as *gradient flows* on the Wasserstein space $\mathcal{W}_2$. This interpretation, because of the displacement convexity of the functionals, is well suited to establishing uniqueness and to studying energy dissipation and convergence to equilibrium. A rigorous setting in which to make sense of the Otto calculus has been developed by Ambrosio, Gigli, and Savaré [5] for a suitable class of functionals. Convexity along geodesics in $\mathcal{W}_2$ also leads to new proofs of various geometric and functional inequalities [167], [242, Chapter 9], [72]. Finally, we mention that, when the space is not flat, qualitative properties of optimal transport can be quantified in terms of how bounds on the Ricci–Curbastro curvature affect the displacement convexity of certain specific functionals [243, Part II].

In passing and for completeness, we note that the *tangent space* of $\mathcal{P}_2(\mathbb{R}^n)$ at a probability measure μ, denoted by $T_\mu\mathcal{P}_2(\mathbb{R}^n)$ [5], may be identified with the closure in L^2_μ of the span of vector fields $\{\nabla\varphi : \varphi \in C_c^\infty\}$, where C_c^∞ is the family of smooth functions with compact support. It is naturally equipped with the inner product of L^2_μ. Several recent papers have contributed to the development of second-order calculus in Wasserstein's space [245, 69, 70, 71], building on [164, 113, 6].

3.3. Optimal Mass Transport as a Stochastic Control Problem. Optimal control, deeply rooted in the classical calculus of variations, seeks to modify the natural (*free*) evolution of a system so as to minimize a suitable cost. This field received a boost in the days of the space race, motivated by aeronautical and astronautical problems such as navigation and the soft moon landing problem [98, p. 21]. Foundational contributions were provided in the late 1950s and early 1960s by Pontryagin and his school in the Soviet Union, and by Bellman, Kalman, and others in the United States. When the system starts from random initial conditions and/or is subject to random disturbances, the problem becomes one of a stochastic nature where the cost is now the expectation of a suitable random functional. A crucial aspect of the problem is the information that is available to design the control action. In this paper, we shall mainly discuss the case where the state variables constitute a fully observable (vector) Markov process with values in a Euclidean space or, in the discrete time setting, in a finite alphabet set. In the continuous time setting, standard references are Lee and Markus [151] and Fleming and Rishel [98]. For Markov decision processes, standard references are [200, 26]. As we shall see, both the OMT problem and its regularized version (Schrödinger bridge) can be viewed as stochastic optimal control problems: It is possible to derive suitable Hamilton–Jacobi–Bellman (HJB) equations. The difficulty lies in selecting the appropriate solution of the HJB as the boundary term is missing. Our first step in introducing the optimal control viewpoint on the topics of this paper consists in rederiving the Benamou–Brenier formulation of OMT using elementary control considerations.

[9]More precisely, a *weak Riemannian structure* (see [6, 2.3.2]), a limitation being that the tangent space about singular distributions is not "rich enough," leading toward all other nearby distributions.

Let us start by observing that the square of the Euclidean distance can be expressed as the infimum of an action integral, namely,

$$\frac{1}{2}\|x-y\|^2 = \min_{x\in\mathcal{X}_{xy}} \int_0^1 \frac{1}{2}\|\dot{x}\|^2 dt, \tag{3.6}$$

where $\mathcal{X}_{xy}$ is the family of $C^1([0,1];\mathbb{R}^n)$ paths with $x(0)=x$ and $x(1)=y$. The minimum in (3.6) is evidently achieved by

$$x^*(t) = (1-t)x + ty,$$

namely, the straight line joining x and y. Since $x^*(t)$ is a Euclidean geodesic, any probabilistic average of the lengths of C^1 trajectories starting at x at time 0 and ending at y at time 1 necessarily gives a higher value. Thus, the probability measure on $C^1([0,1];\mathbb{R}^n)$ concentrated on the path $\{x^*(t); 0\le t\le 1\}$ solves the problem

$$\inf_{P_{xy}\in\mathbb{D}^1(\delta_x,\delta_y)} \mathbb{E}_{P_{xy}}\left\{\int_0^1 \frac{1}{2}\|\dot{x}\|^2 dt\right\}, \tag{3.7}$$

where $\mathbb{D}^1(\delta_x,\delta_y)$ are the probability measures on $C^1([0,1];\mathbb{R}^n)$ whose initial and final one-time marginals are Dirac deltas concentrated at x and y, respectively. Since (3.7) provides us with yet another representation for $\frac{1}{2}\|x-y\|^2$, in view of (3.4), we also find that

$$\inf_{\pi\in\Pi(\nu_0,\nu_1)} \int \frac{1}{2}\|x-y\|^2 d\pi(x,y) = \inf_{\pi\in\Pi(\nu_0,\nu_1)} \int \inf_{P_{xy}\in\mathbb{D}^1(\delta_x,\delta_y)} \mathbb{E}_{P_{xy}}\left\{\int_0^1 \frac{1}{2}\|\dot{x}\|^2 dt\right\} d\pi.$$

Now observe that if $P_{xy}\in\mathbb{D}^1(\delta_x,\delta_y)$ and $\pi\in\Pi(\nu_0,\nu_1)$, then

$$P = \int_{\mathbb{R}^n\times\mathbb{R}^n} P_{xy} d\pi(x,y)$$

is a probability measure in $\mathbb{D}^1(\nu_0,\nu_1)$, namely, a measure on $C^1([0,1];\mathbb{R}^n)$ with initial and final marginals ν_0 and ν_1, respectively. On the other hand, the disintegration of any measure $P\in\mathbb{D}^1(\nu_0,\nu_1)$ with respect to the initial and final positions[10] yields $P_{xy}\in\mathbb{D}^1(\delta_x,\delta_y)$ and $\pi\in\Pi(\nu_0,\nu_1)$. Thus, we get that the original optimal transport problem is equivalent to

$$\inf_{P\in\mathbb{D}^1(\nu_0,\nu_1)} \mathbb{E}_P\left\{\int_0^1 \frac{1}{2}\|\dot{x}\|^2 dt\right\}. \tag{3.8}$$

So far, we have followed [157, pp. 2–3]. Instead of the "particle" picture, we can also consider the hydrodynamic version of (3.6), namely, the optimal control problem

$$\begin{aligned} \frac{1}{2}\|x-y\|^2 &= \inf_{v\in\mathcal{V}_y} \int_0^1 \frac{1}{2}\|v(t,x^v(t))\|^2 dt, \\ \dot{x}^v(t) &= v(t,x^v(t)), \quad x(0)=x, \end{aligned} \tag{3.9}$$

[10]Disintegration can be viewed here as the opposite process to the construction of a product measure.

where the admissible feedback control laws $v(\cdot,\cdot)$ in $\mathcal{V}_y$ are continuous and such that $x^v(1)=y$. Following the same steps as before, we get that the optimal transport problem is equivalent to the following stochastic control problem with atypical boundary constraints:

$$\inf_{v\in\mathcal{V}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\|v(t,x^v(t))\|^2 dt\right\}, \tag{3.10a}$$

$$\dot{x}^v(t) = v(t,x^v(t)) \quad \text{a.s.}, \quad x(0)\sim\nu_0, \quad x(1)\sim\nu_1. \tag{3.10b}$$

Finally, suppose $d\nu_0(x)=\rho_0(x)dx$, $d\nu_1(y)=\rho_1(y)dy$, and $x^v(t)\sim\rho(t,x)dx$. Then, necessarily, ρ satisfies (weakly) the continuity equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(v\rho) = 0, \tag{3.11}$$

expressing the conservation of probability mass. Moreover,

$$\mathbb{E}\left\{\int_0^1 \frac{1}{2}\|v(t,x^v(t))\|^2 dt\right\} = \int_{\mathbb{R}^n}\int_0^1 \frac{1}{2}\|v(t,x)\|^2\rho(t,x)dtdx.$$

Hence (3.10) turns into the Benamou–Brenier problem (3.5),

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n}\int_0^1 \frac{1}{2}\|v(t,x)\|^2\rho(t,x)dtdx, \tag{3.12a}$$

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(v\rho) = 0, \tag{3.12b}$$

$$\rho(0,x)=\rho_0(x), \quad \rho(1,y)=\rho_1(y). \tag{3.12c}$$

The variational analysis for (3.10) or, equivalently, for (3.12) can be carried out in many different ways. For instance, let $\mathcal{P}_{\rho_0\rho_1}$ be the family of flows of probability densities $\rho=\{\rho(t,\cdot); 0\le t\le 1\}$ satisfying (3.12c) and let $\mathcal{V}$ be the family of continuous feedback control laws $v(\cdot,\cdot)$. Consider the unconstrained minimization of the Lagrangian over $\mathcal{P}_{\rho_0\rho_1}\times\mathcal{V}$,

$$\mathcal{L}(\rho,v) = \int_{\mathbb{R}^n}\int_0^1 \left[\frac{1}{2}\|v(t,x)\|^2\rho(t,x) + \lambda(t,x)\left(\frac{\partial\rho}{\partial t}+\nabla\cdot(v\rho)\right)\right] dtdx, \tag{3.13}$$

where λ is a C^1 Lagrange multiplier. Integrating by parts, assuming that limits for $\|x\|\to\infty$ are zero, we get

$$\begin{aligned} &\int_{\mathbb{R}^n}\int_0^1 \left[\frac{1}{2}\|v(t,x)\|^2 + \left(-\frac{\partial\lambda}{\partial t} - \nabla\lambda\cdot v\right)\right]\rho(t,x)dtdx \\ &\qquad + \int_{\mathbb{R}^n}\left[\lambda(1,x)\rho_1(x) - \lambda(0,x)\rho_0(x)\right]dx. \end{aligned} \tag{3.14}$$

The last integral is constant over $\mathcal{P}_{\rho_0\rho_1}$ for a fixed λ and can therefore be discarded. We are left to minimize

$$\int_{\mathbb{R}^n}\int_0^1 \left[\frac{1}{2}\|v(t,x)\|^2 + \left(-\frac{\partial\lambda}{\partial t} - \nabla\lambda\cdot v\right)\right]\rho(t,x)dtdx \tag{3.15}$$

over $\mathcal{P}_{\rho_0\rho_1}\times\mathcal{V}$. We consider doing this in two stages, starting from minimization with respect to v for a fixed flow of probability densities $\rho=\{\rho(t,\cdot); 0\le t\le 1\}$ in $\mathcal{P}_{\rho_0\rho_1}$. Pointwise minimization of the integrand at each time $t\in[0,1]$ gives that

$$v_\rho^*(t,x) = \nabla\lambda(t,x), \tag{3.16}$$

which is continuous. Substituting this expression for the optimal control into (3.15), we obtain

$$J(\rho) = -\int_{\mathbb{R}^n}\int_0^1 \left[\frac{\partial\lambda}{\partial t} + \frac{1}{2}\|\nabla\lambda\|^2\right]\rho(t,x)dtdx. \tag{3.17}$$

In view of this, if λ satisfies the Hamilton–Jacobi equation

$$\frac{\partial\lambda}{\partial t} + \frac{1}{2}\|\nabla\lambda\|^2 = 0, \tag{3.18}$$

then $J(\rho)$ is identically zero over $\mathcal{P}_{\rho_0\rho_1}$ and any $\rho \in \mathcal{P}_{\rho_0\rho_1}$ minimizes the Lagrangian (3.13) together with the feedback control (3.16). We have therefore established the following proposition [22].

Proposition 3.2. *Let $\rho^*(t,x)$ with $t \in [0,1]$ and $x \in \mathbb{R}^n$ satisfy*

$$\frac{\partial\rho^*}{\partial t} + \nabla\cdot(\rho^*\nabla\lambda) = 0, \quad \rho^*(0,x) = \rho_0(x), \tag{3.19}$$

where λ is a solution of the Hamilton–Jacobi equation

$$\frac{\partial\lambda}{\partial t} + \frac{1}{2}\|\nabla\lambda\|^2 = 0 \tag{3.20}$$

for some boundary condition $\lambda(1,x) = \lambda_1(x)$. If $\rho^*(1,x) = \rho_1(x)$, then the pair (ρ^*, v^*) with $v^*(t,x) = \nabla\lambda(t,x)$ is a solution of (3.5).

The stochastic nature of the Benamou–Brenier formulation (3.12) stems from the fact that the initial and final densities are specified. Accordingly, the above requires solving a two-point boundary value problem and the resulting control dictates the local velocity field. In general, one cannot expect to have a classical solution of (3.20) and has to be content with a viscosity solution [104]. See [232] for a recent contribution in the case when only samples of ρ_0 and ρ_1 are known.

3.4. Optimal Mass Transport with a "Prior." The stochastic control formulation (3.10) of OMT casts this as a problem to steer the dynamical system $\dot{x} = u$, where u is the control input, between specified marginal distributions for the state. The generalization to nontrivial underlying dynamics of the form $\dot{x} = f(t,x)+u$ leads in a similar manner to

$$\inf_{u\in\mathcal{V}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\|u(t,x^u(t))\|^2 dt\right\}, \tag{3.21a}$$

$$\dot{x}^u(t) = f(t,x^u(t)) + u(t,x^u(t)) \quad \text{a.s.}, \quad x(0)\sim\nu_0, \quad x(1)\sim\nu_1. \tag{3.21b}$$

Once again, this is a nonstandard minimum energy control-effort problem due to the constraint on the final state distribution. A fluid dynamic reformulation proceeds as follows. Suppose we are given two end-point marginal probability densities ρ_0 and ρ_1, and suppose we are also given a model

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(f\rho) = 0 \tag{3.22}$$

for the flow of probability densities $\{\rho(t,x); 0 \le t \le 1\}$, for a continuous vector field $f(\cdot,\cdot)$, which, however, is not consistent with the given end-point marginals. Then

(3.22) represents a "prior" evolution to serve as a reference when seeking an update in the vector field to minimize the quadratic cost,

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2}\|v(t,x) - f(t,x)\|^2 \rho(t,x) dt dx, \tag{3.23a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{3.23b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y). \tag{3.23c}$$

Clearly, if the prior flow satisfies $\rho(0,x) = \rho_0(x)$ and $\rho(1,y) = \rho_1(y)$, then it solves the problem and $v^* = f$. Moreover, the standard OMT problem is recovered when the prior evolution is constant, i.e., $f \equiv 0$.

Remark 3.3. From a different angle, problem (3.23) can be motivated as follows. It seeks a correction v to a transportation plan f that has already been computed from obsolete data (e.g., marginal distributions for transportation of resources in $\rho_{0,\text{old}}$ to meet demands in $\rho_{1,\text{old}}$) as this data is being updated to a new set of marginals ρ_0 and ρ_1, respectively.

The particle version of (3.23) takes the form of a more familiar OMT problem, namely,

$$\inf_{\pi \in \Pi(\nu_0,\nu_1)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y) d\pi(x,y), \tag{3.24}$$

where $d\nu_0(x) = \rho_0(x)dx$, $d\nu_1(y) = \rho_1(y)dy$, and

$$c(x,y) = \inf_{x \in \mathcal{X}_{xy}} \int_0^1 L(t, x(t), \dot{x}(t)) dt, \quad L(t,x,\dot{x}) = \|\dot{x} - f(t,x)\|^2. \tag{3.25}$$

The explicit calculation of the function $c(x,y)$ when $f \not\equiv 0$ is nontrivial. Moreover, the zero-noise limit results of [156, section 3], based on a large deviations principle [81], although very general in other ways, seem to cover here only the case where $c(x,y) = c(x-y)$ is strictly convex originating from a Lagrangian $L(t,x,\dot{x}) = c(\dot{x})$. We mention that [55] deals with OMT problems where the Lagrangian is *not strictly convex* with respect to $\dot{x}$. Finally, we feel that the present formulation is a most natural one in which to study zero-noise limits of Schrödinger bridges with a general Markovian prior evolution. References [54, 55] discuss this same problem in the case of a Gaussian prior and show directly the convergence of the solution to the HJB equation to the solution of a Hamilton–Jacobi equation.

The variational analysis for (3.23) can be carried out as in section 3.3, leading to the following result.

PROPOSITION 3.4. *If λ satisfies the Hamilton–Jacobi equation*

$$\frac{\partial \lambda}{\partial t} + f \cdot \nabla \lambda + \frac{1}{2}\|\nabla \lambda\|^2 = 0, \tag{3.26}$$

and is such that the solution ρ^* to

$$\frac{\partial \rho^*}{\partial t} + \nabla \cdot [(f + \nabla \lambda)\rho^*] = 0, \quad \rho^*(0,x) = \rho_0(x), \tag{3.27}$$

satisfies the end-point condition $\rho^*(1,x) = \rho_1(x)$ as well, then the pair $(\rho^*(t,x), v^*(t,x))$, with

$$v^*(t,x) = f(t,x) + \nabla \lambda(t,x),$$

solves (3.23), *provided $\lambda(t,x)\rho^*(t,x)$ vanishes as $\|x\| \to \infty$ for each fixed t.*

4. Schrödinger's Bridges.

Two excellent surveys on this topic are [247, 157]. See also [62] for an exposition of the Schrödinger bridge problem in the context of control engineering.

4.1. The Hot Gas Gedankenexperiment.

In 1931–1932, Erwin Schrödinger considered the following Gedankenexperiment [215, 216]: We have a cloud of N independent Brownian particles evolving in time.[11] This cloud of particles has been observed having at the initial time $t = 0$ an empirical distribution approximately equal to $\rho_0(x)dx$. At time $t = 1$, an empirical distribution approximately equal to $\rho_1(x)dx$ is observed, which differs considerably from what it should be according to the law of large numbers (N is large, say, of the order of Avogadro's number), namely,

$$\rho_1(y) \neq \int_{\mathbb{R}^n} p(0, x, 1, y)\rho_0(x)dx,$$

where

$$p(s, x, t, y) = [2\pi(t-s)]^{-\frac{n}{2}} \exp\left[-\frac{\|x-y\|^2}{2(t-s)}\right], \quad s < t, \tag{4.1}$$

is the transition density of the Wiener process (heat kernel). It is apparent that the particles have been transported in an unlikely way. But of the many unlikely ways in which this could have happened, which one is the most likely? In modern probabilistic terms, this is a problem of *large deviations of the empirical distribution* as observed by Föllmer [103]. Thus, at the outset, Schrödinger's motivation and the context of his question had no connection to OMT that we have just discussed—the confluence of the two that we highlight in what follows may be seen as a deep and lucky coincidence.

4.2. Large Deviations and Maximum Entropy.

The area of large deviations is concerned with the probabilities of rare events. In light of Sanov's theorem [211], Schrödinger's problem can be seen as a large deviations (maximum entropy) problem for distributions on trajectories, as we proceed to explain.

Let $\Omega = C([0,1]; \mathbb{R}^n)$ be the space of $\mathbb{R}^n$-valued continuous functions, let $\mathcal{D}$ be the space of probability measures on Ω, and let $W_x \in \mathcal{D}$ denote the Wiener measure starting at x at $t = 0$. If, instead assuming a Dirac marginal concentrated at x, we enlarge W_x to subsume the volume measure at $t = 0$, we obtain

$$W := \int W_x \, dx.$$

This is an unbounded nonnegative measure on the path space Ω, called the *stationary Wiener measure* (or sometimes *reversible Brownian motion*); W has marginals at each point in time that coincide with the Lebesgue measure and therefore, while still nonnegative, it is not a probability measure. In fact, it serves as a convenient analogue of the Lebesgue measure on paths that "symmetrizes" and "uniformizes" the Wiener measure with respect to the time arrow.

Alternatively, we can enlarge W_x to

$$W_\rho := \int W_x \, \rho(x)dx$$

[11] To put Schrödinger's 1931 work into perspective, one has to recall what science had accomplished on the atomic hypothesis and physical Brownian motion up to that time, besides Boltzmann's work [34]. For this, see [176] and section 6.

so that instead the Dirac marginal at $t = 0$ now has a marginal probability measure ρdx. Clearly, W_ρ is a probability measure on Ω which is absolutely continuous with respect to W. Indeed, if

$$\mathbb{D}(P\|Q) = \begin{cases} \mathbb{E}_P\left\{\log \frac{\mathrm{d}P}{\mathrm{d}Q}\right\} & \text{if } P \ll Q, \\ +\infty & \text{otherwise} \end{cases}$$

denotes the relative entropy functional (divergence, Kullback–Leibler index) between nonnegative measures, it can be seen that

$$\mathbb{D}(W_\rho\|W) = \int \rho \log(\rho) dx \; = -S(\rho),$$

where $S(\rho)$ is the differential entropy of the measure $\rho(x)dx$. Moreover,

$$\mathbb{D}(W_\rho\|W_{\rho_0}) = \int \rho \log\left(\frac{\rho}{\rho_0}\right) dx \; =: \mathbb{D}(\rho\|\rho_0).$$

Now let $X^1, X^2, \ldots$ be i.i.d. Wiener evolutions on $[0, 1]$ with values in $\mathbb{R}^n$ and let starting value x^i be distributed according to ρ. The *empirical distribution* μ_N associated to $X^1, X^2, \ldots, X^N$ is defined by

$$\mu_N(X^1, X^2, \ldots, X^N) := \frac{1}{N}\sum_{i=1}^{N} \delta_{X^i}. \tag{4.2}$$

The expression in (4.2) defines a map from Ω^N to the space $\mathcal{D}$ of probability distributions on $C([0, 1]; \mathbb{R}^n)$. Hence, for $E \subset \mathcal{D}$, we may consider the probability of

$$\{(\omega^1, \ldots, \omega^N) | \mu_N(\cdot) \in E\}$$

in the product measure W_ρ^N on Ω^N. By the ergodic theorem (see, e.g., [90, Theorem A.9.3]), as N tends to infinity, the sequence of distributions μ_N converges weakly to W_ρ. Hence, if $W_\rho \notin E$, it follows that

$$W_\rho^N(\{(\omega^1, \ldots, \omega^N) | \mu_N(\cdot) \in E\}) \searrow 0.$$

In this, large deviations theory provides us with the much finer result that the decay is *exponential* and the exponent may be characterized as solving a *maximum entropy problem* [81].

Specifically, in our setting we let $E = \mathcal{D}(\rho_0, \rho_1)$, namely, the set of distributions on $C([0, 1]; \mathbb{R}^n)$ having marginal densities ρ_0 and ρ_1 at times $t = 0$ and $t = 1$, respectively. Then Sanov's theorem [211], [81, Theorem 6.2.10] asserts that, assuming the "prior" W_ρ does not have the required marginals, the probability of observing an empirical distribution μ_N in a neighborhood of $\mathcal{D}(\rho_0, \rho_1)$ in the weak topology decays as

$$e^{-N \inf\{\mathbb{D}(P\|W_\rho) | P \in \mathcal{D}(\rho_0, \rho_1)\}}$$

as $N \to \infty$. This large deviations statement can be turned around in the spirit of the Gibbs conditioning principle (cf. [81, section 7.3]) to deduce that, given as data the marginal distributions ρ_0, ρ_1, the most likely distribution is the closest to W_ρ in the sense of relative entropy. But, in our setting, ρ is unspecified, and in light of the fact that

$$\mathbb{D}(P\|W_\rho) = \mathbb{D}(P\|W) - \mathbb{D}(\rho_0\|\rho),$$

the most likely random evolution between two given marginals is in fact the solution of the Schrödinger bridge problem that follows.

PROBLEM 4.1.

$$P_{\rm SBP} := \operatorname{argmin}\{\mathbb{D}(P\|W) \mid P \in \mathcal{D}(\rho_0, \rho_1)\}. \tag{4.3}$$

The optimal solution is referred to as the *Schrödinger bridge* between ρ_0 and ρ_1 over W, since its marginal flow $\{\rho(t,\cdot); 0 \le t \le 1\}$, which is the *entropic interpolation* between ρ_0 and ρ_1, is seen as a "bridge" between the two marginals.

Next we discuss the structure of the solution of the above and its reduction to a static problem.

Let $P \in \mathcal{D}$ be a finite-energy diffusion [102]; that is, under P, the canonical coordinate process $X_t(\omega) = \omega(t)$ has a (forward) Itô differential

$$dX_t = \beta_t dt + dW_t, \tag{4.4}$$

where β_t is adapted to $\{\mathcal{F}_t^-\}$ ($\mathcal{F}_t^-$ is the σ-algebra of events up to time t) and

$$\mathbb{E}_P\left[\int_0^1 \|\beta_t\|^2 dt\right] < \infty. \tag{4.5}$$

Conditioning the process on starting at $X_0 = x$ and ending at $X_1 = y$ gives

$$P_{xy} = P\left[\,\cdot \mid X_0 = x, X_1 = y\right], \quad W_{xy} = W\left[\,\cdot \mid X_0 = x, X_1 = y\right].$$

These laws are referred to as the disintegrations of P and W with respect to the initial and final positions [48]. Let ρ_{01}^P and ρ_{01}^W be the joint initial-final time distributions under P and W, respectively. Then we have the following decomposition of relative entropy [103]:

$$\begin{aligned}\mathbb{D}(P\|W) = \mathbb{E}_P\left[\log \frac{dP}{dW}\right] &= \int\int \left[\log \frac{\rho_{01}^P(x,y)}{\rho_{01}^W(x,y)}\right] \rho_{01}^P(x,y)dxdy \\ &+ \int\int\int \left(\log \frac{dP_x^y}{dW_x^y}\right) dP_x^y \rho_{01}^P(x,y)dxdy.\end{aligned} \tag{4.6}$$

Clearly, since $\rho_{01}^P(x,y)$ and dP_x^y can be chosen independently, the choice $P_x^y = W_x^y$, which actually makes the second summand equal to zero, is optimal. Thus, Problem 4.1 reduces to the following "static" problem.

PROBLEM 4.2. *Minimize*

$$\mathbb{D}(\rho_{01}\|\rho_{01}^W) = \int\int \left[\log \frac{\rho_{01}(x,y)}{\rho_{01}^W(x,y)}\right] \rho_{01}(x,y)dxdy \tag{4.7}$$

over the set of densities

$$\Pi(\rho_0, \rho_1) := \left\{\rho_{01} \;\; on\; \mathbb{R}^n \times \mathbb{R}^n \;\middle|\; \int \rho_{01}(x,y)dy = \rho_0(x), \quad \int \rho_{01}(x,y)dx = \rho_1(y)\right\}.$$

One should note that the conditions defining Π are linear; the elements of $\Pi(\rho_0, \rho_1)$ are referred to as *couplings* between ρ_0 and ρ_1. If ρ_{01}^* is the solution to Problem 4.2, i.e., the optimal coupling, then, evidently,

$$P^*(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} W_{xy}(\cdot)\rho_{01}^*(dxdy) \tag{4.8}$$

solves[12] Problem 4.1. The structure of the problem further endows P^* with the same three-point transition density

$$p^*(s,x,t,y,u,z) = \frac{p(s,x,t,y)p(t,y,u,z)}{p(s,x,u,z)}, \quad s<t<u,$$

as the prior—a property that is referred to by saying that it belongs to the same *reciprocal class* as the prior measure [24, 129].

It is natural to consider the case when the prior is a Wiener measure with variance ϵ, denoted by W_ϵ and with transition kernel

$$p_\epsilon(0,x,1,y) = [2\pi\epsilon]^{-\frac{n}{2}} \exp\left[-\frac{\|x-y\|^2}{2\epsilon}\right],$$

and to contemplate the limiting process when $\epsilon \to 0$. Indeed, using $\rho_{01}^{W_\epsilon}(x,y) = \rho_0^{W_\epsilon}(x)p_\epsilon(0,x;1,y)$ and the fact that

$$\int\int \left[\log \rho_0^{W_\epsilon}(x)\right] \rho_{01}(x,y)dxdy = \int \left[\log \rho_0^{W_\epsilon}(x)\right] \rho_0(x)dx$$

is independent of ρ_{01}, we obtain

(4.9)
$$\begin{aligned}\mathbb{D}(\rho_{01}\|\rho_{01}^{W_\epsilon}) &= -\int\int \left[\log \rho_{01}^{W_\epsilon}(x,y)\right] \rho_{01}(x,y)dxdy + \int\int \left[\log \rho_{01}(x,y)\right] \rho_{01}(x,y)dxdy \\ &= \int\int \frac{\|x-y\|^2}{2\epsilon}\rho_{01}(x,y)dxdy - S(\rho_{01}) + \text{constant},\end{aligned}$$

where S is the differential entropy. Thus, Problem 4.2 that considers minimizing $\mathbb{D}(\rho_{01}\|\rho_{01}^{W_\epsilon})$ over the couplings $\Pi(\rho_0,\rho_1)$ is equivalent to

$$\min_{\rho_{01}\in\Pi(\rho_0,\rho_1)} \int \frac{\|x-y\|^2}{2}\rho_{01}(x,y)\mathrm{d}x\mathrm{d}y + \epsilon\int \rho_{01}(x,y)\log\rho_{01}(x,y)\mathrm{d}x\mathrm{d}y. \tag{4.10}$$

It is seen that in the limit as $\epsilon \to 0$, the cost reduces to $\frac{1}{2}\|x-y\|^2$, which is in the form of the Kantorovich functional in (2.1). Thus, *the Schrödinger bridge problem represents a regularization of OMT obtained by subtracting from the cost functional a term proportional to the entropy.*

We have already seen that the Schrödinger bridge problem can be motivated in three different ways: First via the original statistical mechanical thought experiment of Schrödinger (a large deviations problem), and second via Sanov's theorem and Gibbs conditioning principle, as a maximum entropy problem. It is an early and important instance of an inference method that prescribes how to choose a posterior distribution while making the fewest number of assumptions beyond the available information. This approach has been considerably developed over many years by Jaynes, Burg, Dempster, and Csiszár [131, 132, 133, 39, 40, 84, 73, 74, 75]. Both forms of the problem have their roots in Boltzmann's work [34]. The more recent third motivation comes from regularized OMT [168, 169, 170, 156, 157, 46], which mitigates the computational challenges [7, 22, 198].

[12]Note that decomposition (4.6) and the resulting argument remain valid even when the prior measure is induced by any, possibly non-Markovian, finite-energy diffusion $\bar{P}$; see (4.27).

4.3. Derivation of the Schrödinger System. The Lagrangian function of Problem 4.2 has the form

$$\mathcal{L}(\rho_{01},\lambda,\mu)=\int\int\left[\log\frac{\rho_{01}(x,y)}{\rho_{01}^W(x,y)}\right]\rho_{01}(x,y)dxdy$$
$$+\int\lambda(x)\left[\int\rho_{01}(x,y)dy-\rho_0(x)\right]+\int\mu(y)\left[\int\rho_{01}(x,y)-\rho_1(y)\right].$$

Setting the first variation equal to zero, we obtain the (sufficient) optimality condition

$$1+\log\rho_{01}^*(x,y)-\log p(0,x,1,y)-\log\rho_0^W(x)+\lambda(x)+\mu(y)=0,$$

where we have used the expression $\rho_{01}^W(x,y)=\rho_0^W(x)p(0,x,1,y)$ with p as in (4.1). We get

$$\frac{\rho_{01}^*(x,y)}{p(0,x,1,y)}=\exp\left[\log\rho_0^W(x)-1-\lambda(x)-\mu(y)\right]$$
$$=\exp\left[\log\rho_0^W(x)-1-\lambda(x)\right]\exp\left[-\mu(y)\right].$$

Thus, the ratio $\rho_{01}^*(x,y)/p(0,x,1,y)$ factors into a function of x times a function of y. Let us denote them by $\hat\varphi(x)$ and $\varphi(y)$, respectively. The optimal $\rho_{01}^*(\cdot,\cdot)$ then has the form

$$\rho_{01}^*(x,y)=\hat\varphi(x)p(0,x,1,y)\varphi(y), \tag{4.11}$$

with φ and $\hat\varphi$ satisfying

$$\hat\varphi(x)\int p(0,x,1,y)\varphi(y)dy=\rho_0(x), \tag{4.12}$$

$$\varphi(y)\int p(0,x,1,y)\hat\varphi(x)dx=\rho_1(y). \tag{4.13}$$

Let $\hat\varphi(0,x)=\hat\varphi(x)$, $\varphi(1,y)=\varphi(y)$, and

$$\hat\varphi(1,y):=\int p(0,x,1,y)\hat\varphi(0,x)dx,\quad \varphi(0,x):=\int p(0,x,1,y)\varphi(1,y).$$

Then (4.12)–(4.13) is equivalent to the system

$$\hat\varphi(1,y)=\int p(0,x,1,y)\hat\varphi(0,x)dx, \tag{4.14a}$$

$$\varphi(0,x)=\int p(0,x,1,y)\varphi(1,y)dy \tag{4.14b}$$

with the boundary conditions

$$\varphi(0,x)\cdot\hat\varphi(0,x)=\rho_0(x),\quad \varphi(1,y)\cdot\hat\varphi(1,y)=\rho_1(y). \tag{4.14c}$$

The question of existence and uniqueness of positive functions $\hat\varphi$, φ satisfying (4.14a)–(4.14c), left open by Schrödinger, is highly nontrivial and has been settled in various degrees of generality by Fortet, Beurling, Jamison, and Föllmer [106, 27, 130, 103]; see also [157, 68]. Note that both Fortet and Beurling predate Sinkhorn's work [223, 224] on a problem in statistics that turned out to be closely related.

There are basically two ways to deal with the existence of solutions to the Schrödinger system of equations (4.14a)–(4.14c). One is to prove existence for the dual problem of the original convex optimization problem. This was first accomplished by Beurling [27], leading to [130]; see also the recent paper [68]. Alternatively, one can try to prove convergence of a suitable successive approximation scheme. This was first accomplished by Fortet [106]; see also [58]. We outline Fortet's results in section 8.1 and remark that, in the special case where the marginals ρ_0, ρ_1 are Gaussian, the Schrödinger system has a closed-form solution. This was only recently discovered in [50]. The discrete counterpart of Fortet's algorithm is the so-called IPF-Sinkhorn algorithm discussed in section 8.2. Note that in the recent paper [91] (see also [159]), the bulk of Fortet's paper has been rewritten, filling in all the missing steps and explaining the rationale behind his complex approximation scheme.

The pair $(\varphi, \hat{\varphi})$ satisfying (4.14a)–(4.14c) is unique up to multiplication of φ by a positive constant c and division of $\hat{\varphi}$ by the same constant. At each time t, the marginal $\rho(t, \cdot)$ factors as

$$\rho(t,x) = \varphi(t,x) \cdot \hat{\varphi}(t,x). \tag{4.15}$$

The factorization (4.15) resembles Born's relation (in quantum theory)

$$\rho(t,x) = \psi(t,x) \cdot \bar{\psi}(t,x),$$

with ψ and $\bar{\psi}$ satisfying two adjoint equations like φ and $\hat{\varphi}$. Moreover, the solution of Problem 4.1 exhibits the following remarkable *reversibility property*: Swapping the two marginal densities ρ_0 and ρ_1, the new solution is simply the time reversal of the previous one; cf. the title *On the reversal of natural laws* of [215]. These are the remarkable analogies to quantum mechanics which appeared to Schrödinger to be very worthy of reflection. But wait a minute: When Schrödinger posed his question, the very foundations of probability theory were still missing, and the notion of stochastic process had not yet been introduced. Although Wiener had given a rigorous construction of Wiener measure in 1923 [248], hardly any theory of continuous parameter stochastic processes had been developed in the early 1930s. Many relevant results such as Sanov's theorem and multiplicative functional transformations of Markov processes [128] were not available. How could Schrödinger formulate and, to a large extent, solve such an abstract problem in the early 1930s? Schrödinger, in his countryman Boltzmann's style, discretized space to be able to compute by the De Moivre–Stirling formula the most likely joint initial-final distribution, very much like Boltzmann had done in 1877 [34].

In [196], the Schrödinger bridge problem was considered in the case when only samples of the boundary marginals are available. A numerical method was developed which has potential to work in high-dimensional settings employing constrained maximum likelihood in place of the nonlinear boundary coupling and importance sampling to propagate φ and $\hat{\varphi}$. Another paper dealing with a similar problem is [25].

4.4. Stochastic Control Formulation. In 1975 Jamison [130] showed that the solution of the Schrödinger bridge problem is an h-path process in the sense of Doob [87], [88, p. 566]. Indeed, dividing both sides of (4.11) by $\rho_0(x)$ (assumed everywhere positive), we get

$$p^*(0,x,1,y) = \frac{1}{\varphi(x)} p(0,x,1,y)\varphi(y), \tag{4.16}$$

where φ, in Doob's language, is *space time harmonic* satisfying

$$\frac{\partial \varphi}{\partial t} + \frac{1}{2}\Delta\varphi = 0. \tag{4.17}$$

The solution is obtained from the prior distribution via a multiplicative functional transformation of Markov processes [128]. It is worthwhile to mention that Francesco Guerra has connected such a function to the *importance function* of neutron transport theory [117].

The connection between Schrödinger bridges and stochastic control, clearly established in [77], was prepared for by work in the latter field on the so-called logarithmic transformation of parabolic differential equations [99, 100, 101, 122, 142, 172, 30] as well as by work in mathematical physics [117, 253]. The Schrödinger bridge problem can be turned, thanks to *Girsanov's theorem*, into a stochastic calculus of variations problem [175, 246, 79, 1] that, in turn, can be reformulated in the language of stochastic control [77, 78, 192, 96]. Let $P \in \mathcal{D}$ be a finite-energy diffusion with forward differential (4.4). Then, by Girsanov's theorem [143],

$$\begin{aligned} \log\frac{dP}{dW} &= \log\frac{\rho_0^P(X_0)}{\rho_0^W(X_0)} + \int_0^1 \beta_t dX_t - \int_0^1 \frac{1}{2}\|\beta_t\|^2 dt, \quad P \text{ a.s.}, \\ &= \log\frac{\rho_0^P(X_0)}{\rho_0^W(X_0)} + \int_0^1 \beta_t dW_t + \int_0^1 \frac{1}{2}\|\beta_t\|^2 dt, \quad P \text{ a.s..} \end{aligned} \tag{4.18}$$

By the finite-energy condition (4.5),

$$Y_t := \int_0^t \beta_\tau dW_\tau$$

is a martingale and has therefore constant expectation. Since $Y_0 = 0$, we have

$$\mathbb{E}_P\left[\int_0^1 \beta_t dW_t\right] = 0.$$

Then (4.18) yields

$$\mathbb{D}(P\|W) = \mathbb{E}_P\left[\log\frac{dP}{dW}\right] = \mathbb{D}(\rho_0\|\rho_0^W) + \mathbb{E}_P\left[\int_0^1 \frac{1}{2}\|\beta_t\|^2 dt\right]. \tag{4.19}$$

Note that $\mathbb{D}(\rho_0\|\rho_0^W)$ is constant over $\mathcal{D}(\rho_0, \rho_1)$. We then obtain a stochastic control formulation. Problem 4.1 (when the prior has variance ϵ) is equivalent to the following problem.

Problem 4.3.

$$\begin{aligned} &\text{Minimize}_{u\in\mathcal{U}}\ J(u) = \mathbb{E}\left[\int_0^1 \frac{1}{2\epsilon}\|u_t\|^2 dt\right] \\ &\text{subject to } dX_t = u_t dt + \sqrt{\epsilon} dW_t, \quad X_0 \sim \rho_0(x), \quad X_1 \sim \rho_1(y), \end{aligned} \tag{4.20}$$

where the family $\mathcal{U}$ consists of adapted, finite-energy control functions.

The optimal control is of the feedback type

$$u^*(t, x) = \epsilon\nabla\log\varphi(t, x), \tag{4.21}$$

where $(\varphi, \hat{\varphi})$ solve the Schrödinger system

$$\frac{\partial \varphi}{\partial t} + \frac{\epsilon}{2}\Delta\varphi = 0, \tag{4.22a}$$

$$\frac{\partial \hat{\varphi}}{\partial t} - \frac{\epsilon}{2}\Delta\hat{\varphi} = 0, \tag{4.22b}$$

$$\varphi(0,x)\cdot\hat{\varphi}(0,x) = \rho_0(x), \tag{4.22c}$$

$$\varphi(1,y)\cdot\hat{\varphi}(1,y) = \rho_1(y). \tag{4.22d}$$

4.5. Fluid-Dynamic Formulation. Problem 4.3 leads immediately to the following fluid-dynamic problem.

PROBLEM 4.4.

$$\inf_{(\rho,u)} \int_{\mathbb{R}^n}\int_0^1 \frac{1}{2}\|u(t,x)\|^2\rho(t,x)dtdx, \tag{4.23a}$$

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(u\rho) - \frac{\epsilon}{2}\Delta\rho = 0, \tag{4.23b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y), \tag{4.23c}$$

where $u(\cdot,\cdot)$ *varies over continuous functions on* $[0,1]\times\mathbb{R}^n$*.*

Remark 4.5. Contrary to what is often stated in the literature (see, e.g., [153]),[13] Problem 4.4 is not equivalent to Problems 4.1, 4.2, and 4.3 in that it only reproduces the optimal entropic interpolating flow $\{\rho^*(t,\cdot); 0\le t\le 1\}$. Information about correlations at different times and smoothness of the trajectories in the support of the measure P^* is lost here. Indeed, let (ρ^*, u^*) be optimal for Problem 4.4 and define the *current velocity field* [176]

$$v^*(t,x) := u^*(t,x) - \frac{\epsilon}{2}\nabla\log\rho^*(t,x). \tag{4.24}$$

Assume that v^* guarantees existence and uniqueness of the following initial value problem on $[0,1]$ for any deterministic initial condition:

$$\dot{X}_t = v^*(t,X_t), \quad X_0 \sim \rho_0 dx.$$

Then the probability density $\rho(t,x)$ of X_t satisfies the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(v^*\rho) = 0$$

as well as (4.23b) with the same initial condition and therefore coincides with $\rho^*(t,x)$.

It appears that as $\epsilon \searrow 0$, the solution to this problem converges to the solution of the Benamou–Brenier OMT problem [22]. This is indeed the case [168, 169, 170, 157, 156]. The analysis in Remark 4.5, however, suggests that an alternative fluid-dynamic problem characterization of the entropic interpolation flow $\{\rho^*(t,\cdot); 0\le t\le 1\}$ may be possible. Indeed, such an alternative time-symmetric problem was derived in [54]; see also [111].

[13] We contributed to this confusion ourselves in [54, section 5].

PROBLEM 4.6.

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2}\|v(t,x)\|^2 + \frac{\epsilon^2}{8}\|\nabla \log \rho\|^2\right] \rho(t,x)dtdx, \tag{4.25a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{4.25b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y). \tag{4.25c}$$

The two criteria differ by a (scaled) Fisher information functional

$$\mathcal{I}(\rho) = \int \|\nabla \log \rho(t,x)\|^2 \rho(t,x)dx,$$

while the Fokker–Planck equation has been replaced by the continuity equation. Both Problem 4.4 and Problem 4.6 can be thought of as regularizations of the Benamou–Brenier problem [22] and as dynamic counterparts of (4.10) [168, 169, 170, 157, 156, 52, 54].

Remark 4.7. The problem of minimizing the Yasue [250] action (4.25a) under the continuity equation (4.25b) and boundary conditions (4.25c) was formulated by Carlen in [45, p. 131]. He was investigating possible connections between OMT and Nelson's stochastic mechanics. He wrote that "…the Euler-Lagrange equations for it are not easy to understand." The solution to Carlen's problem was provided in [54, section 5] through the fluid-dynamic problem associated to the Schrödinger bridge. Carlen's statement has caused several authors (see, e.g., [152]) to believe that Yasue in [250] had already discussed this problem. This is not true, since what Yasue developed there, using Nelson's integration-by-parts formula for semimartingales [176, Theorem 11.12], was a stochastic calculus of variations that may be viewed as the "particle" form of the Hamilton principle in stochastic mechanics. The fluid dynamic counterpart of this principle was developed in [116] (see also [177]), but with a different action (there is a minus in front of the Fisher information functional)! For a clarification of the connection between Schrödinger bridges, Nelson's stochastic mechanics, and Bohm's stochastic mechanics [31, 33, 32] through calculus of variations in the Wasserstein space, see [71, section VI].

4.6. General Prior. All we have seen in this section can be generalized to a general Markovian prior measure. Indeed, suppose that $\bar{P}$ is a Markov finite energy diffusion [102]. Under $\bar{P}$, the canonical coordinate process has a forward Itô differential

$$dX_t = f(t,X_t)dt + \sqrt{\epsilon}dW_t.$$

Let $\bar{p}(s,x,t,y)$ be the corresponding transition density. Consider the Schrödinger bridge problem with prior $\bar{P}$ as follows.

PROBLEM 4.8.

$$\text{Minimize} \quad \mathbb{D}(P\|\bar{P}) \quad \text{over} \quad P \in \mathcal{D}(\rho_0,\rho_1). \tag{4.26}$$

A decomposition like (4.6) then turns the problem into Problem 4.2. In particular, (4.8) is replaced by

$$P^*(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \bar{P}_{xy}(\cdot)\rho_{01}^*(dxdy). \tag{4.27}$$

The various dynamic formulations are modified accordingly. Problem 4.3 becomes the following problem.

PROBLEM 4.9.

$$\min_{u\in\mathcal{U}} J(u)=\mathbb{E}\left[\int_0^1 \frac{1}{2\epsilon}\|u_t\|^2dt\right], \tag{4.28}$$
$$dX_t=[f(t,X_t)+u_t]\,dt+\sqrt{\epsilon}dW_t,\quad X_0\sim\rho_0(x)dx,\quad X_1\sim\rho_1(y)dy,$$

where the family $\mathcal{U}$ consists of adapted, finite-energy control functions.

Then (4.21) is replaced by

$$u_t^*=\epsilon\nabla\log\varphi(t,X_t), \tag{4.29}$$

where $(\varphi,\hat\varphi)$ now solve the Schrödinger system

$$\frac{\partial\varphi}{\partial t}+f\cdot\nabla\varphi+\frac{\epsilon}{2}\Delta\varphi=0, \tag{4.30a}$$
$$\frac{\partial\hat\varphi}{\partial t}+\nabla\cdot(f\hat\varphi)-\frac{\epsilon}{2}\Delta\hat\varphi=0, \tag{4.30b}$$

with the same boundary conditions. The fluid dynamic formulations also change accordingly. Let

$$\bar v(t,x):=f(t,x)-\frac{\epsilon}{2}\nabla\log\bar\rho(t,x)$$

be the current velocity of the prior process. Then Problem 4.6 becomes

$$\inf_{(\rho,v)}\int_{\mathbb{R}^n}\int_0^1\left[\frac{1}{2}\|v(t,x)-\bar v(t,x)\|^2+\frac{\epsilon^2}{8}\left\|\nabla\log\frac{\rho(t,x)}{\bar\rho(t,(x)}\right\|^2\right]\rho(t,x)dxdt, \tag{4.31a}$$
$$\frac{\partial\rho}{\partial t}+\nabla\cdot(v\rho)=0, \tag{4.31b}$$
$$\rho_0=\nu_0,\quad \rho_1=\nu_1, \tag{4.31c}$$

where the two criteria differ by a *relative Fisher information* term.

The theory can be further extended to the case of a general diffusion coefficient (anisotropic diffusion) and to the situation where the prior Markovian measure features creation and/or killing, so that the probability mass is not preserved at each time [79, 1, 247, 51]. Both cases are treated in section 5.

Remark 4.10. We like to stress here one aspect of (4.8), namely, that this is just a *representation* of the solution P^* even if we have been able to solve Problem 4.2. This representation needs to be supplemented with the multiplicative functional relation between transition densities,

$$p^*(s,x,t,y)=\frac{1}{\varphi(s,x)}\bar p(s,x,t,y)\varphi(t,y),\quad 0\le s<t\le 1, \tag{4.32}$$

and the associated relationship between drifts (4.29). We shall come back to this point when discussing the discrete state space case.

5. Stochastic Control and General Bridge Problems.

The classical probabilistic literature on the Schrödinger bridge problem only considers the case in which the noise and control matrices are constant, diagonal, and nonsingular (often the identity matrix); see Problem 4.9. This permits us, in particular, to employ the Girsanov transformation (4.18). In many applications, however, such a requirement represents a serious limitation. Noise may not affect all components of the state vector like in the controlled oscillator (6.5) and/or the control channel may be dictated by technological constraints such as in some engineering applications. Central in such applications are problems for controlled Gauss–Markov models.[14] Find an adapted control u minimizing

$$J(u) = \mathbb{E}\left\{\int_0^1 \|u(t)\|^2\,dt\right\}$$

among those which achieve the transfer

$$\begin{aligned} dx(t) &= A(t)x(t)dt + B(t)u(t)dt + B_1(t)dW_t, \\ x(0) &\sim \mathcal{N}(m_0, \Sigma_0), \quad x(1) \sim \mathcal{N}(m_1, \Sigma_1). \end{aligned}$$

It is therefore important to pose the minimum energy steering problem for probability distributions in a more general setting where the original large deviations/maximum entropy motivation might be lost. This will make salient the advantage of the control formulation, which always makes perfect sense and in which the free (uncontrolled) evolution plays the role of the prior model. We proceed in this section to see how much of the fluid-dynamic formulation can be extended to the general case [51].

We consider a cloud of particles with density $\rho(t,x)$, $x \in \mathbb{R}^n$, which evolves according to the transport-diffusion equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (f(t,x)\rho) + V(t,x)\rho = \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 (a_{ij}(t,x)\rho)}{\partial x_i \partial x_j}, \tag{5.1}$$

with $\rho(0,\cdot) = \rho_0(\cdot)$ a probability density. Unlike in previous literature on connections to Feynman–Kac [247, 192, 78], and following our desire to be able to model inertial particles as in the previous section, we assume that the matrix $a(t,x) = [a_{ij}(t,x)]_{i,j=1}^m$ is only positive semidefinite of constant rank on $[0,1] \times \mathbb{R}^n$ with

$$a_{ij}(t,x) = \sum_k \sigma_{ik}(t,x)\sigma_{kj}(t,x)$$

for a matrix $\sigma(t,x) = [\sigma_{ik}(t,x)] \in \mathbb{R}^{n\times m}$ of constant rank $m \le n$. Notice that the presence of $V(t,x) \ge 0$ allows for the possibility of loss of mass, so that the integral of $\rho(t,x)$ over $\mathbb{R}^n$ is not necessarily constant. This flexibility permits us to model particles satisfying

$$dX_t = f(t,X_t)dt + \sigma(t,X_t)dW_t, \tag{5.2}$$

which are absorbed at some rate by the medium in which they travel or, if the sign of V is negative, are created out of this same medium [144, p. 272]. We assume that f and σ are smooth and that the operator

$$L = \sum_{i,j=1}^{n} a_{ij}(t,x)\partial_{x_i}\partial_{x_j} + \sum_{j=1}^{n} f_j(t,x)\partial_{x_j} - \partial_t$$

[14] In a twin paper [61], we review the by now vast literature [125, 126, 222, 115, 254, 50, 52, 55, 56, 57, 60, 118, 13, 14, 15, 16, 17, 114, 205, 206, 181, 182, 208, 183, 3, 67] on optimal steering of probability distributions for Gauss–Markov models in continuous and discrete time, over a finite or infinite time horizon, with or without state and/or control constraints and applications.

is *hypoelliptic*, satisfying Hörmander's condition [123, 184]. Hypoelliptic diffusions model important processes in many branches of science: Ornstein–Uhlenbeck stochastic oscillators, Nyquist–Johnson circuits with noisy resistors, image reconstruction based on Petitot's model of neurogeometry of vision [37], etc. The "reweighing" of the original measure of the Markov process (5.2) when V is unbounded is a delicate issue and can be accomplished via the Nagasawa transformation; see [247, section 8B] for details.

Let us now suppose that (5.1) represents a *prior* evolution and that at time $t = 1$ we measure an empirical probability density $\rho_1(\cdot) \neq \rho(1, \cdot)$ as dictated by (5.1). Thus, the model (5.1) is thought not to be consistent with the estimated end-point empirical distribution. However, suppose that one has reasons to believe[15] that the actual evolution is close to the nominal one and that only the actual drift field is different, perturbed by an additive term $\sigma(t, x)u(t, x)$, i.e.,

$$v(t, x) = f(t, x) + \sigma(t, x)u(t, x).$$

Notice that the control variables, of which there may be fewer than n, act through the same channels of the diffusive part. The assumption that stochastic excitation and control enter through the same "channels" is natural in certain applications, as is explained and treated in [50] for linear diffusions. The case where these channels may differ is considered in [52].

Taking (5.1) as a reference evolution and given the terminal probability density ρ_1, we are led to consider the following generalization of Problem 4.4:

$$\inf_{(\rho,u)} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2}\|u\|^2 + V(t, x)\right] \rho(t, x) dt dx, \tag{5.3a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot ((f + \sigma u)\rho) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 (a_{ij}\rho)}{\partial x_i \partial x_j}, \tag{5.3b}$$

$$\rho(0, x) = \rho_0(x), \quad \rho(1, x) = \rho_1(x). \tag{5.3c}$$

When $[a_{ij}]$ does depend on x, the connection to a relative entropy problem on path space is apparently available only under rather restrictive assumptions such as uniform boundedness of a [97, section 5]. Problem (5.3) thus appears as a generalization of Problem 4.4 that is not necessarily connected to a large deviations problem. In [43], a special case of Problem 5.3 was considered ($V \equiv 0$, $\sigma(t, x) = B(t)$). For certain classes of drifts f, a Wasserstein proximal algorithm was used to numerically solve a pair of initial value problems equivalent to (5.10) below. This result has been generalized to the situation in which there are hard state constraints (the diffusion paths have to remain in a bounded domain) in [44]. We mention here that chance state constraints were considered for the covariance control problem in [181]. Moreover, input constraints for discrete and continuous time models have been considered in [14, 183]. In [207, 251], a nonlinear covariance control problem was studied by iteratively solving an approximate linearized problem and by differential dynamic programming, respectively.

In what follows we provide in some detail the variational analysis for (5.3) which can be viewed as a generalization of that of section 3.3. Let $\mathcal{P}_{\rho_0\rho_1}$ be the family of flows of probability densities satisfying (5.3c). Let $\mathcal{U}$ be a family of continuous feedback

[15] Alternative reasoning based on the Gibbs conditioning principle, as before, may be possible.

control laws $u(\cdot,\cdot)$. Consider the unconstrained minimization of the Lagrangian over $\mathcal{P}_{\rho_0\rho_T}\times\mathcal{U}$,

$$\mathcal{L}(\rho,u,\lambda)=\int_{\mathbb{R}^n}\int_0^1\left[\left(\frac{1}{2}\|u(t,x)\|^2+V(t,x)\right)\rho+\lambda(t,x)\left(\frac{\partial\rho}{\partial t}+\nabla\cdot((f+\sigma u)\rho)\right.\right.$$
$$\left.\left.-\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial x_i\partial x_j}\left(a_{ij}(t,x)\rho\right)\right)\right]dtdx,$$

where λ is a C^1 Lagrange multiplier. After integration by parts, assuming that limits for $x\to\infty$ are zero, and observing that the boundary values are constant over $\mathcal{P}_{\rho_0\rho_1}$, we get the problem

$$\inf_{(\rho,u)\in\mathcal{X}_{\rho_0\rho_1}\times\mathcal{U}}\int_{\mathbb{R}^n}\int_0^1\left[\frac{1}{2}\|u\|^2+V-\left(\frac{\partial\lambda}{\partial t}+(f+\sigma u)\cdot\nabla\lambda+\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}\frac{\partial^2\lambda}{\partial x_i\partial x_j}\right)\right]\rho dtdx. \tag{5.4}$$

Pointwise minimization of the integrand with respect to u for each fixed flow of probability densities ρ gives

$$u^*_\rho(t,x)=\sigma'\nabla\lambda(t,x). \tag{5.5}$$

Plugging this form of the optimal control into (5.4), we get the functional of $\rho\in\mathcal{P}_{\rho_0\rho_1}$

$$J(\rho,\lambda)=\int_{\mathbb{R}^n}\int_0^1\left[\frac{\partial\lambda}{\partial t}+f\cdot\nabla\lambda+\frac{1}{2}\nabla\lambda\cdot a\nabla\lambda-V+\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}\frac{\partial^2\lambda}{\partial x_i\partial x_j}\right]\rho dtdx. \tag{5.6}$$

We then have the following result.

PROPOSITION 5.1. *If ρ^* satisfies*

$$\frac{\partial\rho}{\partial t}+\nabla\cdot((f+a\nabla\lambda)\rho)=\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2\left(a_{ij}\rho\right)}{\partial x_i\partial x_j}, \tag{5.7}$$

with λ a solution of the HJB-like equation

$$\frac{\partial\lambda}{\partial t}+f\cdot\nabla\lambda+\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}(t,x)\frac{\partial^2\lambda}{\partial x_i\partial x_j}+\frac{1}{2}\nabla\lambda\cdot a\nabla\lambda-V=0 \tag{5.8}$$

and $\rho^*(1,\cdot)=\rho_1(\cdot)$, then the pair (ρ^*,u^*) with $u^*=\sigma'\nabla\lambda$ is a solution of (5.3).

Of course, the difficulty lies with the nonlinear equation (5.8), for which no boundary value is available. Together, $\rho(t,x)$ and $\lambda(t,x)$ satisfy the coupled equations (5.7)–(5.8) and the split boundary conditions for $\rho(t,x)$ in (5.3c). Let us, however, define

$$\varphi(t,x)=\exp\left[\lambda(t,x)\right],\quad (t,x)\in[0,1]\times\mathbb{R}^n.$$

If λ satisfies (5.8), we find that φ satisfies the *linear* equation

$$\frac{\partial\varphi}{\partial t}+f\cdot\nabla\varphi+\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}(t,x)\frac{\partial^2\varphi}{\partial x_i\partial x_j}=V\varphi. \tag{5.9}$$

Moreover, for ρ satisfying (5.7) and φ satisfying (5.9), let us define

$$\hat{\varphi}(t,x) = \frac{\rho(t,x)}{\varphi(t,x)}, \quad (t,x) \in [0,1] \times \mathbb{R}^n.$$

Then a long but straightforward calculation shows that $\hat{\varphi}$ satisfies the original equation (5.1). Thus, we have the system of linear PDEs

$$\frac{\partial \varphi}{\partial t} + f \cdot \nabla \varphi + \frac{1}{2} \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} = V \varphi, \tag{5.10a}$$

$$\frac{\partial \hat{\varphi}}{\partial t} + \nabla \cdot (f \hat{\varphi}) - \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 (a_{ij} \hat{\varphi})}{\partial x_i \partial x_j} = -V \hat{\varphi}, \tag{5.10b}$$

nonlinearly coupled through their boundary values as

$$\varphi(0,\cdot)\hat{\varphi}(0,\cdot) = \rho_0(\cdot), \quad \varphi(1,\cdot)\hat{\varphi}(1,\cdot) = \rho_1(\cdot). \tag{5.10c}$$

Equations (5.10a)–(5.10c) constitute a *generalized Schrödinger system.* We have therefore established the following result.

THEOREM 5.2. *Let* $(\varphi(t,x), \hat{\varphi}(t,x))$ *be nonnegative functions satisfying* (5.10a)–(5.10c) *for* $(t,x) \in ([0,1] \times \mathbb{R}^n)$. *Suppose* φ *is everywhere positive. Then the pair* (ρ^*, u^*) *with*

$$u^*(t,x) = \sigma' \nabla \log \varphi(t,x), \tag{5.11a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot ((f + a \nabla \log \varphi)\rho) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 (a_{ij} \rho)}{\partial x_i \partial x_j} \tag{5.11b}$$

is a solution of (5.3).

Establishing existence and uniqueness (up to multiplication/division of the two functions by a positive constant) of the solution of the Schrödinger system is extremely challenging even when the diffusion coefficient matrix a is constant and nonsingular. Particular care is required in the case when V is unbounded or singular [1]. Nevertheless, if the fundamental solution p of (5.1) is everywhere positive on $([0,1] \times \mathbb{R}^n)$, which is to be expected in the hypoelliptic case [123, 146], existence and uniqueness follow from a deep result of Beurling [27] suitably extended by Jamison [129, Theorem 3.2], [247, section 10].

Remark 5.3. It is interesting to note that although (5.3) is not convex in (ρ, u), it can be turned into a convex problem in a new set of coordinates (ρ, m) where $m = \rho u$, in which case it becomes

$$\inf_{(\rho,m)} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2} \frac{\|m\|^2}{\rho(t,x)} + V(t,x)\rho(t,x) \right] dt dx, \tag{5.12a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (f\rho + \sigma m) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 (a_{ij} \rho)}{\partial x_i \partial x_j}, \tag{5.12b}$$

$$\rho(0,\cdot) = \rho_0(\cdot), \quad \rho(1,\cdot) = \rho_1(\cdot). \tag{5.12c}$$

This type of coordinate transformation has been effectively used in [22] in the context of OMT.

6. The Atomic Hypothesis, Microscopes, and Stochastic Oscillators.

Richard Feynman, in his first lecture of the famous Caltech series, stated, "If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis that all things are made of atoms—little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another."

The atomic hypothesis apparently originated with Democritus of Abdera (a colony of Miletus in what is now Greek Thrace) and his mentor Leucippus. Democritus, according to the Greek historian Diogenes Laërtius (Democritus, Vol. IX, 44), states:

$$\alpha\rho\chi\acute{\alpha}\varsigma\ \epsilon\iota\nu\alpha\iota\ \tau\omega\nu\ \acute{o}\lambda\omega\nu\ \alpha\tau\acute{o}\mu o\upsilon\varsigma\ \kappa\alpha\acute{\iota}\ \kappa\epsilon\nu\acute{o}\nu,\ \tau\alpha\ \delta'\acute{\alpha}\lambda\lambda\alpha\ \pi\acute{\alpha}\nu\tau\alpha\ \nu\epsilon\nu o\mu\acute{\iota}\sigma\theta\alpha\iota$$

which (Robert Drew Hicks (1925)) can be translated as, "The first principles of the universe are atoms and empty space; everything else is merely thought to exist."

Democritus had (correctly) imagined that the wearing down of a wheel and the drying of clothes were due to small particles of wood and water, respectively, flying out of them. Then he made the following philosophical argument (according to Aristotle's report): If matter were infinitely divisible, only points with no extension would remain. But putting together an arbitrary number of them, we would still get things without extension.[16] Democritus was largely ignored in ancient Athens. He is said to have been disliked so much by Plato that the latter wished all of his books burned. It should also be stressed that, unlike the subsequent Plato and Aristotle, the atomists Leucippus and Democritus, following the scientific rationalist philosophy associated with the Miletus school, wanted to investigate in the 5th century BCE the *causes* of natural phenomena rather than their *significance*!

The long history of the atomic hypothesis intersects twice with the history of the microscope. The *first intersection* simply occurred because the invention of this instrument at the beginning of the 17th century made it possible to observe the very irregular motion of particles immersed in a fluid. These observations of "animated" or "irritable" particles were made by, among others, van Leeuwenhoek, Buffon, and Spallanzani long before the British botanist Robert Brown.[17] Many other important contributions to the atomistic theory were made before Brown from, among others, Dalton and Avogadro. By 1877 the kinetic theory asserting that Brownian motion of particles is caused by bombardment by the molecules of the fluid was rather well established.

In 1877 [34], Boltzmann posed and solved the first large deviation and relative maximum entropy problem in history, in which his "loaded dice" were actually molecules! Nevertheless, the theory was still not open to experimental verification as the velocity of Brownian particles could not be measured accurately. At the beginning of the 20th century, there were still prominent scientists such as Ostwald and Mach who were not convinced of the existence of atoms and molecules due to their

[16] This kind of subtle argument has its roots in the Elean philosophical school of Parmenides and Zeno (he of the turtle-Achilles paradox). According to Diogenes Laërtius, Leucippus was a pupil of Zeno. Elea, nowadays called Velia, located approximatively 90 miles southeast of Naples, was a Greek colony flourishing in the 5th century BCE.

[17] In [176, Chapter 2], Edward Nelson writes about Robert Brown, "His contribution was to establish Brownian motion as an important phenomenon, to demonstrate clearly its presence in inorganic as well as organic matter, and to refute by experiment facile mechanical explanations of the phenomenon."

"positivistic philosophical attitude" (Albert Einstein [213, p. 49]). In 1905 Einstein (and independently in 1906 Smoluchowski) proposed a PDE model open to the experimental check of measuring the diffusion coefficient, thereby circumventing the need to measure the velocity of Brownian particles. This was accomplished in 1908 by Perrin [197, section 3], some 2,350 years after Democritus' argument! Meanwhile, in 1908, a fundamental step in the direction of a large body of modern science was taken by Paul Langevin in [149]. He argued that the equation of motion has the form

$$m\frac{d^2x}{dt^2} = -\gamma\frac{dx}{dt} + F, \quad \gamma > 0, \tag{6.1}$$

where F is a complementary force which sustains the agitation of the particle in the presence of viscous resistance. This is the first stochastic differential equation in history, written down long before the relative probabilistic foundations and concepts (Wiener process) were introduced! After important contributions by a number of theoretical physicists and engineers such as Fokker and Planck, and the Nyquist–Johnson model for RLC networks with noisy resistors in 1928 [135, 179], we come in 1930–1931 to the accepted model for physical Brownian motion in a conservative force field [239, 141, 47],[18] [176, Chapter 10] given by the stochastic oscillator

$$dx(t) = \quad v(t)\,dt, \tag{6.2a}$$

$$dv(t) = -\beta v(t)\,dt - \frac{1}{m}\nabla V(x(t))dt + \sigma dW_t, \tag{6.2b}$$

with $x(t_0) = x_0$ and $v(t_0) = v_0$ a.s., where $w(t)$ is a standard three-dimensional Wiener process and Einstein's *fluctuation-dissipation relation* holds:

$$\sigma^2 = 2kT\beta. \tag{6.3}$$

Here k is Boltzmann's constant and T is the absolute temperature of the fluid. The original Einstein–Smoluchowski theory is the high-friction limit of this model [176, Theorem 10.1].[19] Condition (6.3) guarantees the existence and the Boltzmann–Gibbs nature of an invariant measure for (6.2) with density

$$\rho_{BG}(x,v) = Z^{-1}\exp\left[-\frac{H(x,v)}{kT}\right] \text{ for } H(x,v) = \frac{1}{2}m\|v\|^2 + V(x), \tag{6.4}$$

and Z is a suitable normalizing constant (partition function); see [120, 53] for a generalization of this result.

These models have since played a central role in many areas of science beside microphysics, such as electric circuits [179], astronomy [47], mathematical finance since [11], biology, chemistry, etc. In more recent times, stochastic oscillators have played a central role in *cold damping feedback*, where (6.2) is replaced by

$$dx(t) = \quad v(t)\,dt, \tag{6.5a}$$

$$dv(t) = -\beta v(t)\,dt + u(x(t),v(t))dt - \frac{1}{m}\nabla V(x(t))dt + \sigma dW_t, \tag{6.5b}$$

[18] Ornstein and Uhlenbeck only considered the case of a quadratic potential leading to a Gauss–Markov model in phase space.

[19] Like the Aristotelian $F = mv$, often observed in nature, is the high friction limit of the Newtonian $F = ma$.

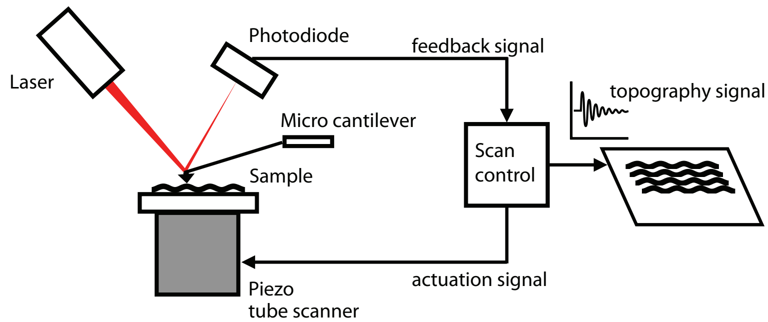


Fig. 4 *Surface topography: Velocity-dependent feedback control used to reduce thermal noise of a cantilever in atomic force microscopy.*

The purpose of this feedback control action is to reduce the effect of thermal noise on the motion of an oscillator by applying a viscous-like force, which was the very first feedback control action mathematically analyzed [166]. James Clerk Maxwell writes in that reference: "In one class of regulators of machinery, which we may call *moderators*, the resistance is increased by a quantity depending on the velocity." The first implementation on electrometers [171] dates back to the 1950s. Since then, it has been successfully employed in a variety of areas such as atomic force microscopy (AFM) [162] (*second intersection*!) (see[20] Figure 4), polymer dynamics [86, 38], and nano- to meter-sized resonators [93, 165, 218, 244, 199]. For (6.5), the feedback control action $u(t) = -\alpha v(t)$, $\alpha > 0$, asymptotically steers the phase space distribution to the steady state

$$\bar{\rho}(x, v) = \bar{Z}^{-1}\exp\left[-\frac{H(x, v)}{kT_{\text{eff}}}\right], \tag{6.6}$$

where the *effective temperature* T_{eff} satisfies

$$T_{\text{eff}} = \frac{\beta}{\beta + \alpha}T < T.$$

These new applications also pose new *physics* questions as the system is driven to a nonequilibrium *steady state* [201, 145, 35, 193]. In [95], a suitable *efficiency measure* for these diffusion-mediated devices was introduced which involves a class of stochastic control problems. Stochastic oscillators also play an important role in accelerating convergence of stochastic gradient descent for neural networks [191, Chapter 6], [49].

In [53], the problem of asymptotically driving system (6.5) to a desired *steady state* corresponding to reduced thermal noise was considered. Among the feedback controls achieving the desired asymptotic transfer, it was found that the one with the least energy is characterized by *time-reversibility*. This problem has its roots in the classical covariance control of Skelton, Grigoriadis, and collaborators [125, 126, 222, 115, 254].

The problem of steering such a system in *finite time* with minimum effort to a target steady state distribution was also solved in [53] as a generalized Schrödinger bridge problem. The system can then be maintained in the desired state through

[20]Notice that the experimental apparatus here is partially inspired by that of Kappler [141], with a light being shone onto a small mirror and the angle being measured through the position of the reflected spot a large distance away.

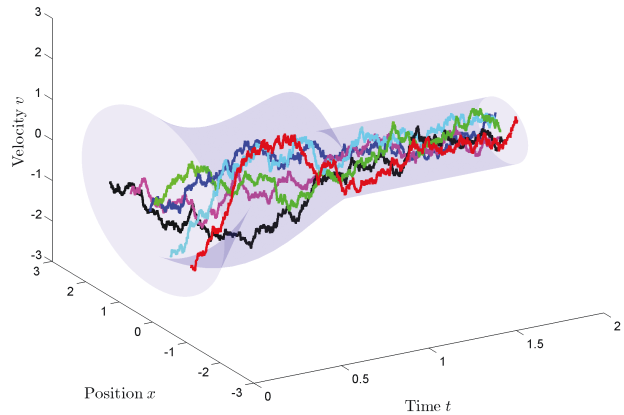


Fig. 5 *Inertial particles: trajectories in phase space.*

the optimal steady state feedback control. The solution, in the finite-horizon case, involves a space-time harmonic function φ satisfying

$$\frac{\partial \varphi}{\partial t} + v \cdot \nabla_x \varphi + \left(-\beta v - \frac{1}{m} \nabla_x V \right) \cdot \nabla_v \varphi + \frac{\sigma^2}{2} \Delta_v \varphi = 0. \tag{6.7}$$

Here $-\log \varphi$ plays the role of an artificial, time-varying potential under which the desired evolution takes place. This two-step control strategy is effectively illustrated by the following simple Gaussian example. The system

$$\begin{aligned} dx(t) &= \quad v(t)dt, \\ dv(t) &= -v(t)dt + u(t)dt - x(t)dt + dW_t \end{aligned}$$

is first optimally steered from time $t = 0$ to time $t = 1$ between the initial and final Gaussian marginals $\mathcal{N}(0, (1/2)I)$ and $\mathcal{N}(0, (1/2^4)I)$, respectively. The latter distribution is then maintained through constant feedback in this terminal desired state; see Figure 5, where the transparent tube represents the 3-standard deviation region of the state distribution.

7. Minimizing the Free Energy.

We now proceed to clarify how the problems considered by Sinkhorn [223, 224] are connected to Schrödinger bridges and to thermodynamic free energy. To achieve this in the most transparent way, we turn to the discrete setting.

7.1. Regularized Transport Problems.

Let us first recall the notion of the simplex of probability distributions on a finite set. Let V be a vector space and $A \subseteq V$. The *convex hull* [209] of A, written *con*A, is the intersection of all convex subsets of V containing A. The convex hull of $n+1$ *affinely independent*[21] points of a Euclidean space is called an *n-simplex*. For example, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, and so on. Let $\mathcal{D}(\mathcal{X})$ denote the family

[21] The points $x_1, x_2, \ldots, x_{n+1}$ are called affinely independent if every point x in their convex hull admits a *unique* representation as a convex combination of the points.

of all probability distributions on the sample space $\mathcal{X} = \{1, 2, \ldots, n\}$. Then $\mathcal{D}(\mathcal{X})$ is an $(n-1)$-simplex whose vertices are the distributions $p_i(j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta.

The discrete OMT problem [202] has been popularized in the following form. Suppose there are n mines with mine i producing the fraction p_i of the total production. There are also n factories which need the raw material from the mines. To operate, factory j needs the fraction q_j of the total available supply. Let $C = (c_{ij})_{i,j=1}^n$ be a matrix of "transportation costs"[22] with nonnegative elements. On $\mathcal{D}(\mathcal{X})$, we can then define a metric in the following way: Given the two probability distributions $p, q \in \mathcal{D}(\mathcal{X})$, let $\Pi(p, q)$ be the family of probability distributions on $\mathcal{X} \times \mathcal{X}$ that are "couplings" of p and q, namely, $\pi \in \Pi(p, q)$ has marginals p and q, respectively. Any $\pi \in \Pi(p, q)$ represents a feasible transport plan, the quantity π_{ij} representing the amount of the demand of factory j which is satisfied by mine i. Then the discrete OMT problem of minimizing the total cost of transportation while respecting the constraints leads to the *optimal transport distance* between p and q defined by

$$d_C(p, q) := \min_{\pi \in \Pi(p,q)} \sum_{i,j} c_{ij}\pi_{ij}. \tag{7.1}$$

When $c_{ij} = d(i, j)^2$, where $d(\cdot, \cdot)$ is a distance on $\mathcal{X} \times \mathcal{X}$,

$$W_2(p, q) := \left(d_C(p, q)\right)^{1/2}$$

is called the *earth mover distance (Wasserstein* 2*-distance)*. It can be shown [198, Proposition 2.2] that W_2 is a *bona fide* distance on $\mathcal{D}(\mathcal{X})$. This distance has recently found important applications in many diverse fields of science such as economics, physics, engineering, probability, and, in particular, in information engineering for problems of imaging (DTI, multimodal, color, etc.), robust-efficient transport over networks, spectral analysis, collective dynamics, etc. A regularized version of (7.1), which features important algorithmic and computational advantages [76, 198], is obtained by subtracting a term proportional to the entropy,

$$\inf_{\pi \in \Pi(p,q)} \left[\sum_{i,j} c_{ij}\pi_{ij} - \epsilon S(\pi)\right], \quad S(\pi) = -\sum_{ij} \pi_{ij} \log(\pi_{ij}), \tag{7.2}$$

for $\epsilon > 0$. Notice, in particular, that the resulting functional

$$J(\pi) = \sum_{i,j} c_{ij}\pi_{ij} + \epsilon \sum_{ij} \pi_{ij} \log(\pi_{ij}) \tag{7.3}$$

is strictly convex in π.

7.2. Thermodynamic Systems: Statics. We consider a physical system with state space $\mathcal{X} = \{1, 2, \ldots, M\}$. We can think of this mesoscopic description as originating from a microscopic description where the *phase space*, in Boltzmann's style, has undergone a "coarse graining" through subdivision into small cells, which is what we typically observe. Each of the cells represents a mesoscopic state. While the microscopic states are equally likely, this is no longer true for the macroscopic states, which correspond to different numbers of microstates.

[22]c_{ij} is the cost of transporting one unit of material from mine i to factory j.

For each macroscopic state x we consider its *energy* $E_x \geq 0$. The function $H : x \mapsto E_x$ is referred to as the *Hamiltonian*. The thermodynamic states of the system are given by probability distributions on $\mathcal{Z}$ reflecting how many microscopic states correspond to the macroscopic states, namely, by $\mathcal{D}(\mathcal{X})$. On $\mathcal{D}(\mathcal{X})$, we define the *internal energy* as the expected value of the energy *observable* in state π, namely,

$$U(\pi) = \mathbb{E}_\pi\{H\} = \sum_x E_x \pi_x = \langle E, \pi \rangle, \tag{7.4}$$

where E denotes the n-dimensional vector with components E_x. Let us also introduce the *Gibbs entropy*

$$S_G(\pi) = kS(\pi) = -k \sum_x \pi_x \log \pi_x, \tag{7.5}$$

where k is Boltzmann's constant. As is well known, S_G is nonnegative and strictly concave on $\mathcal{D}(\mathcal{X})$. Let $\bar{E}$ be a constant satisfying

$$\min_x E_x \leq \bar{E} \leq \frac{1}{n} \sum_x E_x. \tag{7.6}$$

We can think of $\bar{E}$ as the energy of the underlying conservative microscopic system (the upper bound $\frac{1}{n}\sum_l E_l$ in (7.6) guarantees existence of a positive multiplier; see below). We now consider the following *maximum entropy* problem:

$$\max \quad \{S_G(\pi) \mid \pi \in \mathcal{D}(\mathcal{X})\} \tag{7.7a}$$
$$\text{subject to} \quad U(\pi) = \bar{E}. \tag{7.7b}$$

This is an (important) instance of a class of maximum entropy problems originating with Boltzmann [34] (see [195] for a survey), where entropy is maximized over probability distributions that give the correct expectation of certain observables in accordance with known macroscopic quantities. The *Lagrangian function* is then given by

$$\mathcal{L}(\pi, \lambda) := S_G(\pi) + \lambda(\bar{E} - U(\pi)), \tag{7.8}$$

where the Lagrange multiplier λ is positive, corresponding to positive "absolute temperatures" $T = \lambda^{-1}$. The problem is then equivalent to minimizing over $\mathcal{D}(\mathcal{X})$ the *Helmholtz free energy* functional

$$F(\pi, T) = U(\pi) - TS_G(\pi) = \sum_x E_x \pi_x + kT \sum_x \pi_x \log \pi_x. \tag{7.9}$$

Since F is strictly convex on $\mathcal{D}(\mathcal{X})$, the first-order optimality conditions are sufficient and determine the unique minimizer in the form of the *Boltzmann distribution*[23]

$$\pi_B(x) = Z(T)^{-1} \exp\left[-\frac{E_x}{kT}\right], \text{ where } Z(T) = \sum_x \exp\left[-\frac{E_x}{kT}\right]; \tag{7.10}$$

see, e.g., [210]. Alternatively, it suffices to observe that

$$F(\pi, T) = T\mathbb{D}(\pi \| \pi_B) - T \log Z(T) \tag{7.11}$$

[23]The letter Z for the *partition function* was chosen by Boltzmann to indicate "zuständige Summe" (relevant sum).

and invoke the properties of relative entropy. As is well known, the Boltzmann distribution (7.10) tends to the uniform (maximum entropy) distribution as $T \nearrow +\infty$ and tends to concentrate on the set of minimal energy states as $T \searrow 0$. Hence, for $0 < T < +\infty$, the Boltzmann distribution represents in a precise way a *compromise* between minimizing energy and maximizing entropy.

7.3. Schrödinger and Sinkhorn, Redux. The fact that the Boltzmann distribution minimizes the free energy may be viewed as an elementary version of what is often called *Gibbs' variational principle*. Notice that the minimization of F in (7.9) is *unconstrained*. Nevertheless, we are often interested in minimizing the free energy under additional constraints. This is usually the case with natural evolutions which tend to maximal entropy configurations[24] while respecting certain constraints. In particular, we now consider a constrained version of the minimization of (7.9). In the notation of section 7.1, let $\mathcal{Z} = \mathcal{X} \times \mathcal{X}$ and consider the problem

$$\inf_{\pi \in \Pi(p,q)} F(\pi, T). \tag{7.12}$$

Then, letting $\epsilon = kT$, comparing (7.9) with (7.3) shows that (7.12) coincides with the regularized optimal transport problem (7.2). In particular, up to constants, *the negative Lagrangian* (7.8) *for problem* (7.7a)–(7.7b) *coincides with the functional* (7.3) *to be minimized in regularized optimal transport*. On the other hand, because of (7.11), problem (7.2) is equivalent to

$$\min_{\pi \in \Pi(p,q)} \mathbb{D}(\pi \| \pi_B), \tag{7.13}$$

where

$$\pi_B(i,j) = Z(T)^{-1} \exp\left[-\frac{c_{ij}}{kT}\right], \quad Z(T) = \sum_{ij} \exp\left[-\frac{c_{ij}}{kT}\right], \tag{7.14}$$

which is a discrete counterpart of Problem 4.2. Naturally, this and the other maximum entropy problems of this section also admit the large deviations interpretation of section 4.2.

Let us now write the joint probability $\pi_B(i,j)$ as

$$\pi_B(i,j) = p_B(i) p_B(i,j),$$

where $p_B(i,j)$ is the conditional probability. Introducing Lagrange multipliers for the linear constraints,

$$\sum_j \pi_{ij} = p_i, \quad i = 1, 2, \ldots, n, \tag{7.15}$$

$$\sum_i \pi_{ij} = q_j, \quad j = 1, 2, \ldots, n, \tag{7.16}$$

and proceeding precisely as in section 4.3, we readily get the following expression for the optimal π:

$$\pi^*_{ij} = \hat{\varphi}(i) p_B(i,j) \varphi(j), \tag{7.17}$$

[24]According to Planck, nature seems to favor high entropy states.

where the nonnegative functions $\hat{\varphi}$ and φ satisfy the system

$$\hat{\varphi}(i)\left[\sum_j p_B(i,j)\varphi(j)\right] = p_i, \tag{7.18}$$

$$\varphi(j)\left[\sum_i p_B(i,j)\hat{\varphi}(i)\right] = q_j. \tag{7.19}$$

Defining $\hat{\varphi}(0,i) = \hat{\varphi}(i)$, $\varphi(1,j) = \varphi(j)$, we see that (7.18)–(7.19) can be replaced by the system

$$\varphi(0,i) = \sum_j p_B(i,j)\varphi(1,j), \tag{7.20a}$$

$$\hat{\varphi}(1,j) = \sum_i p_B(i,j)\hat{\varphi}(0,i), \tag{7.20b}$$

$$\varphi(0,i)\cdot\hat{\varphi}(0,i) = p_i, \tag{7.20c}$$

$$\varphi(1,j)\cdot\hat{\varphi}(1,j) = q_j. \tag{7.20d}$$

Let us write

$$\pi^*_{ij} = p_i \cdot p^*(i,j)$$

and assume $p_i > 0$ for all i. Dividing both sides of (7.17) by p_i we get, in view of (7.20c),

$$p^*(i,j) = \frac{1}{\varphi(0,i)} p_B(i,j)\varphi(1,j), \tag{7.21}$$

which should be compared to (4.16). It is interesting to write (7.21) in matricial form. Let $P^* = (p^*(i,j))$ and $P_B = (p_B(i,j))$. Then (7.21) gives

$$P^* = \operatorname{diag}\left(\frac{1}{\varphi(0,1)}, \dots, \frac{1}{\varphi(0,n)}\right) P_B \operatorname{diag}\left(\varphi(1,1), \dots, \varphi(1,n)\right). \tag{7.22}$$

System (7.20) represents a *discrete counterpart of the Schrödinger system* (4.14a)–(4.14c). Existence for the latter, as previously observed after (4.14c), is extremely challenging, with the first solution being provided by Robert Fortet in 1940 [106] through a complex iterative scheme. The same problem is much simpler in the discrete setting with the first convergence proof for the classical IPF procedure being provided in a special case by Sinkhorn in 1964 [223]. Indeed, consider the special case where both marginals are uniform distributions so that $p_i = q_i = 1/n, i = 1, 2, \dots, n$. Let $C = (c_{ij})$ be the matrix of transportation costs. Then $\pi \in \mathcal{D}(\mathcal{X} \times \mathcal{X})$ belongs to $\Pi(p,q)$ if and only if it satisfies the constraints

$$\sum_j \pi_{ij} = \frac{1}{n}, \quad i = 1, 2, \dots, n, \tag{7.23}$$

$$\sum_i \pi_{ij} = \frac{1}{n}, \quad j = 1, 2, \dots, n. \tag{7.24}$$

Thus, the matrix $n\pi$ must be *doubly stochastic*, which was the original Sinkhorn problem.[25]

[25] Sinkhorn [223]: "It is not the intent of this paper to obtain properties of this estimate." Sinkhorn appears only concerned with establishing convergence of the iterative method to a doubly stochastic matrix without clearly connecting the latter to an optimization problem.

8. The Fortet-IPF-Sinkhorn Algorithm.

8.1. Continuous Case. In 1938–1940 Robert Fortet, a fine French analyst and former student of Maurice Fréchet, set out to solve the problem of existence and uniqueness for the Schrödinger system (4.14a)–(4.14c), left open by Schrödinger[26] as well as by Bernstein in [24]. Fortet's proof in [105, 106] is, to this day, the only *algorithmic* one and, in a rather general setting, establishes convergence of successive approximations. More explicitly, let $g(x,y)$ be a nonnegative, continuous function on $\mathbb{R}\times\mathbb{R}$ bounded from above. Suppose $g(x,y)>0$ except possibly for a zero measure set for each fixed value of x or of y. Suppose that $\rho_0(x)$ and $\rho_1(y)$ are continuous, nonnegative functions such that

$$\int \rho_0(x)dx = \int \rho_1(y)dy.$$

Suppose, moreover, that the integral

$$\int \frac{\rho_1(y)}{\int g(z,y)\rho_0(z)dz}dy$$

is finite (this is Fortet's crucial hypothesis). Then the system [106, Theorem 1]

$$\hat{\varphi}(x)\int g(x,y)\varphi(y)dy = \rho_0(x), \tag{8.1}$$

$$\varphi(y)\int g(x,y)\hat{\varphi}(x)dx = \rho_1(y) \tag{8.2}$$

admits a solution $(\varphi(x),\hat{\varphi}(y))$ with $\varphi\geq 0$ continuous and $\hat{\varphi}\geq 0$ measurable. Moreover, $\hat{\varphi}(x)=0$ only where $\rho_0(x)=0$ and $\varphi(y)=0$ only where $\rho_1(y)=0$.

The result is proven by setting up a complex approximation scheme to show that $\mathcal{C}(h)=h$ with

$$\mathcal{C}(h)(\cdot) := \int g(\cdot,y)\frac{\rho_1(y)dy}{\int g(z,y)\frac{\rho_0(z)}{h(z)}dz} \tag{8.3}$$

has a fixed point. The map $\mathcal{C}$ is considered on functions of class (K), namely, functions $h:\mathbb{R}\to(\mathbb{R}\cup+\infty)$ which satisfy the following properties:

(i) h is measurable;
(ii) there exists $\alpha>0$ such that $h(x)\geq\alpha\ \forall x\in\mathbb{R}$;
(iii) for almost every $x\in\mathbb{R}$, $h(x)<+\infty$.

If h_0 and h_1 are of class (K) and $h_0\leq h_1$ a.e., then $\mathcal{C}(h_0)\leq\mathcal{C}(h_1)$. Moreover, on class (K) functions, $\mathcal{C}$ is positively homogeneous of degree one. Unfortunately, the map $\mathcal{C}$ does not map class (K) functions into class (K) functions as it does not preserve the property of being bounded away from zero. This is a fundamental difficulty of the continuous case which Fortet circumvents through his brilliant but complex approximation scheme involving three sequences of functions. This difficulty can be avoided altogether in the discrete case through a suitable positivity assumption; see Theorem 8.1 below. Notice that the one-dimensional heat kernel

$$g(x,y)=p(0,x,1,y)=\frac{1}{\sqrt{2\pi\gamma}}\exp\left[-\frac{|x-y|^2}{2\gamma}\right]$$

[26]Schrödinger thought existence and uniqueness should hold since the problem looked to him so natural except, possibly, in the case of very nasty marginals.

satisfies all assumptions of Fortet's theorem. Uniqueness, in the sense described after (4.14c), namely, uniqueness of rays, is much easier to establish. In [91], most of Fortet's paper has been revisited, filling in all the gaps and explaining the meaning of the various steps of his elaborate approach. Another recent paper in this direction is [159].

8.2. Discrete Case. In 1940, an *iterative proportional fitting* (IPF) procedure was proposed in the statistical literature on contingency tables [83]. Convergence for the IPF algorithm was first established (in a special case) by Richard Sinkhorn in 1964 [223]. The iterates were shortly afterwards shown to converge to a "minimum discrimination information" [127, 94, 73], namely, to a minimum entropy distance. This line of research, usually called *Sinkhorn algorithms*, continues to this day; see, e.g., [76, 4, 234].

We now state and, later on, establish the following fundamental result.

THEOREM 8.1. *Let $\mathcal{X} = \{1, 2, \ldots, n\}$ and $p, q \in \mathcal{D}(\mathcal{X})$. Assume that the $n \times n$ matrix $G = (g_{ij})$ has all positive elements. Then there exist vectors $\varphi(0, \cdot)$, $\hat{\varphi}(1, \cdot)$ with positive entries such that*

$$\varphi(0, i) = \sum_j g_{ij}\varphi(1, j), \tag{8.4a}$$

$$\hat{\varphi}(1, j) = \sum_i g_{ij}\hat{\varphi}(0, i), \tag{8.4b}$$

$$\varphi(0, i) \cdot \hat{\varphi}(0, i) = p_i, \tag{8.4c}$$

$$\varphi(1, j) \cdot \hat{\varphi}(1, j) = q_j. \tag{8.4d}$$

The pair $\varphi(0, \cdot)$, $\hat{\varphi}(1, \cdot)$ is unique up to multiplication of $\varphi(0, \cdot)$ by a positive constant α and division of $\hat{\varphi}(1, \cdot)$ by the same constant α.

We first set up a natural iterative scheme (Fortet-IPF-Sinkhorn) for system (8.4). We introduce the following linear maps on $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0\}$, the positive orthant of $\mathbb{R}^n$:

$$\begin{aligned} \mathcal{E} \ &: x \mapsto y = \sum_j g_{ij}x_j, \\ \mathcal{E}^\dagger \ &: x \mapsto y = \sum_i g_{ij}x_i. \end{aligned}$$

Here and in what follows $\dagger$ denotes adjoint.[27] We also define the following nonlinear maps on the interior of the positive orthant int $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i > 0\}$:

$$\begin{aligned} \mathcal{D}_0 \ &: x \mapsto y = \frac{p}{x}, \\ \mathcal{D}_1 \ &: x \mapsto y = \frac{q}{x}, \end{aligned}$$

where division of vectors is *componentwise*. On $\mathbb{R}^n_+$, also consider the composition of the four maps

$$\mathcal{C} := \mathcal{E} \circ \mathcal{D}_1 \circ \mathcal{E}^\dagger \circ \mathcal{D}_0. \tag{8.5}$$

[27]Our use of the adjoint for the map $\mathcal{E}$ is consistent with the standard notation in diffusion processes, where the Fokker–Planck (forward) equation involves the adjoint of the *generator* appearing in the backward Kolmogorov equation.

This is just the discrete counterpart of map (8.3). Consider the vector iteration

$$\varphi_{k+1}(0) = \mathcal{C}(\varphi_k(0)), \quad \varphi_0(0) = \mathbb{1}, \tag{8.6}$$

where $\mathbb{1}^\dagger = (1, 1, \ldots, 1)$. Observe first that (8.6) stays in the interior of $\mathbb{R}^n_+$ even when the marginals p and/or q have some zero components. Indeed, since $g_{ij} > 0\ \forall(i, j)$, maps $\mathcal{E}$ and $\mathcal{E}^\dagger$ map $\mathbb{R}^n_+$ into int $\mathbb{R}^n_+$. It follows that the componentwise divisions of $\mathcal{D}_0$ and $\mathcal{D}_1$ are well defined and $\mathcal{C}$: int $\mathbb{R}^n_+ \mapsto$ int $\mathbb{R}^n_+$. Next, we want to show that the sequence generated by (8.6) converges to a fixed point of $\mathcal{C}$, thereby proving Theorem 8.1. Theorem 8.1 asserts, in particular, that *uniqueness* for the Schrödinger system (8.4) concerns *rays* in the positive orthant. This suggests that contractivity of the map (8.5) should be established on the set of rays endowed with a *projective metric*. This was accomplished in [112, Theorem 3], which extends the approach of Franklin and Lorenz [108] dealing with the scaling of nonnegative matrices. We devote the next subsection to some key results on Hilbert's projective metric in which the rays' convergence can be proved.

8.3. Hilbert's Projective Metric. This metric was introduced by Hilbert in 1895 [121]. In 1957 [28], Garrett Birkhoff proved a crucial contractivity result in this metric that enabled the establishment of existence of solutions of linear equations on cones (such as the Perron–Frobenius theorem (see Theorem 9.6 below)). This result was extended to certain nonlinear maps by Bushell [41, 42]. Besides the ergodic theory for Markov chains, the Birkhoff–Bushell results have been applied to positive integral operators and to positive definite matrices [42, 155]. In recent times, this geometry has been proven useful in problems concerning communication and computations over networks (see [238]). Other significant applications have been developed by Sepulchre and collaborators [219, 36, 12] that concern consensus in noncommutative spaces and metrics for spectral densities. We also mention applications to quantum information theory [204]. On the more mathematical side, a survey on the applications in analysis is [155]. The use of the Hilbert metric is crucial in the nonlinear Frobenius–Perron theory [154]. A further extension of the Perron–Frobenius theory beyond linear positive systems and monotone systems has been recently proposed in [104].

Applications of the Birkhoff–Bushell contractivity results to the topics of this paper were apparently initiated in 1989 with the paper [108], which deals with scaling of nonnegative matrices. In [112], we showed that the Schrödinger bridge for Markov chains and quantum channels can be obtained efficiently from the fixed point of a map that contracts the *Hilbert metric*. In [58], a similar approach was taken in the context of diffusion processes leading to a new proof of a classical result of Jamison on existence and uniqueness for the Schrödinger bridge and also providing an efficient computational scheme for both Schrödinger bridges and OMT. This new computational approach can be effectively employed, for instance, in image interpolation. There are, however, some fundamental difficulties in using this approach in Schrödinger's original setting. These are outlined in Remark 8.7 below.

Following [42], we recall some basic concepts and results of this theory.

Let $\mathcal{S}$ be a real Banach space and let $\mathcal{K}$ be a closed solid cone in $\mathcal{S}$, i.e., $\mathcal{K}$ is closed with nonempty interior int $\mathcal{K}$ and is such that $\mathcal{K} + \mathcal{K} \subseteq \mathcal{K}$, $\mathcal{K} \cap -\mathcal{K} = \{0\}$ as well as $\lambda\mathcal{K} \subseteq \mathcal{K}$ for all $\lambda \geq 0$. Define the partial order

$$x \preceq y \Leftrightarrow y - x \in \mathcal{K}, \quad x < y \Leftrightarrow y - x \in \operatorname{int}\mathcal{K},$$

and for $x, y \in \mathcal{K}_0 := \mathcal{K}\backslash\{0\}$, define

$$M(x,y) := \inf\{\lambda \mid x \preceq \lambda y\},$$
$$m(x,y) := \sup\{\lambda \mid \lambda y \preceq x\}.$$

Then the Hilbert metric is defined on $\mathcal{K}_0$ by

$$d_H(x,y) := \log\left(\frac{M(x,y)}{m(x,y)}\right).$$

Strictly speaking, it is a *projective* metric since it is invariant under scaling by positive constants, i.e., $d_H(x,y) = d_H(\lambda x, \mu y)$ for any $\lambda > 0, \mu > 0$, and $x, y \in \operatorname{int}\mathcal{K}$. Thus, it is actually a distance between rays. If U denotes the unit sphere in $\mathcal{S}$, $(\operatorname{int}\mathcal{K} \cap U, d_H)$ is a metric space.

Example 8.2. *Let $\mathcal{K} = \mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0\}$ be the positive orthant of $\mathbb{R}^n$. Then, for $x, y \in \operatorname{int}\mathbb{R}^n_+$, i.e., with all positive components,*

$$d_H(x,y) = \log\max\{x_i y_j / y_i x_j\}.$$

Another very important example for applications in many diverse areas of statistics, information theory, control, etc., is the cone of Hermitian, positive semidefinite matrices.

Example 8.3. *Let $\mathcal{S} = \{X = X^\dagger \in \mathbb{C}^{n\times n}\}$, where here $\dagger$ denotes transposition plus conjugation and, more generally, adjoint. Let $\mathcal{K} = \{X \in \mathcal{S} : X \geq 0\}$ be the positive semidefinite matrices. Then, for $X, Y \in \operatorname{int}\mathcal{K}$, namely, positive definite, we have*

$$d_H(X,Y) = \log\frac{\lambda_{\max}\left(XY^{-1}\right)}{\lambda_{\min}\left(XY^{-1}\right)} = \log\frac{\lambda_{\max}\left(Y^{-1/2}XY^{-1/2}\right)}{\lambda_{\min}\left(Y^{-1/2}XY^{-1/2}\right)},$$

which is closely connected to the Riemannian (Fisher-information) metric

$$d_R(X,Y) = \left\|\log\left(Y^{-1/2}XY^{-1/2}\right)\right\|_F = \sqrt{\sum_{i=1}^n [\log\lambda_i\left(Y^{-1/2}XY^{-1/2}\right)]^2}.$$

Notice that, in the two examples above, Hilbert's pseudometric puts the boundary of the cone at infinite distance from any interior point.

A map $\mathcal{E} : \mathcal{K} \to \mathcal{K}$ is called *nonnegative*. It is called *positive* if $\mathcal{E} : \operatorname{int}\mathcal{K} \to \operatorname{int}\mathcal{K}$. If $\mathcal{E}$ is positive and $\mathcal{E}(\lambda x) = \lambda^p\mathcal{E}(x)$ for all $x \in \operatorname{int}\mathcal{K}$ and positive λ, then $\mathcal{E}$ is called *positively homogeneous of degree p* in $\operatorname{int}\mathcal{K}$. For a positive map $\mathcal{E}$, the *projective diameter* is defined by

$$\Delta(\mathcal{E}) := \sup\{d_H(\mathcal{E}(x), \mathcal{E}(y)) \mid x, y \in \operatorname{int}\mathcal{K}\}$$

and the *contraction ratio* by

$$\kappa(\mathcal{E}) := \inf\{\lambda \mid d_H(\mathcal{E}(x), \mathcal{E}(y)) \leq \lambda d_H(x,y)\ \forall x, y \in \operatorname{int}\mathcal{K}\}.$$

Finally, a map $\mathcal{E} : \mathcal{S} \to \mathcal{S}$ is called *monotone increasing* if $x \leq y$ implies $\mathcal{E}(x) \leq \mathcal{E}(y)$.

Theorem 8.4 ([42]). *Let $\mathcal{E}$ be a monotone increasing positive mapping which is positive homogeneous of degree p in $\operatorname{int}\mathcal{K}$. Then the contraction $\kappa(\mathcal{E})$ does not exceed p. In particular, if $\mathcal{E}$ is a positive linear mapping, $\kappa(\mathcal{E}) \leq 1$.*

THEOREM 8.5 ([28, 42]). *Let $\mathcal{E}$ be a positive linear map. Then*

$$\kappa(\mathcal{E}) = \tanh\left(\frac{1}{4}\Delta(\mathcal{E})\right). \tag{8.7}$$

THEOREM 8.6 ([42]). *Let $\mathcal{E}$ be either*

(a) *a monotone increasing positive mapping which is positive homogeneous of degree $p(0 < p < 1)$ in $\operatorname{int}\mathcal{K}$, or*

(b) *a positive linear mapping with finite projective diameter.*

Suppose the metric space $Y = (\operatorname{int}\mathcal{K} \cap U, d_H)$ is complete. Then in case (a), there exists a unique $x \in \operatorname{int}\mathcal{K}$ such that $\mathcal{E}(x) = x$; in case (b), there exists a unique positive eigenvector of $\mathcal{E}$ in Y.

This result provides a far-reaching generalization of the celebrated Perron–Frobenius theorem [29] (see Theorem 9.6, below). Notice that in both Examples 8.2 and 8.3, the space $Y = (\operatorname{int}\mathcal{K} \cap U, d_H)$ is indeed complete [42]. We also note that there are other metrics as well that are contracted by positive monotone maps, for instance, the closely related *Thompson metric* [235] $d_T(x, y) = \log\max\{M(x, y), m^{-1}(x, y)\}$. The Thompson metric is a *bona fide* metric on $\mathcal{K}$ and has been, for example, employed in [163, 66, 12].

Remark 8.7. If we try to use a similar approach to prove existence for the Schrödinger system (8.1)–(8.2), we may expect that in an infinite-dimensional setting questions of boundedness or integrability might become delicate. The main difficulty, however, lies in two other issues. To introduce them, let us observe that in the Birkhoff–Bushell theory we have linear or nonlinear iterations that remain in the *interior of a cone*. For example, in the application of the Perron–Frobenius theorem to the ergodic theory of Markov chains, the assumption that there exists a power of the transition matrix with all strictly positive entries ensures that the evolution of the probability distribution occurs in the *interior* of the positive orthant (intersected with the simplex). The first difficulty is that the natural function space cones such as L^1_+ (L^2_+), namely, integrable (square integrable) nonnegative functions on $\mathbb{R}^d$, have *empty interior*! The second difficulty is that even if we manage to somehow define a suitable function space cone with nonempty interior,[28] the nonlinear map Ω defined in (8.3) cannot map the interior of the cone into itself. To overcome precisely this difficulty, in [58] the two marginals were assumed to have compact support. We see here once more, from a slightly different angle, how much more challenging the continuous case is.

8.4. Proof of Theorem 8.1.

We begin with three preliminary results.

LEMMA 8.8. *Consider the maps $\mathcal{E}$ and $\mathcal{E}^\dagger$. We have the following bounds on their contraction ratios:*

$$\kappa(\mathcal{E}) = \kappa(\mathcal{E}^\dagger) = \tanh\left(\frac{1}{4}\Delta(\mathcal{E})\right) < 1. \tag{8.8}$$

Proof. Observe that $\mathcal{E}$ is a positive *linear* map and its projective diameter is

$$\begin{aligned}\Delta(\mathcal{E}) &= \sup\{d_H(\mathcal{E}(x), \mathcal{E}(y)) \mid x_i > 0,\ y_i > 0\}\\ &= \sup\{\log\left(\frac{g_{ij}g_{k\ell}}{g_{i\ell}g_{kj}}\right) \mid 1 \le i, j, k, \ell \le n\}.\end{aligned}$$

[28]This was indeed accomplished in [58].

It is finite, since all entries g_{ij} are positive. It now follows from Theorem 8.5 that its contraction ratio satisfies (8.8). A similar result holds for the adjoint map $\mathcal{E}^\dagger$. ∎

LEMMA 8.9.

$$\kappa(\mathcal{D}_0) \le 1, \quad \kappa(\mathcal{D}_1) \le 1.$$

Proof. Observe that when p and q have positive entries, both $\mathcal{D}_0$ and $\mathcal{D}_1$ are isometries in the Hilbert metric. Indeed, for vectors x, y in the interior of $\mathbb{R}^n_+$, inversion and elementwise scaling are both isometries for the Hilbert metric as can be shown by the calculations

$$\begin{aligned} d_H(x,y) &= \log\left((\max_i(x_i/y_i))\frac{1}{\min_i(x_i/y_i)}\right) \\ &= \log\left(\frac{1}{\min_i((x_i)^{-1}/(y_i)^{-1})}\max_i((x_i)^{-1}/(y_i)^{-1})\right) \\ &= d_H(x^{-1}, y^{-1}), \end{aligned}$$

where x^{-1} and y^{-1} are obtained from x and y, respectively, through componentwise inversion. Moreover, let px and py be the vectors with components $p_i x_i$ and $p_i y_i$, respectively. Then

$$\begin{aligned} d_H(px, py) &= \log\frac{\max_i((p_i x_i)/(p_i y_i))}{\min_i((p_i x_i)/(p_i y_i))} \\ &= \log\frac{\max_i(x_i/y_i)}{\min_i(x_i/y_i)} = d_H(x,y). \end{aligned}$$

If p has zero entries, then the second equality above needs to be replaced by the inequality "$\le$". ∎

LEMMA 8.10. *The composition*

$$\mathcal{C} = \mathcal{E} \circ \mathcal{D}_1 \circ \mathcal{E}^\dagger \circ \mathcal{D}_0 \tag{8.9}$$

contracts the Hilbert metric with contraction ratio $\kappa(\mathcal{C}) < 1$*, namely,*

$$d_H(\mathcal{C}(x), \mathcal{C}(y)) < d_H(x,y) \quad \forall x, y \in \operatorname{int}\mathbb{R}^n_+.$$

Proof. The result follows at once from Lemmas 8.8 and 8.9. ∎

We now complete the proof of Theorem 8.1. The set of rays in $\operatorname{int}\mathbb{R}^n_+$ with the Hilbert metric is complete (this can be proven by intersecting $\operatorname{int}\mathbb{R}^n_+$ with the unit sphere; see [42, section 4]). By Lemma 8.10, $\mathcal{C}$ contracts the Hilbert metric. By the Banach–Caccioppoli Contraction Mapping Theorem, there exists a unique ray in $\operatorname{int}\mathbb{R}^n_+$ to which the iteration (8.6) (starting from any vector in $\operatorname{int}\mathbb{R}^n_+$) converges. Next, we prove that, due to (8.4a)–(8.4d), the iteration (8.6) actually converges to a fixed vector. Since the iteration has a fixed ray, we have that for some positive constant α,

$$\alpha \cdot \varphi(0) = \mathcal{C}(\varphi(0)),$$

where the composed map is $\mathcal{C} := \mathcal{E} \circ \mathcal{D}_1 \circ \mathcal{E}^\dagger \circ \mathcal{D}_0$. From this we can obtain

$$\begin{aligned} \hat\varphi(1) &= \mathcal{E}^\dagger(\hat\varphi(0)), \\ \varphi(0) &= \mathcal{E}(\varphi(1)), \end{aligned}$$

while

$$\hat{\varphi}(1)\varphi(1) = q \text{ and}$$
$$\alpha\hat{\varphi}(0)\varphi(0) = p,$$

where, as usual, multiplication is componentwise. Let $\langle\cdot,\cdot\rangle$ denote the scalar product in $\mathbb{R}^n$. Since p and q are probability distributions, we have

$$\begin{aligned} 1 &= \alpha\langle\hat{\varphi}(0), \varphi(0)\rangle \\ &= \alpha\langle\hat{\varphi}(0), \mathcal{E}(\varphi(1))\rangle \\ &= \alpha\langle\mathcal{E}^\dagger(\hat{\varphi}(0)), \varphi(1)\rangle \\ &= \alpha\langle\hat{\varphi}(1), \varphi(1)\rangle \\ &= \alpha. \end{aligned}$$

Thus $\alpha = 1$, the iteration (8.6) converges to a fixed point (vector in $\mathbb{R}^n_+$), and the four vectors $(\varphi(0), \hat{\varphi}(0), \varphi(1), \hat{\varphi}(1))$ satisfy the Schrödinger system (8.4). The vectors $\varphi(0)$ and $\hat{\varphi}(1)$ have all positive components.

9. Efficient vs. Robust Routing for Network Flows.

While problem (7.13), and the corresponding equivalent regularized OMT (7.2), are the discrete counterparts of Problem 4.2, it is apparent that the discrete counterpart of the "dynamic" Schrödinger bridge problem, Problem 4.1, is still missing. Before we turn to dynamic problems with discrete state space, let us mention that work has also been done on discrete time and continuous state space [249, 21, 13, 114, 14, 17]. This literature mostly deals with Gaussian distributions, in the finite and infinite horizon cases, with and without noise in the dynamics, and with and without constraints. The case of regularized transport on discrete metric graphs was studied by Léonard in [158].

9.1. Generalized Bridge Problems.

Consider a directed, strongly connected (i.e., with at least one path joining each pair of vertices), aperiodic graph $\mathbf{G} = (\mathcal{X}, \mathcal{E})$ with vertex set $\mathcal{X} = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{X} \times \mathcal{X}$. Time is discrete and taken in $\mathcal{T} = \{0, 1, \ldots, N\}$.

Let $\mathcal{F}P_0^N \subseteq \mathcal{X}^{N+1}$ denote the family of feasible paths $x = (x_0, \ldots, x_N)$ of length N, namely, paths such that $x_i x_{i+1} \in \mathcal{E}$ for $i = 0, 1, \ldots, N-1$. We seek a probability distribution $\mathfrak{P}$ on $\mathcal{FP}_0^N$ with prescribed initial and final marginal probability distributions $\nu_0(\cdot)$ and $\nu_N(\cdot)$, respectively, and such that the resulting random evolution is closest to a "prior" measure $\mathfrak{M}$ on $\mathcal{FP}_0^N$ in a suitable sense.

The prior law for our problem is induced by the Markovian evolution

$$\mu_{t+1}(x_{t+1}) = \sum_{x_t \in \mathcal{X}} \mu_t(x_t) m_{x_t x_{t+1}}(t) \tag{9.1}$$

with nonnegative distributions $\mu_t(\cdot)$ over $\mathcal{X}$, $t \in \mathcal{T}$, and weights $m_{ij}(t) \geq 0$ for all indices $i, j \in \mathcal{X}$ and all times. Moreover, to respect the topology of the graph, $m_{ij}(t) = 0$ for all t whenever $ij \notin \mathcal{E}$. Often, but not always, the matrix

$$M(t) = [m_{ij}(t)]_{i,j=1}^n \tag{9.2}$$

may not depend on t. The rows of the transition matrix $M(t)$ do not necessarily sum up to one, so that the "total transported mass" is not necessarily preserved. It occurs, for instance, when M simply encodes the topological structure of the network with m_{ij} being zero or one, depending on whether a certain link exists (i.e., when M represents the *adjacency matrix* of the graph). The evolution (9.1), together with measure $\mu_0(\cdot)$, which we assume positive on $\mathcal{X}$, i.e.,

$$\mu_0(x) > 0 \text{ for all } x \in \mathcal{X}, \tag{9.3}$$

induces a measure $\mathfrak{M}$ on $\mathcal{FP}_0^N$ as follows. It assigns to a path $x = (x_0, x_1, \ldots, x_N) \in \mathcal{FP}_0^N$ the value

$$\mathfrak{M}(x_0, x_1, \ldots, x_N) = \mu_0(x_0) m_{x_0 x_1} \cdots m_{x_{N-1} x_N} \tag{9.4}$$

and gives rise to a flow of *one-time marginals*

$$\mu_t(x_t) = \sum_{x_{\ell \neq t}} \mathfrak{M}(x_0, x_1, \ldots, x_N), \quad t \in \mathcal{T}.$$

Definition 9.1. *We denote by $\mathcal{P}(\nu_0, \nu_N)$ the family of probability distributions on $\mathcal{FP}_0^N$ having the prescribed marginals $\nu_0(\cdot)$ and $\nu_N(\cdot)$.*

We seek a distribution in this set which is closest to the prior $\mathfrak{M}$ in *relative entropy* where, for P and Q measures on $\mathcal{X}^{N+1}$, the relative entropy (divergence, Kullback–Leibler index) $\mathbb{D}(P\|Q)$ is

$$\mathbb{D}(P\|Q) := \begin{cases} \sum_x P(x) \log \frac{P(x)}{Q(x)}, & \mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q), \\ +\infty, & \mathrm{Supp}(P) \not\subseteq \mathrm{Supp}(Q). \end{cases}$$

Here, by definition, $0 \cdot \log 0 = 0$. Naturally, while the value of $\mathbb{D}(P\|Q)$ may turn out to be negative due to a mismatch in the scaling (in the case $Q = \mathfrak{M}$ is not a probability measure), the relative entropy is always jointly convex. Thus, we are led to the *Schrödinger bridge problem*.

Problem 9.2. *Determine*

$$\mathfrak{M}^*[\nu_0, \nu_N] := \mathrm{argmin}\{\mathbb{D}(P\|\mathfrak{M}) \mid P \in \mathcal{P}(\nu_0, \nu_N)\}. \tag{9.5}$$

The following result may be proven in the usual way; cf. section 7.3 and [194, 63, 64].

Theorem 9.3. *Assume that the entries of the matrix product*

$$G := M(0)M(1) \cdots M(N-2)M(N-1)$$

are all positive. Then there exist nonnegative functions $\varphi(\cdot)$ and $\hat{\varphi}(\cdot)$ on $\mathcal{T} \times \mathcal{X}$ satisfying

$$\varphi(t, i) = \sum_j m_{ij}(t) \varphi(t+1, j), \tag{9.6a}$$

$$\hat{\varphi}(t+1, j) = \sum_i m_{ij}(t) \hat{\varphi}(t, i) \tag{9.6b}$$

for $t \in \{0, 1, \ldots, N-1\}$, along with the (nonlinear) boundary conditions

$$\varphi(0, x_0)\hat{\varphi}(0, x_0) = \nu_0(x_0), \tag{9.6c}$$

$$\varphi(N, x_N)\hat{\varphi}(N, x_N) = \nu_N(x_N) \tag{9.6d}$$

for $x_0, x_N \in \mathcal{X}$. *Moreover, the solution* $\mathfrak{M}^*[\nu_0, \nu_N]$ *to Problem* 9.2 *is unique and obtained by*

$$\mathfrak{M}^*(x_0, \dots, x_N) = \nu_0(x_0)\pi_{x_0 x_1}(0) \cdots \pi_{x_{N-1} x_N}(N-1),$$

where the one-step transition probabilities

$$\pi_{ij}(t) := m_{ij}(t) \frac{\varphi(t+1, j)}{\varphi(t, i)} \tag{9.7}$$

are well defined.

As usual, factors φ and $\hat{\varphi}$ are unique up to multiplication of φ by a positive constant and division of $\hat{\varphi}$ by the same constant. Let $\varphi(t)$ and $\hat{\varphi}(t)$ denote the column vectors with components $\varphi(t, i)$ and $\hat{\varphi}(t, i)$, respectively, with $i \in \mathcal{X}$. In matricial form, (9.6a), (9.6b), and (9.7) read

$$\varphi(t) = M(t)\varphi(t+1), \quad \hat{\varphi}(t+1) = M(t)^T \hat{\varphi}(t), \tag{9.8}$$

and

$$\Pi(t) := [\pi_{ij}(t)] = \operatorname{diag}(\varphi(t))^{-1} M(t) \operatorname{diag}(\varphi(t+1)). \tag{9.9}$$

We see that the scheduling of the transport plan amounts to modifying the prior transition mechanism. Therefore, this brings us to an alternative interpretation of the Schrödinger problem, as a special Markov decision process problem [200, 26]. This can be accomplished, once more, through a generalized multiplicative functional transformation (4.16) even when the prior is not a probability measure.

9.2. Invariance of Most Probable Paths. In [77, section 5], Dai Pra established an interesting path space property of the Schrödinger bridge for diffusion processes, namely, that the "most probable path" [89, 233] of the prior and the solution are the same. Loosely speaking, a most probable path is similar to a *mode* for the path space measure P. More precisely, if both the drift $b(\cdot, \cdot)$ and the diffusion coefficient $\sigma(\cdot, \cdot)$ of the Markov diffusion process

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$$

are smooth and bounded, with $\sigma(t, x)\sigma(t, x)' > \eta I$, $\eta > 0$, and $x(t)$ is a path of class C^2, then there exists an asymptotic estimate of the probability P of a small tube around $x(t)$ of radius ϵ. It follows from this estimate that the most probable path is the minimizer in a deterministic calculus of variations problem where the Lagrangian is an *Onsager–Machlup functional*; see [128, p. 532] for the full story.[29]

The concept of most probable path is, of course, much simpler in our discrete setting. We now define this for general positive measures on paths.

Given a positive measure $\mathfrak{M}$ as in section 9.1 on the feasible paths of our graph $\mathbf{G}$, we say that $x = (x_0, \dots, x_N) \in \mathcal{FP}_0^N$ is of *maximal mass* if for all other feasible paths $y \in \mathcal{FP}_0^N$ we have $\mathfrak{M}(y) \le \mathfrak{M}(x)$. Likewise, we consider paths of *maximal mass* connecting particular nodes. It is apparent that paths of maximal mass always exist but are, in general, not unique. If $\mathfrak{M}$ is a probability measure, then the maximal mass

[29]The Onsager–Machlup functional was introduced in [187, 188] to develop a theory of fluctuations in equilibrium and nonequilibrium thermodynamics.

paths—most probable paths—are simply the modes of the distribution. We establish below that the maximal mass paths joining two given nodes under the solution of the Schrödinger bridge problem as in the previous section are the same as for the prior measure.

PROPOSITION 9.4. *Consider marginals ν_0 and ν_N in Problem* 9.2. *Assume that $\nu_0(x) > 0$ on all nodes $x \in \mathcal{X}$ and that the product $M(0)M(1)\cdots M(N-2)M(N-1)$ of transition matrices of the prior has all positive elements (cf. with M as in* (9.2)*). Let x_0 and x_N be any two nodes. Then, under the solution $\mathfrak{M}^*[\nu_0, \nu_N]$ of the Schrödinger bridge problem, the family of maximal mass paths joining x_0 and x_N in N steps is the same as under the prior measure $\mathfrak{M}$.*

Proof. Suppose path $y = (y_0 = x_0, y_1, \ldots, y_{N-1}, y_N = x_N)$ has maximal mass under the prior $\mathfrak{M}$. In view of (9.4) and (9.7) and assumption (9.3), we have

$$\begin{aligned}\mathfrak{M}^*[\nu_0, \nu_N](y) &= \nu_0(y_0)\pi_{y_0 y_1}(0)\cdots \pi_{y_{N-1}y_N}(N-1)\\ &= \frac{\nu_0(x_0)}{\mu_0(x_0)}\frac{\varphi(N, x_N)}{\varphi(0, x_0)}\mathfrak{M}(y_0, y_1, \ldots, y_N).\end{aligned}$$

Since the quantity

$$\frac{\nu_0(x_0)}{\mu_0(x_0)}\frac{\varphi(N, x_N)}{\varphi(0, x_0)}$$

is positive and does not depend on the particular path joining x_0 and x_N, the conclusion follows. ∎

The above calculation in fact establishes the following stronger result.

PROPOSITION 9.5. *Let x_0 and x_N be any two nodes in $\mathcal{X}$. Then, under the assumptions of Proposition* 9.4*, the measures $\mathfrak{M}$ and $\mathfrak{M}^*[\nu_0, \nu_N]$, restricted on the set of paths that begin at x_0 at time* 0 *and end at x_N at time N, are identical.*

9.3. Robust Network Routing. We now discuss yet another possible usage and interpretation of Schrödinger bridges, motivated by a concept of robustness of a transportation plan for a given network.

Network robustness is typically understood as the ability of a network to maintain connectivity, or to be insensitive (observables), in the event of node or link failures, or a disturbance. Maintaining connectivity may be seen as an *inverse percolation problem* [2, 19]. There exist several other notions of robustness such the one defined through a *fluctuation-dissipation relation* involving the *topological entropy rate*. This notion captures the behavior while relaxing back to equilibrium after a perturbation; see [10, 82, 220, 221]. Also, robust network design to meet demands in a given uncertainty set was studied in [186]. Finally, a concept of *resilience* of a routing policy in the presence of *cascading failures* was introduced and studied in [212]. Following the latter rationale, linking robustness to resilience, equilibrium considerations play no role. Indeed, this rationale motivates maximal utilization of all available options equally, as much as possible, so as to preempt failures.

Thus, we are led to formulate the following problem: Given times $t = 0, 1, \ldots, N$ and a directed, strongly connected graph, find a transportation plan from a *source node* to a *sink node*[30] such that most of the mass arrives by time N even in the

[30] We always have a loop on the sink node to allow part of the mass to arrive there earlier than the planned time horizon.

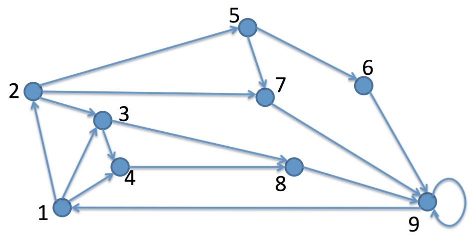


Fig. 6 *Transportation network.*

presence of failures (of course, more general initial and final distributions can also be similarly treated). For instance, consider the graph of Figure 6, with the total mass residing at node 1 at time $t = 0$, and a requirement to transport the total mass to node 9 in $N = 3$, or $N = 4$, time steps. A natural idea is to spread the mass as far as the topology of the graph permits, before reassembling the mass at the target end-point distribution (here at node 9). The property of "spreading" of distributions along interpolation between end-point marginals, whether in OMT or Schrödinger bridge problems, is often referred to as "lazy gas" [242, 243]. Thus, it is natural to consider transportation plans that share such a property via utilizing the framework of Schrödinger bridges!

However, the current task to transport provides no "prior" measure: It is simply a problem in transportation. Can we select a suitable measure? Perhaps select as prior a prespecified plan that no longer meets desired transportation requirements? Or modify a transportation plan by adding costs to mediate congestion? All of these directions are possible and can be profitably pursued as engineering problems. Yet, herein we motivate a different prior, one that *maximally* spreads mass utilizing available options. Indeed, such a prior must be a sort of *uniform distribution* on paths; but what exactly does that mean in this setting? Fortunately, such a notion already exists, as we find in the interesting paper [80]. Let us first recall a most famous result in linear algebra [124].

THEOREM 9.6 (Perron–Frobenius). *Let $A = (a_{ij})$ be an $n \times n$ matrix with nonnegative entries. Let $\lambda_A = \max\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|\}$ be its spectral radius. Suppose there exists N such that A^N has only positive entries. Then the following hold:*

(i) *$\lambda_A > 0$ is an eigenvalue of A.*
(ii) *λ_A is a simple eigenvalue.*
(iii) *There exists an eigenvector v corresponding to λ_A with strictly positive entries.*
(iv) *v is the only nonnegative eigenvector of A.*
(v) *Let $B = [b_{ij}]$ be an $n \times n$ matrix with nonnegative entries. If $a_{ij} \leq b_{ij}$ $\forall i, j \leq n$ and $A \neq B$, then $\lambda_A < \lambda_B$.*

Returning to graphs we discuss first the notion of topological entropy, namely, the rate by which the cardinality of the set of paths of length N increases as $N \to \infty$. To this end, we consider a strongly connected directed graph $\mathbf{G} = (\mathcal{X}, \mathcal{E})$ as before. The *topological entropy rate* is

$$H_{\mathbf{G}} = \limsup_{N\to\infty} [\log |\{\text{paths of length } N\}|/N].$$

If A denotes the *adjacency matrix* of the graph, it is easy to show[31] that

$$H_{\mathbf{G}} = \log(\lambda_A).$$

Since the graph is strongly connected, A^N has only positive entries. Let $\hat{\varphi}$ and φ be its left and right eigenvectors with positive entries corresponding to λ_A (see Theorem 9.6), so that

$$A^T\hat{\varphi} = \lambda_A\hat{\varphi}, \quad A\varphi = \lambda_A\varphi,$$

and select/scale them such that

$$\langle\hat{\varphi}, \varphi\rangle := \sum_i \hat{\varphi}_i\varphi_i = 1.$$

Then

$$\nu_{RB}(i) = \hat{\varphi}_i\varphi_i \tag{9.10}$$

defines a probability distribution on $\mathcal{X}$ which is invariant under the transition matrix

$$R = [r_{ij}], \quad r_{ij} = \frac{1}{\lambda_A}\frac{\varphi_j}{\varphi_i}a_{ij}, \tag{9.11}$$

that is,

$$R^T\nu_{RB} = \nu_{RB}.$$

The transition matrix R in (9.11), together with stationary measure ν_{RB} in (9.10), defines the *Ruelle–Bowen* (*Markovian*) *path measure*

$$\mathfrak{M}_{\mathrm{RB}}(x_0, x_1, \dots, x_N) := \nu_{RB}(x_0)r_{x_0x_1}\cdots r_{x_{N-1}x_N}.$$

Equation (9.11) brings up the unmistakable links to the structure of Schrödinger bridges that maximize entropy. But here a deeper fact is at play, in that the Ruelle–Bowen distribution [190, 210] represents a uniform distribution on paths, made precise by the following remarkable proposition.

PROPOSITION 9.7. *The measure* $\mathfrak{M}_{\mathrm{RB}}$ *assigns probability* $\lambda_A^{-t}\hat{\varphi}_i\varphi_j$ *to any path of length* t *from node* i *to node* j.

Proof. Starting from the stationary distribution (9.10), and in view of (9.11), the probability of a path ij is

$$\hat{\varphi}_i\varphi_i\left(\frac{1}{\lambda_A}\varphi_i^{-1}\varphi_j\right) = \frac{1}{\lambda_A}\hat{\varphi}_i\varphi_j,$$

assuming that node j is accessible from node i in one step. Likewise, if node k is accessible from j, the probability of the path ijk is

$$\hat{\varphi}_i\varphi_i\left(\frac{1}{\lambda_A}\varphi_i^{-1}\varphi_j\right)\left(\frac{1}{\lambda_A}\varphi_j^{-1}\varphi_k\right) = \frac{1}{\lambda_A^2}\hat{\varphi}_i\varphi_k,$$

independent of the intermediate state j, and so on. ∎

[31] This follows from the fact that the ijth entry of A^N (a positive integer) enumerates the number of distinct paths from vertex i to vertex j in N steps.

Shannon entropy of paths of length N grows like $N \log \lambda_A$. Thus, since the *entropy rate* of this particular distribution is $\log \lambda_A = H_{\mathbf{G}}$, it is indeed the maximum possible!

We note that the stationary measure ν_{RB} of this law was used in [80] in order to obtain a *centrality measure* (entropy ranking) similar to the Google Page ranking but more robust and discriminating. In what follows, instead, $\mathfrak{M}_{\mathrm{RB}}$ is a natural choice as a prior distribution in the Schrödinger bridge problem to achieve the spreading of the mass. Moreover, as noted earlier, the Ruelle–Bowen measure $\mathfrak{M}_{\mathrm{RB}}$ on paths may itself be seen as the solution of a Schrödinger bridge problem where the "prior" transition matrix is the adjacency matrix A and the two marginals are $\nu_0 = \nu_N = \nu_{RB}$; see [63, section 4].

Returning to the transportation problem, we seek

$$\mathfrak{M}^*[\delta_1, \delta_n] = \operatorname{argmin}\{\mathbb{D}(P\|\mathfrak{M}_{\mathrm{RB}}) \mid P \in \mathcal{P}(\delta_1, \delta_n)\}.$$

By Theorem 9.3, the solution is the Markovian evolution starting at $t = 0$ with the distribution δ_1 and with transition matrix

$$\Pi^*(t) = \operatorname{diag}(\varphi(t))^{-1} R \operatorname{diag}(\varphi(t+1)),$$

where

$$\varphi(t) = R\varphi(t+1),\ \hat{\varphi}(t+1) = R^T \hat{\varphi}(t),$$

with the boundary conditions

$$\varphi(0, x)\hat{\varphi}(0, x) = \delta_1(x),\ \varphi(N, x)\hat{\varphi}(N, x) = \delta_n(x) \quad \forall x \in \mathcal{X}.$$

Thus, the solution $\mathfrak{M}^*[\delta_1, \delta_n]$ is a bridge over $\mathfrak{M}_{\mathrm{RB}}$, which is itself a bridge. In other words, the solution is a *bridge over a bridge*.

We conclude with a remarkable *iterated bridge property* of the Schrödinger bridges (not to be confused with an iterated I-projection property [73, 74]). Suppose first that the prior is a probability distribution; this defines a *reciprocal class* [129, 160] of distributions. The solution to the Schrödinger bridge problem is in fact the unique Markovian evolution in the same reciprocal class as the prior (the same three times transition probabilities). In other words, if we take the solution as a prior for a new bridge problem, the reciprocal class stays the same. Notice that this is the case even when there is loss/creation of mass in the prior evolution; see [64, section IIA] for the details. Hence, the new distribution on paths $\mathfrak{M}^*[\delta_1, \delta_n]$ can be obtained by solving a *unique* bridge problem with prior transition the adjacency matrix A and marginals δ_1 and δ_n; see [63]. The solution may be computed through an iterative algorithm like the one described in section 8.2, where $G = A^N$. We illustrate the steps and rationale above with the following academic exercise.

EXAMPLE 9.8. *Consider Figure* 6 *and let* $\nu_0 = \delta_1$, $\nu_N = \delta_9$. *First take* $N = 3$. *The shortest path from node* 1 *to* 9 *is of length* 3 *and there are three such paths:* $1-2-7-9$, $1-3-8-9$, *and* $1-4-8-9$. *Using* $\mathfrak{M}_{\mathrm{RB}}$ *as the prior, we then obtain a transport plan with* equal probabilities *for all three paths. The evolution of the mass distribution is given by the rows of the following matrix, where row* i *represents the distribution on the nodes at time* $t = i, i = 0, 1, 2, 3$*:*

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For $N = 4$, the mass spreads even more before reassembling at node 9:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4/7 & 2/7 & 1/7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/7 & 1/7 & 2/7 & 0 & 1/7 & 2/7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/7 & 1/7 & 2/7 & 3/7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

9.4. Optimal Flows on Weighted Graphs. With the same notation as in section 9.1, we now suppose that to each edge ij is associated a length $l_{ij} \geq 0$. If $ij \notin \mathcal{E}$, we set $l_{ij} = +\infty$. The length may represent distance, cost of transport, cost of communication, inverse capacity of the link, and so on. As an example, suppose a relief organization, operating in an area where a natural disaster or an epidemic has occurred or in a war zone, needs to transport resources. At the initial time $t = 0$, there is a distribution $\nu_0(x)$ of available relief goods in sites $x \in \mathcal{X}$. Using the available road network, the goods must reach certain other locations after N units of time to be distributed according to a desired distribution $\nu_N(x)$. On the one hand, since the feasibility of the various possible routes is uncertain, it is desirable that the goods spread as much as the road network allows before reaching the target nodes. But at the same time, it is also important that shorter paths are used to keep the fuel consumption within the available budget. In such a scenario it is possible to repeat the construction of section 9.3, replacing the adjacency matrix A with a weighted adjacency matrix B

$$B = [b_{ij}] = [\exp\left(-l_{ij}\right)],$$

again assuming that B^N has positive entries representing cost, thereby allowing us to employ the Perron–Frobenius theorem.

The measure $\mathfrak{M}_L$ that replaces the Ruelle–Bowen measure is, of course, no longer "uniform." Given the development of section 7.2, we already know what we can expect. Rather than maximizing entropy, it is natural to seek a compromise between the latter goal and that of minimizing length/energy/cost. Indeed, let $\hat{\varphi}$ and φ be left and right eigenvectors with positive entries of the matrix B corresponding to the spectral radius λ_B of B, so that

$$B^T\hat{\varphi} = \lambda_B\hat{\varphi}, \quad B\varphi = \lambda_B\varphi.$$

Suppose once again that $\hat{\varphi}$ and φ are chosen so that $\langle\hat{\varphi}, \varphi\rangle = \sum_i \hat{\varphi}_i\varphi_i = 1$. Then μ_L given by

$$\mu_L(i) = \hat{\varphi}_i\varphi_i \tag{9.12}$$

is a probability distribution which is invariant for the transition matrix

$$R_L = \lambda_B^{-1}\operatorname{diag}(\hat{\varphi}_i\varphi_i)^{-1}B\operatorname{diag}(\hat{\varphi}_i\varphi_i), \tag{9.13}$$

namely,

$$R_L^T\mu_L = \mu_L. \tag{9.14}$$

As expected, the corresponding path measure $\mathfrak{M}_L$ is no longer uniform on paths of equal length joining two specific nodes. Indeed, the probability of the path $(i = x_0, x_1, \ldots, x_{t-1}, j = x_t)$ is

$$\lambda_B^{-t}\exp\left(-\sum_{k=0}^{t-1} l_{x_k x_{k+1}}\right)\hat{\varphi}_i\varphi_j.$$

It is, in other words, the minimum *free energy rate* distribution (topological pressure in thermodynamics) attaining the minimum value, $-\log \lambda_B$, and has therefore the form of a *Boltzmann distribution* (7.10); see [80, section IV] for details. Indeed, for a path $x = (x_0, \ldots, x_N) \in \mathcal{X}^{N+1}$, define the length of x to be

$$l(x) = \sum_{t=0}^{N-1} l_{x_t x_{t+1}},$$

and for any distribution P on $\mathcal{X}^{N+1}$, define the *average path length* to be

$$L(P) = \sum_{x \in \mathcal{X}^{N+1}} l(x) P(x). \tag{9.15}$$

This plays the same role as the internal energy in the state P of section 7.2, corresponding to the length $l(x)$ of x with the energy E_x. Clearly, $L(P)$ is finite if and only if P is supported on actual, existing paths of $\mathbf{G}$. The Boltzmann distribution (7.10) on X^{N+1} is then

$$p_B(x) = \mathfrak{M}_L(x) = Z(T)^{-1} \exp\left[-\frac{l(x)}{kT}\right] \text{ for } Z(T) = \sum_{x \in \mathcal{X}} \exp\left[-\frac{l(x)}{kT}\right]. \tag{9.16}$$

Note that the the support of the Boltzmann distribution $\mathrm{Supp}(p_B)$ is contained in $\mathcal{FP}_0^N$. By statement (v) in Theorem 9.6, we then have that $\log \lambda_A < \log \lambda_B$, namely, the topological entropy increases in a way that is consistent with intuition.

We can take $\mathfrak{M}_L$ as the prior distribution in a maximum entropy problem as in section 9.3, obtaining again through the solution $\mathfrak{M}_L^*[\delta_1, \delta_n]$ a robust-efficient transportation plan from node 1 to node n. We are now ready to prove a striking result that generalizes Proposition 9.7.

THEOREM 9.9. *$\mathfrak{M}_L^*[\delta_1, \delta_n](x)$ assigns equal probability to paths $x \in \mathcal{X}^{N+1}$ of equal cost. In particular, it assigns maximum and equal probability to minimum length paths.*

Proof. For a path $x = (x_0, x_1, \ldots, x_N)$, we have

$$\begin{aligned} \mathfrak{M}_L^*[\delta_1, \delta_n](x) &= \delta_1(x_0) \frac{\varphi_v(N, x_N)}{\varphi_v(0, x_0)} \prod_{t=0}^{N-1} b_{x_t x_{t+1}} \\ &= \delta_1(x_0) \frac{\varphi_v(N, x_N)}{\varphi_v(0, x_0)} \exp\left[-\sum_{t=0}^{N-1} l_{x_t x_{t+1}}\right]. \end{aligned} \tag{9.17}$$

Observe once more that $\delta_1(x_0) \frac{\varphi_v(N, x_N)}{\varphi_v(0, x_0)}$ does not depend on the particular path joining x_0 and x_N. Since $\sum_{t=1}^{N-1} l_{x_t x_{t+1}} = l(x)$ is the total length of the path, the conclusion now follows. ∎

Remark 9.10. In the discrete (OMT) problem [202] introduced in section 7.1, one first seeks to identify the shortest path(s) $(x_0, x_1^*, \ldots, x_{N-1}^*, x_N)$ from any starting node $x_0 \in \mathcal{X}$ to any ending node x_N,

$$l_{\min}(x_0 x_N) = \min_{x_1^*, \ldots, x_{N-1}^*} \sum_{t=0}^{N-1} l_{x_t^* x_{t+1}^*}. \tag{9.18}$$

This is a *combinatorial problem* but can also be cast as a linear program [20]. It is apparent that the computational complexity of such a problem becomes rapidly unbearable as the number of nodes n and the length of the path N increase. Having a solution to this first problem, the OMT problem can then be recast as the linear program in (7.1), where the cost of a path is its length. Alternatively, the OMT problem can be cast directly as a linear program in as many variables as there are edges [20]. The transport provided by Theorem 9.9, which readily generalizes to any two marginals ν_0 and ν_N, provides an attractive alternative to the OMT approach: Minimum length paths all have maximum probability, but some of the mass is also transported on alternative paths, thereby ensuring a certain amount of robustness of the transportation plan. Also notice that Theorem 9.9 provides an alternative paradigm to find the minimum length paths through simulation!

We conclude this subsection with a few observations on the role of the temperature parameter, referring the reader to [64, section V] for the proofs.

Remark 9.11. Consider the solution $\mathfrak{M}^*_{L,T}[\delta_{x_0}, \delta_{x_N}] =: \mathfrak{M}^*_T$ to the maximum entropy problem

$$\mathfrak{M}^*_{L,T}[\delta_1, \delta_n] = \mathrm{argmin}\{\mathbb{D}(P\|p_B(x;T)) \mid P \in \mathcal{P}(\delta_1, \delta_n)\}$$

with prior (9.16), where we have emphasized the dependence on the parameter T. Let $l_{\min}(x_0 x_N)$ be as in (9.18).

(i) For $T \searrow 0$, $\mathfrak{M}^*_T$ tends to concentrate itself on the set of feasible, minimum length paths joining x_0 and x_N in N steps. Namely, if $y = (y_0 = x_0, y_1, \ldots, y_{N-1}, y_N = x_N)$ is such that $l(y) > l_{\min}(x_0 x_N)$, then $\mathfrak{M}^*_T(y) \searrow 0$ as $T \searrow 0$.

(ii) For $T \nearrow +\infty$, $\mathfrak{M}^*_T$ tends to the uniform distribution on all feasible paths joining x_0 and x_N in N steps.

We notice that, as in the diffusion case [168, 169, 170, 157, 156, 52, 54, 46], when the "heat bath" temperature T is close to 0, the solution of the Schrödinger bridge problem is close to the solution of the discrete OMT problem. Since for the former an efficient iterative algorithm is available (8.6), we see that also in this discrete setting the Schrödinger bridge problem provides a valuable computational approach to solving OMT problems. We illustrate this, as well as Theorem 9.9 and the invariance of the most probable paths, in a simple example.

EXAMPLE 9.12. *Consider again Figure* 6 *and let* $\nu_0 = \delta_1$ *and* $\nu_N = \delta_9$. *Let the time horizon for the transport be* $N = 3$ *or* $N = 4$. *We first set the lengths of all edges equal to* 1 *except* $l_{99} = 0$. *The shortest path from node* 1 *to* 9 *is of length* 3, *and there are three such paths:* $1-2-7-9$, $1-3-8-9$, *and* $1-4-8-9$. *If we want to transport the mass with a minimum number of steps, we may end up using one of these three paths. We use the results of section* 9.3 *to compute a robust routing policy. Since all three feasible paths have equal length, we obtain a transport plan with equal probabilities using all three paths, regardless of the choice of temperature* T. *The evolution of mass distribution is given by*

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where the four rows of the matrix show the mass distribution at time step $t = 0, 1, 2, 3$, *respectively, while columns correspond to vertices. As we can see, the mass spreads* *out first and then goes to node 9. When we allow for more steps $N = 4$, we get, for "temperature" $T = 1$,*

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4705 & 0.3059 & 0.2236 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0823 & 0.0823 & 0.1645 & 0 & 0.2236 & 0.4473 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0823 & 0.0823 & 0.1645 & 0.6709 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

There are seven feasible paths of length 4: $1-2-7-9-9$, $1-3-8-9-9$, $1-4-8-9-9$, $1-2-5-6-9$, $1-2-5-7-9$, $1-3-4-8-9$, and $1-2-3-8-9$. The amounts of mass traveling along these paths are

$$0.2236, \quad 0.2236, \quad 0.2236, \quad 0.0823, \quad 0.0823, \quad 0.0823, \quad 0.0823,$$

respectively. The first three are the most probable paths. This is consistent with Proposition 9.4, since they are the paths with minimum length. If we change the temperature T, the flow changes. The set of most probable paths, however, remains invariant. In particular, when $T = 0.1$, the flow concentrates on the most probable set (effecting OMT-like transport):

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3334 & 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.3334 & 0.6666 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now we change the graph by setting the length of edge $(7, 9)$ as 2, that is, $l_{79} = 2$. When $N = 3$ steps are allowed to transport a unit mass from node 1 to node 9, the evolution of mass distribution for the optimal transport plan, for $T = 1$, is given by

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1554 & 0.4223 & 0.4223 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1554 & 0.8446 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The mass is transported through paths $1-2-7-9$, $1-3-8-9$, and $1-4-8-9$, but unlike the first case, the transport plan doesn't equalize probability for these three paths. Since the length of the edge $(7, 9)$ is larger, the probability that the mass takes this path becomes smaller. The plan does, however, assign equal probability to the two paths $1-3-8-9$ and $1-4-8-9$ with minimum length; that is, these are the most probable paths. The evolutions of mass for $T = 0.1$ and $T = 100$ are

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3311 & 0.3344 & 0.3344 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.3311 & 0.6689 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

respectively. We observe that when $T = 100$*, the flow assigns almost equal mass to the three available paths, while when* $T = 0.1$ *(OMT-like transport), the flow concentrates on the most probable paths* $1-3-8-9$ *and* $1-4-8-9$*. This is clearly a consequence of the properties seen in Remark* 9.11.

Remark 9.13. As in the continuous case, it is possible to transform the dynamic problems considered in this section into static problems using a decomposition for the relative entropy similar to (4.6). Indeed, let P and Q be two probability distributions on $\mathcal{X}^{N+1}$. For $x = (x_0, x_1, \dots, x_N) \in \mathcal{X}^{N+1}$, consider the multiplicative decomposition

$$P(x) = P_{x_0,x_N}(x) p_{0N}(x_0, x_N),$$

where

$$P_{\bar{x}_0,\bar{x}_N}(x) = P(x|x_0 = \bar{x}_0, x_n = \bar{x}_N)$$

and we have assumed that the joint initial-final distribution p_{0N} is everywhere positive on $\mathcal{X} \times \mathcal{X}$, and similarly for Q. We get

$$\begin{aligned}\mathbb{D}(P\|Q) = &\sum_{x_0 x_N} p_{0N}(x_0, x_N) \log \frac{p_{0N}(x_0, x_N)}{q_{0N}(x_0, x_N)} \\ &+ \sum_{x \in \mathcal{X}^{N+1}} P_{x_0,x_N}(x) \log \frac{P_{x_0,x_N}(x)}{Q_{x_0,x_N}(x)} p_{0N}(x_0, x_N).\end{aligned}$$

This is the sum of two nonnegative quantities. The second becomes zero if and only if $P_{x_0,x_N}(x) = Q_{x_0,x_N}(x)$ for all $x \in \mathcal{X}^{N+1}$.

Thus, expanding on the above remark, the problem

$$\min\{\mathbb{D}(P\|p_B(x;T)) \mid P \in \mathcal{P}(\nu_0, \nu_N)\}, \tag{9.19}$$

for instance, can be reduced to the following one.

PROBLEM 9.14. *Minimize*

$$J(p_{0N}) := \mathbb{D}(p_{0N}\|p_{B;0N}) \tag{9.20}$$

over

$$\left\{ p_{0N} \textit{ probability distribution on } \mathcal{X} \times \mathcal{X} : \sum_{x_N} p_{0N}(\cdot, x_N) = \nu_0(\cdot), \sum_{x_0} p_{0N}(x_0 \cdot) = \nu_N(\cdot) \right\}. \tag{9.21}$$

If p_{0N}^* solves the above problem, then

$$P^*(x) = p_{B;x_0,x_N}(x) p_{0N}^*(x_0, x_N) \tag{9.22}$$

solves problem (9.19). As in the continuous setting, the solution lies in the same reciprocal class of the prior. Observe, however, that contrary to (9.7), the solution in the form (9.22) does not yield immediate information on the new transition probabilities and on what paths the optimal mass flow selects (cf. Remark 4.10 in the continuous setting). It is therefore less suited for many network routing applications.

10. Closing Comments.

We have reviewed some of the essential theoretical features of the Kantorovich relaxed formulation of OMT with quadratic cost and of the theory of the Schrödinger bridge problem. We have tried to do this from a somewhat unusual angle, namely, that of stochastic control. We have explained that such a viewpoint opens the way to natural generalizations and important applications, particularly in physics (such as cooling), aerospace engineering (such as guidance for spacecrafts), robotics (such as controlling swarms), and many other areas. Both problems, OMT and the Schrödinger bridge problem, in their Eulerian formulations, lead to stochastic control steering problems for probability distributions. In this respect, we have highlighted the nonequivalence between Schrödinger's original problem and the related problems of steering for one-time densities; cf. Remark 4.5. We have also attempted to clarify the relationship between Yasue's action, Carlen's problem, and the fluid-dynamic Problem 4.6 introduced in [54] that involves a Fisher information functional; see Remark 4.7.

In order to keep the paper within a reasonable length, we have avoided touching on a large number of related topics such as noncommutative OMT, gradient flows on Wasserstein spaces, geometry of displacement and entropic interpolation, multi-marginal transport, functional inequalities, mismatched transport, barycenters, and connections to mean-field games, information theory, stochastic mechanics, and differential games, to name but a few.

In the discrete setting, we have shown in some detail the connection between the relaxed OMT, Schrödinger bridges, and the statistical mechanical free energy. It turns out (section 7) that the regularized OMT is just a discrete Schrödinger bridge problem equivalent to a constrained minimization of the free energy for distributions on path space. The latter may be viewed as a constrained *Gibbs variational principle* recovering Schrödinger's original large deviations motivation. Moreover, the IPF-Sinkhorn algorithm is just a discrete counterpart of Fortet's algorithm (section 8), for which Fortet had already proven convergence under rather general assumptions in the much more challenging continuous case in 1940. In section 8, we have also proven convergence of the algorithm as a consequence of convergence in a projective metric of rays in the positive orthant of $\mathbb{R}^n$.

This paper also has the ambition to act as a *navigation chart* for the topics it discusses. We felt that this was needed since, on the one hand, probabilists considered the theory of Schrödinger bridges to be pretty much complete by the early 1990s (see Wakolbinger's survey [247] that was published in 1992), but as we argued at the beginning of section 5, this is far from being true. On the other hand, some very fine analysts who contributed to the fantastic development of OMT of the past twenty or so years have very little interest in the probabilistic motivation and the statistical physics underlying the Schrödinger bridge theory in spite of the fine work of Mikami, Thieullen, and Léonard [168, 169, 170, 157, 156]. Moreover, many first-class scientists working in the field have a more computational background and interests, having been driven to these problems by the effectiveness of the *earth mover distance* in several modern applications such as those (we quote from [198]) "in imaging sciences (such as color or texture processing), computer vision and graphics (for shape manipulation) or machine learning (for regression, classification and density fitting)." Finally, control engineers have been interested since 1985 in a special bridge problem on an infinite horizon problem, called covariance control, starting from the seminal work of Skelton and collaborators [125, 126, 222, 115, 254]. The field was considered exhausted more than twenty years ago, but this turned out to be far from the truth; it is currently experiencing another phase of fast development.

On top of all of these reasons for incomprehension and difficult communication, there is, of course, a Babelian confusion of tongues as scientists working in this field have had a plethora of backgrounds in various areas such as pure and applied mathematics, statistical physics, statistics, computer graphics, control, mechanical and aerospace engineering, numerical analysis, machine learning, etc. Nevertheless, considering the spectacular flourishing of OMT, Schrödinger bridges, and relative applications, one feels tempted to end the paper with a Chinese quote: "Great is the confusion under the sky. The situation is therefore excellent."

REFERENCES

[1] R. Aebi and M. Nagasawa, *Large deviations and the propagation of chaos for Schrödinger processes*, Probab. Theory Related Fields, 94 (1992), pp. 53–68. (Cited on pp. 269, 272, 276)

[2] R. Albert, H. Jeong, and A.-L. Barabási, *Error and attack tolerance of complex networks*, Nature, 406 (2000), pp. 378–382. (Cited on p. 294)

[3] D. Alpago, Y. Chen, T. T. Georgiou, and M. Pavon, *On optimal steering for a non-Markovian Gaussian process*, in 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 2556–2561. (Cited on p. 273)

[4] J. Altschuler, J. Weed, and P. Rigollet, *Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration*, in Proc. 31st International Conference on Neural Information Processing Systems, Curran Associates, 2017, pp. 1964–1974. (Cited on pp. 251, 286)

[5] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures in Math. ETH Zürich, Birkhäuser Verlag, Basel, 2008. (Cited on pp. 255, 256, 257, 258)

[6] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, in Modeling and Optimisation of Flows on Networks, Lecture Notes in Math. 2062, B. Piccoli and M. Rascle, eds., Springer, 2013. (Cited on p. 258)

[7] S. Angenent, S. Haker, and A. Tannenbaum, *Minimizing flows for the Monge–Kantorovich problem*, SIAM J. Math. Anal., 35 (2003), pp. 61–97, https://doi.org/10.1137/S0036141002410927. (Cited on pp. 257, 266)

[8] P. Appel, *Mémoires sur les déblais et les rémblais*, Mém. Acad. Sci. Paris, 29 (1887). (Cited on p. 254)

[9] M. Arjovsky, S. Chintala, and L. Bottou, *Wasserstein generative adversarial networks*, in Proc. 34th International Conference on Machine Learning, 2017, pp. 214–223. (Cited on p. 256)

[10] L. Arnold, V. M. Gundlach, and L. Demetrius, *Evolutionary formalism for products of positive random matrices*, Ann. Appl. Probab., 4 (1994), pp. 859–901. (Cited on p. 294)

[11] L. Bachelier, *Théorie de la spéculation*, Ann. Sci. École Norm. Sup., 3 (1900), pp. 21–86. (Cited on p. 278)

[12] G. Baggio, A. Ferrante, and R. Sepulchre, *Finslerian Metrics in the Cone of Spectral Densities*, preprint, https://arxiv.org/abs/1708.02818, 2017. (Cited on pp. 287, 289)

[13] E. Bakolas, *Optimal covariance control for discrete-time stochastic linear systems subject to constraints*, in 2016 IEEE 55th Conference on Decision and Control (CDC), 2016, pp. 1153–1158. (Cited on pp. 273, 291)

[14] E. Bakolas, *Finite-horizon covariance control for discrete-time stochastic linear systems subject to input constraints*, Automatica, 91 (2018), pp. 61–68. (Cited on pp. 273, 274, 291)

[15] E. Bakolas, *Covariance control for discrete-time stochastic linear systems with incomplete state information*, in 2017 American Control Conference (ACC), 2017, pp. 432–437. (Cited on p. 273)

[16] E. Bakolas, *Finite-horizon separation-based covariance control for discrete-time stochastic linear systems*, in 2018 IEEE 57th Conference on Decision and Control (CDC), 2018, pp. 3299–3304. (Cited on p. 273)

[17] E. Bakolas, *Dynamic output feedback control of the Liouville equation for discrete-time SISO linear systems*, IEEE Trans. Automat. Control, 64 (2019), pp. 4268–4275. (Cited on pp. 273, 291)

[18] R. Bapat, *D_1AD_2 theorems for multidimensional matrices*, J. Linear Algebra Appl., 48 (1982), pp. 437–442. (Cited on p. 251)

[19] A.-L. Barabàsi, *Network Science*, Cambridge University Press, 2014. (Cited on p. 294)

[20] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*, John Wiley & Sons, 2011. (Cited on pp. 299, 300)

[21] A. Beghi, *On the relative entropy of discrete-time Markov processes with given end-point densities*, IEEE Trans. Inform. Theory, 42 (1996), pp. 1529–1535. (Cited on p. 291)

[22] J.-D Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393. (Cited on pp. 257, 261, 266, 270, 271, 276)

[23] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, *Iterative Bregman projections for regularized transportation problems*, SIAM J. Sci. Comput., 37 (2015), pp. A1111–A1138, https://doi.org/10.1137/141000439. (Cited on pp. 251, 256)

[24] S. Bernstein, *Sur les liaisons entre le grandeurs aleatoires*, in Proc. Int. Congr. Math., Zürich, Switzerland, 1932, pp. 288–309.

[25] E. Bernton, J. Heng, A. Doucet, and P. E. Jacob, *Schrödinger Bridge Samplers*, preprint, https://arxiv.org/abs/1912.13170, 2019. (Cited on pp. 251, 266, 285)

[26] D. Bertsekas, *Dynamic Programming and Optimal Control*, Vol. 2, Athena Scientific, 1995. (Cited on pp. 251, 268)

[27] A. Beurling, *An automorphism of product measures*, Ann. of Math. (2), 72 (1960), pp. 189–200. (Cited on pp. 258, 293)

[28] G. Birkhoff, *Extensions of Jentzch's theorem*, Trans. Amer. Math. Soc., 85 (1957), pp. 219–227. (Cited on pp. 251, 267, 268, 276)

[29] G. Birkhoff, *Uniformly semi-primitive multiplicative processes*, Trans. Amer. Math. Soc., 104 (1962), pp. 37–51. (Cited on pp. 251, 287, 289)

[30] A. Blaquière, *Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control*, Dynam. Control, 2 (1992), pp. 235–253. (Cited on p. 289)

[31] D. Bohm, *A suggested interpretation of the quantum theory in terms of "hidden" variables.* I, Phys. Rev., 85 (1952), pp. 166–179. (Cited on p. 269)

[32] D. Bohm and B. J. Hiley, *Non-locality and locality in the stochastic interpretation of quantum mechanics*, Phys. Rep., 172 (1989), pp. 93–122. (Cited on p. 271)

[33] D. Bohm and J. P. Vigier, *Model of the causal interpretation of quantum theory in terms of a fluid with irregular fluctuations*, Phys. Rev., 96 (1954), pp. 208–216. (Cited on p. 271) (Cited on p. 271)

[34] L. Boltzmann, *Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. den Sätzen über das Wärmegleichgewicht*, Wiener Berichte, 76 (1877), pp. 373–435; reprinted in Wissenschaftliche Abhandlungen, Vol. 2, J. A. Barth, 1909, pp. 164–223. (Cited on pp. 251, 263, 266, 268, 277, 282)

[35] M. Bonaldi, L. Conti, P. De Gregorio et al., *Nonequilibrium steady-state fluctuations in actively cooled resonators*, Phys. Rev. Lett., 103 (2009), art. 010601. (Cited on p. 279)

[36] S. Bonnabel, A. Astolfi, and R. Sepulchre, *Contraction and observer design on cones*, in Proc. Decision and Control and European Control Conference (CDC-ECC), IEEE, 2011, pp. 12–15. (Cited on p. 287)

[37] U. Boscain, R. A. Chertovskih, J. P. Gauthier, and A. O. Remizov, *Hypoelliptic diffusion and human vision: A semidiscrete new twist*, SIAM J. Imaging Sci., 7 (2014), pp. 669–695, https://doi.org/10.1137/130924731. (Cited on p. 274)

[38] Y. Braiman, J. Barhen, and V. Protopopescu, *Control of friction at the nanoscale*, Phys. Rev. Lett., 90 (2003), art. 094301. (Cited on p. 279)

[39] J. P. Burg, *Maximum entropy spectral analysis*, in Proc. 37th Meet. Society of Exploration Geophysicists, 1967; reprinted in Modern Spectrum Analysis, D. G. Childers, ed., IEEE Press, 1978, pp. 34–41. (Cited on p. 266)

[40] J. Burg, D. Luenberger, and D. Wenger, *Estimation of structured covariance matrices*, Proc. IEEE, 70 (1982), pp. 963–974. (Cited on p. 266)

[41] P. Bushell, *On the projective contraction ratio for positive linear mappings*, J. London Math. Soc., 2 (1973), pp. 256–258. (Cited on p. 287)

[42] P. Bushell, *Hilbert's metric and positive contraction mappings in a Banach space*, Arch. Rational Mech. Anal., 52 (1973), pp. 330–338. (Cited on pp. 251, 287, 288, 289, 290)

[43] K. F. Caluya and A. Halder, *Wasserstein Proximal Algorithms for the Schrödinger Bridge Problem: Density Control with Nonlinear Drift*, preprint, https://arxiv.org/abs/1912.01244, 2019. (Cited on p. 274)

[44] K. F. Caluya and A. Halder, *Reflected Schrödinger Bridge: Density Control with Path Constraints*, preprint, https://arxiv.org/abs/2003.13895, 2020. (Cited on p. 274)

[45] E. Carlen, *Stochastic mechanics: A look back and a look ahead*, in Diffusion, Quantum Theory and Radically Elementary Mathematics, W. G. Faris, ed., Math. Notes 47, Princeton University Press, 2006, pp. 117–139. (Cited on p. 271)

[46] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM J. Math. Anal., 49 (2017), pp. 1385–1418, https://doi.org/10.1137/15M1050264. (Cited on pp. 257, 266, 300)

[47] S. Chandrasekhar, *Stochastic problems in physics and astronomy*, Rev. Modern Phys., 15 (1943), pp. 1–89. (Cited on p. 278)

[48] J. T. Chang and David Pollard, *Conditioning as disintegration*, Statist. Neerlandica, 51 (1997), pp. 287–317. (Cited on p. 265)

[49] P. Chaudhari, A. Oberman, S. Osher, S. Soatto, and G. Carlier, *Deep relaxation: Partial differential equations for optimizing deep neural networks*, Res. Math. Sci., 5 (2018), art. 30. (Cited on p. 279)

[50] Y. Chen, T. T. Georgiou, and M. Pavon, *Optimal steering of a linear stochastic system to a final probability distribution, Part* I, IEEE Trans. Automat. Control, 61 (2016), pp. 1158–1169. (Cited on pp. 257, 268, 273, 274)

[51] Y. Chen, T. T. Georgiou, and M. Pavon, *Optimal steering of inertial particles diffusing anisotropically with losses*, in Proc. 2015 American Control Conference (ACC), 2015, pp. 1252–1257. (Cited on pp. 257, 272, 273)

[52] Y. Chen, T. T. Georgiou, and M. Pavon, *Optimal steering of a linear stochastic system to a final probability distribution, Part* II, IEEE Trans. Automat. Control, 61 (2016), pp. 1170–1180. (Cited on pp. 257, 271, 273, 274, 300)

[53] Y. Chen, T. T. Georgiou, and M. Pavon, *Fast cooling for a system of stochastic oscillators*, J. Math. Phys., 56 (2015), art. 113302. (Cited on pp. 257, 278, 279)

[54] Y. Chen, T. T. Georgiou, and M. Pavon, *On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint*, J. Optim. Theory Appl., 169 (2016), pp. 671–691. (Cited on pp. 257, 262, 270, 271, 300, 303)

[55] Y. Chen, T. T. Georgiou, and M. Pavon, *Optimal transport over a linear dynamical system*, IEEE Trans. Automat. Control, 62 (2017), pp. 2137–2152. (Cited on pp. 257, 262, 273)

[56] Y. Chen, T. T. Georgiou, and M. Pavon, *Optimal control of the state statistics for a linear stochastic system*, in Proc. 2015 54th IEEE Conference on Decision and Control (CDC), 2015, pp. 6508–6515. (Cited on pp. 257, 273)

[57] Y. Chen, T. T. Georgiou, and M. Pavon, *Steering state statistics with output feedback*, in Proc. 2015 54th IEEE Conference on Decision and Control (CDC), 2015, pp. 6502–6507. (Cited on pp. 257, 273)

[58] Y. Chen, T. T. Georgiou, and M. Pavon, *Entropic and displacement interpolation: A computational approach using the Hilbert metric*, SIAM J. Appl. Math., 76 (2016), pp. 2375–2396, https://doi.org/10.1137/16M1061382. (Cited on pp. 257, 268, 287, 289)

[59] Y. Chen, T. T. Georgiou, and M. Pavon, *Stochastic control, entropic interpolation and gradient flows on Wasserstein product spaces*, in Proc. 22nd International Symposium on the Mathematical Theory of Networks and Systems, 2016, http://hdl.handle.net/11299/178610. (Cited on p. 257)

[60] Y. Chen, T. T. Georgiou, and M. Pavon, *Optimal steering of a linear stochastic system to a final probability distribution, Part* III, IEEE Trans. Automat. Control, 63 (2018), pp. 3112–3118. (Cited on p. 273)

[61] Y. Chen, T. T. Georgiou, and M. Pavon, *Optimal transport in systems and control*, Ann. Rev. Control Robotics Autonomous Syst., 4 (2021), to appear. (Cited on p. 273)

[62] Y. Chen, T. T. Georgiou, and M. Pavon, *Controlling uncertainty: Schrödinger's inference method and the optimal steering of probability distributions*, IEEE Control Syst. Mag., to appear. (Cited on p. 263)

[63] Y. Chen, T. T. Georgiou, M. Pavon, and A. Tannenbaum, *Robust transport over networks*, IEEE Trans. Automat. Control, 62 (2017), pp. 4675–4682. (Cited on pp. 292, 297)

[64] Y. Chen, T. T. Georgiou, M. Pavon, and A. Tannenbaum, *Efficient-robust routing for single commodity network flows*, IEEE Trans. Automat. Control, 63 (2018), pp. 2287–2294. (Cited on pp. 257, 292, 297, 300)

[65] Y. Chen, T. T. Georgiou, M. Pavon, and A. Tannenbaum, *Relaxed Schrödinger bridges and robust network routing*, IEEE Trans. Control Network Syst., 7 (2020), pp. 923–931. (Cited on p. 257)

[66] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Scaling algorithms for unbalanced optimal transport problems*, Math. Comp., 87 (2018), pp. 2563–2609. (Cited on pp. 251, 289)

[67] V. Ciccone, Y. Chen, T. T. Georgiou, and M. Pavon, *Regularized transport between singular covariance matrices*, IEEE Trans. Automat. Control, 2020, https://doi.org/10.1109/TAC.2020.3017714. (Cited on p. 273)

[68] C. Clason, D. A. Lorenz, H. Mahler, and B. Wirth, *Entropic regularization of continuous optimal transport problems*, J. Math. Anal. Appl., 494 (2021), art. 124432. (Cited on pp. 267, 268)

[69] G. Conforti, *A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost*, Probab. Theory Related Fields, 174 (2019), pp. 1–47. (Cited on p. 258)

[70] G. Conforti and M. Pavon, *Extremal curves in Wasserstein space*, in Proc. Geometric Science of Information: Third International Conference, GSI 2017, Paris, 2017, F. Nielsen and F. Barbaresco, eds., Springer-Verlag, 2017, pp. 92–99. (Cited on p. 258)

[71] G. Conforti and M. Pavon, *Extremal flows in Wasserstein space*, J. Math. Phys., 59 (2018), art. 063502. (Cited on pp. 258, 271)

[72] G. Conforti and M. Von Renesse, *Couplings, gradient estimates and logarithmic Sobolev inequality for Langevin bridges*, Probab. Theory Related Fields, 172 (2018), pp. 493–524. (Cited on p. 258)

[73] I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probab., 3 (1975), pp. 146–158. (Cited on pp. 266, 286, 297)

[74] I. Csiszár, *Sanov property, generalized I-projections, and a conditional limit theorem*, Ann. Probab., 12 (1984), pp. 768–793. (Cited on pp. 266, 297)

[75] I. Csiszár, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Ann. Statist., 19 (1991), pp. 2032–2066. (Cited on p. 266)

[76] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, in Advances in Neural Information Processing Systems, 2013, pp. 2292–2300. (Cited on pp. 251, 281, 286)

[77] P. Dai Pra, *A stochastic control approach to reciprocal diffusion processes*, Appl. Math. Optim., 23 (1991), pp. 313–329. (Cited on pp. 269, 293)

[78] P. Dai Pra and M. Pavon, *On the Markov processes of Schroedinger, the Feynman-Kac formula and stochastic control*, in Realization and Modeling in System Theory (Proc. 1989 International Symposium on the Mathematical Theory of Networks and Systems), M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, eds., Birkhäuser, Boston, 1990, pp. 497–504. (Cited on pp. 269, 273)

[79] D. Dawson, L. Gorostiza, and A. Wakolbinger, *Schrödinger processes and large deviations*, J. Math. Phys., 31 (1990), pp. 2385–2388. (Cited on pp. 269, 272)

[80] J.-C. Delvenne and A.-S. Libert, *Centrality measures and thermodynamic formalism for complex networks*, Phys. Rev., 83 (2011), art. 046117. (Cited on pp. 295, 297, 299)

[81] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston, 1993. (Cited on pp. 262, 264)

[82] L. Demetrius and T. Manke, *Robustness and network evolution: An entropic principle*, Phys. A, 346 (2005), pp. 682–696. (Cited on p. 294)

[83] W. E. Deming and F. F. Stephan, *On a least squares adjustment of a sampled frequency table when the expected marginal totals are known*, Ann. Math. Statist., 11 (1940), pp. 427–444. (Cited on pp. 251, 286)

[84] A. P. Dempster, *Covariance selection*, Biometrics, 28 (1972), pp. 157–175. (Cited on p. 266)

[85] S. Di Marino and A. Gerolin, *An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm*, J. Sci. Comput., 85 (2020), pp. 1–28. (Cited on p. 251)

[86] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Oxford University Press, New York, 1988. (Cited on p. 279)

[87] J. L. Doob, *A Markov chain theorem*, in Probability and Statistics: The Harold Cramer Volume, John Wiley, 1959, pp. 50–57. (Cited on p. 268)

[88] J. L. Doob, *Classical Potential Theory and Its Probabilistic Counterpart*, Springer-Verlag, Berlin, 1984. (Cited on p. 268)

[89] D. Durr and A. Bach, *The Onsager-Machlup functional as Lagrangian for the most probable path of a diffusion process*, Comm. Math. Phys., 60 (1978), pp. 153–170. (Cited on p. 293)

[90] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer-Verlag, New York, 1985. (Cited on p. 264)

[91] M. Essid and M. Pavon, *Traversing the Schrödinger Bridge strait: Robert Fortet's marvelous proof redux*, J. Optim. Theory Appl., 181 (2019), pp. 23–60. (Cited on pp. 268, 286)

[92] L. C. Evans, *Partial differential equations and Monge-Kantorovich mass transfer*, in Current Developments in Mathematics, Int. Press, Boston, 1999, pp. 65–126. (Cited on pp. 255, 256)

[93] I. Favero and K. Karrai, *Optomechanics of deformable optical cavities*, Nat. Photon., 3 (2009), pp. 201–205. (Cited on p. 279)

[94] S. E. Fienberg, *An iterative procedure for estimation in contingency tables*, Ann. Math. Statist., 41 (1970), pp. 907–917. (Cited on pp. 251, 286)

[95] R. Filliger and M.-O. Hongler, *Relative entropy and efficiency measure for diffusion-mediated transport processes*, J. Phys. A, 38 (2005), pp. 1247–1255. (Cited on p. 279)

[96] R. Filliger, M.-O. Hongler, and L. Streit, *Connection between an exactly solvable stochastic optimal control problem and a nonlinear reaction-diffusion equation*, J. Optim. Theory Appl., 137 (2008), pp. 497–505. (Cited on p. 269)

[97] M. Fischer, *On the form of the large deviation rate function for the empirical measures of weakly interacting systems*, Bernoulli, 20 (2014), pp. 1765–1801. (Cited on p. 274)

[98] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, Berlin, 1975. (Cited on pp. 252, 258)

[99] W. H. Fleming, *Exit probabilities and optimal stochastic control*, Appl. Math. Optim., 4 (1978), pp. 329–346. (Cited on p. 269)

[100] W. H. Fleming, *Logarithmic transformation and stochastic control*, in Advances in Filtering and Optimal Stochastic Control, W. H. Fleming and L. G. Gorostiza, eds., Lect. Notes Control Inform. Sci. 42, Springer-Verlag, 1982, pp. 131–141. (Cited on p. 269)

[101] W. Fleming and S. Sheu, *Stochastic variational formula for fundamental solutions of parabolic PDE*, Appl. Math. Optim., 13 (1985), pp. 193–204. (Cited on p. 269)

[102] H. Föllmer, *Time reversal on Wiener space*, in Stochastic Processes: Mathematics and Physics, Lecture Notes in Math. 1158, Springer-Verlag, New York, 1986, pp. 119–129. (Cited on pp. 265, 271)

[103] H. Föllmer, *Random fields and diffusion processes*, in Ècole d'Ètè de Probabilitès de Saint-Flour XV-XVII, P. L. Hennequin, ed., Lecture Notes in Math. 1362, Springer-Verlag, New York, 1988, pp. 102–203. (Cited on pp. 251, 263, 265, 267)

[104] F. Forni and R. Sepulchre, *Differentially positive systems*, IEEE Trans. Automat. Control, 61 (2016), pp. 346–359. (Cited on pp. 261, 287)

[105] R. Fortet, *Résolution d'un système d'equations de M. Schrödinger*, Comptes Rendus, 206 (1938), pp. 721–723. (Cited on pp. 251, 285)

[106] R. Fortet, *Résolution d'un système d'equations de M. Schrödinger*, J. Math. Pures Appl., 19 (1940), pp. 83–105. (Cited on pp. 251, 267, 268, 284, 285)

[107] J. B. J. Fourier, *Solution d'une question particuliére du calcul des inégalités*, Nouveau Bull. Sci. Soc. Philomatique Paris (1826), pp. 99–100; reprinted in Oeuvres de Fourier, Tome II, J. G. Darboux, ed., Gauthier, 1890. (Cited on p. 253)

[108] J. Franklin and J. Lorenz, *On the scaling of multidimensional matrices*, Linear Algebra Appl., 114 (1989), pp. 717–735. (Cited on pp. 251, 287)

[109] A. Galichon, *Optimal Transport Methods in Economics*, Princeton University Press, 2016. (Cited on p. 256)

[110] W. Gangbo and R. McCann, *The geometry of optimal transportation*, Acta Math., 177 (1996), pp. 113–161. (Cited on p. 256)

[111] I. Gentil, C. Léonard, and L. Ripani, *About the analogy between optimal transport and minimal entropy*, Ann. Fac. Sci. Toulouse Math. (6), 26 (2017), pp. 569–600. (Cited on p. 270)

[112] T. T. Georgiou and M. Pavon, *Positive contraction mappings for classical and quantum Schrödinger systems*, J. Math. Phys., 56 (2015), art. 033301. (Cited on p. 287)

[113] N. Gigli, *Second order analysis on* $(\mathcal{P}_2(M), W_2)$, Mem. Amer. Math. Soc., 216 (2012), no. 1018. (Cited on p. 258)

[114] M. Goldshtein and P. Tsiotras, *Finite-horizon covariance control of linear time-varying systems*, in 2017 IEEE 56th Conference on Decision and Control (CDC), 2017, pp. 3606–3611. (Cited on pp. 273, 291)

[115] K. M. Grigoriadis and R. E. Skelton, *Minimum-energy covariance controllers*, Automatica, 33 (1997), pp. 569–578. (Cited on pp. 273, 279, 303)

[116] F. Guerra and L. Morato, *Quantization of dynamical systems and stochastic control theory*, Phys. Rev. D, 27 (1983), pp. 1774–1786. (Cited on p. 271)

[117] F. Guerra, *Unpublished notes, March* 1987*, and talk delivered at the workshop "Stochastic Dynamical Systems in Physics,"* Padua, 1988. (Cited on p. 269)

[118] A. Halder and E. D. Wendel, *Finite horizon linear quadratic Gaussian density regulator with Wasserstein terminal cost*, in Proc. IEEE Amer. Control Conf., 2016, pp. 7249–7254. (Cited on p. 273)

[119] H. von Helmholtz, *Physical Memoirs Selected and Translated from Foreign Sources*, Taylor and Francis, 1882. (Cited on p. 251)

[120] D. B. Hernandez and M. Pavon, *Equilibrium description of a particle system in a heat bath*, Acta Appl. Math., 14 (1989), pp. 239–256. (Cited on p. 278)

[121] D. Hilbert, *Über die gerade linie als kürzeste verbindung zweier punkte*, Math. Ann., 46 (1895), pp. 91–96. (Cited on p. 287)

[122] C. J. Holland, *A new energy characterization of the smallest eigenvalue of the Schrödinger equation*, Comm. Pure Appl. Math., 30 (1977), pp. 755–765. (Cited on p. 269)

[123] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math., 119 (1967), pp. 147–171. (Cited on pp. 274, 276)

[124] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 2012. (Cited on p. 295)

[125] A. F. Hotz and R. E. Skelton, *A covariance control theory*, in 24th IEEE Conference on Decision and Control (CDC), 1985, pp. 552–557. (Cited on pp. 273, 279, 303)
[126] A. Hotz and R. E. Skelton, *Covariance control theory*, Internat. J. Control, 46 (1987), pp. 13–32. (Cited on pp. 273, 279, 303)
[127] C. T. Ireland and S. Kullback, *Contingency tables with given marginals*, Biometrika, 55 (1968), pp. 179–188. (Cited on pp. 251, 286)
[128] K. Itô and S. Watanabe, *Transformation of Markov processes by multiplicative functionals*, Ann. Inst. Fourier (Grenoble), 15 (1965), pp. 13–30. (Cited on pp. 251, 268, 269, 293)
[129] B. Jamison, *Reciprocal processes*, Z. Wahrsch. Verw. Gebiete, 30 (1974), pp. 65–86. (Cited on pp. 251, 266, 276, 297)
[130] B. Jamison, *The Markov processes of Schrödinger*, Z. Wahrsch. Verw. Gebiete, 32 (1975), pp. 323–331. (Cited on pp. 251, 267, 268)
[131] E. T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. (2), 106 (1957), pp. 620–630. (Cited on p. 266)
[132] E. T. Jaynes, *Information theory and statistical mechanics* II, Phys. Rev. (2), 108 (1957), pp. 171–190. (Cited on p. 266)
[133] E. T. Jaynes, *On the rationale of maximum-entropy methods*, Proc. IEEE, 70 (1982), pp. 939–952. (Cited on p. 266)
[134] X. Jiang, Z. Luo, and T. T. Georgiou, *Geometric methods for spectral analysis*, IEEE Trans. Signal Process., 60 (2012), pp. 1064–1074. (Cited on p. 257)
[135] J. Johnson, *Thermal agitation of electricity in conductors*, Phys. Rev., 32 (1928), pp. 97–109. (Cited on p. 278)
[136] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17, https://doi.org/10.1137/S0036141096303359. (Cited on p. 258)
[137] L. V. Kantorovich, *Mathematical Methods in the Organization and Planning of Production*, Leningrad University, 1939 (in Russian); Management Sci., 6 (1960), pp. 363–422 (in English). (Cited on p. 254)
[138] L.V. Kantorovich, *On translocation of masses*, C. R. (Doklady) Acad. Sci. URSS (N.S.), 37 (1942), pp. 227–229 (in Russian); J. Math. Sci., 133 (2006), pp. 1381–1382 (in English). (Cited on pp. 254, 255)
[139] L. V. Kantorovich, *On a problem of Monge*, Uspekhi Mat. Nauk, 3 (1948), pp. 225–226 (in Russian); J. Math. Sci. (N.Y.), 133 (2006), art. 1383 (in English). (Cited on p. 255)
[140] L. V. Kantorovich and G. S. Rubinshtein, *On a space of totally additive functions*, Vestn. Leningrad. Univ., 13 (1958), pp. 52–59 (in Russian). (Cited on p. 255)
[141] E. Kappler, *Versuche zur Messung der Avogadro-Loschmidtschen Zahl aus der Brownschen Bewegung einer Drehwaage*, Ann. Phys., 11 (1931), pp. 233–256. (Cited on pp. 278, 279)
[142] I. Karatzas, *On a stochastic representation for the principal eigenvalue of a second-order differential equation*, Stochastics, 3 (1980), pp. 305–321. (Cited on p. 269)
[143] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988. (Cited on p. 269)
[144] S. Karlin and H. Taylor, *A Second Course in Stochastic Processes*, Academic Press, 1981. (Cited on p. 273)
[145] K. H. Kim and H. Qian, *Entropy production of Brownian macromolecules with inertia*, Phys. Rev. Lett., 93 (2004), art. 120602. (Cited on p. 279)
[146] W. Kliemann, *Recurrence and invariant measures for degenerate diffusions*, Ann. Probab., 15 (1987), pp. 690–707. (Cited on p. 276)
[147] A. Kolmogorov, *Zur Umkehrbarkeit der statistischen Naturgesetze*, Math. Ann., 113 (1937), pp. 766–772. (Cited on p. 252)
[148] J. Kruithof, *Telefoonverkeersrekening*, De Ingenieur, 52 (1937), pp. E15–E25. (Cited on p. 251)
[149] P. Langevin, *Sur la théorie du mouvement Brownien*, C. R. Acad. Sci. (Paris), 146 (1908), pp. 530–533. (Cited on p. 278)
[150] H. Lavenant, S. Claici, E. Chien, and J. Solomon, *Dynamical optimal transport on discrete surfaces*, ACM Trans. Graph., 37 (2018), art. 250. (Cited on pp. 256, 257)
[151] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, New York, Wiley, 1967. (Cited on p. 258)
[152] F. Léger, *A geometric perspective on regularized optimal transport*, J. Dynam. Differential Equations, 31 (2019), pp. 1777–1791. (Cited on p. 271)
[153] F. Léger and W. Li, *Hopf-Cole transformation and Schrödinger problems*, in Geometric Science of Information (GSI 2019), Lecture Notes in Comput. Sci. 11712, F. Nielsen and F. Barbaresco, eds., Springer, 2019, pp. 733–738. (Cited on p. 270)

[154] B. Lemmens and R. Nussbaum, *Nonlinear Perron-Frobenius Theory*, Cambridge Tracts in Math. 189, Cambridge University Press, 2012. (Cited on p. 287)

[155] B. Lemmens and R. Nussbaum, *Birkhoff's version of Hilbert's metric and its applications in analysis*, in Handbook of Hilbert Geometry, G. Besson, A. Papadopoulos, and M. Troyanov, eds., European Mathematical Society, Zürich, 2014, pp. 275–306. (Cited on p. 287)

[156] C. Léonard, *From the Schrödinger problem to the Monge-Kantorovich problem*, J. Funct. Anal., 262 (2012), pp. 1879–1920. (Cited on pp. 257, 262, 266, 270, 271, 300, 303)

[157] C. Léonard, *A survey of the Schroedinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst. A, 34 (2014), pp. 1533–1574. (Cited on pp. 257, 259, 263, 266, 267, 270, 271, 300, 303)

[158] C. Léonard, *Lazy random walks and optimal transport on graphs*, Ann. Probab., 44 (2016), pp. 1864–1915. (Cited on p. 291)

[159] C. Léonard, *Revisiting Fortet's Proof of Existence of a Solution to the Schrödinger System*, preprint, https://arxiv.org/abs/1904.13211, 2019. (Cited on pp. 268, 286)

[160] B. C. Levy, R. Frezza, and A. J. Krener, *Modeling and estimation of discrete-time Gaussian reciprocal processes*, IEEE Trans. Automat. Control, 35 (1990), pp. 1013–1023. (Cited on p. 297)

[161] W. Li, P. Yin, and S Osher, *Computations of optimal transport distance with Fisher information regularization*, J. Sci. Comput., 75 (2018), pp. 1581–1595. (Cited on p. 251)

[162] S. Liang, D. Medich, D. M. Czajkowsky, S. Sheng, J. Yuan, and Z. Shao, *Thermal noise reduction of mechanical oscillators by actively controlled external dissipative forces*, Ultramicroscopy, 84 (2000), pp. 119–125. (Cited on p. 279)

[163] C. Liverani and M. P. Wojtkowski, *Generalization of the Hilbert metric to the space of positive definite matrices*, Pacific J. Math., 166 (1994), pp. 339–355. (Cited on p. 289)

[164] J. Lott, *Some geometric calculations on Wasserstein space*, Comm. Math. Phys., 277 (2008), pp. 423–437. (Cited on p. 258)

[165] F. Marquardt and S. M. Girvin, *Optomechanics*, Physics, 2 (2009), art. 40. (Cited on p. 279)

[166] J. C. Maxwell, *On governors*, Proc. Roy. Soc., 100 (1868), pp. 270–282. (Cited on p. 279)

[167] R. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179. (Cited on p. 258)

[168] T. Mikami, *Monge's problem with a quadratic cost by the zero-noise limit of h-path processes*, Probab. Theory Related Fields, 129 (2004), pp. 245–260. (Cited on pp. 257, 266, 270, 271, 300, 303)

[169] T. Mikami and M. Thieullen, *Duality theorem for the stochastic optimal control problem*, Stochastic Process. Appl., 116 (2006), pp. 1815–1835. (Cited on pp. 257, 266, 270, 271, 300, 303)

[170] T. Mikami and M. Thieullen, *Optimal transportation problem by stochastic optimal control*, SIAM J. Control Optim., 47 (2008), pp. 1127–1139, https://doi.org/10.1137/050631264. (Cited on pp. 257, 266, 270, 271, 300, 303)

[171] J. M. W. Milatz, J. J. Van Zolingen, and B. B. Van Iperen, *The reduction in the Brownian motion of electrometers*, Physica, 19 (1953), pp. 195–202. (Cited on p. 279)

[172] S. Mitter, *Non-linear filtering and stochastic mechanics*, in Stochastic Systems: The Mathematics of Filtering and Identification and Applications, M. Hazewinkel and J. C. Willems, eds., Reidel, 1981, pp. 479–503. (Cited on p. 269)

[173] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, in Histoire de l'Academie Royale des Sciences avec les Mémoires de Mathématique et de Physique, De l'Imprimerie Royale, 1781, pp. 666–704. (Cited on p. 251)

[174] P. Monteiller, S. Claici, E. Chien, F. Mirzazadeh, J. Solomon, and M. Yurochkin, *Alleviating label switching with optimal transport*, Adv. Neural Inform. Process. Syst., 32 (2019), pp. 13612–13622. (Cited on pp. 256, 257)

[175] M. Nagasawa, *Stochastic variational principle of Schrödinger processes*, in Seminar on Stochastic Processes, Birkhäuser, 1989, pp. 165–175. (Cited on p. 269)

[176] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, NJ, 1967. (Cited on pp. 252, 263, 270, 271, 277, 278)

[177] E. Nelson, *Quantum Fluctuations*, Princeton University Press, Princeton, NJ, 1985. (Cited on p. 271)

[178] E. Nelson, *Afterword*, in Diffusion, Quantum Theory and Radically Elementary Mathematics, W. G. Faris, ed., Math. Notes 47, Princeton University Press, Princeton, NJ, 2006, pp. 229–232. (Cited on p. 252)

[179] H. Nyquist, *Thermal agitation of electric charge in conductors*, Phys. Rev., 32 (1928), pp. 110–113. (Cited on p. 278)

[180] J. J. O'Connor and E. F. Robertson, *Jean Baptiste Joseph Fourier*, MacTutor (Fourier), http://www-history.mcs.st-andrews.ac.uk/Biographies/Fourier.html. (Cited on p. 252)

[181] K. Okamoto, M. Goldshtein, and P. Tsiotras, *Optimal covariance control for stochastic systems under chance constraints*, IEEE Control Syst. Lett., 2 (2018), pp. 266–271. (Cited on pp. 273, 274)

[182] K. Okamoto and P. Tsiotras, *Optimal stochastic vehicle path planning using covariance steering*, IEEE Robotics Automat. Lett., 4 (2019), pp. 2276–2281. (Cited on p. 273)

[183] K. Okamoto and P. Tsiotras, *Input hard constrained optimal covariance steering*, in 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 3497–3502. (Cited on pp. 273, 274)

[184] O. A. Olejnik and E. V. Radkevic, *Second Order Equations with Nonnegative Characteristic Form*, AMS, Providence, RI, 1973. (Cited on p. 274)

[185] Y. Ollivier, H. Pajot, and C. Villani, *Optimal Transportation. Theory and Applications*, London Math. Soc. Lect. Notes Ser. 413, Cambridge University Press, 2014. (Cited on pp. 255, 256)

[186] N. K. Olver, *Robust Network Design*, Ph.D. Dissertation, Department of Mathematics and Statistics, McGill University, 2010. (Cited on p. 294)

[187] L. Onsager and S. Machlup, *Fluctuations and irreversible processes*, I, Phys. Rev., 91 (1953), pp. 1505–1512. (Cited on p. 293)

[188] L. Onsager and S. Machlup, *Fluctuations and irreversible processes*, II, Phys. Rev., 91 (1953), pp. 1512–1515. (Cited on p. 293)

[189] F. Otto, *The geometry of dissipative evolution equations: The porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174. (Cited on p. 258)

[190] W. Parry, *Intrinsic Markov chains*, Trans. Amer. Math. Soc., 112 (1964), pp. 55–66. (Cited on p. 296)

[191] G. A. Pavliotis, *Stochastic Processes and Applications*, Springer, Berlin, 2014. (Cited on p. 279)

[192] M. Pavon and A. Wakolbinger, *On free energy, stochastic control, and Schroedinger processes*, in Modeling, Estimation and Control of Systems with Uncertainty, G. B. Di Masi, A. Gombani, A. Kurzhanski, eds., Birkhäuser, Boston, 1991, pp. 334–348. (Cited on pp. 269, 273)

[193] M. Pavon and F. Ticozzi, *On entropy production for controlled Markovian evolution*, J. Math. Phys., 47 (2006), art. 06330. (Cited on p. 279)

[194] M. Pavon and F. Ticozzi, *Discrete-time classical and quantum Markovian evolutions: Maximum entropy problems on path space*, J. Math. Phys., 51 (2010), art. 042104. (Cited on p. 292)

[195] M. Pavon and A. Ferrante, *On the geometry of maximum entropy problems*, SIAM Rev., 55 (2013), pp. 415–439, https://doi.org/10.1137/120862843. (Cited on p. 282)

[196] M. Pavon, E. G. Tabak, and G. Trigila, *The data-driven Schrödinger bridge*, Comm. Pure Appl. Math., 2021, https://doi.org/10.1002/cpa.21975. (Cited on p. 268)

[197] J. Perrin, *Brownian Movement and Molecular Reality*, translated from Annales de Chimie et de Physique, 8me Series, 1909, by F. Soddy, Taylor and Francis, London, 1910. (Cited on p. 278)

[198] G. Peyré and M. Cuturi, *Computational optimal transport*, Found. Trends Mach. Learn., 11 (2019), pp. 1–257. (Cited on pp. 251, 266, 281, 303)

[199] M. Poot and H. S. J. van der Zant, *Mechanical systems in the quantum regime*, Phys. Rep., 511 (2012), pp. 273–335. (Cited on p. 279)

[200] M. L. Puterman, *Markov Decision Processes*, John Wiley, 1994. (Cited on pp. 258, 293)

[201] H. Qian, *Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations*, Phys. Rev. E, 63 (2001), art. 042103. (Cited on p. 279)

[202] T. Rachev and L. Rüschendorf, *Mass Transportation Problems, Vol.* I: *Theory*, Probab. Appl. (N. Y.), Springer-Verlag, New York, 1998. (Cited on pp. 255, 256, 281, 299)

[203] T. Rachev and L. Rüschendorf, *Mass Transportation Problems, Vol.* II: *Applications*, Probab. Appl. (N. Y.), Springer-Verlag, New York, 1998. (Cited on pp. 255, 256)

[204] D. Reeb, M. J. Kastoryano, and M. M. Wolf, *Hilbert's projective metric in quantum information theory*, J. Math. Phys., 52 (2011), art. 082201. (Cited on p. 287)

[205] J. Ridderhof and P. Tsiotras, *Uncertainty quantification and control during Mars powered descent and landing using covariance steering*, in AIAA Guidance, Navigation, and Control Conference, Kissimmee, FL, 2018. (Cited on pp. 256, 273)

[206] J. Ridderhof and P. Tsiotras, *Minumum-fuel powered descent in the presence of random disturbances*, in AIAA Guidance, Navigation, and Control Conference, San Diego, CA, 2019. (Cited on pp. 256, 273)

[207] J. Ridderhof, K. Okamoto, and P. Tsiotras, *Nonlinear uncertainty control with iterative covariance steering*, in 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 3484–3490. (Cited on p. 274)

[208] J. Ridderhof, K. Okamoto, and P. Tsiotras, *Chance constrained covariance control for linear stochastic systems with output feedback*, in Proceedings of the 59th IEEE Conference on Decision and Control (CDC), IEEE, 2020, pp. 1758–1763, https://doi.org/10.1109/CDC42340.2020.9303731. (Cited on p. 273)

[209] T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970. (Cited on p. 280)

[210] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, Cambridge University Press, 2004. (Cited on pp. 282, 296)

[211] I. S. Sanov, *On the probability of large deviations of random magnitudes*, Mat. Sb. N. S., 42 (1957), pp. 11–44 (in Russian); Select. Transl. Math. Statist. Probab., 1 (1961), pp. 213–244 (in English). (Cited on pp. 263, 264)

[212] K. Savla, G. Como, and M. A. Dahleh, *Robust network routing under cascading failures*, IEEE Trans. Network Sci. Engrg., 1 (2014), pp. 53–66. (Cited on p. 294)

[213] P. A. Schilpp, ed., *Albert Einstein: Philosopher-Scientist*, in The Library of Living Philosophers, Vol. VII, The Library of Living Philosophers, Inc., Evanston, IL, 1949. (Cited on p. 278)

[214] E. Schrödinger, *Collected Papers on Wave Mechanics*, translated by J. F. Shearer and W. M. Deans, Blackie & Son Ltd., London, 1928. (Cited on p. 252)

[215] E. Schrödinger, *Über die Umkehrung der Naturgesetze*, Sitzungsberichte der Preuss Akad. Wissen. Berlin, Phys. Math. Klasse, 10 (1931), pp. 144–153. (Cited on pp. 251, 252, 263, 268)

[216] E. Schrödinger, *Sur la théorie relativiste de l'électron et l'interpretation de la mécanique quantique*, Ann. Inst. H. Poincaré, 2 (1932), pp. 269–310. (Cited on pp. 251, 252, 263)

[217] E. Schrödinger, *The meaning of wave mechanics*, in Louis de Broglie Physicien et Penseur, A. George, ed., éditions Albin Michel, Paris, 1953, pp. 16–30. (Cited on p. 252)

[218] K. C. Schwab and M. L. Roukes, *Putting mechanics into quantum mechanics*, Phys. Today, 58 (2005), pp. 36–42. (Cited on p. 279)

[219] R. Sepulchre, A. Sarlette, and P. Rouchon, *Consensus in non-commutative spaces*, in the 49th IEEE Conference on Decision and Control (CDC), 2010, pp. 6596–6601. (Cited on p. 287)

[220] R. Sandhu, T. T. Georgiou, E. Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tannenbaum, *Graph curvature for differentiating cancer networks*, Sci. Rep., 5 (2015), pp. 1–13. (Cited on p. 294)

[221] R. Sandhu, T. T. Georgiou, and A. R. Tannenbaum, *Ricci curvature: An economic indicator for market fragility and systemic risk*, Sci. Adv., 2 (2016), art. e1501495. (Cited on p. 294)

[222] R. Skelton and M. Ikeda, *Covariance controllers for linear continuous-time systems*, Internat. J. Control, 49 (1989), pp. 1773–1785. (Cited on pp. 273, 279, 303)

[223] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist., 35 (1964), pp. 876–879. (Cited on pp. 251, 267, 280, 284, 286)

[224] R. Sinkhorn, *Diagonal equivalence to matrices with prescribed row and column sums*, Amer. Math. Monthly, 74 (1967), pp. 402–405. (Cited on pp. 251, 267, 280)

[225] R. Sinkhorn and P. Knopp, *Concerning nonnegative matrices and doubly stochastic matrices*, Pacific J. Math., 21 (1967), pp. 343–348. (Cited on p. 251)

[226] J. Solomon, L. Guibas, and A, Butscher, *Dirichlet energy for analysis and synthesis of soft maps*, Comput. Graphics Forum, 32 (2013), pp. 197–206. (Cited on p. 256)

[227] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher, *Earth mover's distances on discrete surfaces*, ACM Trans. Graphics, 33 (2014), art. 67. (Cited on p. 256)

[228] J. Solomon, R. Rustamov, G. Leonidas, and A. Butscher, *Wasserstein propagation for semi-supervised learning*, in Proc. 31st International Conference on Machine Learning (ICML-14), 2014, pp. 306–314. (Cited on p. 256)

[229] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas, *Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains*, ACM Trans. Graphics, 34 (2015), art. 66. (Cited on pp. 256, 257)

[230] J. Solomon, G. Peyré, V. G. Kim, and S. Sra, *Entropic metric alignment for correspondence problems*, ACM Trans. Graphics, 35 (2016), art. 72. (Cited on p. 256)

[231] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher, *Continuous-Flow Graph Transportation Distances*, preprint, https://arxiv.org/abs/1603.06927, 2016. (Cited on p. 256)

[232] E. G. Tabak and G. Trigila, *Explanation of variability and removal of confounding factors from data through optimal transport*, Comm. Pure Appl. Math., 71 (2018), pp. 163–199. (Cited on p. 261)

[233] Y. Takahashi and S. Watanabe, *The probability functionals (Onsager-Machlup functions) of diffusion processes*, in Stochastic Integrals, Lecture Notes in Math. 851, Springer-Verlag, Berlin, 1980, pp. 433–463. (Cited on p. 293)

[234] A. Thibault, L. Chizat, C. Dossal, and N. Papadakis, *Overrelaxed Sinkhorn-Knopp Algorithm for Regularized Optimal Transport*, Technical report, presented at the NIPS 2017 Workshop on Optimal Transport & Machine Learning, 2017, https://arxiv.org/abs/1711.01851. (Cited on pp. 251, 286)

[235] A. Thompson, *On certain contraction mappings in a partially ordered vector space*, Proc. Amer. Math. Soc., 14 (1963), pp. 438–443. (Cited on p. 289)

[236] A. N. Tolstoi, *Metody nakhozhdeniya naimen shego summovogo kilometrazha pri planirovanii perevozok v prostranstve ("Methods of finding the minimal total mileage in cargo transportation planning in space")*, TransPress Natl. Commissariat Transport., 1 (1930), pp. 23–55. (Cited on p. 254)

[237] A. N. Tolstoi, *Metody ustraneniya neratsional nykh perevozok priplanirovanii ("Methods of removing irrational transportation in planning")*, Sotsialisticheski Transport, 9 (1939), pp. 28–51. (Cited on p. 254)

[238] J. Tsitsiklis, D. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Trans. Automat. Control, 31 (1986), pp. 803–812. (Cited on p. 287)

[239] G. E. Uhlenbeck and L. S. Ornstein, *On the theory of Brownian motion*, Phys. Rev., 36 (1930), pp. 823–841. (Cited on p. 278)

[240] L. N. Vasershtein, *Markov processes over denumerable products of spaces describing large system of automata*, Problems Inform. Transmission, 5 (1969), pp. 47–52. (Cited on p. 255)

[241] A. M. Vershik, *Long history of the Monge-Kantorovich problem*, Math. Intelligencer, 35 (2013), pp. 1–9. (Cited on p. 255)

[242] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, AMS, 2003. (Cited on pp. 251, 255, 256, 257, 258, 295)

[243] C. Villani, *Optimal Transport. Old and New*, Grundlehren Math. Wiss. 338, Springer-Verlag, 2009. (Cited on pp. 255, 256, 257, 258, 295)

[244] A. Vinante, M. Bignotto, M. Bonaldi et al., *Feedback cooling of the normal modes of a massive electromechanical system to submillikelvin temperature*, Phys. Rev. Lett., 101 (2008), art. 033601. (Cited on p. 279)

[245] M. K. von Renesse, *An optimal transport view on Schrödinger's equation*, Canad. Math. Bull., 55 (2012), pp. 858–869. (Cited on p. 258)

[246] A. Wakolbinger, *A simplified variational characterization of Schrödinger processes*, J. Math. Phys., 30 (1989), pp. 2943–2946. (Cited on p. 269)

[247] A. Wakolbinger, *Schrödinger Bridges from* 1931 *to* 1991, in Proc. 4th Latin American Congress in Probability and Mathematical Statistics, Mexico City, 1990, Contribuciones en Probabilidad y Estadistica Matematica 3, 1992, pp. 61–79. (Cited on pp. 263, 272, 273, 274, 276, 303)

[248] N. Wiener, *Differential space*, J. Math. Phys., 2 (1923), pp. 132–174. (Cited on p. 268)

[249] J.-H. Xu and R. E. Skelton, *An improved covariance assignment theory for discrete systems*, IEEE Trans. Automat. Control, 37 (1992), pp. 1588–1591. (Cited on p. 291)

[250] K. Yasue, *Stochastic calculus of variations*, J. Funct. Anal., 41 (1981), pp. 327–340. (Cited on p. 271)

[251] Z. Yi, Z. Cao, E. Theodorou, and Y. Chen, *Nonlinear covariance control via differential dynamic programming*, in Proc. 2020 American Control Conference (ACC), IEEE, 2020, pp. 3571–3576. (Cited on p. 274)

[252] G. U. Yule, *On the methods of measuring association between two attributes*, J. Roy. Statist. Soc., 75 (1912), pp. 579–652. (Cited on p. 251)

[253] J. C. Zambrini, *Stochastic mechanics according to E. Schrödinger*, Phys. Rev. A, 33 (1986), pp. 1532–1548. (Cited on p. 269)

[254] G. Zhu, K. M. Grigoriadis, and R. E. Skelton, *Covariance control design for Hubble space telescope*, J. Guidance Control Dynam., 18 (1995), pp. 230–236. (Cited on pp. 273, 279, 303)