

The most likely evolution of diffusing and vanishing particles: Schrödinger Bridges with unbalanced marginals

Yongxin Chen, Tryphon T. Georgiou and Michele Pavon

Abstract

Stochastic flows of an advective-diffusive nature are ubiquitous in biology and the physical sciences. Of particular interest is the problem to reconcile observed marginal distributions with a given prior posed by E. Schrödinger in 1932/32 and known as the Schrödinger Bridge Problem (SBP). It turns out that Schrödinger’s problem can be viewed both as a modeling as well as a control problem. Due to the fundamental significance of this problem, interest in SBP and in its deterministic (zero-noise limit) counterpart of Optimal Mass Transport (OMT) has in recent years enticed a broad spectrum of disciplines, including physics, stochastic control, computer science, probability theory, and geometry. Yet, while the mathematics and applications of SBP/OMT have been developing at a considerable pace, accounting for marginals of unequal mass has received scant attention; the problem to interpolate between “unbalanced” marginals has been approached by introducing source/sink terms into the transport equations, in an adhoc manner, chiefly driven by applications in image registration.

Nevertheless, losses are inherent in many physical processes and, thereby, models that account for lossy transport may also need to be reconciled with observed marginals following Schrödinger’s dictum; that is, to adjust the probability of trajectories of particles, including those that do not make it to the terminal observation point, so that the updated law represents the *most likely way that particles may have been transported, or vanished, at some intermediate point*. Thus, the purpose of this work is to develop such a natural generalization of the SBP for stochastic evolution with losses, whereupon particles are “killed” (jump into a coffin/extinction state) according to a probabilistic law, and thereby mass is gradually lost along their stochastically driven flow. Through a suitable embedding we turn the problem into an *SBP for stochastic processes that combine diffusive and jump characteristics*. Then, following a large-deviations formalism in the style of E. Schrödinger, given a prior law that allows for losses, we ask for the most probable evolution of particles along with the most likely killing rate as the particles transition between the specified marginals. Our approach differs sharply from previous work involving a Feynman-Kac multiplicative reweighing of the reference measure: The latter, as we argue, is far from Schrödinger’s quest. An iterative scheme, generalizing the celebrated Fortet-IPF-Sinkhorn algorithm, permits to compute the new drift *and* the new killing rate of the path-space solution measure. We finally formulate and solve a related fluid-dynamic control problem for the flow of one-time marginals where both the drift and the new killing rate play the role of control variables.

I. INTRODUCTION

Consider the problem of estimating the velocity field of oceanic currents by releasing into the water a cloud of tracer particles and by sampling their distribution at a later time. The diffusion coefficient is assumed known and the original cloud that is released at time $t = 0$ consists of N particles. These are expected to remain in suspension for a duration of time while they diffuse and drift with the current. At time $t = 1$, their distribution is sampled again. Some of the particles in the meantime have sunk, so that the number of found particles is less than N . Suppose this experiment is performed several times, treating the model originating from previous experiments as a “prior”. Is it conceivable to “improve” a prior model in a rational way? More explicitly, by relying on a prior model and the new sampling result,

Y. Chen is with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
 T.T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697, USA
 M. Pavon is with the Department of Mathematics “Tullio Levi-Civita”, Università di Padova, 35121 Padova, Italy

is it possible to determine an updated model that represents the most probable way that the tracer cloud may have been transported?

At first sight, this problem appears to be of a different nature than those treated in the theory of Large Deviations [1], [2], [3], in that the sought path-space measure is not a probability measure per se. Nevertheless, in spite of the paucity of the available data, it is possible to solve this inverse problem by a natural embedding technique. A byproduct is a physically motivated framework to interpolate distributions of unequal mass (integrals). The blueprint for the rationale in our work has been provided by the celebrated duo of papers by E. Schrödinger in 1931/32 [4], [5] where he considered the problem of reconciling marginal distributions with a prior stochastic evolution.

The original Schrödinger Bridge Problem (SBP) asks for the most likely evolution of stochastic particles as they travel between marginal probability densities ρ_0 and ρ_1 , specified at two points in time (taken as $t_0 = 0$ and $t_1 = 1$ without loss of generality), when these marginals fail to be consistent with a known *prior law*. Interestingly, Schrödinger considered this abstract problem before a theory of continuous parameter stochastic processes was in place, and had only been preceded by Ludwig Boltzmann [6]. Schrödinger attacks the problem very much in his countryman's style, through coarse graining, and applying the De Moivre-Stirling formula and Lagrange multipliers. In spite of the lack of proper tools (Sanov's theorem [7] will be published in Russian only in 1957), he arrives at the correct answer [4], [5], that the most likely evolution is obtained by solving a certain two-point boundary value problem (Schrödinger system of equations). Important contributions to this theory are then provided by Fortet, Beurling and Jamison [8], [9], [10], [11]. It took more than half a century before Föllmer [12], recovering Schrödinger's original motivation, properly cast the problem within the framework of Large-Deviations theory. The field has since seen several other significant contributions, a partial selection being [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. Here [19], [25], [27] are survey papers. Observe that, in view of Sanov's theorem [3], the SBP amounts to seeking a new probability law on the path space of the stochastic particles that is consistent with the given marginals and, at the same time, is the closest to the prior probability law in the relative entropy sense.

Schrödinger's Bridge Problem (SBP), as well as its zero-noise limit of Optimal Mass Transport (OMT), continue to impact a growing range of disciplines and applications. In this expanding mathematical landscape, the problem to account for variable mass along the transport path received attention from early on. It was chiefly motivated by the need to interpolate distributions of unequal mass for times series spectral analysis and image registration [29]. The viewpoint that is being pursued herein is closer in spirit to the original rationale of E. Schrödinger as we build on a Large Deviations formalism. To this end, we consider below a diffusion process with killing and seek the closest update of the corresponding law that is in agreement with the marginal data. Thus, we ask for the most likely evolution of stochastic particles which are known to obey a given *prior law with potential for losses* ("killing rate") while they transition between two marginal distributions ρ_0 and ρ_1 as before. The two distributions are not necessarily consistent with the prior law and neither is the loss of mass necessarily consistent with the prior killing rate.

In our formulation of the unbalanced Schrödinger Bridge Problem (uSBP), the marginals cannot be assumed to be probability distributions as their integrals differ due to losses. To this end, we embed the distributions into a frame that includes a coffin/extinction state, leading to a probability law on a continuum together with a discrete state. Thereupon, we find the updated law and killing rate that minimize the relative entropy to the prior with losses, and are consistent with the two marginals. In the special case when the marginals are already consistent with the prior, naturally, the solution coincides with the lossy prior, differently from what happens in other formulations of SBP with killing which are based on Feynman-Kac functionals [18], [15], [16], [17], [20], [30], [31], [32], and unbalanced transport [33], [34], [35], [29], as discussed in Section V.

The structure of the paper is as follows. In Section II we revisit classical Schrödinger bridge problems. The main framework on Schrödinger bridges with unbalanced marginals is presented in Section III. We also present a fluid dynamic formulation of the main framework in Section IV. A comprehensive comparison between Schrödinger bridges with unbalanced marginals and existing results on Schrödinger bridges with killing is provided in Section V. This is followed by a numerical example in Section VI and a concluding remark in Section VII.

II. PRELIMINARIES ON SCHRÖDINGER BRIDGE PROBLEM

We briefly review elements of the theory of Schrödinger Bridge Problem (SBP). To this end, consider a diffusion process

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad (1)$$

over the Euclidean space \mathbb{R}^n . In Schrödinger's original thought experiment, a large number N of trajectories over the time interval $[0, 1]$ are independently sampled from (1) with probability distribution of X_t at the initial time $t = 0$ being ρ_0 . The law of large numbers dictates that the terminal distribution at time $t = 1$ must be (approximately)

$$\int_{\mathbb{R}^n} q(0, x, 1, \cdot) \rho_0(x) dx, \quad (2)$$

where $q(t, x, s, y)$, $t < s$ denotes the kernel of transition rates from state x at time t to state y at time s . Now suppose the observed marginal distribution at time $t = 1$, denoted by ρ_1 , is inconsistent with (2) and the prior kernel $q(0, x, 1, y)$, that is,

$$\rho_1(\cdot) \neq \int_{\mathbb{R}^n} q(0, x, 1, \cdot) \rho_0(x) dx.$$

Schrödinger's problem then seeks the most likely evolution that the particles may have taken between the specified marginals. That is, SBP seeks a suitable update of the law of the diffusion process that reconciles the two marginals ρ_0, ρ_1 . In the sequel and for notational simplicity, we use the same symbol ρ to denote both, the probability density, as well as the corresponding measure $d\rho = \rho dx$, depending on the context.

As first noted by Föllmer [12], SBP can be more clearly expressed in the language of the theory of *large deviations* [3]. Specifically, let $\Omega = C([0, 1], \mathbb{R}^n)$ denote the space of continuous functions on $[0, 1]$ with values in \mathbb{R}^n , and $\mathcal{P}(\Omega)$ denote the space of probability laws over Ω . Given any two probability measures P, Q , the *relative entropy* of P with respect to Q is

$$H(P | Q) = \begin{cases} \int dP \log \frac{dP}{dQ} & \text{if } P \ll Q \\ +\infty & \text{otherwise.} \end{cases} \quad (3)$$

Now consider N independent trajectories $X_t^1, X_t^2, \dots, X_t^N \in \Omega$ of a diffusion having law $R \in \mathcal{P}(\Omega)$, and let L_N denote their empirical distribution. Then, asymptotically as $N \rightarrow \infty$, Sanov's theorem¹ gives the exponential rate of decay for the probability of occurrence of an empirical distribution that differs from the law R [3] as

$$\text{Prob}(L_N \in A) \approx \exp(-N \inf_{P \in A} H(P | R)), \quad \forall A \subset \mathcal{P}(\Omega). \quad (4)$$

Thus, Sanov's result expresses the likelihood of observing an empirical distribution approximated by P in terms of the relative entropy $H(P | R)$. Thence, SBP can be formulated as follows:

¹Sanov's theorem holds when the process takes values in any Polish space.

Problem 1. Let $R \in \mathcal{P}(\Omega)$ be the probability measure on Ω induced by the prior process (1) with initial distribution ρ_0 . Determine

$$P^* := \arg \min_{P \in \mathcal{P}(\Omega)} \{H(P | R) \mid P_0 = \rho_0, P_1 = \rho_1\}, \quad (5)$$

where P_t denotes the marginal distribution of P at time t (i.e., the push forward $X_t \# P = P_t$).

The entropy functional is strictly convex which ensures uniqueness of the minimizer (when it exists). Further, if $R_0 = \rho_0$ as well as $R_1 = \rho_1$, the solution to SBP coincides (trivially) with the prior law R , i.e., $P^* = R$, achieving the minimal value $H(P^* | R) = 0$. When $R_1 \neq \rho_1$, the SBP thus seeks an updated law P^* that is closest to the prior in the sense of relative entropy and restores consistency with the marginals (which fails for the prior R). Next we briefly discuss the solution to SBP. For an in depth exposition see [25] and the review articles [36], [27].

By disintegration of measure,

$$\begin{aligned} R(\cdot) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} R^{xy}(\cdot) R_{01}(dxdy), \text{ and} \\ P(\cdot) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} P^{xy}(\cdot) P_{01}(dxdy), \end{aligned}$$

where R_{01} (P_{01}) denotes the joint marginal distribution of R (P) of X_t for $t \in \{0, 1\}$ (i.e., $R_{01} = X_{01} \# R$, and similarly, for P), and R^{xy} (P^{xy}) denotes the measure induced by R (P) conditioned on $(X_0 = x, X_1 = y)$. It follows that

$$H(P | R) = H(P_{01} | R_{01}) + \int H(P^{xy} | R^{xy}) P_{01}(dxdy). \quad (6)$$

Clearly, when $P^{xy} = R^{xy}$ for any $x, y \in \mathbb{R}^n$, the second term on the right assumes the minimal value 0. An immediate consequence is the following *static* formulation of the SBP.

Problem 2. Determine

$$\pi^* := \arg \min_{\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)} \{H(\pi | R_{01}) \mid \pi_0 = \rho_0, \pi_1 = \rho_1\}. \quad (7)$$

To distinguish between the two formulations (5) and (7), we refer to (5) as the *dynamic* SBP. The two formulations are equivalent in the sense that solving one provides a solution to the other, as noted next.

Theorem 3 ([25]). Suppose P^* is a solution to the dynamic SBP (5), then P_{01}^* solves the static SBP (7). On the other hand, if π^* is a solution to (7), then setting $P^* = \int_{\mathbb{R}^n \times \mathbb{R}^n} R^{xy}(\cdot) \pi^*(dxdy)$ solves (5), while $P_{01}^* = \pi^*$.

Proof. It follows readily from (6). □

A direct consequence of Theorem 3 is that the Radon-Nikodym ratios between solutions and priors for the two problems, the static (18) and the dynamic (17), coincide, namely,

$$\frac{dP^*}{dR} = \frac{d\pi^*}{dR_{01}}(X_0, X_1). \quad (8)$$

In fact, this ratio can be factored into two parts, one that depends only on X_0 and one that depends on X_1 , as follows.

Theorem 4 ([25]). *Assume that $R_{01} \ll R_0 \otimes R_1$, and that there exists $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ such that $\pi_0 = \rho_0$, $\pi_1 = \rho_1$ (i.e., feasible), for which $H(\pi | R_{01}) < +\infty$. Then the static problem (7) admits a unique solution π^* and there exist two measurable functions $f, g : \mathcal{X} \rightarrow \mathbb{R}_+$ such that*

$$\pi^* = f(X_0)g(X_1)R_{01}. \quad (9)$$

The two factors f, g are solutions to the Schrödinger system

$$\frac{d\rho_0}{dR_0}(x) = f(x)R(g(X_1) | X_0 = x), \quad (10a)$$

$$\frac{d\rho_1}{dR_1}(y) = g(y)R(f(X_0) | X_1 = y). \quad (10b)$$

Moreover, the unique solution to the dynamic problem (5) is

$$P^* = f(X_0)g(X_1)R. \quad (11)$$

The above theorem provides an abstract construction of the sought probability law(s) via the solution of the Schrödinger system (10). The local characteristics and the modified stochastic differential equation for the process with law P^* follow. Computationally, these can be expressed most succinctly in terms of a pair of two, forward and backward in time (and identical to that of the prior Fokker-Planck and its adjoint) equations, that are nonlinearly coupled through boundary conditions. We explain this next.

First recall that the marginals $R_t(x)$ of the prior law R for the diffusion (1) satisfy (weakly) the Fokker-Planck equation

$$\partial_t R_t + \nabla \cdot (bR_t) = \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij} R_t)}{\partial x_i \partial x_j}. \quad (12)$$

In what follows, $a(t, x) = \sigma(t, x)\sigma(t, x)'$ is assumed to be everywhere positive definite. Let the two endpoint marginals be absolutely continuous with densities ρ_0 and ρ_1 , respectively. The Schrödinger system (10) can be reparametrized in terms of

$$\hat{\varphi}(0, x) := f(x)R_0(x) \quad (13a)$$

$$\varphi(1, y) := g(y), \quad (13b)$$

and takes the form

$$\partial_t \hat{\varphi} = -\nabla \cdot (b\hat{\varphi}) + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij} \hat{\varphi})}{\partial x_i \partial x_j} \quad (14a)$$

$$\partial_t \varphi = -b \cdot \nabla \varphi - \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \quad (14b)$$

$$\rho_0 = \varphi(0, \cdot) \hat{\varphi}(0, \cdot) \quad (14c)$$

$$\rho_1 = \varphi(1, \cdot) \hat{\varphi}(1, \cdot). \quad (14d)$$

Theorem 5. *Let R be the law of (1) with $a(t, x) = \sigma(t, x)\sigma(t, x)'$ being positive definite for all $(t, x) \in \mathbb{R} \times \mathbb{R}^n$, and assume that ρ_0, ρ_1 are absolutely continuous with respect to the Lebesgue measure. There exists a unique (up to a constant positive scaling) pair $(\hat{\varphi}(t, x), \varphi(t, x))$ of non-negative functions that satisfies the Schrödinger system (14). Moreover, the law P^* for the dynamic problem (5) is law of the diffusion*

$$dX_t = (b(t, X_t) + a(t, X_t)\nabla \log \varphi(t, X_t))dt + \sigma(t, X_t)dW_t, \quad (15)$$

with distribution of X_0 being ρ_0 , and at any time $t \in [0, 1]$, the marginal density for P^* satisfies the identity $p_t(x) = \varphi(t, x)\hat{\varphi}(t, x)$.

Existence and uniqueness of solutions for the Schrödinger system have been provided in various degrees of generality by Fortet, Beurling and Jamison [8], [9], [10], [11]. For a detailed exposition of the theory of Schrödinger's problem we refer to Leonard [25], in particular, [25, Theorems 2.8, 2.9, 3.4]. A more recent account along with a proof that is based on the contractiveness of suitable maps in the Hilbert metric was given in [37]. In the present paper, we follow a similar approach as in [37] when analyzing the more general Schrödinger system for diffusions with losses and, therefore, we sketch key steps for this more general case that we consider. An added benefit in recasting the Schrödinger system as in (14) is that it leads, after discretization, to an efficient algorithm for computing $(\hat{\varphi}(t, x), \varphi(t, x))$, and thereby, f, g as well as P^* . The discretized version of the Schrödinger system (14) amounts to the celebrated Sinkhorn algorithm for matrix scaling [27].

III. UNBALANCED STOCHASTIC TRANSPORT

We now analyze stochastic flows between unequal marginals following E. Schrödinger original rationale that is rooted in large deviations theory. To this end, we consider a diffusion process with killing and seek the closest update of the corresponding prior law that restores agreement with marginal data.

Once again consider the diffusion process (1) but, this time, with a nonnegative killing rate $V(t, x)$ (assume $V(\cdot, \cdot)$ is continuous and not constantly zero). A thought experiment similar to Schrödinger's, calls for a large number N of trajectories over a time interval $[0, 1]$, that are independently sampled from (1) with initial probability distribution ρ_0 , and a recorded empirical distribution for the surviving particles at time $t = 1$ approximated by ρ_1 , which is inconsistent with the prior law, that is,

$$\rho_1(\cdot) \neq \int_{\mathbb{R}^n} q(0, x, 1, \cdot) \rho_0(x) dx.$$

The kernel $q(0, x, 1, y)$ is no longer a probability kernel in that $\int_{\mathbb{R}^n} q(0, x, 1, y) dy \neq 1$, in general, and thus, neither $\int_{\mathbb{R}^n} q(0, x, 1, \cdot) \rho_0(x) dx$ nor ρ_1 are necessarily probability densities, due to killing. In particular, $\int \rho_1(x) dx = N_s/N \leq 1$ where N_s denotes the number of survival particles at time 1. Just as in the standard SBP, we consider continuous distributions, assuming that N is large, and seek to identify the most likely behavior of the particles. By behavior we mean the most likely evolution of the particles along with the most likely times that the particles may have gotten killed (or, absorbed by an underlying medium).

As in the standard Schrödinger bridge, the problem arising from the above thought experiment can be formally stated using the theory of *large deviations* [3]. However, in this case, the space of trajectories needs to be modified to accommodate for possible killing of particles. To this end, we augment the state space of the diffusion \mathbb{R}^n with a "coffin state" \mathfrak{c} , resulting in the state space

$$\mathcal{X} = \mathbb{R}^n \cup \{\mathfrak{c}\}.$$

Let $\Omega = D([0, 1], \mathcal{X})$ be the Skorokhod space over \mathcal{X} , that is, each element in Ω is a càdlàg over \mathcal{X} [38]. Denote by $\mathcal{P}(\Omega)$ and $\mathcal{P}(\mathcal{X})$ the spaces of probability distributions over Ω and \mathcal{X} , respectively. Each diffusion process X_t ($t \in [0, 1]$) on \mathbb{R}^n with killing corresponds to a process \mathbf{X}_t taking values in \mathcal{X} , and thereby, to a law in $\mathcal{P}(\Omega)$.

Evidently, \mathcal{X} is a Polish space. The space Ω of càdlàg over \mathcal{X} is thus, with the appropriate topology, also a Polish space [39], [40]. Sanov's theorem applies to measures on Polish spaces and, therefore, the likelihood function is once again expressed in terms of the relative entropy between probability laws. In

our unbalanced SBP setting, the set of probability laws over path space $\mathcal{P}(\Omega)$ that are in alignment with the observations is

$$\{\mathbf{P} \in \mathcal{P}(\Omega) \mid \mathbf{P}_0 = p_0, \mathbf{P}_1 = p_1\},$$

where p_0, p_1 are the natural augmentation of ρ_0, ρ_1 so that they belong in $\mathcal{P}(\mathcal{X})$, respectively. Specifically, assuming that $\int_{\mathbb{R}^n} \rho_1(x) dx = 1$, we set

$$p_0 = (\rho_0(\cdot), 0) \tag{16a}$$

and

$$p_1 = (\rho_1(\cdot), 1 - \int_{\mathbb{R}^n} \rho_1(x) dx). \tag{16b}$$

Thus, we arrive at the following recasting of uSBP as an ordinary SBP.

Problem 6 (Unbalanced Schrödinger Bridge Problem (uSBP)). *Determine*

$$\mathbf{P}^* := \arg \min_{\mathbf{P} \in \mathcal{P}(\Omega)} \{H(\mathbf{P} \mid \mathbf{R}) \mid \mathbf{P}_0 = p_0, \mathbf{P}_1 = p_1\}. \tag{17}$$

As before, verbatim, $\mathbf{R}(\cdot) = \int_{\mathcal{X}^2} \mathbf{R}^{xy}(\cdot) \mathbf{R}_{01}(dxdy)$ and $\mathbf{P}(\cdot) = \int_{\mathcal{X}^2} \mathbf{P}^{xy}(\cdot) \mathbf{P}_{01}(dxdy)$, where now \mathbf{R}_{01} (\mathbf{P}_{01}) denotes the joint marginal distribution of \mathbf{R} (\mathbf{P}) over the marginal $\mathbf{X}_{0,1}$, and \mathbf{R}^{xy} (\mathbf{P}^{xy}) denotes the law conditioned on $\mathbf{X}_0 = x \in \mathcal{X}$ and $\mathbf{X}_1 = y \in \mathcal{X}$. As before, the relation to the static SBP emerges.

Problem 7. *Determine*

$$\boldsymbol{\pi}^* := \arg \min_{\boldsymbol{\pi} \in \mathcal{P}(\mathcal{X}^2)} \{H(\boldsymbol{\pi} \mid \mathbf{R}_{01}) \mid \boldsymbol{\pi}_0 = p_0, \boldsymbol{\pi}_1 = p_1\}. \tag{18}$$

The two formulations are once again equivalent, as it readily follows from the identity $H(\mathbf{P} \mid \mathbf{R}) = H(\mathbf{P}_{01} \mid \mathbf{R}_{01}) + \int_{\mathcal{X}^2} H(\mathbf{P}^{xy} \mid \mathbf{R}^{xy}) \mathbf{P}_{01}(dxdy)$.

Theorem 8. *Suppose \mathbf{P}^* solves the dynamic uSBP (17), then \mathbf{P}_{01}^* also solves the static uSBP (18). On the other hand, if $\boldsymbol{\pi}^*$ solves (18), then setting $\mathbf{P}^* = \int_{\mathcal{X}^2} \mathbf{R}^{xy}(\cdot) \boldsymbol{\pi}^*(dxdy)$ solves (17), while $\mathbf{P}_{01}^* = \boldsymbol{\pi}^*$.*

The Radon-Nikodym ratios between solutions and priors for the two problems, analogous to (8), applies here too, and the analogous expressions for the Schrödinger system in Theorem 4 follow as well. More explicitly, the solutions to (17) and (18) are of the form

$$\mathbf{P}^* = f(\mathbf{X}_0)g(\mathbf{X}_1)\mathbf{R} \tag{19}$$

and

$$\boldsymbol{\pi}^* = f(\mathbf{X}_0)g(\mathbf{X}_1)\mathbf{R}_{01} \tag{20}$$

respectively. The divergence between the standard SBP and the present uSBP becomes noticeable when we seek explicit solutions via analogues of system (14) and of the corresponding Fokker-Plank equation in Theorem 5, since now, we need to specify the update on the prior killing rate. We detail this next.

A. Generalized Schrödinger system

The Fokker-Planck equation for a diffusion (1) with killing rate $V(t, x)$ is

$$\partial_t R_t + \nabla \cdot (bR_t) + VR_t = \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij} R_t)}{\partial x_i \partial x_j}. \quad (21)$$

As before, $a(t, X) = \sigma(t, X)\sigma(t, X)'$ is assumed to be positive definite throughout. The corresponding Schrödinger system and its relation to the law of \mathbf{P}^* can be expressed after reparametrizing the pair (f, g) of functions on \mathcal{X} as follows

$$f(x)\mathbf{R}_0(x) = \begin{cases} \hat{\varphi}(0, x) & \text{if } x \in \mathbb{R}^n \\ \hat{\psi}(0) & \text{if } x = \mathbf{c}, \end{cases} \quad (22a)$$

$$g(y) = \begin{cases} \varphi(1, y) & \text{if } y \in \mathbb{R}^n \\ \psi(1) & \text{if } y = \mathbf{c}. \end{cases} \quad (22b)$$

Comparing with (14), the Schrödinger system along with the non-linear coupling constraints now becomes

$$\partial_t \hat{\varphi} = -\nabla \cdot (b\hat{\varphi}) - V\hat{\varphi} + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij}\hat{\varphi})}{\partial x_i \partial x_j} \quad (23a)$$

$$\frac{d\hat{\psi}}{dt} = \int V\hat{\varphi}(t, x) dx \quad (23b)$$

$$\partial_t \varphi = -b \cdot \nabla \varphi + V\varphi - \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} - V\psi \quad (23c)$$

$$\frac{d\psi}{dt} = 0 \quad (23d)$$

$$\rho_0 = \varphi(0, \cdot)\hat{\varphi}(0, \cdot) \quad (23e)$$

$$\rho_1 = \varphi(1, \cdot)\hat{\varphi}(1, \cdot) \quad (23f)$$

$$\hat{\psi}(0) = 0 \quad (23g)$$

$$\psi(1)\hat{\psi}(1) = 1 - \int \rho_1. \quad (23h)$$

Theorem 9. *Let R be the law of a diffusion (1) with nontrivial killing rate $V(t, x)$ and $a(t, x) = \sigma(t, x)\sigma(t, x)'$ being positive definite for all $(t, x) \in \mathbb{R} \times \mathbb{R}^n$, and assume that ρ_0, ρ_1 are absolutely continuous with respect to the Lebesgue measure. There exists a unique (up to a constant positive scaling) 4-tuple $(\hat{\varphi}(t, x), \hat{\psi}(t), \varphi(t, x), \psi(t))$ of non-negative functions that satisfies the Schrödinger system (23).*

The proof of the theorem, given in Appendix IX, is based on the contractiveness of the iterative scheme that consists in alternating between evaluation of $(\hat{\varphi}(1, \cdot), \hat{\psi}(1))$ from $(\varphi(0, \cdot), \psi(0))$ using (23e-23a-23g-23b), and then evaluating $(\varphi(0, \cdot), \psi(0))$ in a followup cycle from $(\hat{\varphi}(1, \cdot), \hat{\psi}(1))$ using the backward in time integration, via the remaining equations. Specifically, we prove that the iteration

$$\begin{pmatrix} \hat{\varphi}(1, \cdot) \\ \hat{\psi}(1) \end{pmatrix} \mapsto \begin{pmatrix} \varphi(0, \cdot) \\ \psi(0) \end{pmatrix} \mapsto \begin{pmatrix} \hat{\varphi}(1, \cdot) \\ \hat{\psi}(1) \end{pmatrix}_{\text{next}} \quad (24)$$

is strictly contractive in the Hilbert metric.

As in the ordinary SBP the discretized Schrödinger system (23) leads to an efficient algorithm to compute $(\hat{\varphi}(t, x), \hat{\psi}(t), \varphi(t, x), \psi(t))$, and thereby, \mathbf{P}^* as well as the corresponding Fokker-Planck equation for the corresponding marginals, that is explained next.

B. Dynamic formulation

In general, a multiplicative transformation such as $\mathbf{R} \rightarrow \mathbf{P}^* = f(\mathbf{X}_0)g(\mathbf{X}_1)\mathbf{R}$, preserves the Markovian character. Moreover, the generators of the respective semi-groups that relate in this way, herein, $\mathcal{L}_t^{\mathbf{R}}, \mathcal{L}_t^{\mathbf{P}^*}$, can be evaluated from one another directly by utilizing the multiplicative factors and the so-called carré du champ operator

$$\Gamma_t(u, v) := \mathcal{L}_t(uv) - u\mathcal{L}_t(v) - v\mathcal{L}_t(u).$$

Specifically (see [25, Equation (3.6)], and also [30], [41]),

$$\mathcal{L}_t^{\mathbf{P}^*} u(x) = \mathcal{L}_t^{\mathbf{R}} u(x) + \Gamma_t^{\mathbf{R}}(g_t, u)(x)/g_t(x), \quad (25)$$

where

$$g_t(y) = \begin{cases} \varphi(t, y) & \text{if } y \in \mathbb{R}^n \\ \psi(t) & \text{if } y = \mathbf{c}. \end{cases} \quad (26)$$

In light of Theorem 9 we now establish an explicit characterization of the dynamic unbalanced Schrödinger bridge problem (17). We denote by P_t the marginal of \mathbf{X}_t restricted to the first component in \mathcal{X} , and by q_t the probability of the coffin state. Thus, we use the vectorial notation

$$\mathbf{P}_t =: (P_t, q_t).$$

Accordingly, for the marginals $\mathbf{R}_t = (R_t, s_t)$ of the prior, R_t satisfies the Fokker-Planck equation (21) while $s_t = 1 - \int_{\mathbb{R}^n} R_t(x)dx$. The solution \mathbf{P}^* to (17) is then characterized by the following theorem.

Theorem 10. *The solution \mathbf{P}^* to (17) corresponds to a diffusion process*

$$dX_t = (b(t, X_t) + a(t, X_t)\nabla \log \varphi(t, X_t))dt + \sigma(t, X_t)dW_t \quad (27)$$

with killing rate $\psi V/\varphi$, where φ is obtained from the solution of the generalized Schrödinger system (23). Accordingly,

$$\partial_t P_t + \nabla \cdot ((b + a\nabla \log \varphi)P_t) = \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij}P_t)}{\partial x_i \partial x_j} - \frac{\psi}{\varphi} V P_t. \quad (28)$$

Proof. The generator of \mathbf{R} , is of the form

$$\mathcal{L}_t^{\mathbf{R}} : \begin{bmatrix} \varphi \\ \psi \end{bmatrix} \mapsto \begin{bmatrix} b \cdot \nabla \varphi - V\varphi + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} + V\psi \\ 0 \end{bmatrix}. \quad (29)$$

The carré du champ operator becomes

$$\Gamma_t^{\mathbf{R}} \left(\begin{bmatrix} \varphi_1 \\ \psi_1 \end{bmatrix}, \begin{bmatrix} \varphi_2 \\ \psi_2 \end{bmatrix} \right) = \begin{bmatrix} V\varphi_1\varphi_2 + a\nabla\varphi_1 \cdot \nabla\varphi_2 + V\psi_1\psi_2 - V\psi_1\varphi_2 - V\varphi_1\psi_2 \\ 0 \end{bmatrix}. \quad (30)$$

We readily obtain that

$$\begin{aligned} \mathcal{L}_t^{\mathbf{P}^*} \begin{bmatrix} v \\ \eta \end{bmatrix} &= \mathcal{L}_t^{\mathbf{R}} \begin{bmatrix} v \\ \eta \end{bmatrix} + \Gamma_t^{\mathbf{R}} \left(\begin{bmatrix} \varphi \\ \psi \end{bmatrix}, \begin{bmatrix} v \\ \eta \end{bmatrix} \right) / \begin{bmatrix} \varphi \\ \psi \end{bmatrix} \\ &= \begin{bmatrix} b \cdot \nabla v - Vv + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 v}{\partial x_i \partial x_j} + V\eta + Vv + a\nabla \log \varphi \cdot \nabla v + V\frac{\psi}{\varphi}\eta - V\frac{\psi}{\varphi}v - V\eta \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} (b + a\nabla \log \varphi) \cdot \nabla v - \frac{\psi}{\varphi}Vv + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 v}{\partial x_i \partial x_j} + \frac{\psi}{\varphi}V\eta \\ 0 \end{bmatrix}, \end{aligned} \quad (31)$$

where the division in (31) is carried out componentwise. The generator is that of the diffusion process (27) with killing rate $\psi V/\varphi$ over the extended state space \mathcal{X} . The Fokker-Planck equation (28) can be obtained by taking the dual of $\mathcal{L}_t^{\mathbf{P}^*}$. \square

Theorem 11. *The marginal distribution \mathbf{P}_t^* on the first component of \mathcal{X} is $P_t = \varphi(t, \cdot)\hat{\varphi}(t, \cdot)$, and on the second component of \mathcal{X} is $q_t = \psi(t)\hat{\psi}(t)$.*

Proof. We verify that P_t as above satisfies the Fokker-Planck equation associated with the diffusion (27) with killing rate $\psi V/\varphi$. To this end, let $P_t(\cdot) := \varphi(t, \cdot)\hat{\varphi}(t, \cdot)$. Then by (23a) and (23c) we obtain

$$\begin{aligned} 0 &= \partial_t P_t + \nabla \cdot ((b + a\nabla \log \varphi)P_t) - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2(a_{ij}P_t)}{\partial x_i \partial x_j} + \hat{\varphi}\psi V \\ &= \partial_t P_t + \nabla \cdot ((b + a\nabla \log \varphi)P_t) - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2(a_{ij}P_t)}{\partial x_i \partial x_j} + \frac{\psi}{\varphi} V P_t, \end{aligned}$$

which is exactly the desired Fokker-Planck equation (28). Similarly, by (23b) and (23d),

$$\frac{dq_t}{dt} = \psi(t) \int V \hat{\varphi}(t, x) dx = \int \frac{\psi}{\varphi} V P_t dx,$$

which is consistent with P_t and the new killing rate $\psi V/\varphi$. \square

IV. FLUID DYNAMIC FORMULATION

The original Schrödinger bridge problem, when there is no killing, is known to be equivalent to the stochastic control problem of minimizing control energy subject to the marginal two end-point constraints [26], or equivalently, to a fluid dynamic formulation whereby the velocity field $u(t, \cdot)$ effecting the flow minimizes this action integral, namely,

$$\min_{P_t(\cdot), u(t, \cdot)} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \|u(t, x)\|^2 P_t dx dt \quad (32a)$$

$$\partial_t P_t + \nabla \cdot ((b + \sigma u)P_t) - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2(a_{ij}P_t)}{\partial x_i \partial x_j} = 0 \quad (32b)$$

$$P_0 = \rho_0, \quad P_1 = \rho_1. \quad (32c)$$

The optimization takes place over the *feedback control policy-flow field* $u(t, x)$ together with the corresponding density flow $P_t(x)$. Below, in this section, we derive an analogous formulation for the Schrödinger bridge problems with unbalanced marginals.

Along the flow, the killing rate may deviate from the prior V and is to be determined. To quantify the deviation of the posterior killing rate from the prior, we introduce an entropic cost inside the action integral, to penalize changes in the ratio $\alpha(t, x)$ between the posterior and the prior killing rate. That is, α is an added optimization variable which is $\alpha(t, x) \geq 0$, and with the posterior killing rate being αV . To penalize differences between the posterior and the prior killing rates we introduce the factor

$$\alpha \log \alpha - \alpha + 1 \quad (33)$$

inside the action integral, which is convex and achieves the minimal value 0 at $\alpha = 1$. This entropy cost has been used in [25], [42] to study Schrödinger bridge problem over graphs. It is associated with the

large deviation principle for continuous-time Markov chain with discrete state. Combining this entropic cost term for the ratio of killing rates with (32) we arrive at

$$\min_{P,u,\alpha} \int_0^1 \int_{\mathbb{R}^n} \left[\frac{1}{2} \|u(t,x)\|^2 P_t + (\alpha \log \alpha - \alpha + 1) V P_t \right] dx dt \quad (34a)$$

$$\partial_t P_t + \nabla \cdot ((b + \sigma u) P_t) + \alpha V P_t - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij} P_t)}{\partial x_i \partial x_j} = 0 \quad (34b)$$

$$P_0 = \rho_0, \quad P_1 = \rho_1. \quad (34c)$$

Note that the control strategy has now two components, a drift term $u(t,x)$ and a correcting term $\alpha(t,x)$ for the killing rate.

Theorem 12. *Let $(\hat{\varphi}(t,x), \hat{\psi}(t), \varphi(t,x), \psi(t))$ be the solution to the Schrödinger system (23), then the solution to (34) is given by the choice*

$$u^*(t,x) = \sigma(t,x)' \nabla \log \varphi(t,x) \quad (35a)$$

$$\alpha^*(t,x) = \frac{\psi(t)}{\varphi(t,x)} \quad (35b)$$

$$P_t(x) = \varphi(t,x) \hat{\varphi}(t,x). \quad (35c)$$

Proof. We verify that conditions (35) ensure stationarity of the Lagrangian for (34). Introducing the Lagrange multiplier $\lambda(t,x)$, the Lagrangian for (34) is

$$\begin{aligned} \mathcal{L} = \int_0^1 \int \left[\frac{1}{2} \|u\|^2 P_t + (\alpha \log \alpha - \alpha + 1) V P_t \right. \\ \left. + \lambda \left(\partial_t P_t + \nabla \cdot ((b + \sigma u) P_t) + \alpha V P_t - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij} P_t)}{\partial x_i \partial x_j} \right) \right] dx dt. \end{aligned}$$

Applying integration by part we obtain

$$\begin{aligned} \mathcal{L} = \int_0^1 \int \left[\frac{1}{2} \|u\|^2 P_t + (\alpha \log \alpha - \alpha + 1) V P_t - P_t \partial_t \lambda - \nabla \lambda \cdot (b + \sigma u) P_t + \alpha V \lambda P_t \right. \\ \left. - \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \lambda}{\partial x_i \partial x_j} P_t \right] dx dt + \int \lambda(1,x) P_1(x) dx - \int \lambda(0,x) P_0(x) dx. \end{aligned} \quad (36)$$

Minimizing the above over u yields

$$u^*(t,x) = \sigma' \nabla \lambda. \quad (37a)$$

Similarly, minimization over α yields

$$\alpha^*(t,x) = \exp(-\lambda). \quad (37b)$$

Substituting (37) into (36) we obtain

$$\begin{aligned} \mathcal{L} = \int_0^1 \int P_t \left(-\frac{1}{2} a \nabla \lambda \cdot \nabla \lambda - b \cdot \nabla \lambda - \partial_t \lambda - \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \lambda}{\partial x_i \partial x_j} + V(1 - \exp(-\lambda)) \right) dx dt \\ + \int \lambda(1,x) P_1(x) dx - \int \lambda(0,x) P_0(x) dx. \end{aligned}$$

The optimality condition

$$\partial_t \lambda + b \cdot \nabla \lambda + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \lambda}{\partial x_i \partial x_j} + \frac{1}{2} a \nabla \lambda \cdot \nabla \lambda - V(1 - \exp(-\lambda)) = 0 \quad (38)$$

follows. Now, let

$$\lambda(t, x) = \log \frac{\varphi(t, x)}{\psi(t)}, \quad (39)$$

then (38) becomes

$$\partial_t \varphi - \frac{d\psi}{dt} \frac{\varphi}{\psi} + b \cdot \nabla \varphi - V\varphi + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} + V\psi = 0, \quad (40)$$

and by setting $d\psi/dt = 0$, the above reduces to (23c). Finally, plugging (39) into (37) yields (35). \square

Substituting the optimal control (35) into (34b) yields the closed loop dynamics under optimal control strategy. Clearly, it is the same as (28) associated with the solution \mathbf{P}^* to the uSBP (17).

V. SBP OVER REWEIGHTED PROCESSES

Some early attempts to formulate the Schrödinger Bridge Problem for diffusions with losses date back to Nagasawa and Wakolbinger [18], [15]. These focused on processes that are suitably reweighed via a Feynman-Kac multiplicative functional to model losses. Earlier relevant work on Schrödinger Bridges over reweighed processes includes [18], [15], [16], [17], [20], [30], [31]. In particular, e.g., [15, Section 8], and more recently, [30] discuss Feynman-Kac reweighing of the prior measure R , into $f(X_0) \exp\left(-\int_0^1 V(t, X_t) dt\right) g(X_1) R$. Such a process, with this special Radon-Nikodym derivative, is referred to as the h -transform of R . To distinguish this prior work from our uSBP formulation, we refer to the earlier formulation as *SBP over reweighted processes*.

Let $\hat{\rho}_1$ be a normalized version of ρ_1 so that $\hat{\rho}_1$ is a probability distribution, then the classical Schrödinger bridge problem over reweighted processes can be formulated as

$$\min_{P \in \mathcal{P}(\Omega)} \left\{ H(P | \hat{R}) \mid P_0 = \rho_0, P_1 = \hat{\rho}_1 \right\}, \quad (41)$$

where

$$\hat{R} = \exp\left(-\int_0^1 V(t, X_t) dt\right) R \quad (42)$$

is the (unnormalized) distribution induced by the survival trajectories of the diffusion process (1) with killing rate V . The solution to this problem reads

$$P^* = f(X_0) g(X_1) \hat{R} = f(X_0) \exp\left(-\int_0^1 V(t, X_t) dt\right) g(X_1) R \quad (43)$$

where the two multipliers f, g are chosen such that P^* satisfies the constraints $P_0 = \rho_0, P_1 = \hat{\rho}_1$. These two multipliers can again be obtained by solving a Schrödinger system. More specifically, let $\varphi, \hat{\varphi}$ be the solution to

$$\partial_t \hat{\varphi} = -\nabla \cdot (b\hat{\varphi}) - V\hat{\varphi} + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij}\hat{\varphi})}{\partial x_i \partial x_j} \quad (44a)$$

$$\partial_t \varphi = -b \cdot \nabla \varphi + V\varphi - \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \quad (44b)$$

$$\rho_0 = \varphi(0, \cdot) \hat{\varphi}(0, \cdot) \quad (44c)$$

$$\hat{\rho}_1 = \varphi(1, \cdot) \hat{\varphi}(1, \cdot), \quad (44d)$$

then $\varphi, \hat{\varphi}$ relate to f, g as

$$\hat{\varphi}(0, x) = f(x)\hat{R}_0(x) \quad (45a)$$

$$\varphi(1, y) = g(y). \quad (45b)$$

Unlike the solution \mathbf{P}^* to the uSBP (17), the solution P^* to (41) is a probability measure over $\Omega = C([0, 1], \mathbb{R}^n)$. Indeed, it is associated with the diffusion process

$$dX_t = (b(t, X_t) + a(t, X_t)\nabla \log \varphi(t, X_t))dt + \sigma(t, X_t)dW_t$$

without losses. The marginal distribution of it equals $P_t = \varphi(t, \cdot)\hat{\varphi}(t, \cdot)$ and is a probability measure over \mathbb{R}^n for all $t \in [0, 1]$. We argue that the SBP over weighted process doesn't address Schrödinger's original problem as described in the thought experiment in Section III. The prior \hat{R} describes the distribution of the surviving trajectories and the problem (41) can be interpreted as finding the most likely evolution of surviving trajectories that are compatible with the two marginals $\rho_0, \hat{\rho}_1$. However, the mechanism of how the particles that did not survive got killed is completely ignored in this formulation.

The importance of explicitly considering a possible update of the killing rate becomes salient when the end-point marginals are consistent with the prior law. Such a case highlights a dichotomy between our formulation of uSBP, and the rationale behind SBP over reweighted processes. To see this, consider a scenario where the two marginals are already consistent with the prior law, that is

$$\rho_1(\cdot) = \int_{\mathbb{R}^n} q(0, x, 1, \cdot)\rho_0(x)dx.$$

One would expect the solution to be the prior \hat{R} , *since the prior is consistent* with the end-point marginals. *This is, however, not the case!* Indeed, \hat{R}_0 represents the distribution at $t = 0$ of those particles that are destined to survive, and this differs from ρ_0 , the distribution of all particles. Thus, \hat{R}_0 is not the solution to (41).

One could attempt to modify Schrödinger's thought experiment by postulating that ρ_0 is precisely the distribution at $t = 0$ of those particles that eventually survive. With this modification, it is easy to see that the prior \hat{R} solves (41). *This modification, however, is not physical:* It is not possible to measure at time $t = 0$ the marginal of the survival trajectories!

Finally, we note that the Schrödinger bridge problem over reweighted processes has the following fluid dynamic (stochastic control) formulation

$$\min_{P_t(\cdot), u(t, \cdot)} \int_0^1 \int_{\mathbb{R}^n} [\frac{1}{2}\|u(t, x)\|^2 + V(t, x)]P_t dx dt \quad (46a)$$

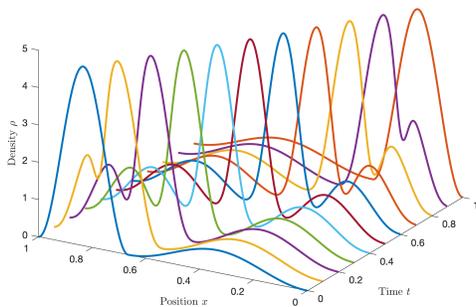
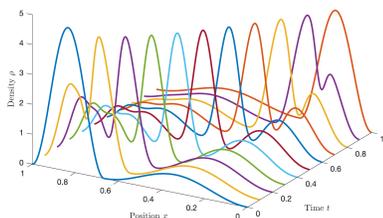
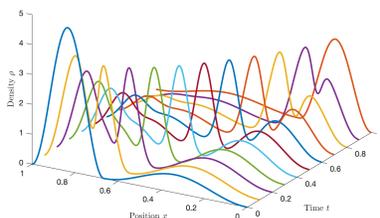
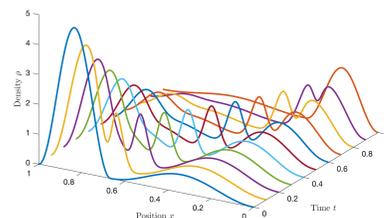
$$\partial_t P_t + \nabla \cdot ((b + \sigma u)P_t) - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 (a_{ij}P_t)}{\partial x_i \partial x_j} = 0 \quad (46b)$$

$$P_0 = \rho_0, \quad P_1 = \hat{\rho}_1. \quad (46c)$$

This stochastic control problem is over the diffusion process without losses

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

and the control $u(t, x)$ only enters the system through the drift. The prior killing rate V serves as a cost term. This is substantially different from the control formulation (34) of the uSBP where the control has a drift term $u(t, x)$ and a correcting term $\alpha(t, x)$, and the killing rate V appears in the dynamics instead of the cost function.

Fig. 1: Marginal flow of uSBP for $s = 1$ (a) $s = 0.8$ (b) $s = 0.6$ (c) $s = 0.4$ Fig. 2: Marginal flow of uSBP for $s = 0.8, 0.6, 0.4$

VI. NUMERICAL EXAMPLE

We conclude by highlighting the uSBP formalism with an academic/numerical example. To this end, we consider the diffusion process

$$dX_t = \sigma dW_t,$$

with $X_t, W_t \in \mathbb{R}$ (i.e., in a 1-dimensional state space), $\sigma = 0.05$, and killing rate

$$V(t, x) \equiv 1.$$

We work out the solution of the unbalanced Schrödinger bridge problem (uSBP) with initial marginal density

$$\rho_0(x) = \begin{cases} 0.3 - 0.3 \cos(3\pi x) & \text{if } 0 \leq x < 2/3 \\ 2.4 - 2.4 \cos(6\pi x - 4\pi) & \text{if } 2/3 \leq x \leq 1, \end{cases}$$

and target marginal density

$$\rho_1(x) = s\rho_0(1 - x),$$

where $s \leq 1$ denotes the percentage of survival particles.

Figures 1 and 2 display the marginal flow of the uSBP for different values of s . When $s < 1$, only a portion of the particles survive until the end and many particles vanish along the way. Thus, the total mass of the particles is a decreasing function of time, as can be seen from Figure 3. Note that the terminal percentage of surviving particles is consistent with the chosen value for s , in each case.

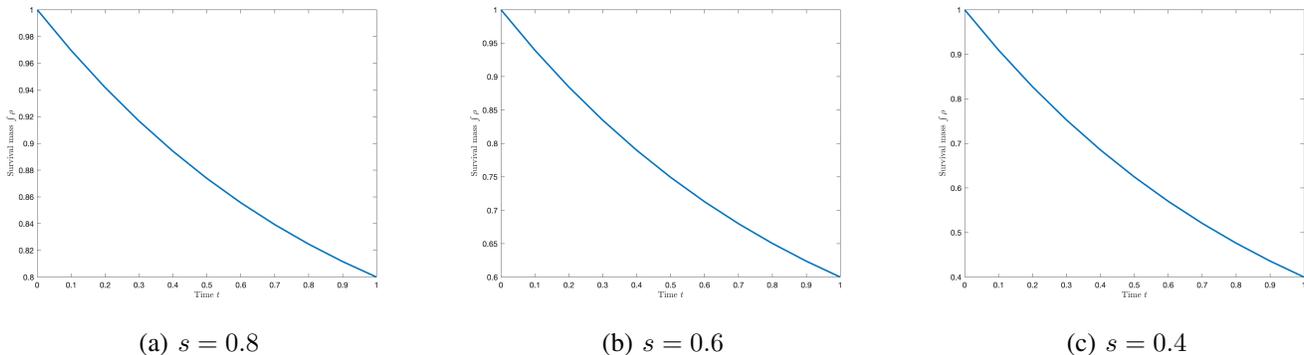


Fig. 3: Survival mass of uSBP for $s = 0.8, 0.6, 0.4$

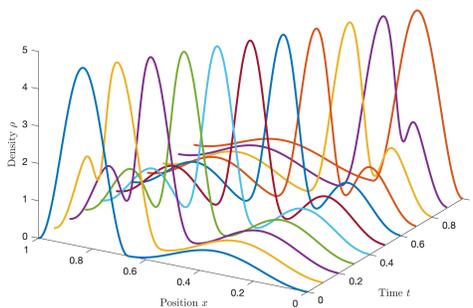


Fig. 4: Marginal flow of SBP over reweighted processes

For comparison, we also display the solution to the SBP over reweighted processes in Figure 4. Note that its solution is independent of s . The solution describes the posterior distribution of the survived particles only, and thus the marginal flow remains a probability measure at all times. In fact, it coincides with uSBP for $s = 1$.

VII. CONCLUDING REMARKS

We introduced Schrödinger bridges between unbalanced marginals in the spirit of E. Schrödinger's original rationale (that led to the standard SBP), aimed to reconcile a given prior law, that now includes a killing rate, with marginal observations. We formulated the problem as maximum entropy problem over an augmented state space that includes a coffin state representing the state of vanishing particles. The solution is characterized by a Schrödinger-type system, different to the classical one, that yields a diffusion process whose drift as well as killing rate are suitable adjusted as compared to the prior. Just like in the standard SBP, this new unbalanced Schrödinger bridge problem (uSBP) can be formulated as a stochastic control problem. Naturally, departing from the standard SBP, the control variable in this control formulation includes both the drift and killing rate. We underscore an apparent dichotomy between our formulation of the uSBP and earlier work on SBP over reweighted processes with Feynman-Kac functionals. Though both pertain to SBP's for diffusions with losses, we argued that our uSBP is a natural formulation in the spirit of Schrödinger's original quest to reconcile probabilistic models with observations. The nature of the zero-noise limit of the uSBP and its relation to a corresponding optimal transport flow between unbalanced marginals is left as a topic of future research.

VIII. APPENDIX A: HILBERT'S PROJECTIVE METRIC

Herein we discuss Hilbert's projective metric and highlight some important contraction theorems due to Garrett Birkhoff and P. J. Bushell [43], [44], [45] that we use in this work. A first application of the Birkhoff-Bushell contractive maps to scaling of nonnegative matrices, a topic closely connected to Schrödinger bridges, was presented in [46]. In [47], it was shown that the Schrödinger bridge for Markov chains and quantum channels can be efficiently obtained from the fixed-point of a map which contracts the Hilbert metric. We refer to [37], [27] for more detailed information and further applications of this metric. Below, following [45], we recall some basic concepts and results of this theory.

Let \mathcal{B} be a real Banach space and let \mathcal{K} be a closed solid cone in \mathcal{B} , i.e., \mathcal{K} is closed with nonempty interior \mathcal{K}_0 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

$$\mathbf{x} \preceq \mathbf{y} \Leftrightarrow \mathbf{y} - \mathbf{x} \in \mathcal{K}, \quad \mathbf{x} \prec \mathbf{y} \Leftrightarrow \mathbf{y} - \mathbf{x} \in \mathcal{K}_0$$

and for $\mathbf{x}, \mathbf{y} \in \mathcal{K}^+ := \mathcal{K} \setminus \{0\}$, define

$$\begin{aligned} M(\mathbf{x}, \mathbf{y}) &:= \inf \{ \lambda \mid \mathbf{x} \preceq \lambda \mathbf{y} \} \\ m(\mathbf{x}, \mathbf{y}) &:= \sup \{ \lambda \mid \lambda \mathbf{y} \preceq \mathbf{x} \}. \end{aligned}$$

Then, the Hilbert metric is defined on \mathcal{K}^+ by

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

It is easily seen that $d_H(\cdot, \cdot)$ is symmetric, i.e., that $d_H(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{y}, \mathbf{x})$, and invariant under scaling by positive constants, since $d_H(\mathbf{x}, \mathbf{y}) = d_H(\lambda \mathbf{x}, \lambda \mathbf{y})$ for any $\lambda > 0$ and $\mathbf{x}, \mathbf{y} \in \mathcal{K}_0$. Therefore $d_H(\lambda \mathbf{x}, \mathbf{x}) = 0$. It can also be shown that the triangular inequality holds and, therefore, $d_H(\cdot, \cdot)$ is a *projective* metric that represents distance between rays.

In our analysis we encounter two types of maps. We encounter inversion

$$\mathcal{E}_{\text{inv}} : \mathbf{x} \mapsto \mathbf{x}^{-1}, \tag{47}$$

of elements $\mathbf{x} \in \mathcal{K}_0$, and also linear maps that are *positive*, namely,

$$\mathcal{E} : \mathcal{K}^+ \rightarrow \mathcal{K}^+.$$

For both types of maps we are interested in determining their *contraction ratio*,

$$\kappa(\mathcal{E}) := \inf \{ \lambda \mid d_H(\mathcal{E}(\mathbf{x}), \mathcal{E}(\mathbf{y})) \leq \lambda d_H(\mathbf{x}, \mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathcal{K}_0 \}.$$

It turns out that the former are isometries in the Hilbert metric whereas the latter are contractions. Thus, the composition of a combination of both types turns out to be a contraction.

That (47) is an isometry, i.e., $\kappa(\mathcal{E}_{\text{inv}}) = 1$, follows immediately from

$$M(\mathbf{x}, \mathbf{y}) = \frac{1}{m(\mathbf{x}^{-1}, \mathbf{y}^{-1})}.$$

Then, by G. Birkhoff's theorem [43], [45], any positive linear map \mathcal{E} is contractive and the contraction ratio can be expressed in terms of the *projective diameter*

$$\Delta(\mathcal{E}) := \sup \{ d_H(\mathcal{E}(\mathbf{x}), \mathcal{E}(\mathbf{y})) \mid \mathbf{x}, \mathbf{y} \in \mathcal{K}_0 \}$$

of the range of \mathcal{E} . Specifically, under these conditions, G. Birkhoff's theorem states that

$$\kappa(\mathcal{E}) = \tanh\left(\frac{1}{4}\Delta(\mathcal{E})\right). \quad (48)$$

Thus, a positive linear map is strictly contractive if its projective diameter $\Delta(\mathcal{E})$ is finite. A further useful observation, that follows from the triangular inequality, is that for any $\mathbf{x}_0 \in \mathcal{K}_0$,

$$\Delta(\mathcal{E}) \leq 2 \sup\{d_H(\mathcal{E}(\mathbf{x}), \mathbf{x}_0) \mid \mathbf{x} \in \mathcal{K}_0\}. \quad (49)$$

This allows bounding $\Delta(\mathcal{E})$ to ensure strict contraction for \mathcal{E} .

Important examples are provided by the positive orthant of \mathbb{R}^n , the cone of Hermitian, positive semidefinite matrices, spaces of bounded positive functions, and so on. Notice that, in all cases, the boundary of the cone lies at an infinite distance from any interior point.

IX. APPENDIX B: PROOF OF THEOREM 9 ON THE GENERALIZED SCHRÖDINGER SYSTEM

We herein establish existence and uniqueness of solution (up to constant positive scaling) for the system (23). The steps mimic the analogous case for the SBP where the marginals are supported on a Euclidean space [37]. The difference at present lies in that the support of functions includes an added point that represents the coffin state.

We assume throughout that the marginal measures ρ_0, ρ_1 are absolutely continuous with respect to the Lebesgue measure, in that $\rho_0(dx) = \rho_0(x)dx$ and $\rho_1(dx) = \rho_1(x)dx$ for density functions ρ_0, ρ_1 with support $S_0, S_1 \subseteq \mathbb{R}^n$, respectively, and that ρ_0 is a probability measure while ρ_1 is a nonnegative measure with $\int_{S_1} \rho_1(x)dx \leq 1$. The case $\int_{S_1} \rho_1(x)dx = 1$ reduces to the standard SBP and is easy to handle. Thus, without loss of generality, we assume

$$\int_{S_1} \rho_1(x)dx > 1.$$

As before, we let $q(0, x_0, t, x)$ for $0 < t \leq 1$ denote the fundamental solution of equation (23a) and assume that it is continuous and strictly positive on compact subsets. This is guaranteed by sufficient smoothness of the coefficients b, V, a , positivity of V and positive definiteness on the whole domain of the matrix $a = (a_{ij})$. Under these assumptions we rewrite the Schrödinger system (23) as follows,

$$\hat{\varphi}(t, x) = \int_{\mathbb{R}^n} q(0, x_0, t, x) \hat{\varphi}(0, x_0) dx_0, \quad (50a)$$

$$\hat{\psi}(1) = \int_0^1 \int_{\mathbb{R}^n} V(t, x) \hat{\varphi}(t, x) dx dt \quad (50b)$$

$$\varphi(0, x_0) = \int_{\mathbb{R}^n} q(0, x_0, 1, x_1) \varphi(1, x_1) dx_1 + \int_0^1 \int_{\mathbb{R}^n} q(0, x_0, t, x) V(t, x) \psi(t) dx dt \quad (50c)$$

$$\psi(t) = \text{constant}, \quad 0 \leq t \leq 1, \quad (50d)$$

$$\rho_0(x_0) = \varphi(0, x_0) \hat{\varphi}(0, x_0) \quad (50e)$$

$$\rho_1(x_1) = \varphi(1, x_1) \hat{\varphi}(1, x_1) \quad (50f)$$

$$\psi(0) \hat{\psi}(0) = 1 - \int_{S_0} \rho_0 = 0 \quad (50g)$$

$$\psi(1) \hat{\psi}(1) = 1 - \int_{S_1} \rho_1 =: c_1 > 0. \quad (50h)$$

We consolidate the system of equations (50) into

$$\hat{\varphi}(t, x) = \int_{S_0} q(0, x_0, t, x) \rho_0(x_0) \frac{1}{\varphi(0, x_0)} dx_0 \quad (51a)$$

$$\begin{aligned} \hat{\psi}(1) &= \int_0^1 \int_{\mathbb{R}^n} V(t, x) \hat{\varphi}(t, x) dx dt \\ &= \int_{S_0} \frac{1}{\varphi(0, x_0)} \rho_0(x_0) \underbrace{\int_0^1 \int_{\mathbb{R}^n} q(0, x_0, t, x) V(t, x) dx dt}_{r(x_0)} dx_0 \end{aligned} \quad (51b)$$

$$\varphi(0, x_0) = \int_{S_1} q(0, x_0, 1, x_1) \rho_1(x_1) \frac{1}{\hat{\varphi}(1, x_1)} dx_1 + \frac{1}{\hat{\psi}(1)} c_1 \underbrace{\int_0^1 \int_{\mathbb{R}^n} q(0, x_0, t, x) V(t, x) dx dt}_{r(x_0)} \quad (51c)$$

$$\psi(0) = \psi(1) = \frac{1}{\hat{\psi}(1)} c_1. \quad (51d)$$

These four equations, that encapsulate the Schrödinger system, suggest considering the composition of maps

$$\begin{pmatrix} \hat{\varphi}(1, \cdot) \\ \hat{\psi}(1) \end{pmatrix} \xrightarrow{\mathcal{E}_1} \begin{pmatrix} \frac{1}{\hat{\varphi}(1, \cdot)} \\ \frac{1}{\hat{\psi}(1)} \end{pmatrix} \xrightarrow{\mathcal{E}_2} \begin{pmatrix} \varphi(0, \cdot) \\ \psi(0) \end{pmatrix} \xrightarrow{\mathcal{E}_3} \begin{pmatrix} \frac{1}{\varphi(0, \cdot)} \\ \frac{1}{\psi(0)} \end{pmatrix} \xrightarrow{\mathcal{E}_4} \begin{pmatrix} \hat{\varphi}(1, \cdot) \\ \hat{\psi}(1) \end{pmatrix}_{\text{next}} \quad (52)$$

in order to analyze existence of solutions. Indeed, we utilize the theory of the Hilbert metric (outlined in Appendix VIII) to show that the composition is a strict contraction along rays, resulting in a unique fixed point.

To this end, we consider the Banach space $\mathcal{B} = \mathcal{L}^\infty(\mathcal{X})$ of *real-valued* functions $h(\cdot)$ on $\mathcal{X} = S \cup \{\mathfrak{c}\}$, where $S \in \mathbb{R}^n$ satisfies that $S_0 \cup S_1 \subset S$. For notational convenience we use the vectorial notation $\begin{pmatrix} h(x) \\ h(\mathfrak{c}) \end{pmatrix}$ to specify the values of h on the two constituents of its support, for $x \in \mathbb{R}^n$ and $\mathfrak{c} \in \{\mathfrak{c}\}$. Thus, the norm of h is

$$\|h\| := \max\{\|h|_S\|_\infty, |h(\mathfrak{c})|\}.$$

We consider the cone of positive functions

$$\mathcal{K} = \{h \in \mathcal{B} \mid h(\mathfrak{c}) \geq 0 \text{ and } h(x) \geq 0 \text{ a.e. } x \in S\}$$

and the corresponding partial order $h_1 \preceq h_2 \Leftrightarrow h_2 - h_1 \in \mathcal{K}$ as usual. We observe that \mathcal{K} is closed, solid and has a non-empty interior (of strictly positive a.e. functions) that we denote \mathcal{K}_0 ; we also denote $\mathcal{K}^+ := \mathcal{K} \setminus \{0\}$.

Note that in the on-going development, the components of functions $h \in \mathcal{B}$, that are (possibly time-dependent) functions on \mathbb{R}^n and $\{\mathfrak{c}\}$, respectively, are differentiated as φ, ψ , or $\hat{\varphi}, \hat{\psi}$, respectively, e.g., $\begin{pmatrix} h(x) \\ h(\mathfrak{c}) \end{pmatrix} = \begin{pmatrix} \varphi(t, x) \\ \psi(t) \end{pmatrix}$. We proceed to consider the composition of maps in (52) and establish first the following weaker version of Theorem 9:

Theorem 13. *Assuming that the support sets S_0, S_1 of the two marginals ρ_0, ρ_1 of the uSBP are compact, the claim in Theorem 9 holds true.*

Recall the notation $M(\cdot, \cdot), m(\cdot, \cdot), \kappa(\cdot)$ and $\Delta(\cdot)$ from Appendix VIII. As noted in the appendix, since $M(h_1, h_2) = m(h_1^{-1}, h_2^{-1})^{-1}$ for $h_1, h_2 \in \mathcal{K}_0$, both \mathcal{E}_1 and \mathcal{E}_3 are isometries. They are readily extended to isometries on \mathcal{K}^+ as well.

The map \mathcal{E}_2 is linear (homogeneous of degree 1) and therefore, by Birkhoff's theorem given in the appendix, contractive on \mathcal{K}^+ . For the same reason, \mathcal{E}_4 is contractive. Unfortunately, neither map is strictly contractive. To see this, note that since, e.g., $\mathcal{E}_2\left(\begin{smallmatrix} \star \\ 0 \end{smallmatrix}\right) = \begin{smallmatrix} \star \\ 0 \end{smallmatrix}$, with \star denoting nonzero entries, certain elements on the boundary of \mathcal{K}^+ map onto the boundary and not the interior.

In order to establish the theorem we proceed as follows. Let $z \in S_0$ be an arbitrary fixed point in S_0 . We modify equation (51d) of the Schrödinger system (51), replacing it with

$$\tilde{\psi}(0) = \varphi(0, z) = \int_{S_1} q(0, z, 1, x_1) \rho_1(x_1) \frac{1}{\hat{\varphi}(1, x_1)} dx_1 + \frac{1}{\hat{\psi}(1)} c_1 r(z), \quad (51d')$$

and, accordingly, replace \mathcal{E}_2 with a corresponding map that we refer to as \mathcal{E}'_2 . We then show the existence and uniqueness of solution for the modified system. Interestingly, except for the last of the elements in the 4-tuple $(\hat{\varphi}(1, x), \hat{\psi}(1), \varphi(0, x), \psi(0))$, namely, $\psi(0)$, the remaining dictate the sought solution of the original Schrödinger system (23). This last entry plays no role in the original Schrödinger system. In particular,

$$\mathcal{E}_4 \circ \mathcal{E}_3 \circ \mathcal{E}_2 \circ \mathcal{E}_1 = \mathcal{E}_4 \circ \mathcal{E}_3 \circ \mathcal{E}'_2 \circ \mathcal{E}_1 =: \mathcal{C}. \quad (53)$$

We now consider $\mathcal{E}'_2 : h \mapsto g$ and show that it is strictly contractive in the Hilbert metric. From (49), taking as \mathbf{x}_0 the function which is identically equal to 1 on S as well as on $\{\mathbf{c}\}$, we deduce that

$$\Delta(\mathcal{E}'_2) \leq 2 \sup \left\{ \log \left(\frac{\max\{\sup_x g(x), g(\mathbf{c})\}}{\min\{\inf_x g(x), g(\mathbf{c})\}} \right) \mid g = \mathcal{E}'_2(h) \text{ and } h \in \mathcal{K}_0 \right\} \quad (54)$$

Since ρ_0, ρ_1 are supported on compact sets S_0, S_1 of \mathbb{R}^n , respectively, we can choose S to be compact as well. Since the kernel q is positive and continuous, the kernel is bounded from below and above on $S \times S$. I.e., there exist $0 < \alpha_1 \leq \beta_1 < \infty$ such that

$$\alpha_1 \leq q(0, x, 1, y) \leq \beta_1, \quad (55)$$

for all $(x, y) \in S \times S$. Similarly, there exist $0 < \alpha_2 \leq \beta_2 < \infty$ such that

$$\alpha_2 \leq r(x) \leq \beta_2 \quad (56)$$

for all $x \in S$.

Let $h(x) = \frac{1}{\hat{\varphi}(1, x)}$ and $h(\mathbf{c}) = \frac{1}{\hat{\psi}(1)}$, then

$$\alpha_1 \int_{S_1} \rho_1(x_1) h(x_1) dx_1 \leq \int_{S_1} q(0, x_0, 1, x_1) \rho_1(x_1) h(x_1) dx_1 \leq \beta_1 \int_{S_1} \rho_1(x_1) h(x_1) dx_1, \quad \forall x_0 \in S.$$

It follows that, in view of (51c),

$$\frac{\sup_x g(x)}{\inf_x g(x)} \leq \frac{\max_{i \in \{1, 2\}} \beta_i}{\min_{i \in \{1, 2\}} \alpha_i} < \infty.$$

Thanks to the modification (51d'), $g(\mathbf{c}) = g(z)$ and therefore

$$\frac{\max\{\sup_x g(x), g(\mathbf{c})\}}{\min\{\inf_x g(x), g(\mathbf{c})\}} \leq \frac{\max_{i \in \{1, 2\}} \beta_i}{\min_{i \in \{1, 2\}} \alpha_i} < \infty.$$

Thus, from (54) and using Birkhoff's theorem (48),

$$\kappa(\mathcal{E}'_2) < 1.$$

As a consequence, the composition $\mathcal{E}_4 \circ \mathcal{E}_3 \circ \mathcal{E}'_2 \circ \mathcal{E}_1$ is strictly contractive, i.e.,

$$\kappa(\mathcal{E}_4 \circ \mathcal{E}_3 \circ \mathcal{E}'_2 \circ \mathcal{E}_1) < 1.$$

It follows from (53) that

$$\kappa(\mathcal{C}) = \kappa(\mathcal{E}_4 \circ \mathcal{E}_3 \circ \mathcal{E}_2 \circ \mathcal{E}_1) < 1.$$

The above condition ensures that \mathcal{C} has a unique fixed point in terms of the Hilbert metric [37]. Since Hilbert metric is a projective metric, the uniqueness is up to a constant scaling. Denote the fixed point on the unit sphere U by h , then

$$\mathcal{C}(h) = \lambda h$$

for some positive number λ . We next show $\lambda = 1$. To this end, we introduce a different factorization of \mathcal{C} as

$$\mathcal{C} = \mathcal{E}^\dagger \circ \mathcal{E}_{p_0} \circ \mathcal{E} \circ \mathcal{E}_{p_1},$$

where

$$\begin{aligned} \mathcal{E}(u) &= \begin{bmatrix} \int_{S_1} q(0, x, 1, x_1) u(x_1) dx_1 + r(x) u(\mathbf{c}) \\ u(\mathbf{c}) \end{bmatrix} \\ \mathcal{E}_{p_0}(u) &= \begin{bmatrix} \frac{\rho_0(x)}{u(x)} \\ 0 \end{bmatrix} \\ \mathcal{E}_{p_1}(u) &= \begin{bmatrix} \frac{\rho_1(x)}{u(x)} \\ \frac{c_1}{u(\mathbf{c})} \end{bmatrix}, \end{aligned}$$

and \mathcal{E}^\dagger is the adjoint operator of \mathcal{E} . Clearly,

$$\langle u, \mathcal{E}_{p_0}(u) \rangle = \langle \mathcal{E}_{p_1}(u), u \rangle = 1, \quad \forall u \in \mathcal{K}_0.$$

It follows that

$$\begin{aligned} 1 &= \langle \mathcal{E} \circ \mathcal{E}_{p_1}(h), \mathcal{E}_{p_0} \circ \mathcal{E} \circ \mathcal{E}_{p_1}(h) \rangle \\ &= \langle \mathcal{E}_{p_1}(h), \mathcal{E}^\dagger \circ \mathcal{E}_{p_0} \circ \mathcal{E} \circ \mathcal{E}_{p_1}(h) \rangle \\ &= \langle \mathcal{E}_{p_1}(h), \mathcal{C}(h) \rangle \\ &= \langle \mathcal{E}_{p_1}(h), \lambda h \rangle = \lambda. \end{aligned}$$

Once the fixed point h is computed, the 4-tuple $(\hat{\varphi}(1, x), \hat{\psi}(1), \varphi(0, x), \psi(0))$ can be recovered by

$$\hat{\varphi}(1, x) = h(x), \quad \hat{\psi}(1) = h(\mathbf{c}),$$

and

$$\begin{bmatrix} \varphi(0, \cdot) \\ \psi(0) \end{bmatrix} = \mathcal{E}_2 \circ \mathcal{E}_1(h).$$

The uniqueness of the 4-tuple $(\hat{\varphi}(1, x), \hat{\psi}(1), \varphi(0, x), \psi(0))$ follows from the uniqueness of the fixed point h . This completes the proof of Theorem 13. A standard argument [37, Theorem 3.5] can be used to extend the proof to the setting where S_0, S_1 are not necessarily compact for Theorem 9.

REFERENCES

- [1] S. S. Varadhan, "Asymptotic probabilities and differential equations," *Communications on Pure and Applied Mathematics*, vol. 19, no. 3, pp. 261–286, 1966.
- [2] S. S. Varadhan, *Large deviations and applications*. SIAM, 1984.
- [3] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.
- [4] E. Schrödinger, "Über die Umkehrung der Naturgesetze," *Sitzungsberichte der Preuss Akad. Wissen. Phys. Math. Klasse, Sonderausgabe*, vol. IX, pp. 144–153, 1931.
- [5] E. Schrödinger, "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique," in *Annales de l'institut Henri Poincaré*, vol. 2, no. 4. Presses universitaires de France, 1932, pp. 269–310.
- [6] L. Boltzmann, "Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. dem Sätzen über das Wärmegleichgewicht," *Wiener Berichte*, vol. 76, pp. 373–435, 1877.
- [7] I. N. Sanov, "On the probability of large deviations of random variables," *Mat. Sb. N. S. / Selected Translations in Mathematical Statistics and Probability*, vol. 42 / 1, pp. 11–44 / 213–244, 1957 / 1961.
- [8] R. Fortet, "Résolution d'un système d'équations de M. Schrödinger," *J. Math. Pure Appl. IX*, vol. 1, pp. 83–105, 1940.
- [9] A. Beurling, "An automorphism of product measures," *Annals of Mathematics*, pp. 189–200, 1960.
- [10] B. Jamison, "Reciprocal processes," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 30, pp. 65–86, 1974.
- [11] B. Jamison, "The Markov processes of Schrödinger," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 32, no. 4, pp. 323–331, 1975.
- [12] H. Föllmer, "Random fields and diffusion processes," in *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*. Springer, 1988, pp. 101–203.
- [13] J. Zambrini, "Stochastic mechanics according to E. Schrödinger," *Physical review A*, vol. 33, no. 3, p. 1532, 1986.
- [14] M. Nagasawa, "Transformations of diffusion and Schrödinger processes," *Probability theory and related fields*, vol. 82, no. 1, pp. 109–136, 1989.
- [15] A. Wakolbinger, "A simplified variational characterization of Schrödinger processes," *Journal of mathematical physics*, vol. 30, no. 12, pp. 2943–2946, 1989.
- [16] A. Blaquièrre, "Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version)," *Dynamics and Control*, vol. 2, no. 3, pp. 235–253, 1992.
- [17] D. Dawson, L. Gorostiza, and A. Wakolbinger, "Schrödinger processes and large deviations," *Journal of mathematical physics*, vol. 31, no. 10, pp. 2385–2388, 1990.
- [18] M. Nagasawa, "Stochastic variational principle of Schrödinger processes," in *Seminar on Stochastic Processes, 1989*. Springer, 1990, pp. 165–175.
- [19] A. Wakolbinger, "Schrödinger bridges from 1931 to 1991," in *Proc. of the 4th Latin American Congress in Probability and Mathematical Statistics, Mexico City, 1990*, pp. 61–79.
- [20] R. Aebi and M. Nagasawa, "Large deviations and the propagation of chaos for Schrödinger processes," *Probability Theory and Related Fields*, vol. 94, no. 1, pp. 53–68, 1992.
- [21] T. Mikami, "Monge's problem with a quadratic cost by the zero-noise limit of h-path processes," *Probability theory and related fields*, vol. 129, no. 2, pp. 245–260, 2004.
- [22] T. Mikami and M. Thieullen, "Optimal transportation problem by stochastic optimal control," *SIAM Journal on Control and Optimization*, vol. 47, no. 3, pp. 1127–1139, 2008.
- [23] Y. Chen., T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part I," *IEEE Trans. on Automatic Control*, vol. 61, no. 5, pp. 1158–1169, 2016.
- [24] C. Léonard, "From the Schrödinger problem to the Monge–Kantorovich problem," *Journal of Functional Analysis*, vol. 262, no. 4, pp. 1879–1920, 2012.

- [25] C. Léonard, “A survey of the Schrödinger problem and some of its connections with optimal transport,” *Discrete Contin. Dyn. Syst. A*, vol. 34, no. 4, pp. 1533–1574, 2014.
- [26] Y. Chen, T. T. Georgiou, and M. Pavon, “On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint,” *Journal of Optimization Theory and Applications*, vol. 169, no. 2, pp. 671–691, 2016.
- [27] Y. Chen, T. T. Georgiou, and M. Pavon, “Stochastic control liaisons: Richard Sinkhorn Meets Gaspard Monge on a Schrödinger bridge,” *SIAM Review*, vol. 63, no. 2, pp. 249–313, 2021.
- [28] G. Conforti, “A second order equation for schrödinger bridges with applications to the hot gas experiment and entropic transportation cost,” *Probability Theory and Related Fields*, vol. 174, no. 1, pp. 1–47, 2019.
- [29] P. Koehl, M. Delarue, and H. Orland, “Physics approach to the variable-mass optimal-transport problem,” *Physical Review E*, vol. 103, no. 1, p. 012113, 2021.
- [30] C. Léonard, “Stochastic derivatives and generalized h-transforms of Markov processes,” *arXiv preprint arXiv:1102.3172*, 2011.
- [31] Y. Chen, T. T. Georgiou, and M. Pavon, “Optimal steering of inertial particles diffusing anisotropically with losses,” in *Proc. American Control Conf.*, 2015, pp. 1252–1257.
- [32] Y. Chen, T. Georgiou, and M. Pavon, “Optimal steering of a linear stochastic system to a final probability distribution, Part III,” *arXiv:1608.03622, IEEE Trans. on Automatic Control*, to appear, 2017.
- [33] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, “Scaling algorithms for unbalanced optimal transport problems,” *Mathematics of Computation*, vol. 87, no. 314, pp. 2563–2609, 2018.
- [34] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, “Unbalanced optimal transport: Dynamic and Kantorovich formulations,” *Journal of Functional Analysis*, vol. 274, no. 11, pp. 3090–3123, 2018.
- [35] Y. Chen, T. T. Georgiou, and A. Tannenbaum, “Interpolation of matrices and matrix-valued densities: The unbalanced case,” *European Journal of Applied Mathematics*, vol. 30, no. 3, pp. 458–480, 2019.
- [36] Y. Chen, T. T. Georgiou, and M. Pavon, “Optimal transport in systems and control,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 89–113, 2021.
- [37] Y. Chen, T. Georgiou, and M. Pavon, “Entropic and displacement interpolation: a computational approach using the Hilbert metric,” *SIAM Journal on Applied Mathematics*, vol. 76, no. 6, pp. 2375–2396, 2016.
- [38] P. Billingsley, *Convergence of probability measures*. John Wiley & Sons, 1999.
- [39] D. Pollard, *Convergence of stochastic processes*. Springer Science & Business Media, 2012.
- [40] S. N. Ethier and T. G. Kurtz, *Markov processes: characterization and convergence*. John Wiley & Sons, 2009, vol. 282.
- [41] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*. Springer Science & Business Media, 2013, vol. 293.
- [42] C. Léonard, “Lazy random walks and optimal transport on graphs,” *The annals of Probability*, vol. 44, no. 3, pp. 1864–1915, 2016.
- [43] G. Birkhoff, “Extensions of Jentzsch’s theorem,” *Transactions of the American Mathematical Society*, vol. 85, no. 1, pp. 219–227, 1957.
- [44] P. Bushell, “On the projective contraction ratio for positive linear mappings,” *Journal of the London Mathematical Society*, vol. 2, no. 2, pp. 256–258, 1973.
- [45] P. J. Bushell, “Hilbert’s metric and positive contraction mappings in a Banach space,” *Archive for Rational Mechanics and Analysis*, vol. 52, no. 4, pp. 330–338, 1973.
- [46] J. Franklin and J. Lorenz, “On the scaling of multidimensional matrices,” *Linear Algebra and its applications*, vol. 114, pp. 717–735, 1989.
- [47] T. T. Georgiou and M. Pavon, “Positive contraction mappings for classical and quantum Schrödinger systems,” *Journal of Mathematical Physics*, vol. 56, no. 3, p. 033301, 2015.