On the Contraction Coefficient of the Schrödinger Bridge for Stochastic Linear Systems

Alexis M. H. Teter, *Student Member, IEEE*, Yongxin Chen, *Senior Member, IEEE*, and Abhishek Halder, *Senior Member, IEEE*

Abstract—Schrödinger bridge is a stochastic optimal control problem to steer a given initial state density to another, subject to controlled diffusion and deadline constraints. A popular method to numerically solve the Schrödinger bridge problems, in both classical and in the linear system settings, is via contractive fixed point recursions. These recursions can be seen as dynamic versions of the well-known Sinkhorn iterations, and under mild assumptions, they solve the so-called Schrödinger systems with guaranteed linear convergence. In this letter, we study a priori estimates for the contraction coefficients associated with the convergence of respective Schrödinger systems. We provide new geometric and control-theoretic interpretations for the same. Building on these newfound interpretations, we point out the possibility of improved computation for the worst-case contraction coefficients of linear SBPs by preconditioning the endpoint support sets.

Index Terms—Stochastic optimal control, stochastic systems, Markov processes.

I. INTRODUCTION

THE SCHRÖDINGER bridge problem (SBP) for a stochastic linear system concerns with minimum effort steering of a controlled stochastic process $x^u(t)$ satisfying Itô diffusion

$$\mathrm{d}x^u(t) = \left(A(t)x^u + B(t)u\right)\mathrm{d}t + \sqrt{2\varepsilon}B(t)\mathrm{d}w(t), \quad (1)$$

over a fixed time horizon $[0, 1]$, from a given initial state PDF $\rho_0(\cdot) := \rho^u(t = 0, \cdot)$ to another given terminal state PDF $\rho_1(\cdot) := \rho^u(t = 1, \cdot)$.

Alexis M. H. Teter is with the Department of Applied Mathematics, University of California at Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: amteter@ucsc.edu).

Yongxin Chen is with the Department of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: yongchen@gatech.edu).

Abhishek Halder is with the Department of Aerospace Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: ahalder@iastate.edu).

In (1), $x^u \in \mathbb{R}^n$ is the controlled state, $u \in \mathbb{R}^m$ is the (to-be-designed) control, and $w \in \mathbb{R}^m$ is the standard m-dimensional Wiener process. The constant $\varepsilon > 0$ denotes the strength of the process noise, and is not necessarily small.

Notice that the process noise in (1) enters through the same "channels" as the input, which is the case in many practical applications, e.g., in noisy actuators, in external disturbances such as wind gust affecting the system state via forcing, and in unmodeled dynamics.

We assume that

A1. the matricial trajectory pair $(A(t), B(t))$ is continuous and bounded for all $t \in [0, 1]$,

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

$$M_{10} := \int_0^1 \Phi_{1\tau} B(\tau)B^\top(\tau)\Phi_{1\tau}^\top \mathrm{d}\tau \quad (2)$$

is symmetric positive definite, where $\Phi_{t\tau} := \Phi(t, \tau)$ for $0 \leq \tau \leq t \leq 1$ denotes the state transition matrix associated with the state matrix $A(t)$,

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

The minimum effort objective translates to the following stochastic optimal control problem:

$$\arg\inf_u \ \mathbb{E}_{x^u} \int_0^1 \|u\|_2^2 \, \mathrm{d}t \quad (3a)$$

$$\text{subject to } (1), \ x^u(t = 0) \sim \rho_0, \ x^u(t = 1) \sim \rho_1, \quad (3b)$$

where the expectation in (3a) is w.r.t. the controlled stochastic state $x^u \sim \rho^u(t, \cdot)$. The problem (3) is to be solved over the feasible set $\mathcal{U}$ that comprises of finite energy Markovian control policies, i.e.,

$$\mathcal{U} := \{u : [0, 1] \times \mathbb{R}^n \mapsto \mathbb{R}^m \mid \mathbb{E}\int_0^1 \|u\|_2^2 \mathrm{d}t < \infty\}.$$

With slight abuse of nomenclature, we will refer to problem (3) as the "linear SBP".

The classical SBP [1], [2], [3], [4] is the following special case of problem (3): $A(t) \equiv 0, B(t) \equiv I_n$. For recent works elaborating connections between the SBP and stochastic optimal control, see [5], [6].
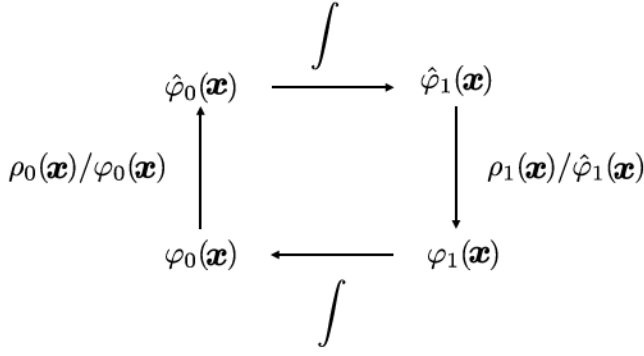
Fig. 1. An SBP over time horizon [0, 1] is solved via contractive fixed point recursion shown over the pair $(\hat{\varphi}_0(\cdot), \varphi_1(\cdot))$. The recursion solves the associated Schrödinger system (4). The top (resp. bottom) horizontal arrow computes the integral in (4b) (resp. (4a)). The vertical arrows apply given boundary conditions ρ_0, ρ_1.

1) Schrödinger System, Schrödinger Factors, and the Solution of SBP: As is well-known [6, Sec. 8], [7, Sec. II], SBPs such as (3) can be solved by computing the function pair $(\hat{\varphi}_0(\cdot), \varphi_1(\cdot))$ satisfying a system of nonlinear integral equations, referred to as the *Schrödinger system*:

$$\rho_0(x) = \hat{\varphi}_0(x) \int_{\mathbb{R}^n} q(0, x, 1, y)\varphi_1(y)\mathrm{d}y, \tag{4a}$$

$$\rho_1(x) = \varphi_1(x) \int_{\mathbb{R}^n} q(0, y, 1, x)\hat{\varphi}_0(y)\mathrm{d}y, \tag{4b}$$

where q is the *uncontrolled* Markov kernel associated with (1), i.e., the Markov kernel with $u \equiv 0$. The system (4) can in turn be solved via a fixed point recursion over $(\hat{\varphi}_0(\cdot), \varphi_1(\cdot))$ shown in Fig. 1, which is provably contractive [8, Sec. III] in Hilbert's projective metric [9], [10]. See also [11].

The $(\hat{\varphi}_0(\cdot), \varphi_1(\cdot))$, thus computed, are used to find the *Schrödinger factors*

$$\hat{\varphi}(t, x) := \int_{\mathbb{R}^n} q(0, y, t, x)\hat{\varphi}_0(y)\mathrm{d}y, \quad t \geq 0, \tag{5a}$$

$$\varphi(t, x) := \int_{\mathbb{R}^n} q(t, x, 1, y)\varphi_1(y)\mathrm{d}y, \quad t \leq 1, \tag{5b}$$

which then yields the optimally controlled state PDF $\rho_{\mathrm{opt}}^u(t, x) = \hat{\varphi}(t, x)\varphi(t, x)$, and the optimal control $u_{\mathrm{opt}}(t, x) = 2\varepsilon B(t)^\top \nabla_x \log \varphi(t, x)$ for all $t \in [0, 1]$. In particular, (5) shows that the factor $\hat{\varphi}(t, \cdot)$ (resp. $\varphi(t, \cdot)$) is *q-harmonic* solving a backward Kolmogorov PDE (resp. *q-coharmonic* solving a forward Kolmogorov PDE).

This result, on one hand, guarantees the existence and uniqueness of solution for the SBP. On the other hand, it offers a practical algorithm in the form of a cone-preserving fixed point recursion shown in Fig. 1. Thanks to the Banach contraction mapping theorem, this recursion has guaranteed *linear* convergence with a *contraction coefficient* $\kappa \in (0, 1)$. The smaller the κ, the faster the convergence.

For fixed SBP data $\mathcal{X}_0, \mathcal{X}_1, \varepsilon, A(t), B(t)$, the contraction coefficient κ in general, depends on the specific choices of ρ_0, ρ_1. In this letter, we focus on the *worst-case contraction coefficient* γ satisfying $\kappa \leq \gamma$. The "worst-case" is understood over all possible ρ_0, ρ_1 supported on given compact sets $\mathcal{X}_0, \mathcal{X}_1 \subset \mathbb{R}^n$. In other words, γ is the tightest upper bound

on κ that only depends on $\mathcal{X}_0, \mathcal{X}_1, \varepsilon, A(t), B(t)$, i.e., on the geometry of the endpoint supports, and the drift and diffusion parameters in (1).

2) Support Function: Let $\langle \cdot, \cdot \rangle$ denote the standard Euclidean inner product. We will need the notion of support function $h_{\mathcal{K}}(\cdot)$ of closed convex set $\mathcal{K}$, defined as

$$h_{\mathcal{K}}(y) := \sup_{x \in \mathcal{K}} \langle y, x \rangle, \quad y \in \mathbb{R}^n, \tag{6}$$

which measures the distance from the origin to a supporting hyperplane of $\mathcal{K}$ with normal along y. From (6), the support function is positive homogeneous of degree one, i.e., $h_{\mathcal{K}}(ay) = ah_{\mathcal{K}}(y)$ for $a > 0$. Since only the direction of y matters, it is customary [12, p. 209] to consider y as a unit vector, i.e., to restrict the domain of $h_{\mathcal{K}}(\cdot)$ to $\mathbb{S}^{n-1}$.

From (6), the support function is finite for a bounded set. It uniquely characterizes a convex set $\mathcal{K}$ since $h_{\mathcal{K}}(\cdot)$ equals the Legendre-Fenchel conjugate of the indicator function [13, Th. 13.2] of $\mathcal{K}$. Definition (6) can be extended to nonconvex $\mathcal{K}$ in the sense $h_{\mathcal{K}}(\cdot)$ is invariant under closure of convexification of $\mathcal{K}$.

The function $h_{\mathcal{K}}(\cdot)$ has nice properties that will find use in Section IV. The function is distributive under Minkowski sum: $h_{\mathcal{K}_1 + \mathcal{K}_2}(\cdot) = h_{\mathcal{K}_1}(\cdot) + h_{\mathcal{K}_2}(\cdot)$. Given convex set $\mathcal{K}$ and $T \in \mathbb{R}^{n \times n}$, $\tau \in \mathbb{R}^n$, the support function of the affine transformed set $T\mathcal{K} + \tau$ is

$$h_{T\mathcal{K}+\tau}(y) = h_{\mathcal{K}}\left(T^\top y\right) + \langle \tau, y \rangle. \tag{7}$$

3) Contributions: In practice, the contraction coefficient κ is numerically observed to be small (i.e., fast convergence) even in more general settings [14, Fig. 4], [7], [15], [16] than classical or linear SBPs. The recent work [17] investigates classical SBP convergence from a sample complexity perspective. To the best of the authors' knowledge, this is the first work analyzing and interpreting the worst-case contraction coefficient in terms of the SBP data.

Our specific contributions are threefold:
- derivation of a formula for the worst-case contraction coefficient for linear SBP in terms of the problem data $\mathcal{X}_0, \mathcal{X}_1, \varepsilon, A(t), B(t)$,
- novel control-theoretic as well as geometric interpretations for the aforesaid formula,
- highlighting how pre-conditioning the supports $\mathcal{X}_0, \mathcal{X}_1$ can help reduce the worst-case contraction coefficient, thereby improving the convergence of the linear SBP.

4) Organization: The layout of the material is as follows. In Section II, we express the worst-case contraction coefficient for classical SBP in a way that permits generalization. Building on this, Section III presents a formula for the worst-case contraction coefficient for linear SBP (Thm. 1), and provides a control-theoretic interpretation (Section III-B) for the same. Section IV focuses on the case when the compact supports $\mathcal{X}_0, \mathcal{X}_1$ are convex, and gives geometric interpretations (Thm. 2) for the quantities discussed earlier. The usage of preconditioning to improve the computation of the worst-case contraction coefficient is also discussed in Section IV with an illustrative example. Section V concludes this letter.

II. CONTRACTION COEFFICIENT FOR CLASSICAL SBP

To prepare ground for analyzing the contraction coefficient in linear SBP, we recall corresponding ideas for the classical SBP, and express them in a way to help generalization.

For a given $\varepsilon > 0$, consider the (scaled) standard Wiener process $\sqrt{2\varepsilon}\mathrm{d}w(t)$ in $\mathbb{R}^n$ with the associated Markov kernel

$$q_B(0, x_0, 1, x_1) := (4\pi\varepsilon)^{-n/2} \exp\left(-\frac{\|x_0 - x_1\|_2^2}{4\varepsilon}\right), \quad (8)$$

where the subscript B stands for the Brownian a.k.a. standard Wiener process. For classical SBP, the $q \equiv q_B$ in (4)-(5).

We note that q_B is continuous for all $(x_0, x_1) \in \mathbb{R}^n \times \mathbb{R}^n$. Furthermore, for $\mathcal{X}_0, \mathcal{X}_1$ compact, there exist constants α_B, β_B such that $0 < \alpha_B \le \beta_B < \infty$, and

$$\alpha_B \le q_B(0, x_0, 1, x_1) \le \beta_B \quad \forall (x_0, x_1) \in \mathcal{X}_0 \times \mathcal{X}_1. \quad (9)$$

In [8, eq. (17)], the rate of convergence for the Schrödinger system associated with the classical SBP was related to the quantity

$$\gamma_B := \tanh^2\left(\frac{1}{2}\log\left(\frac{\beta_B}{\alpha_B}\right)\right) \in (0, 1). \quad (10)$$

Specifically, γ_B was shown [8, Lemma 5] be to the *worst-case contraction coefficient* for a single pass of the recursion shown in Fig. 1.

For ensuing development, it is helpful to define

$$\widetilde{\alpha}_B := \max_{x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1} \|x_0 - x_1\|_2^2, \quad (11a)$$

$$\widetilde{\beta}_B := \min_{x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1} \|x_0 - x_1\|_2^2, \quad (11b)$$

wherein the maximum and minimum are guaranteed to exist due to the compactness of $\mathcal{X}_0, \mathcal{X}_1$. From (8) and (9), it then follows that the α_B, β_B in (10) can be expressed as

$$\alpha_B = \frac{\exp(-\widetilde{\alpha}_B/(4\varepsilon))}{\sqrt{(4\pi\varepsilon)^n}}, \quad \beta_B = \frac{\exp(-\widetilde{\beta}_B/(4\varepsilon))}{\sqrt{(4\pi\varepsilon)^n}}. \quad (12)$$

Consequently, we can rewrite (10) as

$$\gamma_B = \tanh^2\left(\frac{\widetilde{\alpha}_B - \widetilde{\beta}_B}{8\varepsilon}\right) \in (0, 1). \quad (13)$$

Remark 1: Notice from (11) and (12) that $0 \le \widetilde{\beta}_B < \widetilde{\alpha}_B < \infty$ but $0 < \alpha < \beta < \infty$. In particular, $\widetilde{\beta}_B = 0$ if and only if $\mathcal{X}_0$ and $\mathcal{X}_1$ overlap, i.e., $\mathcal{X}_0 \cap \mathcal{X}_1 \ne \emptyset$.

Remark 2: Formula (13) provides an explicit relation between the worst-case contraction coefficient γ_B for the classical Schrödinger system and the problem data given by the tuple $(\mathcal{X}_0, \mathcal{X}_1, \varepsilon)$. In the following, we investigate how the worst-case contraction coefficient γ_L for the problem (3) depend on its problem data given by the tuple $(\mathcal{X}_0, \mathcal{X}_1, \varepsilon, A(t), B(t))$.

III. CONTRACTION COEFFICIENT FOR LINEAR SBP

In this Section, we seek to generalize the development in Section II for the linear SBP (3). Under the stated assumptions A1-A3, the existence-uniqueness for the solution of this problem are guaranteed, and can be computed by solving the associated Schrödinger system (4) with $q \equiv q_L$ where

$$q_L(0, x_0, 1, x_1)$$
$$:= \det(M_{10})^{-1/2} q_B\left(0, M_{10}^{-1/2}\Phi_{10}x_0, 1, M_{10}^{-1/2}x_1\right). \quad (14)$$

A. Contraction Coefficient

Notice that the problem data for the linear SBP involves both the endpoint PDFs (with their compact supports) as well as the dynamical coefficients $A(t), B(t), \varepsilon$ in (1). Intuition suggests that the rate of convergence will differ for different choices of controllable pair $(A(t), B(t))$ while keeping $(\mathcal{X}_0, \mathcal{X}_1, \varepsilon)$ fixed. For instance, faster (resp. slower) convergence is expected for linear systems which are "easier (resp. harder) to control" than others. So we anticipate that the worst-case contraction coefficient γ_L in this case will depend on the controllability Gramian (2).

We have the following result.

Theorem 1: Consider the linear SBP (3) with assumptions A1-A3. The associated Schrödinger system (4) with $q \equiv q_L$ has worst-case contraction coefficient $\gamma_L \in (0, 1)$, given by

$$\gamma_L = \tanh^2\left(\frac{\widetilde{\alpha}_L - \widetilde{\beta}_L}{8\varepsilon}\right), \quad (15)$$

where

$$\widetilde{\alpha}_L := \max_{x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1} (\Phi_{10}x_0 - x_1)^\top M_{10}^{-1}(\Phi_{10}x_0 - x_1), \quad (16a)$$

$$\widetilde{\beta}_L := \min_{x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1} (\Phi_{10}x_0 - x_1)^\top M_{10}^{-1}(\Phi_{10}x_0 - x_1). \quad (16b)$$

Proof: For a given $\varepsilon > 0$, consider the uncontrolled Ornstein-Uhlenbeck (OU) process with time-varying coefficients:

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

with the associated Markov kernel (14) as

$$q_L(0, x_0, 1, x_1)$$
$$= \frac{\exp\left(-\dfrac{(\Phi_{10}x_0 - x_1)^\top M_{10}^{-1}(\Phi_{10}x_0 - x_1)}{4\varepsilon}\right)}{\sqrt{(4\pi\varepsilon)^n \det(M_{10})}}. \quad (17)$$

Recall that M_{10} being symmetric positive definite, so are its inverse M_{10}^{-1} and the principal square root $M_{10}^{-1/2}$. Thus (17) is well-defined.

As was the case for the kernel q_B in (8), the kernel q_L too is continuous and positive for all $(x_0, x_1) \in \mathcal{X}_0 \times \mathcal{X}_1$. Compactness of $\mathcal{X}_0, \mathcal{X}_1$ implies that there exist constants α_L, β_L given by (16) such that $0 < \alpha_L \le \beta_L < \infty$, and

$$\alpha_L \le q_L(0, x_0, 1, x_1) \le \beta_L \quad \forall (x_0, x_1) \in \mathcal{X}_0 \times \mathcal{X}_1. \quad (18)$$

From (16), (17) and (18), we then obtain the following analogue of (12):

$$\alpha_L = \frac{\exp(-\widetilde{\alpha}_L/(4\varepsilon))}{\sqrt{(4\pi\varepsilon)^n \det(M_{10})}}, \quad \beta_L = \frac{\exp(-\widetilde{\beta}_L/(4\varepsilon))}{\sqrt{(4\pi\varepsilon)^n \det(M_{10})}}. \quad (19)$$

Consequently, the worst-case contraction coefficient

$$\gamma_L = \tanh^2\left(\frac{1}{2}\log\left(\frac{\beta_L}{\alpha_L}\right)\right) = \tanh^2\left(\frac{\widetilde{\alpha}_L - \widetilde{\beta}_L}{8\varepsilon}\right). \quad (20)$$

■

Remark 3: Due to the sub-multiplicative nature of 2 norm, the objective in (16), in general, satisfies the bound

$$\frac{\|\mathbf{\Phi}_{10}x_0 - x_1\|_2^2}{\lambda_{\max}(M_{10})} \le \left\|M_{10}^{-1/2}\mathbf{\Phi}_{10}x_0 - M_{10}^{-1/2}x_1\right\|_2^2$$

$$\le \frac{\|\mathbf{\Phi}_{10}x_0 - x_1\|_2^2}{\lambda_{\min}(M_{10})}, \quad (21)$$

where $\lambda_{\max}, \lambda_{\min}$ denote the maximum and minimum eigenvalue of M_{10}, respectively.

B. Control-Theoretic Interpretation

We note that the objective in (16) is precisely the minimum cost for the deterministic optimal control problem:

$$\min_{\boldsymbol{u}} \int_0^1 \|\boldsymbol{u}\|_2^2 \, \mathrm{d}t \quad (22a)$$

$$\text{subject to } \dot{x}^u = A(t)x^u + B(t)u, \quad (22b)$$

$$x^u(t=0) = x_0, \quad x^u(t=1) = x_1, \quad (22c)$$

i.e., the cost for minimum effort steering of a controllable LTV system from a fixed initial state x_0 to a fixed terminal state x_1 over the given time horizon [0, 1]. See, e.g., [18, p. 194].

For fixed $(A(t), B(t))$, and therefore fixed $\mathbf{\Phi}_{10}, M_{10}$, the optimal cost (22a) varies with the variation in endpoints $x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1$. Thus, $\widetilde{\alpha}_L$ (resp. $\widetilde{\beta}_L$) equals the worst-case (resp. best-case) optimal state transfer cost for the source and target supports $\mathcal{X}_0, \mathcal{X}_1$. Recall that $\tanh^2(\cdot)$ is increasing over positive real. Hence γ_L in (15) is an increasing function of the *range* of optimal state transfer cost: $\widetilde{\alpha}_L - \widetilde{\beta}_L$.

The ε in the denominator in (15) implies that a stronger process noise helps to reduce γ_L with other parameters held fixed, thus improving the contraction coefficient, as expected.

In the following, we provide geometric insights for (11) and (16). We then discuss the computation of γ_L.

IV. GEOMETRIC INTERPRETATION AND COMPUTING γ_L FOR CONVEX $\mathcal{X}_0, \mathcal{X}_1$

In (11) and (16), the $\widetilde{\alpha}_B, \widetilde{\alpha}_L$ (resp. $\widetilde{\beta}_B, \widetilde{\beta}_L$) can be seen as the maximal (resp. minimal) separation between the sets $\mathcal{X}_0, \mathcal{X}_1$ or their linear transforms. While (11a) and (16a) are invariant under convexification of the supports, (11b) and (16b) can decrease or stay the same under convexification. So the contraction rate (13) (resp. (15)) under convexification is greater than or equal to (meaning slower or same rate) the corresponding rate without convexification of the supports.

When the compact sets $\mathcal{X}_0, \mathcal{X}_1$ are also convex, then computing the minimum values in (11b) and (16b), in general, reduce to solving the "best approximation pair" problem; see, e.g., [19]. Then, (11b) and (16b) can be numerically computed using the Gilbert-Johnson-Keerthi (GJK) algorithm or its improved variants [20], [21], [22].

On the other hand, the maximum values in (11a) and (16a) correspond to the *squared diameters* of the Cartesian products

$\mathcal{X}_0 \times \mathcal{X}_1$ and $M_{10}^{-1/2}\mathbf{\Phi}_{10}\mathcal{X}_0 \times M_{10}^{-1/2}\mathcal{X}_1$, respectively. When $\mathcal{X}_0, \mathcal{X}_1$ are compact and convex, so are these Cartesian products. Therefore, the maximum values in (11a) and (16a) must be attained at the boundaries of the sets $\mathcal{X}_0 \times \mathcal{X}_1$ and $M_{10}^{-1/2}\mathbf{\Phi}_{10}\mathcal{X}_0 \times M_{10}^{-1/2}\mathcal{X}_1$, respectively. However, numerical computation of these maximum values can be cumbersome depending on what kind of description for the convex sets $\mathcal{X}_0, \mathcal{X}_1$ are available.

In Theorem 2 next, we point out that for $\mathcal{X}_0, \mathcal{X}_1$ convex, (11) and (16) can be expressed in terms of the support functions (see (6)) of these sets, thus offering geometric insights on these quantities.

Theorem 2: Consider compact convex $\mathcal{X}_0, \mathcal{X}_1$ with respective support functions $h_{\mathcal{X}_0}(\cdot), h_{\mathcal{X}_1}(\cdot)$. Let $\mathbb{S}^{n-1}$ denote the Euclidean unit sphere in $\mathbb{R}^n$. Then (11) can be expressed as

$$\widetilde{\alpha}_B = \left\{\max_{y \in \mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}(y) + h_{\mathcal{X}_1}(-y)\right)\right\}^2, \quad (23a)$$

$$\widetilde{\beta}_B = \left\{\min_{y \in \mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}(y) + h_{\mathcal{X}_1}(-y)\right)\right\}^2. \quad (23b)$$

Furthermore, (16) can be expressed as

$$\widetilde{\alpha}_L = \left\{\max_{y \in \mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}\left(\mathbf{\Phi}_{10}^\top M_{10}^{-1/2}y\right) + h_{\mathcal{X}_1}\left(-M_{10}^{-1/2}y\right)\right)\right\}^2, (24a)$$

$$\widetilde{\beta}_L = \left\{\min_{y \in \mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}\left(\mathbf{\Phi}_{10}^\top M_{10}^{-1/2}y\right) + h_{\mathcal{X}_1}\left(-M_{10}^{-1/2}y\right)\right)\right\}^2. (24b)$$

Proof: Consider the set difference

$$\mathcal{X}_0 - \mathcal{X}_1 := \mathcal{X}_0 + (-\mathcal{X}_1) = \{x_0 - x_1 \mid x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1\}.$$

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

From (11a), we have

$$\widetilde{\alpha}_B = \left\{\max_{x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1} \|x_0 - x_1\|_2\right\}^2$$

$$= \left\{\max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \|x\|_2\right\}^2$$

$$= \left\{\max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \left\langle \frac{x}{\|x\|_2}, x \right\rangle\right\}^2$$

$$= \left\{\max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \max_{y \in \mathbb{S}^{n-1}} \langle y, x \rangle\right\}^2$$

$$= \left\{\max_{y \in \mathbb{S}^{n-1}} \max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \langle y, x \rangle\right\}^2$$

$$= \left\{\max_{y \in \mathbb{S}^{n-1}} h_{\mathcal{X}_0 - \mathcal{X}_1}(y)\right\}^2 \quad (25)$$

$$= \left\{\max_{y \in \mathbb{S}^{n-1}} \left(h_{\mathcal{X}_0}(y) + h_{-\mathcal{X}_1}(y)\right)\right\}^2, \quad (26)$$

where (25) follows from the definition (6), and (26) holds because support function is distributive over Minkowski sum. Using (7), we get $h_{-\mathcal{X}_1}(y) = h_{\mathcal{X}_1}(-y)$, and therefore (26) equals (23a).

Likewise, from (11b), we have

$$\widetilde{\beta}_B = \left\{ \min_{x_0 \in \mathcal{X}_0, x_1 \in \mathcal{X}_1} \|x_0 - x_1\|_2 \right\}^2$$

$$= \left\{ -\max_{x \in \mathcal{X}_0 - \mathcal{X}_1} (-\|x\|_2) \right\}^2$$

$$= \left\{ \max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \left\langle \frac{-x}{\|x\|_2}, x \right\rangle \right\}^2$$

$$= \left\{ \max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \min_{y \in \mathbb{S}^{n-1}} \langle y, x \rangle \right\}^2 \tag{27}$$

$$= \left\{ \max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \min_{y \in \mathbb{B}^n} \langle y, x \rangle \right\}^2, \tag{28}$$

where the last line is due to the linear objective which allows lossless convexification for the inner minimization in (27). This can be seen explicitly from the Cauchy-Schwarz inequality: $-\|y\|_2\|x\|_2 \leq \langle y, x \rangle$ where the equality is achieved when y is an unit vector pointing opposite to x.

Since the sets $\mathcal{X}_0 - \mathcal{X}_1, \mathbb{B}^n$ are both compact convex, applying the Von Neumann minimax theorem [23], [24], we rewrite (28) as

$$\widetilde{\beta}_B = \left\{ \min_{y \in \mathbb{B}^n} \max_{x \in \mathcal{X}_0 - \mathcal{X}_1} \langle y, x \rangle \right\}^2 = \left\{ \min_{y \in \mathbb{B}^n} h_{\mathcal{X}_0 - \mathcal{X}_1}(y) \right\}^2. \tag{29}$$

We next revert back the feasible set of the minimization in (29) to $\mathbb{S}^{n-1}$. To justify this, note that since the origin is within $\mathbb{B}^n$, the minimum in (29) cannot be positive. If this minimum value is zero, we can scale the arg min to lie on the unit sphere. So it remains to consider the case when the minimum value $h_{\mathcal{X}_0 - \mathcal{X}_1}(y^{\text{opt}}) < 0$, achieved by y^{opt} with $0 < \|y^{\text{opt}}\| = \delta < 1$. In other words, y^{opt} is in the interior of $\mathbb{B}^n$. Now consider a vector $\widetilde{y} := y^{\text{opt}}/\delta \in \mathbb{S}^{n-1}$, which is feasible. Thanks to the positive homogeneity of the support function, we have $h_{\mathcal{X}_0 - \mathcal{X}_1}(\widetilde{y}) = \frac{1}{\delta} h_{\mathcal{X}_0 - \mathcal{X}_1}(y^{\text{opt}})$, which yields

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

contradicting the supposition that y^{opt} is a minimizer. So the minimizer must lie on the boundary of the feasible set $\mathbb{B}^n$, i.e., on $\mathbb{S}^{n-1}$. Therefore, we can express (29) as

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

which is indeed (23b) since $h_{-\mathcal{X}_1}(y) = h_{\mathcal{X}_1}(-y)$.

To derive (24), we start by rewriting (16) as

$$\widetilde{\alpha}_L = \max_{x_0 \in M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0, x_1 \in M_{10}^{-1/2}\mathcal{X}_1} \|x_0 - x_1\|_2^2$$

$$= \max_{x \in M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0 - M_{10}^{-1/2}\mathcal{X}_1} \|x\|_2^2, \tag{30a}$$

$$\widetilde{\beta}_L = \min_{x_0 \in M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0, x_1 \in M_{10}^{-1/2}\mathcal{X}_1} \|x_0 - x_1\|_2^2$$

$$= \min_{x \in M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0 - M_{10}^{-1/2}\mathcal{X}_1} \|x\|_2^2. \tag{30b}$$

We then follow the same steps as before with the additional usage of the formula (7) relating the support functions of affine

transformed sets in terms of the support functions of their pre-images. This completes the proof. ∎

Remark 4: The equalities (30a) and (30b) are particularly insightful. They highlight that $\widetilde{\alpha}_L, \widetilde{\beta}_L$ can respectively be seen as the maximal and minimal separation between the linear transformed sets $M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0$ and $M_{10}^{-1/2}\mathcal{X}_1$. Specializing (30a)-(30b) for the classical SBP with $A(t) \equiv 0$, $B(t) \equiv I$, and thus with $M_{10} = \Phi_{10} = I$, recovers (11), which is the maximal and minimal separation between the original supports $\mathcal{X}_0, \mathcal{X}_1$, as expected.

A. Improved Computation via Preconditioning

Previous works such as [25] have explored the use of preconditioning to improve the performance of optimal transport algorithms. The preconditioning procedure described in [25] transforms the measures and corresponding support sets through a deterministic map such that the preconditioned measures are moved closer together, by creating new measures with the same (zero) mean and diagonal covariance matrix. The solutions (e.g., optimal transport map, optimal coupling) to the optimal transport problem, before and after preconditioning, are related to each other in a certain way according to the preconditioning. Such a strategy can be extended to the SBP because the SBP is an entropy-regularized optimal transport problem [26], [27].

We explore the application of such a preconditioning procedure for improved computation of γ_L. The following example illustrates the main idea.

Example 1: Consider an instance of the linear SBP (3) with time-invariant coefficients

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

i.e., (1) is noisy double integrator

$$dx_1^u = x_2^u \, dt, \quad dx_2^u = u \, dt + \sqrt{2\varepsilon} \, dw.$$

In this case,

$$\Phi_{10} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad M_{10}^{-1} = \begin{bmatrix} 12 & -6 \\ -6 & 4 \end{bmatrix}.$$

We consider ellipsoidal supports

$$\mathcal{X}_i = \mathcal{E}_i(c_i, S_i) := \{x \in \mathbb{R}^2 \mid (x - c_i)^\top S_i^{-1}(x - c_i) \leq 1\}$$

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

$$c_0 := \Phi_{10}^{-1} M_{10}^{1/2} \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \quad c_1 := M_{10}^{1/2} \begin{pmatrix} 3 \\ 0 \end{pmatrix},$$

and respective positive definite shape matrices

$$S_0 := \Phi_{10}^{-1} M_{10} \Phi_{10}^{-\top}, \quad S_1 := M_{10} = \begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix}.$$

Then

$$M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + (y - 3)^2 \leq 1\}, \tag{31a}$$

$$M_{10}^{-1/2}\mathcal{X}_1 = \{(x, y) \in \mathbb{R}^2 \mid (x - 3)^2 + y^2 \leq 1\}. \tag{31b}$$

Without the use of preconditioning, we determine γ_L from Theorem 1 by considering the maximum and minimum
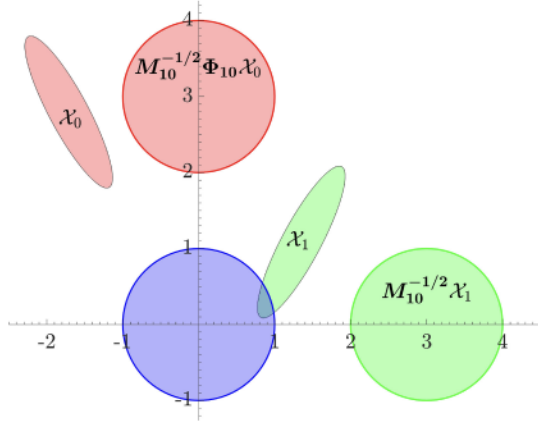
Fig. 2. The sets $\mathcal{X}_0$, $\mathcal{X}_1$ and $M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0$, $M_{10}^{-1/2}\mathcal{X}_1$ in **Example 1**. The preconditioned supports coincide with the origin-centered unit circular disk (in *blue*).

separation between the sets $M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0$ and $M_{10}^{-1/2}\mathcal{X}_1$, which in our example, are two disjoint circular disks. We obtain $\widetilde{\alpha}_{\mathrm{L}} = 2+2\sqrt{3}$ and $\widetilde{\beta}_{\mathrm{L}} = -2+2\sqrt{3}$. From Theorem 1, we determine $\gamma_{\mathrm{L}} = \tanh^2(1) \approx 0.580$.

If the pushforwards $(M_{10}^{-1/2}\Phi_{10})_\sharp \rho_0$, $(M_{10}^{-1/2})_\sharp \rho_1$, i.e.,

$$\frac{\sqrt{\det(M_{10})}}{\det(\Phi_{10})}\rho_0\left(\Phi_{10}^{-1}M_{10}^{1/2}(\cdot)\right), \; \sqrt{\det(M_{10})}\rho_1\left(M_{10}^{1/2}(\cdot)\right),$$

have identical, diagonal covariance matrices, then applying the preconditioning procedure as in [25, Sec. 5] amounts to translating the means of the supports $M_{10}^{-1/2}\Phi_{10}\mathcal{X}_0$ and $M_{10}^{-1/2}\mathcal{X}_1$ to the origin. In our example, the preconditioned supports (31) are both origin-centered unit disks (Fig. 2). Consequently, $\widetilde{\alpha}_{\mathrm{L}}^{\mathrm{precond}} = 2$, $\widetilde{\beta}_{\mathrm{L}}^{\mathrm{precond}} = 0$, and we get $\gamma_{\mathrm{L}}^{\mathrm{precond}} = \tanh^2(0.5) = 0.214$, which is an improvement on the original $\gamma_{\mathrm{L}} \approx 0.580$.

Remark 5: One usage of Theorem 1 is thus to demonstrate the effectiveness of proposed preconditioning procedures in reducing γ_{L}. Additionally, the application of such preconditioning procedures can transform the supports to allow for improved calculation of $\gamma_{\mathrm{L}}^{\mathrm{precond}}$, as was the case in **Example 1**. How to optimally construct such a preconditioning procedure for a given SBP remains an open question.

V. CONCLUSION

This letter advances systems-control-theoretic underpinnings at the intersection of Schrödinger bridge and stochastic control problems by deriving a formula for the worst-case contraction rate for a linear SBP in terms of the problem data. The formula takes the form of squared hyperbolic tangent of a scaled range, which has clean geometric as well as optimal control-theoretic interpretations. These interpretations also suggest the possibility of preconditioning the endpoint supports for improved computation. We illustrate the same through an example, and conclude with an open question on optimal preconditioning for a given SBP.

REFERENCES

[1] E. Schrödinger, "Über die umkehrung der naturgesetze," *Sitzungsberichte der Preuss. Phys. Math. Klasse*, vol. 10, pp. 144–153, Jul. 1931.

[2] E. Schrödinger, "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique," *Annales de l'institut Henri Poincaré*, vol. 2, no. 4, pp. 269–310, 1932.

[3] A. Wakolbinger, "Schrödinger bridges from 1931 to 1991," in *Proc. 4th Latin Amer. Congr. Probability Math. Statist., Mex. City*, 1990, pp. 61–79.

[4] C. Léonard, "From the Schrödinger problem to the Monge–Kantorovich problem," *J. Funct. Anal.*, vol. 262, no. 4, pp. 1879–1920, 2012.

[5] 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.*, vol. 169, pp. 671–691, May 2016.

[6] Y. Chen, T. T. Georgiou, and M. Pavon, "Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge," *SIAM Rev.*, vol. 63, no. 2, pp. 249–313, 2021.

[7] K. F. Caluya and A. Halder, "Wasserstein proximal algorithms for the Schrödinger bridge problem: Density control with nonlinear drift," *IEEE Trans. Autom. Control*, vol. 67, no. 3, pp. 1163–1178, Mar. 2022.

[8] Y. Chen, T. Georgiou, and M. Pavon, "Entropic and displacement interpolation: A computational approach using the hilbert metric," *SIAM J. Appl. Math.*, vol. 76, no. 6, pp. 2375–2396, 2016.

[9] D. Hilbert, "über die gerade linie als kürzeste verbindung zweier punkte: Aus einem an herrn f. klein gerichteten briefe," *Mathematische Annalen*, vol. 46, no. 1, pp. 91–96, 1895.

[10] P. J. Bushell, "Hilbert's metric and positive contraction mappings in a banach space," *Arch. Ration. Mech. Anal.*, vol. 52, pp. 330–338, Dec. 1973.

[11] J. Franklin and J. Lorenz, "On the scaling of multidimensional matrices," *Linear Algebra Appl.*, vols. 114–115, pp. 717–735, Mar./Apr. 1989.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Fundamentals*. vol. 305. Berlin, Germany: Springer, 1996.

[13] R. Rockafellar, "*Convex analysis*," (Princeton Mathematical Series), vol. 28. Princeton, NJ, USA: Princeton Univ. Press, 1970.

[14] K. F. Caluya and A. Halder, "Reflected Schrödinger bridge: Density control with path constraints," 2020, *arXiv:2003.13895*.

[15] V. De Bortoli, J. Thornton, J. Heng, and A. Doucet, "Diffusion Schrödinger bridge with applications to score-based generative modeling," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 17695–17709.

[16] S. Haddad, K. F. Caluya, A. Halder, and B. Singh, "Prediction and optimal feedback steering of probability density functions for safe automated driving," *IEEE Control Syst. Lett.*, vol. 5, pp. 2168–2173, 2020.

[17] A. Stromme, "Sampling from a Schrödinger bridge," in *Proc. Int. Conf. Artif. Intell. Statist.*. PMLR, 2023, pp. 4058–4067.

[18] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*. vol. 87. New York, NY, USA: Wiley, 1967.

[19] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Finding best approximation pairs relative to two closed convex sets in hilbert spaces," *J. Approx. theory*, vol. 127, no. 2, pp. 178–192, 2004.

[20] E. G. Gilbert, D. W. Johnson, and S. S. Keerthi, "A fast procedure for computing the distance between complex objects in three-dimensional space," *IEEE J. Robot. Autom.*, vol. 4, no. 2, pp. 193–203, Apr. 1988.

[21] S. Cameron, "Enhancing GJK: Computing minimum and penetration distances between convex polyhedra," in *Proc. Int. Conf. Robot. Autom.*, 1997, pp. 3112–3117.

[22] M. Montanari, N. Petrinic, and E. Barbieri, "Improving the GJK algorithm for faster and more reliable distance queries between convex objects," *ACM Trans. Graph. (TOG)*, vol. 36, no. 3, pp. 1–17, 2017.

[23] J. V. Neumann, "Zur theorie der Gesellschaftsspiele," *Mathematische annalen*, vol. 100, no. 1, pp. 295–320, 1928.

[24] D.-Z. Du and P. M. Pardalos, Eds., *Minimax and Applications*, (ser. Nonconvex Optimization and Its Applications). vol. 4. New York, NY, USA: Springer, 1995.

[25] M. Kuang and E. G. Tabak, "Preconditioning of optimal transport," *SIAM J. Sci. Comput.*, vol. 39, no. 4, pp. A1793–A1810, 2017.

[26] T. Mikami, "Monge's problem with a quadratic cost by the zero-noise limit of H-path processes," *Probab. Theory Relat. Fields*, vol. 129, no. 2, pp. 245–260, 2004.

[27] S. D. Marino and A. Gerolin, "An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm," *J. Sci. Comput.*, vol. 85, no. 2, p. 27, 2020.