

Schrödinger Bridge in Inner-Product Gromov–Wasserstein Geometry

Mayank Shrivastava

*Siebel School of Computing and Data Science
University of Illinois Urbana-Champaign*

MAYANKS4@ILLINOIS.EDU

Parag Dawani

*Department of Computer Science and Automation
Indian Institute of Science, Bengaluru*

PARAGDAWANI@IISC.AC.IN

Rohan Deb

*Siebel School of Computing and Data Science
University of Illinois Urbana-Champaign*

RD22@ILLINOIS.EDU

Arindam Banerjee

*Siebel School of Computing and Data Science
University of Illinois Urbana-Champaign*

ARINDAMB@ILLINOIS.EDU

Anant Raj

*Department of Computer Science and Automation
Indian Institute of Science, Bengaluru*

ANANTRAJ@IISC.AC.IN

Editors: Surbhi Goel and Nikita Zhivotovskiy

Abstract

The Schrödinger bridge (SB) problem is a variational framework for transport between given source and target probability distributions. The dynamical Schrödinger problem is typically studied in Wasserstein geometry; the metric tensor defines the kinetic action and Fisher information, which determines the optimal interpolation. However, Wasserstein geometry may be unsuitable for tasks that require preserving the global geometric structure of the data. The inner-product Gromov–Wasserstein (IGW) geometry, which is based on pairwise inner products is a more preferable geometry. We formulate a *dynamical Schrödinger bridge problem* in IGW geometry. To do this, we build on the established gradient-flow structure and Benamou–Brenier formulation of IGW geometry, and define the IGW Fisher information. For IGW Schrödinger bridge (IGW-SB) between centered Gaussians, we show that an optimal bridge has Gaussian marginals. We then derive a closed-form expression for their covariance evolution. Finally, we show that this dynamical formulation differs from standard static entropic IGW, even for centered Gaussian endpoints.

Keywords: Gromov–Wasserstein, Schrödinger Bridges, Optimal Transport

1. Introduction

The space $\mathcal{P}_2(\mathbb{R}^d)$ of probability measures with finite-second moment, equipped with a suitable distance, provides a geometric framework for studying interpolation, gradient flows, and dynamical measure transport. The canonical choice is the Wasserstein metric, for which geodesics and gradient flows were developed by [Otto \(2001\)](#); [Ambrosio et al. \(2008\)](#). However, the Wasserstein distance relies on a point-to-point cost and does not account for the internal relational structure of a distribution. The Gromov–Wasserstein (GW) distance ([Mémoli, 2011](#)) instead compares pairwise relations within each measure. This makes it well suited for aligning data across different spaces, such as word embeddings across languages ([Alvarez-Melis and Jaakkola, 2018](#)) and shapes in computer

vision (Mémoli, 2009). Inner-product Gromov–Wasserstein (IGW) (Vayer, 2020; Dandapanthula et al., 2025; Le et al., 2022) replaces pairwise distances with pairwise inner products and is invariant under orthogonal transformations, providing a framework for studying transport through the relational structure of distributions without penalizing global rotations.

Our starting point is the dynamical formulation of the IGW distance introduced by Zhang et al. (2025), which parallels the classical Benamou-Brenier formulation of Wasserstein transport (Benamou and Brenier, 2000). The latter expresses the squared Wasserstein distance as a minimum kinetic action over curves of probability measures $(\rho_t)_{t \in [0,1]}$ and velocity fields $(v_t)_{t \in [0,1]}$ satisfying the continuity equation with given endpoints. In the IGW analogue, the Euclidean kinetic energy is replaced by the kinetic energy of the IGW metric tensor, which couples velocities across distributions and vanishes on infinitesimal rotations.

In the Wasserstein setting, augmenting the Benamou-Brenier action with Fisher information gives the Schrödinger bridge problem (Chen et al., 2016; Léonard, 2012; Gentil et al., 2017). Motivated by this connection, we define an intrinsic Fisher-information regularization of the dynamical IGW action. We use the same underlying geometry and study the resulting variational problem, which we view as a *dynamical Schrödinger bridge* in IGW geometry (IGW-SB).

In the Wasserstein setting, Bunne et al. (2023) establish that Schrödinger bridges (SB) between Gaussian measures are themselves Gaussian processes. For centered Gaussian endpoints, they show that the full SB can be reduced to an action minimization problem on the space of positive-definite covariance matrices equipped with the Bures-Wasserstein metric Bunne et al. (2023, Theorem 2). We show an analogous result in the IGW geometry. For centered Gaussians: (i) the IGW-SB variational problem has a global minimizer with centered Gaussian marginals, with closed-form expression for the covariance, and (ii) after fixing the endpoint eigenvalue pairing, the problem reduces to d independent scalar variational problems for the covariance eigenvalues similar to the spectral reduction of static Gaussian IGW transport (Le et al., 2022). We summarize our contributions below:

1. We define Fisher information in IGW geometry and use it to formulate a dynamical Schrödinger bridge problem, IGW-SB.
2. We prove that, for centered Gaussian endpoints with positive-definite covariances, IGW-SB admits a global minimizer with centered Gaussian marginals. This reduces optimization over probability measures and velocity fields to a finite-dimensional problem involving covariance matrices.
3. We reduce the covariance problem to d independent scalar problems for the eigenvalue trajectories, once the endpoint eigenvalue pairing is fixed, and derive a closed-form solution.
4. We show how the geometry changes the Fisher information penalty: on Gaussian curves, the IGW penalty depends on inverse *squared* covariance eigenvalues, whereas the Wasserstein penalty in Bunne et al. (2023) depends on inverse covariance eigenvalues.
5. We show that dynamical Fisher regularization and static entropy regularization lead to different IGW transport problems, even for centered Gaussian endpoints, by comparing IGW-SB with the formulation of Le et al. (2022).

Organization. In Section 2, we give a brief overview of related work. In Section 3, we review IGW geometry and the recent results of Zhang et al. (2025), and define the IGW Fisher information and the IGW-SB problem. In Section 4, we derive a closed-form solution to the Gaussian IGW-SB problem, and we conclude in Section 5. All proofs are provided in Appendices B.1-B.3. Appendix D compares IGW-SB with Wasserstein SB and static entropic IGW.

2. Related Work

Gromov-Wasserstein distance Motivated by applications in shape analysis and object matching, Mémoli (2011) proposed Gromov-Wasserstein distance to compare objects represented as metric measure (mm) spaces. GW-based methods have since been used to compare and align heterogeneous datasets across modalities, with applications to word-embedding alignment (Alvarez-Melis and Jaakkola, 2018), shape and graph matching (Mémoli, 2009; Xu et al., 2019b; Im et al., 2026), generative modeling and neural transport (Bunne et al., 2019; Klein et al., 2024), unpaired cross-modal alignment (Schnaus et al., 2026), and disentangled representation learning (Uscidda et al., 2024). Several works have studied computational properties and the development of scalable algorithms (Xu et al., 2019a; Scetbon et al., 2022; Houry et al., 2026) for computing GW distance, including entropic GW formulations (Solomon et al., 2016; Peyré et al., 2016; Zhang et al., 2024). More recently, Zhang et al. (2025) provided the first framework for IGW gradient flows directly in $\mathcal{P}_2(\mathbb{R}^d)$, rather than on the quotient space of gauged measure spaces (Sturm, 2012).

Schrödinger Bridge The Schrödinger bridge problem seeks the most likely evolution of a system of diffusing particles between prescribed endpoints. The classical formulation minimizes relative entropy over path measures relative to a reference process (Léonard, 2014). Since then, Schrödinger bridges have been used in generative modelling (Bortoli et al., 2021; Tang, 2026), alongside the development of efficient algorithms for bridge estimation (Manole et al., 2021; Tong et al., 2023). While several equivalent formulations of the SB exist, (Léonard, 2014; Chen et al., 2016) the formulation most relevant to us is the fluid dynamic formulation of Benamou-Brenier perturbed with Fisher Information (Chen et al., 2016; Monsaingeon et al., 2023). In this work, we adopt the latter approach and study its analogue in the IGW geometry.

Closed-form expressions While static optimal transport and its dynamical counterpart are formulated as general variational problems, closed-form solutions are available in special cases, providing insight into the structure of optimal solutions and benchmarks for evaluating numerical methods. For Gaussian endpoints in the Wasserstein setting, explicit solutions are known for optimal transport and entropy-regularized optimal transport (Janati et al., 2020). At the dynamical level, Bunne et al. (2023) derive closed-form expressions for Gaussian Schrödinger bridges. In the IGW setting, while closed-form solutions have been derived for static transport between centered Gaussian measures (Salmona et al., 2021; Le et al., 2022; Dandapanthula et al., 2025), the corresponding dynamical SB problem remains unexplored. Our work fills this gap and provides the first closed-form expression for the IGW SB problem with centered Gaussian endpoints.

3. Setup and Preliminaries

Notation Matrices are denoted by capital letters. We use boldface letters to denote matrices. The operator ∇ denotes the gradient, while $\nabla \cdot$ denotes the divergence, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean

inner product. We write $\mathcal{P}_2(\mathbb{R}^d)$ for the space of probability measures on $\mathbb{R}^d$ with finite second moment. The symbols $\nabla_{\mathcal{W}}$ and ∇_{IGW} denote the Wasserstein and IGW gradients, respectively. $\|\cdot\|$ denotes the Euclidean norm. For a measurable map f and a measure μ , $f_{\#}\mu$ denotes the pushforward of μ by f . $\mathbb{S}_{++}^d$ denotes the set of real symmetric positive definite $d \times d$ matrices.

We work on the space of probability measures with bounded second moments, $\mathcal{P}_2(\mathbb{R}^d)$. To begin with, we provide a brief introduction to IGW distances and the results obtained by [Zhang et al. \(2025\)](#) regarding the corresponding formal riemannian structure of the IGW geometry.

3.1. IGW distance and Gradient Flows

Given measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, IGW distance is given by:

$$\text{IGW}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_{\mathbb{R}^d \times \mathbb{R}^d} |\langle x, x' \rangle - \langle y, y' \rangle|^2 d\pi \otimes \pi(x, y, x', y') \right)^{\frac{1}{2}} \quad (1)$$

where $\Pi(\mu, \nu)$ is the class of admissible couplings, such that $\pi(x, \mathbb{R}^d) = \mu$, $\pi(\mathbb{R}^d, y) = \nu$ for all $\pi \in \Pi$. Since the IGW distance depends only on the inner products, it is invariant under orthogonal transformations. As such, IGW does not define a metric on $\mathcal{P}_2(\mathbb{R}^d)$. Despite this rotational degeneracy, [Zhang et al. \(2025\)](#) develop gradient flows in $\mathcal{P}_2(\mathbb{R}^d)$ w.r.t the IGW geometry. The corresponding IGW gradient is obtained by transforming the $\mathcal{W}_2$ gradient through the inverse of a distribution-dependent operator, called the *mobility operator* defined below:

Definition 1 (Mobility operator) Let $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and $v \in L^2(\rho; \mathbb{R}^d)$, and denote the second-moment matrix of ρ by $\Sigma_\rho := \int xx^\top d\rho(x)$. The IGW mobility operator is defined by

$$\mathcal{L}_{\Sigma_\rho, \rho}[v](x) := 2 \left(\Sigma_\rho v(x) + \int y \langle v(y), x \rangle d\rho(y) \right). \quad (2)$$

With the mobility operator defined, we summarize the results relevant to our discussion and refer the reader to their work for a detailed account.

Theorem 2 (Informal version of Theorem 4.1, [Zhang et al. \(2025\)](#)) Let $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, $\Sigma_{\mu_0} \succ 0$ and $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ be a rotational invariant functional. The IGW gradient flow of $\mathcal{F}$ satisfies the PDE:

$$\partial_t \rho_t = \nabla \cdot (\rho_t \nabla_{\text{IGW}} \mathcal{F}(\rho_t)), \quad \rho_0 = \mu_0. \quad (3)$$

where, $\nabla_{\text{IGW}} \mathcal{F}(\rho) = \mathcal{L}_{\Sigma_\rho, \rho}^{-1}[\nabla_{\mathcal{W}} \mathcal{F}(\rho)]$. Here, $\mathcal{L}_{\Sigma_\rho, \rho}^{-1}$ denotes the inverse of the mobility operator restricted to the invariant subspace $I_\rho := \{v \in L^2(\rho; \mathbb{R}^d) : \int xv(x)^\top d\rho(x) \text{ is symmetric}\}$.

IGW gradient flow is constructed through a minimizing-movement scheme ([Ambrosio et al., 2008](#)) wherein each successive iterate is orthogonally aligned with the preceding iterate in the $\mathcal{W}_2$ sense. This alignment ensures the existence of an optimal IGW coupling and allows the limiting flow to be represented by a continuous curve in $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$. The mobility operator combines a point-wise term $\Sigma_\rho v(x)$ with a nonlocal position-velocity term $\int y \langle v(y), x \rangle d\rho(y)$ that couples velocities across the distribution. The resulting IGW descent incorporates both the Wasserstein gradient of $\mathcal{F}$ and the global inner-product structure of ρ . For any skew-symmetric matrix $S \in \mathbb{R}^{d \times d}$, the rotational field $v(x) = Sx$ lies in the kernel of the mobility operator, since $\mathcal{L}_{\Sigma_\rho, \rho}[v](x) = 2\Sigma_\rho Sx + 2\Sigma_\rho S^\top x = 0$. Consequently, any purely rotational transformation is nullified by the mobility operator.

3.2. Intrinsic IGW metric and Benamou-Brenier formula

Recall that the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$ is a geodesic length space with formal Riemannian metric tensor $\langle v, w \rangle_{L^2(\rho; \mathbb{R}^d)}$ on tangent velocity fields (Otto, 2001; Ambrosio et al., 2008). The Benamou-Brenier formula (Benamou and Brenier, 2000) gives the $\mathcal{W}_2^2$ distance between two measures as the minimum kinetic action induced by this tensor over transport curves connecting them. The IGW distance, however, does not in general induce a geodesic length space on $\mathcal{P}_2(\mathbb{R}^d)$. A geodesic formulation of transport is obtained through the intrinsic IGW metric d_{IGW} introduced by Zhang et al. (2025), which measures the shortest IGW path connecting two measures within this space. The corresponding Riemannian structure is specified by the *IGW metric tensor*, $g_\rho(v, w) := \langle v, \mathcal{L}_{\Sigma_\rho, \rho}[w] \rangle_{L^2(\rho; \mathbb{R}^d)}$.

The quadratic form $g_\rho(v, v)$ represents the instantaneous IGW kinetic energy. Unlike the Wasserstein kinetic energy $\int \|v(x)\|^2 d\rho(x)$, it vanishes for rotational fields $v(x) = Sx$ with $S^\top = -S$. The corresponding Benamou-Brenier-type formulation is given by:

$$d_{\text{IGW}}^2(\mu_0, \mu_1) = \min_{\tilde{\mu}_1 \in \{\mu_1, \mathbf{I}_\#^- \mu_1\}} \inf_{\substack{(\rho_t, v_t)_{t \in [0,1]} \\ \partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0 \\ \rho_0 = \mu_0, \rho_1 = \tilde{\mu}_1}} \int_0^1 g_{\rho_t}(v_t, v_t) dt \quad (4)$$

where $\mathbf{I}^-$ is any fixed reflection matrix. Next, we use the IGW gradient flows and the associated Benamou-Brenier formulation to define the dynamical IGW Schrödinger problem.

3.3. Dynamical Schrödinger bridges in IGW geometry (IGW-SB)

We adopt the Benamou–Brenier formulation of the classical Schrödinger problem, which minimizes the kinetic action $(\int \|v_t\|^2 d\rho_t)$ perturbed by Fisher information $(\int \|\nabla \log \rho_t\|^2 d\rho_t)$ (Léonard, 2014; Léger, 2019), subject to the continuity equation and the terminal constraints. Monsaingeon et al. (2023) extend this variational viewpoint to abstract metric spaces, expressing the kinetic energy through the squared metric speed and the Fisher regularization through the squared entropy slope. We use this geometric interpretation to construct the IGW counterpart. The kinetic energy is given by the IGW metric tensor; we now develop the corresponding Fisher information from the Boltzmann–Shannon entropy and the IGW metric tensor, before stating the bridge problem.

Definition 3 (IGW Fisher information) *Let $\mathcal{H}(\rho) := \int \log \rho d\rho$ denote the (negative) Boltzmann–Shannon entropy. Following Monsaingeon et al. (2023, Section 2), for an absolutely continuous $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ with $\nabla \log \rho \in L^2(\rho; \mathbb{R}^d)$ and $\Sigma_\rho \succ 0$, define the IGW Fisher information $\mathcal{I}_{\text{IGW}}(\rho)$ as the squared IGW norm of the entropy gradient:*

$$\mathcal{I}_{\text{IGW}}(\rho) := g_\rho(\nabla_{\text{IGW}} \mathcal{H}(\rho), \nabla_{\text{IGW}} \mathcal{H}(\rho)). \quad (5)$$

By Theorem 2 and the definition of the IGW metric tensor,

$$\mathcal{I}_{\text{IGW}}(\rho) = \left\langle \nabla \log \rho, \mathcal{L}_{\Sigma_\rho, \rho}^{-1}[\nabla \log \rho] \right\rangle_{L^2(\rho; \mathbb{R}^d)} \quad (6)$$

Perturbing the IGW kinetic action by this Fisher information yields the dynamical IGW Schrödinger bridge cost, defined for $\varepsilon > 0$ by:

$$\text{SB}_{\text{IGW},\varepsilon}(\mu, \nu) := \min_{\tilde{\nu} \in \{\nu, \Gamma_{\#}^{-1}\nu\}} \inf_{\substack{(\rho_t, v_t) \\ \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0 \\ \rho_0 = \mu, \rho_1 = \tilde{\nu}}} \int_0^1 [g_{\rho_t}(v_t, v_t) + \varepsilon^2 \mathcal{I}_{\text{IGW}}(\rho_t)] dt, \quad (7)$$

with admissibility understood as in the preceding dynamical formulation. We write $\mathcal{A}(\rho, \mathbf{v}) := \int_0^1 [g_{\rho_t}(v_t, v_t) + \varepsilon^2 \mathcal{I}_{\text{IGW}}(\rho_t)] dt$ for the IGW-SB action, so that $\text{SB}_{\text{IGW},\varepsilon}(\mu, \nu)$ is the infimum of $\mathcal{A}$ over admissible pairs. We next study this variational problem for centered Gaussian endpoints with positive definite covariances. We first show that Gaussian curves attain the minimum over all admissible curves, and then solve the resulting problem explicitly.

4. Closed-Form Solution for Gaussian IGW Schrödinger Bridge

In this section, we present our main results for the dynamical IGW Schrödinger bridge problem defined in Eq. (7), with centered Gaussian endpoints $\mu = \mathcal{N}(0, \Sigma_\mu)$ and $\nu = \mathcal{N}(0, \Sigma_\nu)$ with $\Sigma_\mu, \Sigma_\nu \in \mathbb{S}_{++}^d$. Let Σ_μ and Σ_ν admit the following eigendecomposition:

$$\Sigma_\mu = Q_\mu \Lambda_\mu Q_\mu^\top, \quad \Lambda_\mu = \text{diag}(\lambda_{\mu,1}, \dots, \lambda_{\mu,d}), \quad (8)$$

$$\Sigma_\nu = Q_\nu \Lambda_\nu Q_\nu^\top, \quad \Lambda_\nu = \text{diag}(\lambda_{\nu,1}, \dots, \lambda_{\nu,d}), \quad (9)$$

and the eigenvalues are ordered increasingly. Choose the endpoint eigenvector signs so that $\det Q_\mu = \det Q_\nu$, and let $Q_t \in \mathcal{O}(d)$ be any continuously differentiable path satisfying $Q_0 = Q_\mu$ and $Q_1 = Q_\nu$. We pair the endpoint eigenvalues in increasing order:

$$\lambda_{\mu_i} \longleftrightarrow \lambda_{\nu_i}, \quad i = 1, \dots, d, \quad (10)$$

and refer to this as the monotone pairing.

Theorem 4 (Closed-form Gaussian IGW Schrödinger bridge) *Let $\mu = \mathcal{N}(0, \Sigma_\mu)$ and $\nu = \mathcal{N}(0, \Sigma_\nu)$ be centered Gaussian measures with $\Sigma_\mu, \Sigma_\nu \in \mathbb{S}_{++}^d$, and with Q_t and the endpoint eigenvalue pairing specified above, the monotone pairing is optimal, and an optimal Gaussian bridge is given by:*

$$\rho_t = \mathcal{N}(0, \Sigma_t), \quad v_t(x) = B_t x, \quad (11)$$

where,

$$\Lambda_t = \text{diag}(\lambda_1(t), \dots, \lambda_d(t)), \quad \Sigma_t = Q_t \Lambda_t Q_t^\top, \quad (12)$$

$$\lambda_i(t) = \left((1-t)^2 \lambda_{\mu_i}^2 + t^2 \lambda_{\nu_i}^2 + 2t(1-t) \sqrt{\lambda_{\mu_i}^2 \lambda_{\nu_i}^2 + \frac{\varepsilon^2}{4}} \right)^{1/2}, \quad \forall i \in [d], \quad (13)$$

$$B_t = \dot{Q}_t Q_t^\top + \frac{1}{2} Q_t \dot{\Lambda}_t \Lambda_t^{-1} Q_t^\top. \quad (14)$$

Remark 5 *For the optimal eigenvalue trajectories, different admissible choices of Q_t yield the same optimal action. They produce different Gaussian curves and velocity fields, reflecting the rotational nonuniqueness of the bridge.*

Proof Outline The proof proceeds in three steps. First, in Theorem 7 we derive a lower bound on the objective (7). Next, in Lemma 8, we show that for centered gaussian endpoints, family of path restricted to centered gaussian marginals achieves the established lower bound thus proving the optimality. Finally, in Theorem 9 we solve for the lower bound explicitly and derive the closed form solution for the eigenvalue.

Definition 6 (Covariance-Cross moment curve) Let $(\rho_t, v_t)_{t \in [0,1]}$ be an admissible curve satisfying the continuity equation

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0. \quad (15)$$

We define the associated covariance-cross-moment curve as: $(\Sigma_t, \mathbf{A}_t)_{t \in [0,1]}$ by

$$\Sigma_t := \int x x^\top d\rho_t(x), \quad \mathbf{A}_t := \int v_t(x) x^\top d\rho_t(x) \quad (16)$$

Theorem 7 (IGW-SB action lower bound) Let $(\rho, \mathbf{v}) := (\rho_t, v_t)_{t \in [0,1]}$ be any admissible solution to the dynamical Schrödinger Bridge problem (7) and let $(\Sigma_t, \mathbf{A}_t)_{t \in [0,1]}$ denote the corresponding covariance-cross moment curve. Then, we have the following lower bound:

$$\mathcal{A}(\rho, \mathbf{v}) \geq \text{tr} \left(\int_0^1 (2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2) dt + \frac{\varepsilon^2}{4} \int_0^1 \Sigma_t^{-2} dt \right) \quad (17)$$

Proof Proof is given in Appendix B.1. ■

Next we show that, for centered Gaussian endpoints, the preceding lower bound is attained by Gaussian paths. Consequently, restricting the optimization to Gaussian paths leaves the infimum unchanged.

Lemma 8 (Gaussian realization) Consider any admissible pair $(\rho, \mathbf{v})$ and the corresponding $(\Sigma_t, \mathbf{A}_t)_{t \in [0,1]}$. The Gaussian path $(\rho_G, \mathbf{v}_G) = (\mathcal{N}(0, \Sigma_t), \mathbf{A}_t \Sigma_t^{-1} x)_{t \in [0,1]}$ satisfies the endpoint and continuity constraints, and attains the lower bound in Theorem 7, i.e.:

$$\mathcal{A}(\rho_G, \mathbf{v}_G) = \int_0^1 \text{tr} \left(2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2 + \frac{\varepsilon^2}{4} \Sigma_t^{-2} \right) dt.$$

Proof Proof is given in Appendix B.2. ■

Thus, every admissible pair $(\rho, \mathbf{v})$ admits a Gaussian realization $(\rho_G, \mathbf{v}_G)$ with the same covariance-moment curve $(\Sigma_t, \mathbf{A}_t)_{t \in [0,1]}$ and $\mathcal{A}(\rho_G, \mathbf{v}_G) \leq \mathcal{A}(\rho, \mathbf{v})$. As a result, restricting the optimization to centered Gaussian paths leaves the infimum unchanged. In the next theorem, we solve the original objective (7) restricted to Gaussian paths and obtain a closed form solution.

Theorem 9 (Closed-form solution of the covariance lower bound) Let $Q_\mu, \Lambda_\mu, Q_\nu, \Lambda_\nu$ and the orthogonal path $(Q_t)_{t \in [0,1]}$ and the eigenvalue pairing as defined in Eqs. (8-9). Consider the covariance-level variational problem:

$$\inf_{\substack{\Sigma_t \succ 0, \mathbf{A}_t \\ \dot{\Sigma}_t = \mathbf{A}_t + \mathbf{A}_t^\top \\ \Sigma_0 = \Sigma_\mu, \Sigma_1 = \Sigma_\nu}} \int_0^1 \left[\text{tr} \left(2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2 \right) + \frac{\varepsilon^2}{4} \text{tr} (\Sigma_t^{-2}) \right] dt. \quad (18)$$

Then

$$\Sigma_t = Q_t \Lambda_t Q_t^\top, \quad \mathbf{A}_t = \dot{Q}_t \Lambda_t Q_t^\top + \frac{1}{2} Q_t \dot{\Lambda}_t Q_t^\top, \quad (19)$$

where

$$\Lambda_t = \text{diag}(\lambda_1(t), \dots, \lambda_d(t)), \quad (20)$$

$$\lambda_i(t) = \sqrt{(1-t)^2 \lambda_{\mu,i}^2 + t^2 \lambda_{\nu,i}^2 + t(1-t) \sqrt{4\lambda_{\mu,i}^2 \lambda_{\nu,i}^2 + \varepsilon^2}}, \quad i = 1, \dots, d. \quad (21)$$

Proof Proof is given in Appendix B.3. ■

Combining Theorem 7, Lemma 8 and Theorem 9 we obtain the original result.

5. Conclusion and Discussion

In this work, we formulate the dynamical Schrödinger bridge problem in IGW geometry and derive closed-form solutions for centered Gaussian measures. We show that the corresponding variational problem reduces to independent scalar variational problems, which can be solved explicitly. We also contrast the IGW Schrödinger bridge with the static Entropic IGW problem as well as the Wasserstein Schrödinger bridge problem. Future work includes establishing Γ -convergence to the unregularized dynamical transport problem and investigating a path-space formulation.

AI Disclosure

For this work, various AI tools were used for literature survey, formulating different proof strategies, proofreading the manuscript and in writing parts of the final manuscript as well.

Acknowledgments

We thank our colleagues and funding agencies; \documentclass[anon]{alt2027} automatically hides this text.

References

- D. Alvarez-Melis and T. S. Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. *arXiv preprint arXiv:1809.00013*, Aug. 2018.
- Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*. Springer Science & Business Media, 2008.
- J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, Jan. 2000.
- Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL <https://openreview.net/forum?id=9BnCwiXB0ty>.

- Charlotte Bunne, David Alvarez-Melis, Andreas Krause, and Stefanie Jegelka. Learning Generative Models across Incomparable Spaces. In *International Conference on Machine Learning (ICML)*, volume 97, 2019.
- Charlotte Bunne, Ya-Ping Hsieh, Marco Cuturi, and Andreas Krause. The schrödinger bridge between gaussian measures has a closed form. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 5802–5833. PMLR, 2023.
- Rainer E. Burkard, Bettina Klinz, and Rüdiger Rudolf. Perspectives of monge properties in optimization. *Discrete Applied Mathematics*, 70(2):95–161, 1996. doi: 10.1016/0166-218X(95)00103-X.
- Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169:671–691, 2016.
- Sanjit Dandapanthula, Aleksandr Podkopaev, Shiva Prasad Kasiviswanathan, Aaditya Ramdas, and Ziv Goldfeld. Optimal transportation and alignment between gaussian measures, 2025. URL <https://arxiv.org/abs/2512.03579>.
- A. Dembo, T.M. Cover, and J.A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991. doi: 10.1109/18.104312.
- Ivan Gentil, Christian Léonard, and Luigia Ripani. About the analogy between optimal transport and minimal entropy. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 26, pages 569–600, 2017.
- Guillaume Houry, Jean Feydy, and François-Xavier Vialard. Gromov-wasserstein at scale, beyond squared norms. In *Forty-third International Conference on Machine Learning*, 2026. URL <https://openreview.net/forum?id=P9LofJB8fs>.
- Jiin Im, Sisung Liu, and Je Hyeong Hong. Shape-of-you: Fused gromov-wasserstein optimal transport for semantic correspondence in-the-wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.
- Hicham Janati, Boris Muzellec, Gabriel Peyré, and Marco Cuturi. Entropic optimal transport between unbalanced gaussian measures has a closed form. In *Advances in Neural Information Processing Systems*, volume 33, 2020.
- Dominik Klein, Théo Uscidda, Fabian J Theis, and marco cuturi. GENOT: Entropic (gromov) wasserstein flow matching with applications to single-cell genomics. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL <https://openreview.net/forum?id=hjspWd7jvg>.
- Khang Le, Dung Q Le, Huy Nguyen, Dat Do, Tung Pham, and Nhat Ho. Entropic Gromov-Wasserstein between Gaussian distributions. In *International Conference on Machine Learning*, pages 12164–12203. PMLR, 2022.

- Flavien Léger. A geometric perspective on regularized optimal transport. *Journal of Dynamics and Differential Equations*, 31(4):1777–1791, 2019. doi: 10.1007/s10884-018-9684-9. URL <https://doi.org/10.1007/s10884-018-9684-9>.
- Christian Léonard. From the schrödinger problem to the monge–kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012. doi: 10.1016/j.jfa.2011.11.026.
- Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete & Continuous Dynamical Systems*, 34(4):1533, 2014.
- Daniel Liberzon. *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton University Press, 2012.
- Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *arXiv preprint: arXiv 2107.12364*, 2021.
- Facundo Mémoli. Spectral Gromov-Wasserstein distances for shape matching. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 256–263. IEEE, 2009.
- Facundo Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.*, 11(4):417–487, 2011. URL <http://dblp.uni-trier.de/db/journals/focm/focm11.html#Memoli11>.
- Léonard Monsaingeon, Luca Tamanini, and Dmitry Vorotnikov. The dynamical schrödinger problem in abstract metric spaces. *Advances in Mathematics*, 426:109100, 2023. ISSN 0001-8708. doi: <https://doi.org/10.1016/j.aim.2023.109100>. URL <https://www.sciencedirect.com/science/article/pii/S0001870823002438>.
- Felix Otto. The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations*, 26:101–174, 2001.
- Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pages 2664–2672. PMLR, 2016.
- Antoine Salmona, Julie Delon, and Agnès Desolneux. Gromov-Wasserstein distances between gaussian distributions. *arXiv preprint arXiv:2104.07970*, 2021.
- Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, United Kingdom, April 2019. doi: 10.1017/9781108186735.
- Meyer Scetbon, Gabriel Peyré, and Marco Cuturi. Linear-time Gromov-Wasserstein distances using low rank couplings and costs. In *International Conference on Machine Learning*, pages 19347–19365. PMLR, 2022.
- Dominik Schnaus, Thomas Dagès, Daniel Cremers, Xi Wang, and Phillip Isola. Shared geometry as a rosetta stone: Cross-modal alignment without paired data, 2026. URL <https://arxiv.org/abs/2610.09411>.

- Justin Solomon, Gabriel Peyré, Vladimir G Kim, and Suvrit Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (ToG)*, 35(4):1–13, 2016.
- Karl-Theodor Sturm. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *arXiv preprint arXiv:1208.0434*, 2012.
- Sophia Tang. Foundations of schrödinger bridges for generative modeling, 2026. URL <https://arxiv.org/abs/2603.18992>.
- Alexander Tong, Nikolay Malkin, Kilian FATRAS, Lazar Atanackovic, Yanlei Zhang, Guillaume Hugué, Guy Wolf, and Yoshua Bengio. Simulation-free schrödinger bridges via score and flow matching. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023. URL <https://openreview.net/forum?id=adkj23mvB0>.
- Théo Uscidda, Luca Eyring, Karsten Roth, Fabian J Theis, Zeynep Akata, and marco cuturi. Disentangled representation learning through geometry preservation with the gromov-monge gap. In *ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling*, 2024. URL <https://openreview.net/forum?id=5LWt9sdYSo>.
- Titouan Vayer. A contribution to optimal transport on incomparable spaces. *arXiv preprint arXiv:2011.04447*, 2020.
- Hongteng Xu, Dixin Luo, and Lawrence Carin. Scalable Gromov-Wasserstein learning for graph partitioning and matching. *Advances in neural information processing systems*, 32, 2019a.
- Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin Duke. Gromov-Wasserstein learning for graph matching and node embedding. In *International conference on machine learning*, pages 6932–6941. PMLR, 2019b.
- Zhengxin Zhang, Ziv Goldfeld, Youssef Mroueh, and Bharath K Sriperumbudur. Gromov-Wasserstein distances: Entropic regularization, duality and sample complexity. *The Annals of Statistics*, 52(4):1616–1645, 2024.
- Zhengxin Zhang, Ziv Goldfeld, Kristjan Greenewald, Youssef Mroueh, and Bharath K. Sriperumbudur. Gradient flows and riemannian structure in the gromov-wasserstein geometry: Gradient flows and riemannian structure in the gromov-wasserstein geometry. *Found. Comput. Math.*, 26(3):1911–2003, July 2025. ISSN 1615-3375. doi: 10.1007/s10208-025-09722-w. URL <https://doi.org/10.1007/s10208-025-09722-w>.

Appendix A. Preliminaries

A.1. Wasserstein probability spaces and Wasserstein gradient

Consider the space of probability distributions with bounded second moment, $\mathcal{P}_2(\mathbb{R}^d)$ equipped with the $\mathcal{W}_2$ metric, i.e. $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$. Recall, that the given two measures $(\mu, \nu) \in \mathcal{P}_2(\mathbb{R}^d)$, the $W_2(\mu, \nu)$ is defined as:

$$\mathcal{W}_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int \|x - y\|^2 d\pi(x, y) \quad (22)$$

where, $\Pi(\mu, \nu)$ is the set of all admissible couplings π , such that $\pi(\mathbb{R}^d, y) = \nu$ and $\pi(x, \mathbb{R}^d) = \mu$. For any functional $\mathcal{F}(\rho)$, the wasserstein gradient is defined as:

$$\nabla_{\mathcal{W}} \mathcal{F}(\rho) = \nabla \frac{\delta}{\delta \rho} \mathcal{F}(\rho) \quad (23)$$

where, $\frac{\delta}{\delta \rho}(\cdot)$ is the first variation and ∇ is the Euclidean gradient and the corresponding gradient flow equation is given as:

$$\partial_t \rho = \nabla \cdot (\nabla_{\mathcal{W}} \mathcal{F}(\rho) \rho) \quad (24)$$

A.2. Gromov Wasserstein and Inner-product Gromov Wasserstein

Given two metric measure spaces $(\mathcal{X}, \mu, d_x)$ and $(\mathcal{Y}, \nu, d_Y)$ the corresponding Gromov-Wasserstein distance is defined as :

$$\text{GW}_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \int \|d_x(x, x') - d_Y(y, y')\|^2 d\pi(x, y) d\pi(x', y') \quad (25)$$

where, Π is the space of all admissible couplings such that $\pi(\mathcal{X}, y) = \nu$ and $\pi(x, \mathcal{Y}) = \mu$. In general, we dont need a metric to define the cost and can work with any arbitrary cost function $c_x : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. When, $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $c_x = \langle \cdot, \cdot \rangle$ we get what is known as the IGW metric:

$$\text{IGW}(\mu, \nu)^2 := \inf_{\pi \in \Pi(\mu, \nu)} \int \int \|\langle x, x' \rangle - \langle y, y' \rangle\|^2 d\pi(x, y) d\pi(x', y') \quad (26)$$

Unlike the GW metric which is both rotation and translation invariant, IGW is only rotational invariant, i.e. given any $R \in O(d)$, we have $\text{IGW}(R_{\#} \mu, \nu) = \text{IGW}(\mu, \nu)$. Thus, it defines a quotient space over $\mathcal{P}_2(\mathbb{R}^d)$.

Appendix B. Proofs

B.1. Proof of Theorem 7

Theorem 7 (IGW-SB action lower bound) *Let $(\rho, \mathbf{v}) := (\rho_t, v_t)_{t \in [0,1]}$ be any admissible solution to the dynamical Schrödinger Bridge problem (7) and let $(\Sigma_t, \mathbf{A}_t)_{t \in [0,1]}$ denote the corresponding covariance-cross moment curve. Then, we have the following lower bound:*

$$\mathcal{A}(\rho, \mathbf{v}) \geq \text{tr} \left(\int_0^1 (2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^{\top} + 2\mathbf{A}_t^2) dt + \frac{\varepsilon^2}{4} \int_0^1 \Sigma_t^{-2} dt \right) \quad (17)$$

Proof The IGW functional is given by:

$$\mathcal{A}(\rho, \mathbf{v}) = \underbrace{\int_0^1 g_{\rho_t}(v_t, v_t) dt}_{(I)} + \varepsilon^2 \underbrace{\int_0^1 \mathcal{I}_{\text{IGW}}(\rho_t) dt}_{(II)} \quad (27)$$

For any $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and $v \in L^2(\rho; \mathbb{R}^d)$ we bound both the terms separately and then apply the bound pointwise for $t \in [0, 1]$. Recall that the cross-moment matrix is given by:

$$\mathbf{A} = \int v(y) y^{\top} d\rho(y) \quad (28)$$

Bounding (I): We begin by expanding the quadratic form induced by the IGW geometry as:

$$g_\rho(v, v) = \int \langle v(x), \mathcal{L}_{\Sigma, \rho}[v](x) \rangle d\rho(x) \quad (29)$$

$$= 2 \int \langle v(x), \Sigma v(x) \rangle d\rho(x) + 2 \int \left\langle v(x), \int y \langle v(y), x \rangle d\rho(y) \right\rangle d\rho(x) \quad (30)$$

$$= 2 \int \text{tr} \left(\Sigma v(x) v(x)^\top \right) d\rho(x) + 2 \int \left\langle v(x), \left(\int y v(y)^\top d\rho(y) \right) x \right\rangle d\rho(x) \quad (31)$$

$$= 2 \int \text{tr} \left(\Sigma v(x) v(x)^\top \right) d\rho(x) + 2 \int \langle v(x), \mathbf{A}^\top x \rangle d\rho(x) \quad (32)$$

$$= 2 \int \text{tr} \left(\Sigma v(x) v(x)^\top \right) d\rho(x) + 2 \int \text{tr}(\mathbf{A}^\top x v(x)^\top) d\rho(x) \quad (33)$$

$$= \text{tr} \left(2\Sigma \int v(x) v(x)^\top \rho(x) dx + 2\mathbf{A}^2 \right) \quad (34)$$

For a fixed $\Sigma \succ 0$, $\mathbf{A}$, from Lemma 12, we get:

$$\int v(x) v(x)^\top \rho(x) dx \succeq \mathbf{A} \Sigma^{-1} \mathbf{A}^\top \quad (35)$$

$$\text{tr} \left(\Sigma \int v(x) v(x)^\top \rho(x) dx \right) \geq \text{tr} \left(\Sigma \mathbf{A} \Sigma^{-1} \mathbf{A}^\top \right) \quad (36)$$

$$g_\rho(v, v) \geq \text{tr}(2\Sigma \mathbf{A} \Sigma^{-1} \mathbf{A}^\top + 2\mathbf{A}^2) \quad (37)$$

and the equality is attained at $v(x) = \mathbf{A} \Sigma^{-1} x$. Applying (37) pointwise to the left hand size for all $t \in [0, 1]$ in (27), we get:

$$\int_0^1 g_{\rho_t}(v_t, v_t) dt \geq \int_0^1 \text{tr}(2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2) dt \quad (38)$$

Next, we lower bound $\mathcal{I}_{\text{IGW}}(\rho)$:

Bounding (II):

$$\mathcal{I}_{\text{IGW}}(\rho) := \int \langle \nabla \log \rho(x), \mathcal{L}_{\Sigma, \rho}^{-1} \nabla \log \rho(x) \rangle d\rho(x) \quad (39)$$

Using Lemma 10, we can write the FI functional $\mathcal{I}_{\text{IGW}}(\rho)$ as:

$$\mathcal{I}_{\text{IGW}}(\rho) = \int \langle \nabla \log \rho(x), \mathcal{L}_{\Sigma, \rho}^{-1} \nabla \log \rho(x) \rangle \rho(x) dx \quad (40)$$

$$= \int \left\langle \nabla \log \rho(x), \left(\frac{1}{2} \Sigma^{-1} \nabla \log \rho(x) + \frac{1}{4} \Sigma^{-2} x \right) \right\rangle \rho(x) dx \quad (41)$$

$$= \underbrace{\frac{1}{2} \int \langle \nabla \log \rho(x), \Sigma^{-1} \nabla \log \rho(x) \rangle \rho(x) dx}_{(I)} + \underbrace{\frac{1}{4} \int \langle \nabla \log \rho(x), \Sigma^{-2} x \rangle \rho(x) dx}_{(II)} \quad (42)$$

For (II) using Lemma C,

$$\int \langle \nabla \log \rho(x), \Sigma^{-2} x \rangle \rho(x) dx = \text{tr} \left(\Sigma^{-2} \int \nabla \log \rho(x) x^\top \rho(x) dx \right) \quad (43)$$

$$= -\text{tr}(\Sigma^{-2}) \quad (44)$$

For (I), we have:

$$\int \langle \nabla \log \rho(x), \Sigma^{-1} \nabla \log \rho(x) \rangle \rho(x) dy = \text{tr} \left(\Sigma^{-1} \int \nabla \log \rho(x) \nabla \log \rho^\top \rho(x) dx \right) \quad (45)$$

Now, $\int \nabla \log \rho(x) \nabla \log \rho(x)^\top \rho(x) dx$ is the Fisher-Information Matrix and can be lower bounded using the Cramer-rao lower bound (Dembo et al., 1991)(Theorem 20):

$$\int \nabla \log \rho(x) \nabla \log \rho(x)^\top \rho(x) dx \succeq \Sigma^{-1} \quad (46)$$

Combining (I) and (II), we have:

$$\mathcal{I}_{\text{IGW}}(\rho) \geq \text{tr} \left(\frac{1}{4} \Sigma^{-2} \right) \quad (47)$$

Combining (38) and (47), we get :

$$\mathcal{A}(\rho, \mathbf{v}) \geq \int_0^1 \text{tr}(2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2) dt + \frac{\varepsilon^2}{4} \int_0^1 \text{tr}(\Sigma_t^{-2}) dt. \quad (48)$$

■

B.2. Proof of Lemma 8

Lemma 8 (Gaussian realization) Consider any admissible pair $(\rho, \mathbf{v})$ and the corresponding $(\Sigma_t, \mathbf{A}_t)_{t \in [0,1]}$. The Gaussian path $(\rho_G, \mathbf{v}_G) = (\mathcal{N}(0, \Sigma_t), \mathbf{A}_t \Sigma_t^{-1} x)_{t \in [0,1]}$ satisfies the endpoint and continuity constraints, and attains the lower bound in Theorem 7, i.e.:

$$\mathcal{A}(\rho_G, \mathbf{v}_G) = \int_0^1 \text{tr} \left(2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2 + \frac{\varepsilon^2}{4} \Sigma_t^{-2} \right) dt.$$

Proof Since, $\mu_0 = \mathcal{N}(0, \Sigma_0)$ is a gaussian and, for linear dynamics $v_t(x) = \mathbf{A}_t \Sigma_t^{-1} x$ we know (Section 6.1, (Särkkä and Solin, 2019)) gaussianity is preserved, i.e. $X_t = \mathcal{N}(0, \mathbf{C}_t)$ for some covariance that covariance satisfies

$$\dot{C}_t = \mathbf{A}_t \Sigma_t^{-1} C_t + C_t \Sigma_t^{-1} \mathbf{A}_t^\top, \quad C_0 = \Sigma_0.$$

Since Σ_t satisfies the same equation with the same initial condition, uniqueness gives $C_t = \Sigma_t$. Thus $(\rho_G, \mathbf{v}_G)$ satisfies the continuity equation and the prescribed Gaussian endpoint conditions. Finally,

$$\int v_{G,t}(x) v_{G,t}(x)^\top \rho_{G,t}(x) dx = \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top, \quad \int v_{G,t}(x) x^\top \rho_{G,t}(x) dx = \mathbf{A}_t. \quad (49)$$

Substituting these identities into the kinetic action and using $\mathcal{I}_{\text{IGW}}(\rho_{G,t}) = \frac{1}{4} \text{tr}(\Sigma_t^{-2})$ yields

$$\mathcal{A}(\rho_G, \mathbf{v}_G) = \int_0^1 \text{tr} \left(2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2 + \frac{\varepsilon^2}{4} \Sigma_t^{-2} \right) dt$$

which is the lower bound in Theorem 7. ■

B.3. Proof of Theorem 9

Theorem 9 (Closed-form solution of the covariance lower bound) *Let $Q_\mu, \Lambda_\mu, Q_\nu, \Lambda_\nu$ and the orthogonal path $(Q_t)_{t \in [0,1]}$ and the eigenvalue pairing as defined in Eqs. (8-9). Consider the covariance-level variational problem:*

$$\inf_{\substack{\Sigma_t > 0, \mathbf{A}_t \\ \dot{\Sigma}_t = \mathbf{A}_t + \mathbf{A}_t^\top \\ \Sigma_0 = \Sigma_\mu, \Sigma_1 = \Sigma_\nu}} \int_0^1 \left[\text{tr} \left(2\Sigma_t \mathbf{A}_t \Sigma_t^{-1} \mathbf{A}_t^\top + 2\mathbf{A}_t^2 \right) + \frac{\varepsilon^2}{4} \text{tr}(\Sigma_t^{-2}) \right] dt. \quad (18)$$

Then

$$\Sigma_t = Q_t \Lambda_t Q_t^\top, \quad \mathbf{A}_t = \dot{Q}_t \Lambda_t Q_t^\top + \frac{1}{2} Q_t \dot{\Lambda}_t Q_t^\top, \quad (19)$$

where

$$\Lambda_t = \text{diag}(\lambda_1(t), \dots, \lambda_d(t)), \quad (20)$$

$$\lambda_i(t) = \sqrt{(1-t)^2 \lambda_{\mu,i}^2 + t^2 \lambda_{\nu,i}^2 + t(1-t) \sqrt{4\lambda_{\mu,i}^2 \lambda_{\nu,i}^2 + \varepsilon^2}}, \quad i = 1, \dots, d. \quad (21)$$

Proof In the objective, we handle the two terms separately:

$$\underbrace{\text{tr} \left(\int_0^1 2\Sigma \mathbf{A} \Sigma^{-1} \mathbf{A}^\top + 2\mathbf{A}^2 \right)}_{(I)} + \frac{\varepsilon^2}{4} \underbrace{\text{tr} \left(\int_0^1 \Sigma^{-2} dt \right)}_{(II)} \quad (50)$$

For (I), fix $t \in [0, 1]$ and, for notational simplicity, suppress the time dependence. Let

$$\Sigma = Q \Lambda Q^\top \quad (51)$$

be an eigendecomposition of Σ , with $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$ and $Q Q^\top = Q^\top Q = \mathbf{I}$. We express $\mathbf{A}$ and $\dot{\Sigma}$ in the same eigenbasis by defining

$$\tilde{\mathbf{A}} := Q^\top \mathbf{A} Q, \quad \tilde{\Sigma} := Q^\top \dot{\Sigma} Q. \quad (52)$$

Then, we can write (I) as:

$$\text{tr}(2\Sigma \mathbf{A} \Sigma^{-1} \mathbf{A}^\top + 2\mathbf{A}^2) = \text{tr}(2Q \Lambda Q^\top \mathbf{A} Q \Lambda^{-1} Q^\top \mathbf{A}^\top + 2\mathbf{A}^2) \quad (53)$$

$$= \text{tr}(2\Lambda \tilde{\mathbf{A}} \Lambda^{-1} \tilde{\mathbf{A}}^\top + 2\tilde{\mathbf{A}}^2) \quad (54)$$

$$= 2 \sum_{ij} \frac{\lambda_i}{\lambda_j} \tilde{\mathbf{A}}_{ij}^2 + 2 \sum_{ij} \tilde{\mathbf{A}}_{ij} \tilde{\mathbf{A}}_{ji} \quad (55)$$

Next, we simplify the constraints:

$$\Sigma = Q\Lambda Q^\top \quad (56)$$

Differentiating with respect to t , we get:

$$\dot{\Sigma} = \dot{Q}\Lambda Q^\top + Q\dot{\Lambda}Q^\top + Q\Lambda\dot{Q}^\top \quad (57)$$

$$\implies \tilde{\Sigma} = \dot{\Lambda} + Q^\top \dot{Q}\Lambda + \Lambda\dot{Q}^\top Q \quad (58)$$

Further, since Q is an orthogonal matrix, we get:

$$QQ^\top = \mathbf{I} \implies \dot{Q}Q^\top + Q\dot{Q}^\top = 0 \quad (59)$$

$$(60)$$

Thus, $Q^\top \dot{Q}$ is a skew-symmetric matrix. Let

$$(Q^\top \dot{Q})_{ij} = \begin{cases} 0, & i = j \\ q_{ij}, & i \neq j. \end{cases} \quad (61)$$

Now from the constraint, we have:

$$\mathbf{A} + \mathbf{A}^\top = \dot{\Sigma} \quad (62)$$

$$\implies \tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top = \tilde{\Sigma} \quad (63)$$

$$\implies \tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top = \dot{\Lambda} + Q^\top \dot{Q}\Lambda + \Lambda\dot{Q}^\top Q \quad (64)$$

Thus,

$$\tilde{\mathbf{A}}_{ii} = \frac{\dot{\lambda}_i}{2} \quad (65)$$

$$\tilde{\mathbf{A}}_{ij} + \tilde{\mathbf{A}}_{ji} = (\lambda_j - \lambda_i)q_{ij} \quad (66)$$

Thus, our goal is to minimize (55) subject to (66). Setting $\tilde{\mathbf{A}}_{ij} = \lambda_j q_{ij}$, $\forall i \neq j$ and $\tilde{\mathbf{A}}_{ii} = \frac{\dot{\lambda}_i}{2}$, $\forall i = j$ minimizes (55). And (55) reduces to :

$$\sum_i \dot{\lambda}_i^2 \quad (67)$$

For (II), we have:

$$\text{tr}(\Sigma^{-2}) = \sum_{i=1}^d \frac{1}{\lambda_i^2} \quad (68)$$

Thus, combining (I) and (II) for all $t \in [0, 1]$, the problem:

$$\inf_{\substack{Q_t \in O(d), \Lambda_t \succ 0 \\ Q_0 \Lambda_0 Q_0^\top = \Sigma_\mu \\ Q_1 \Lambda_1 Q_1^\top = \Sigma_\nu}} \int_0^1 \sum_{i=1}^d \left(\dot{\lambda}_i(t)^2 + \frac{\varepsilon^2}{4\lambda_i(t)^2} \right) dt. \quad (69)$$

This can be reduced to solving independent Euler-Lagrange equations:

$$\sum_{i=1}^d \inf_{\substack{\lambda_i(0)=\lambda_{\mu,i} \\ \lambda_i(1)=\lambda_{\nu,i}}} \int_0^1 \left(\dot{\lambda}_i(t)^2 + \frac{\varepsilon^2}{4\lambda_i(t)^2} \right) dt \quad (70)$$

along with the feasibility condition:

$$Q_t \in O(d), \quad Q_0 \Lambda_0 Q_0^\top = \Sigma_\mu \quad Q_1 \Lambda_1 Q_1^\top = \Sigma_\nu. \quad (71)$$

Solving the Euler-Lagrange equation (Liberzon, 2012): For each i , define

$$L(\lambda_i, \dot{\lambda}_i) = \dot{\lambda}_i^2 + \frac{\varepsilon^2}{4\lambda_i^2}. \quad (72)$$

Then,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\lambda}_i} - \frac{\partial L}{\partial \lambda_i} = 0, \quad (73)$$

which gives

$$2\ddot{\lambda}_i(t) + \frac{\varepsilon^2}{2\lambda_i(t)^3} = 0. \quad (74)$$

Multiplying (74) by $\dot{\lambda}_i(t)$ and integrating once yields

$$\dot{\lambda}_i(t)^2 - \frac{\varepsilon^2}{4\lambda_i(t)^2} = C_i, \quad (75)$$

for some constant C_i . Setting $z_i(t) = \lambda_i(t)^2$, we obtain from (75)

$$\dot{z}_i(t)^2 = \varepsilon^2 + 4C_i z_i(t), \quad (76)$$

and hence $\ddot{z}_i(t) = 2C_i$. Therefore $z_i(t)$ is quadratic in t . Imposing the boundary conditions

$$\lambda_i(0) = \lambda_{\mu,i}, \quad \lambda_i(1) = \lambda_{\nu,i}, \quad (77)$$

gives

$$\lambda_i(t) = \sqrt{(1-t)^2 \lambda_{\mu,i}^2 + t^2 \lambda_{\nu,i}^2 + t(1-t) \sqrt{4\lambda_{\mu,i}^2 \lambda_{\nu,i}^2 + \varepsilon^2}}. \quad (78)$$

Thus, for a fixed pairing of the endpoint eigenvalues, the finite-dimensional problem decouples into d independent scalar variational problems, each admitting the closed-form solution (78). For $\varepsilon > 0$, the Lagrangian L is strictly convex. The derived Euler-Lagrange solution is then the unique global minimizer of the scalar action.

Selecting the eigenvalue pairing. It remains to determine the optimal pairing between the eigenvalues of Σ_μ and Σ_ν . Define the scalar optimal value

$$c_\varepsilon(a, b) := \inf_{\substack{\lambda(0)=a \\ \lambda(1)=b}} \int_0^1 \left(\dot{\lambda}(t)^2 + \frac{\varepsilon^2}{4\lambda(t)^2} \right) dt. \quad (79)$$

For the optimal trajectory, the endpoint variation formula gives

$$\frac{\partial c_\varepsilon}{\partial b}(a, b) = \frac{\partial L}{\partial \dot{\lambda}} \Big|_{t=1} = 2\dot{\lambda}(1) \quad (80)$$

$$= 2b - \frac{\sqrt{4a^2b^2 + \varepsilon^2}}{b}. \quad (81)$$

Consequently,

$$\frac{\partial^2 c_\varepsilon}{\partial a \partial b}(a, b) = -\frac{4ab}{\sqrt{4a^2b^2 + \varepsilon^2}} < 0. \quad (82)$$

Hence, for $a_1 \leq a_2$ and $b_1 \leq b_2$,

$$c_\varepsilon(a_1, b_1) + c_\varepsilon(a_2, b_2) \leq c_\varepsilon(a_1, b_2) + c_\varepsilon(a_2, b_1). \quad (83)$$

Thus, c_ε satisfies the Monge property. By the standard uncrossing argument for assignment problems with Monge costs (Burkard et al., 1996), the optimal assignment is monotone. Therefore, if the endpoint eigenvalues are ordered as

$$\lambda_{\mu,1} \leq \dots \leq \lambda_{\mu,d}, \quad \lambda_{\nu,1} \leq \dots \leq \lambda_{\nu,d}, \quad (84)$$

an optimal pairing is

$$\pi^*(i) = i, \quad i = 1, \dots, d. \quad (85)$$

Selecting the orthogonal path. Having fixed an optimal pairing π^* , let

$$\Sigma_\mu = Q_\mu \Lambda_\mu Q_\mu^\top, \quad \Sigma_\nu = Q_\nu \Lambda_\nu Q_\nu^\top. \quad (86)$$

Let P_{π^*} denote the permutation matrix associated with π^* . The corresponding terminal eigenframe is $Q_\nu P_{\pi^*}$. Eigenvector signs can be chosen so that $\det Q_\mu = \det(Q_\nu P_{\pi^*})$, without changing either endpoint covariance matrix. The two frames then belong to the same connected component of $O(d)$ and can be joined by a continuously differentiable orthogonal path. For the optimal eigenvalue trajectories, any such path Q_t yields the same optimal action. Thus, the minimization determines the optimal eigenvalue trajectories and their pairing, while the choice of orthogonal path remains non-unique. $\blacksquare$

Appendix C. Auxilliary Lemmas

Lemma 10 (IGW gradient of entropy) *Let ρ be a smooth, positive probability density with positive-definite second-moment matrix Σ and finite Fisher information. For $\mathcal{H}(\rho) = \int \rho \log \rho \, dx$, we have*

$$\mathcal{L}_{\Sigma, \rho}^{-1} \nabla_W \mathcal{H}(\rho)(x) = \frac{1}{2} \Sigma^{-1} \nabla \log \rho(x) + \frac{1}{4} \Sigma^{-2} x, \quad (87)$$

Proof Integration by parts gives the score identity

$$\int \nabla \log \rho(x) x^\top \, d\rho(x) = -\mathbf{I}_d, \quad (88)$$

since, entry-wise,

$$\int \partial_i \rho(x) x_j \, dx = - \int \rho(x) \partial_i x_j \, dx = -\delta_{ij}. \quad (89)$$

Define

$$u(x) = \frac{1}{2} \Sigma^{-1} \nabla \log \rho(x) + \frac{1}{4} \Sigma^{-2} x. \quad (90)$$

Its cross moment is

$$\mathbf{B} := \int u(x) x^\top \, d\rho(x) = -\frac{1}{2} \Sigma^{-1} + \frac{1}{4} \Sigma^{-2} \Sigma \quad (91)$$

$$= -\frac{1}{4} \Sigma^{-1}. \quad (92)$$

In particular, $\mathbf{B}$ is symmetric. Therefore,

$$\mathcal{L}_{\Sigma, \rho}[u](x) = 2\Sigma u(x) + 2 \left(\int y u(y)^\top \, d\rho(y) \right) x \quad (93)$$

$$= \nabla \log \rho(x) + \frac{1}{2} \Sigma^{-1} x + 2\mathbf{B}^\top x \quad (94)$$

$$= \nabla \log \rho(x). \quad (95)$$

■

Lemma 11 (Matrix Score Identity)

$$\int \nabla \log \rho(x) x^\top \, d\rho(x) = -\mathbf{I}_d, \quad (96)$$

Proof For the (i, j) -th entry, using Integration by Parts we have:

$$\int \partial_i \rho(x) x_j \, dx = - \int \rho(x) \partial_i x_j \, dx = -\delta_{ij}. \quad (97)$$

■

Lemma 12 (Matrix Cauchy-Schwarz inequality) *Let $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and $v \in L^2(\rho; \mathbb{R}^d)$. Define*

$$\Sigma = \int x x^\top d\rho(x), \quad \mathbf{A} = \int v(x) x^\top d\rho(x), \quad (98)$$

and assume $\Sigma \succ 0$. Then

$$\int v(x) v(x)^\top d\rho(x) \succeq \mathbf{A} \Sigma^{-1} \mathbf{A}^\top. \quad (99)$$

Equality holds if and only if $v(x) = \mathbf{A} \Sigma^{-1} x$ for ρ -almost every x .

Proof Consider the symmetric block matrix:

$$M = \begin{bmatrix} \Sigma & \mathbf{A}^\top \\ \mathbf{A} & \int v(x) v(x)^\top d\rho(x) \end{bmatrix} \quad (100)$$

As an integral of positive semidefinite matrices, $M \succeq 0$. Since $\Sigma \succ 0$, the Schur complement of Σ in M is positive semidefinite:

$$\int v(x) v(x)^\top d\rho(x) - \mathbf{A} \Sigma^{-1} \mathbf{A}^\top \succeq 0. \quad (101)$$

Moreover, this Schur complement equals

$$\int (v(x) - \mathbf{A} \Sigma^{-1} x) (v(x) - \mathbf{A} \Sigma^{-1} x)^\top d\rho(x). \quad (102)$$

It vanishes if and only if $v(x) = \mathbf{A} \Sigma^{-1} x$ ρ -almost everywhere. ■

Appendix D. Comparison with the Wasserstein Schrödinger Problem ($\mathcal{W}_2$ -SB)

In this section, we formally compare IGW-SB and $\mathcal{W}_2$ -SB. We examine the corresponding Fisher information term, their relationship with static entropy-regularized transport, and the applicability of the standard controlled-SDE formulation. We begin with the dynamical formulation of $\mathcal{W}_2$ -SB, which is the Wasserstein counterpart of (7):

$$\text{SB}_{\mathcal{W}_2}(\mu, \nu) = \inf_{\substack{\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0 \\ \rho_0 = \mu, \rho_1 = \nu}} \int_0^1 \left[\int_{\mathbb{R}^d} \|v_t(x)\|^2 d\rho_t(x) + \varepsilon^2 \mathcal{I}_{\mathcal{W}_2}(\rho_t) \right] dt, \quad (103)$$

where $\mathcal{I}_{\mathcal{W}_2}(\rho) = \int_{\mathbb{R}^d} \|\nabla \log \rho(x)\|^2 d\rho(x)$.

D.1. Gaussian bridges in Wasserstein and IGW geometry

For one-dimensional centered Gaussian measures, writing λ for the variance, the scalar actions are

$$L_{\mathcal{W}_2}(\lambda, \dot{\lambda}) = \frac{\dot{\lambda}^2}{4\lambda} + \frac{\varepsilon^2}{\lambda}, \quad L_{\text{IGW}}(\lambda, \dot{\lambda}) = \dot{\lambda}^2 + \frac{\varepsilon^2}{4\lambda^2}. \quad (104)$$

For endpoint variances $a, b > 0$, the optimal trajectories are

$$\lambda_{\mathcal{W}_2}(t) = (1-t)^2 a + t^2 b + 2t(1-t)\sqrt{ab + \varepsilon^2}, \quad (105)$$

$$\lambda_{\text{IGW}}(t) = \sqrt{(1-t)^2 a^2 + t^2 b^2 + t(1-t)\sqrt{4a^2 b^2 + \varepsilon^2}}. \quad (106)$$

Here the Wasserstein formula uses a Brownian variance rate of 2ε . The Fisher-information terms are $\mathcal{I}_{\mathcal{W}_2} = \lambda^{-1}$ and $\mathcal{I}_{\text{IGW}} = (4\lambda^2)^{-1}$. The two geometries also differ in their dependence on the orientation of the endpoint covariances. In Wasserstein geometry, their relative orientation affects the optimal bridge. For IGW-SB, the optimal value depends only on their ordered eigenvalues. Although the endpoint frames must realize the prescribed covariances, any admissible orthogonal path Q_t connecting compatible frames yields the same optimal action. Thus, the minimization does not determine the orientation of the intermediate covariances.

D.2. Comparison with static entropic IGW

In the $\mathcal{W}_2$ setting, the dynamical Schrödinger bridge problem is equivalent, up to endpoint-dependent terms, to static entropy-regularized optimal transport. In the IGW setting, [Le et al. \(2022\)](#) study the static problem

$$\inf_{\pi \in \Pi(\mu, \nu)} \{ \mathcal{D}(\pi) + 2\varepsilon \text{KL}(\pi \| \mu \otimes \nu) \}, \quad (107)$$

where $\mathcal{D}(\pi)$ is the IGW distortion:

$$\mathcal{D}(\pi) := \iint (\langle x, x' \rangle - \langle y, y' \rangle)^2 d\pi(x, y) d\pi(x', y'). \quad (108)$$

We show that, for centered Gaussian endpoints, the dynamical IGW-SB optimal value admits a static representation with a *modified dependence penalty* in place of this KL penalty and thus the static and the entropic formulation are in general, *not equivalent*. Let $\mu = \mathcal{N}(0, \Sigma_\mu)$ and $\nu = \mathcal{N}(0, \Sigma_\nu)$, with positive definite covariances. For $\pi \in \Pi(\mu, \nu)$, define its cross-covariance and normalized cross-covariance by

$$K_\pi := \int xy^\top d\pi(x, y), \quad C_\pi := \Sigma_\mu^{-1/2} K_\pi \Sigma_\nu^{-1/2}. \quad (109)$$

For these Gaussian marginals, the IGW distortion simplifies to

$$\mathcal{D}(\pi) = \text{tr}(\Sigma_\mu^2) + \text{tr}(\Sigma_\nu^2) - 2\|K_\pi\|_F^2. \quad (110)$$

We define the modified dependence penalty by

$$\mathcal{R}(\pi) := \text{KL}(\pi \| \mu \otimes \nu) - \frac{1}{2} \log \det(I + C_\pi C_\pi^\top), \quad (111)$$

with $\mathcal{R}(\pi) := +\infty$ when the KL divergence is infinite.

Proposition 13 (Static representation of the Gaussian optimal value) *Let $\varepsilon > 0$. Then $\mathcal{R} \geq 0$ and*

$$\text{SB}_{\text{IGW}, \varepsilon}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \{ \mathcal{D}(\pi) + 2\varepsilon \mathcal{R}(\pi) \} + \varepsilon \log \frac{\varepsilon^d}{\det \Sigma_\mu \det \Sigma_\nu} - d\varepsilon. \quad (112)$$

Let $a_i = \lambda_{\mu,i}$ and $b_i = \lambda_{\nu,i}$ be increasingly ordered. An optimal static coupling is Gaussian with cross-covariance

$$K^* = Q_\mu \text{diag}(\sqrt{a_i b_i z_i^*})_{i=1}^d Q_\nu^\top, \quad z_i^* = \frac{\sqrt{4a_i^2 b_i^2 + \varepsilon^2} - \varepsilon}{2a_i b_i} \in (0, 1). \quad (113)$$

Proof We first reduce the static problem to Gaussian couplings. At fixed K_π , replacing π by the Gaussian coupling with the same covariance preserves $\mathcal{D}$ and the log-determinant term, and decreases KL by Gaussian entropy maximization. For this Gaussian coupling,

$$\mathcal{R}(\pi) = -\frac{1}{2} \log \det(I - (C_\pi C_\pi^\top)^2) = -\frac{1}{2} \sum_{i=1}^d \log(1 - z_i^2), \quad (114)$$

where $z_i \in [0, 1)$ are the squared singular values of C_π . This expression also proves nonnegativity for arbitrary π with finite KL. Next, the trace inequality in [Le et al. \(2022, Lemma 3.4\)](#) gives $\|K_\pi\|_F^2 \leq \sum_i a_i b_i z_i$ when z_i are increasingly ordered, with equality for aligned eigenframes. The static infimum therefore reduces to a sum of scalar problems:

$$\min_{0 \leq z < 1} \{a^2 + b^2 - 2abz - \varepsilon \log(1 - z^2)\}. \quad (115)$$

The objective is strictly convex, with minimizer

$$z^* = \frac{\sqrt{4a^2 b^2 + \varepsilon^2} - \varepsilon}{2ab}.$$

These minimizers increase with ab and hence respect the required ordering, yielding (113). Finally, the scalar minimum is

$$c_\varepsilon(a, b) + \varepsilon + \varepsilon \log \frac{ab}{\varepsilon}, \quad c_\varepsilon(a, b) = a^2 + b^2 - \sqrt{4a^2 b^2 + \varepsilon^2} + \varepsilon \operatorname{arsinh} \frac{\varepsilon}{2ab}. \quad (116)$$

Substituting the trajectory from Theorem 9 into $\int_0^1 (\dot{\lambda}^2 + \varepsilon^2 / (4\lambda^2)) dt$ gives $c_\varepsilon(a, b)$. By Theorem 4, $\text{SB}_{\text{IGW}, \varepsilon} = \sum_i c_\varepsilon(a_i, b_i)$. Summing (116) and rearranging proves (112). $\blacksquare$

D.3. Controlled-SDE formulation

In $\mathcal{W}_2$ geometry, the dynamical Schrödinger bridge problem in (103) admits a stochastic control formulation: it minimizes the control effort required to steer a Brownian reference process between the prescribed endpoint distributions ([Chen et al., 2016](#)). We examine whether an analogous formulation can be obtained in IGW geometry by using the entropy gradient flow as the reference dynamics.

Entropy-flow reference dynamics. Akin to $\mathcal{W}_2$, we consider the gradient flow of $\mathcal{H}(\rho)$ functional to define the reference dynamics in IGW geometry. Let $\mathcal{H}(\rho) = \int \log \rho d\rho$ By Lemma 10, its IGW gradient is

$$w_\rho(x) := \nabla_{\text{IGW}} \mathcal{H}(\rho)(x) = \frac{1}{2} \Sigma_\rho^{-1} \nabla \log \rho(x) + \frac{1}{4} \Sigma_\rho^{-2} x. \quad (117)$$

Consequently, the entropy gradient-flow equation $\partial_t \rho = \varepsilon \nabla \cdot (\rho w_\rho)$ is the Fokker–Planck equation associated with

$$dX_t = -\frac{\varepsilon}{4} \Sigma_{\rho_t}^{-2} X_t dt + \sqrt{\varepsilon} \Sigma_{\rho_t}^{-1/2} dW_t, \quad \rho_t = \text{Law}(X_t). \quad (118)$$

This SDE represents the entropy flow at the level of marginal laws; its coefficients depend on the evolving law through Σ_{ρ_t} .

Fixed-reference control cost. Fix a continuous positive definite matrix curve (Σ_t) , and let R^Σ and $P^{\Sigma,u}$ denote the path laws of

$$dX_t = \left(-\frac{\varepsilon}{4} \Sigma_t^{-2} X_t + u_t(X_t) \right) dt + \sqrt{\varepsilon} \Sigma_t^{-1/2} dW_t, \quad (119)$$

with $u = 0$ and with control u , respectively, and with the same initial law. Under the usual Girsanov integrability conditions,

$$4\varepsilon \text{KL}(P^{\Sigma,u} \| R^\Sigma) = 2 \int_0^1 \int u_t(x)^\top \Sigma_t u_t(x) d\rho_t(x) dt. \quad (120)$$

When $\Sigma_t = \int x x^\top d\rho_t$, the full IGW quadratic form is

$$g_{\rho_t}(u_t, u_t) = 2 \int u_t^\top \Sigma_t u_t d\rho_t + 2 \text{tr}(\mathbf{A}_{u,t}^2), \quad \mathbf{A}_{u,t} := \int u_t(x) x^\top d\rho_t(x). \quad (121)$$

Thus, the Girsanov cost omits the cross-moment term and does not, in general, establish an equivalence between IGW-SB and the standard path-space KL minimization problem.