STOCHASTIC INTERPOLANTS IV

Instantiations and Extensions: Diffusive, One-Sided, Mirror, and Schrödinger Bridge Interpolants

Ziseok Lee

April 30, 2026

DOCUMENT IV: ROLE IN SEMESTER I

Documents I–III developed the general stochastic interpolant theory:

$$x_t = I(t, x_0, x_1) + \gamma(t) z.$$

This document studies four major instantiations:

diffusive interpolants : Brownian bridges and point-mass sampling,
one-sided interpolants : $\rho_0 = \mathcal{N}(0, I_d)$,
mirror interpolants : $\rho_0 = \rho_1$,
Schrödinger bridges : optimization over the interpolant itself.

The guiding question:

$$\boxed{\text{What changes when we choose special forms of the bridge?}}$$

SECTION MAP

Section	Mathematical content
3.1	Diffusive interpolants and Brownian bridge construction
3.1	Point-mass conditional sampling theorem
3.2	One-sided interpolants for Gaussian base
3.3	Mirror interpolants
3.4	Schrödinger bridges and max-min formulation

Each construction remains governed by the same principle:

$$\text{interpolant law} \Rightarrow \text{TE/FPE} \Rightarrow \text{generative ODE/SDE} \Rightarrow \text{quadratic objectives.}$$

DIFFUSIVE INTERPOLANTS: MOTIVATION

The original stochastic interpolant uses:

$$x_t = I(t, x_0, x_1) + \gamma(t) z.$$

The noise is sampled independently at each time when training.

A natural pathwise alternative is to use a Brownian bridge:

$$B_0 = B_1 = 0.$$

This gives paths that are continuous but not differentiable:

$$x_t^d = I(t, x_0, x_1) + \sqrt{2a(t)}\, B_t.$$

Key fact:

$$B_t \sim \mathcal{N}(0, t(1-t) I_d).$$

Thus the one-time marginals are still representable by an ordinary stochastic interpolant.

DEFINITION 30: DIFFUSIVE INTERPOLANT

A diffusive interpolant is:

$$x_t^d = I(t, x_0, x_1) + \sqrt{2a(t)}\, B_t, \qquad t \in [0,1],$$

where:

$$(x_0, x_1) \sim \nu, \qquad B_t \perp (x_0, x_1),$$

B_t is a standard Brownian bridge,

and:

$$a \in C^2([0,1]), \qquad a(0) > 0, \qquad a(t) \geq 0.$$

Endpoint check:

$$B_0 = B_1 = 0 \quad \Rightarrow \quad x_0^d = x_0, \qquad x_1^d = x_1.$$

BROWNIAN BRIDGE FACTS

A standard Brownian bridge can be written:

$$B_t = W_t - tW_1.$$

Its one-time distribution is:

$$B_t \sim \mathcal{N}(0, t(1-t)I_d).$$

It also solves the SDE:

$$\mathrm{d}B_t = -\frac{B_t}{1-t}\,\mathrm{d}t + \mathrm{d}W_t, \qquad B_0 = 0.$$

The singular drift forces:

$$B_1 = 0.$$

SAME ONE-TIME MARGINALS

Because:

$$B_t \overset{d}{=} \sqrt{t(1-t)}\, z, \qquad z \sim \mathcal{N}(0, I_d),$$

we have:

$$x_t^d \overset{d}{=} I(t, x_0, x_1) + \sqrt{2a(t)t(1-t)}\, z.$$

Therefore the diffusive interpolant has the same one-time law as:

$$x_t = I(t, x_0, x_1) + \gamma(t) z,$$

with:

$$\boxed{\gamma(t) = \sqrt{2a(t)t(1-t)}.}$$

Thus all marginal results from Documents II–III apply.

PATHWISE VS MARGINAL EQUIVALENCE

The processes:

$$x_t^d = I + \sqrt{2a(t)} B_t$$

and:

$$x_t = I + \sqrt{2a(t)t(1-t)} z$$

are not the same process.

But:

$$\mathcal{L}(x_t^d) = \mathcal{L}(x_t) \qquad \forall t.$$

Generative modeling depends only on the marginal path:

$$\rho(t) = \mathcal{L}(x_t).$$

Therefore the Brownian-bridge version is mainly useful for deriving intuition and alternative FPEs.

CONSTANT a: ITÔ CALCULATION

Assume:

$$a(t) = a > 0.$$

Then:

$$x_t^d = I(t, x_0, x_1) + \sqrt{2a}\, B_t.$$

Using:

$$\mathrm{d}B_t = -\frac{B_t}{1-t}\,\mathrm{d}t + \mathrm{d}W_t,$$

we get:

$$\mathrm{d}x_t^d = \left(\partial_t I - \frac{\sqrt{2a}\, B_t}{1-t}\right)\mathrm{d}t + \sqrt{2a}\,\mathrm{d}W_t.$$

Apply Itô to:

$$e^{ik\cdot x_t^d}.$$

ITÔ FORMULA FOR THE CHARACTERISTIC FUNCTION

Itô gives:

$$\mathrm{d}e^{ik\cdot x_t^d} = ik\cdot\left(\partial_t I - \frac{\sqrt{2a}\,B_t}{1-t}\right)e^{ik\cdot x_t^d}\,\mathrm{d}t - a|k|^2 e^{ik\cdot x_t^d}\,\mathrm{d}t + \sqrt{2a}\,ik\cdot e^{ik\cdot x_t^d}\,\mathrm{d}W_t.$$

Taking expectation:

$$\partial_t\mathbb{E}e^{ik\cdot x_t^d} = ik\cdot\mathbb{E}\left[\left(\partial_t I - \frac{\sqrt{2a}\,B_t}{1-t}\right)e^{ik\cdot x_t^d}\right] - a|k|^2\mathbb{E}e^{ik\cdot x_t^d}.$$

REPLACE B_t BY A GAUSSIAN

Since:

$$B_t \overset{d}{=} \sqrt{t(1-t)}z,$$

and:

$$x_t^d \overset{d}{=} x_t,$$

the derivative becomes:

$$\partial_t \mathbb{E} e^{ik\cdot x_t} = ik \cdot \mathbb{E}\left[\left(\partial_t I - \sqrt{\frac{2at}{1-t}}\, z\right) e^{ik\cdot x_t}\right] - a|k|^2 \mathbb{E} e^{ik\cdot x_t}.$$

This is the Fourier form of a Fokker–Planck equation.

DIFFUSIVE INTERPOLANT FPE

Let:

$$\rho(t) = \mathcal{L}(x_t).$$

Then:

$$\boxed{\partial_t \rho + \nabla \cdot (u\rho) = a\Delta\rho,}$$

where:

$$\boxed{u(t,x) = \mathbb{E}\left[\partial_t I(t, x_0, x_1) - \sqrt{\frac{2at}{1-t}}\, z \mid x_t = x\right].}$$

This differs from the transport velocity:

$$b(t,x) = \mathbb{E}[\partial_t I + \dot\gamma z \mid x_t = x].$$

RELATIONSHIP BETWEEN u, b, s

For:

$$\gamma(t) = \sqrt{2at(1-t)},$$

the score identity gives:

$$s(t,x) = -\frac{1}{\sqrt{2at(1-t)}}\mathbb{E}[z \mid x_t = x].$$

The transport velocity is:

$$b(t,x) = \mathbb{E}\left[\partial_t I + \frac{a(1-2t)}{\sqrt{2at(1-t)}} z \mid x_t = x\right].$$

The diffusive drift is:

$$u(t,x) = \mathbb{E}\left[\partial_t I - \sqrt{\frac{2at}{1-t}} z \mid x_t = x\right].$$

Direct comparison yields:

$$\boxed{u = b + a\, s.}$$

CONSISTENCY OF FPE AND TE

The diffusive FPE is:

$$\partial_t \rho + \nabla \cdot (u\rho) = a\Delta\rho.$$

Using:

$$u = b + as, \qquad s\rho = \nabla\rho,$$

we get:

$$\nabla \cdot (u\rho) = \nabla \cdot (b\rho) + a\nabla \cdot (s\rho) = \nabla \cdot (b\rho) + a\Delta\rho.$$

Therefore:

$$\partial_t \rho + \nabla \cdot (b\rho) + a\Delta\rho = a\Delta\rho,$$

so:

$$\partial_t \rho + \nabla \cdot (b\rho) = 0.$$

CONDITIONAL SAMPLING FROM A POINT MASS

The diffusive drift:

$$u(t,x) = \mathbb{E}\left[\partial_t I - \sqrt{\frac{2at}{1-t}}z \mid x_t = x\right]$$

can remain nonsingular even if ρ_0 is replaced by:

$$\delta_{x_0}.$$

This is impossible for a regular deterministic ODE:

$$\dot{X}_t = b(t, X_t),$$

because uniqueness maps one point to one point.

But an SDE can spread a point mass:

$$\mathrm{d}X_t = u^d(t, X_t, x_0)\,\mathrm{d}t + \sqrt{2a}\,\mathrm{d}W_t.$$

THEOREM 31: POINT-MASS CONDITIONAL SAMPLING

Assume:

$$I(t, x_0, x_1) = x_0 \qquad t \in [0, \delta].$$

Fix $x_0 \in \mathbb{R}^d$. Define:

$$u^d(t, x, x_0) = \mathbb{E}_{x_1, z}\left[\partial_t I(t, x_0, x_1) - \sqrt{\frac{2at}{1-t}} z \mid x_t = x\right],$$

where:

$$x_t = I(t, x_0, x_1) + \sqrt{2at(1-t)} z.$$

Then:

$$u^d(\cdot, \cdot, x_0) \in C^0([0,1]; (C^p(\mathbb{R}^d))^d),$$

and:

$$\boxed{\mathrm{d}X_t^d = u^d(t, X_t^d, x_0)\,\mathrm{d}t + \sqrt{2a}\,\mathrm{d}W_t, \qquad X_0^d = x_0}$$

satisfies:

$$X_1^d \sim \rho_1.$$

PROOF OF THEOREM 31: NEAR $t = 0$

For $t \in [0, \delta]$:

$$I(t, x_0, x_1) = x_0.$$

Thus:

$$x_t = x_0 + \sqrt{2at(1-t)}z.$$

Hence:

$$x_t \sim \mathcal{N}(x_0, 2at(1-t)I_d).$$

The density solves:

$$\partial_t \rho + 2at\nabla \cdot (s\rho) = a\Delta\rho,$$

with:

$$s(t,x) = \nabla \log \rho(t,x) = -\frac{x - x_0}{2at(1-t)}.$$

PROOF: DRIFT REMAINS FINITE

On $t \in (0, \delta]$:

$$2at\, s(t, x) = -\frac{x - x_0}{1 - t}.$$

This has a finite limit at $t = 0$:

$$\lim_{t \downarrow 0} 2at\, s(t, x) = -(x - x_0).$$

At the initial point $x = x_0$:

$$u^d(0, x_0, x_0) = \mathbb{E}_{x_1} \partial_t I(0, x_0, x_1).$$

Under the stronger assumption $I(t, x_0, x_1) = x_0$ near zero:

$$u^d(0, x_0, x_0) = 0.$$

PROOF: CONSISTENCY WITH CONDITIONAL NOISE

Since:

$$x_t = x_0 + \sqrt{2at(1-t)}z,$$

we have:

$$\mathbb{E}[z \mid x_t = x] = \frac{x - x_0}{\sqrt{2at(1-t)}}.$$

Then:

$$-\sqrt{\frac{2at}{1-t}}\mathbb{E}[z \mid x_t = x] = -\frac{x - x_0}{1-t}.$$

This equals:

$$2at\, s(t,x).$$

Therefore the formula defining u^d agrees with the nonsingular FPE drift near $t = 0$.

PROOF: ENDPOINT $t = 1$

At $t = 1$, $x_t \to x_1$. The theorem gives:

$$u^d(1, x, x_0) = \partial_t I(1, x_0, x) + 2a\nabla \log \rho_1(x).$$

This is finite under:

$$\rho_1 > 0, \qquad \rho_1 \in C^2, \qquad \int |\nabla \log \rho_1|^2 \rho_1 < \infty.$$

For $t \in (\delta, 1]$, the usual interpolant proof applies.

Thus the SDE has marginal $\rho(t)$, and:

$$X_1^d \sim \rho_1.$$

OBJECTIVE FOR u^d

The drift u^d is again a conditional expectation, hence a regression minimizer.

Define:

$$L_{u^d}[\hat{u}^d] = \int_0^1 \mathbb{E}_{x_1,z}\left[|\hat{u}^d(t,x_t,x_0)|^2 - 2\left(\partial_t I(t,x_0,x_1) - \sqrt{\frac{2at}{1-t}}z\right)\cdot \hat{u}^d(t,x_t,x_0)\right]\,\mathrm{d}t.$$

Completing the square gives:

$$\hat{u}^d = u^d.$$

WHY POINT-MASS SAMPLING MATTERS

The theorem shows:

$$\delta_{x_0} \quad \rightsquigarrow \quad \rho_1$$

by a finite-time SDE.

This is impossible for an ODE with unique solutions:

$$X_0 = x_0 \quad \Rightarrow \quad X_1 = \Phi_{0,1}(x_0),$$

which is a single point.

Therefore diffusion is not merely numerical noise; it changes what can be generated from singular initial data.

ONE-SIDED INTERPOLANTS: MOTIVATION

A common generative modeling setup takes:

$$\rho_0 = \mathcal{N}(0, I_d), \qquad \rho_1 = \rho_{\text{data}}.$$

Then $x_0 \sim \rho_0$ and the latent Gaussian $z \sim \mathcal{N}(0, I_d)$ can be combined.

This yields a simpler form:

$$x_t^{os} = \alpha(t) z + J(t, x_1).$$

This is the finite-time analogue of diffusion-style data-to-noise constructions.

DEFINITION 32: ONE-SIDED INTERPOLANT

A one-sided stochastic interpolant is:

$$x_t^{os} = \alpha(t)z + J(t, x_1), \qquad t \in [0,1],$$

where:

$$z \sim \mathcal{N}(0, I_d), \qquad x_1 \sim \rho_1, \qquad z \perp x_1.$$

Boundary conditions:

$$J(0, x_1) = 0, \qquad J(1, x_1) = x_1,$$

$$\alpha(0) = 1, \qquad \alpha(1) = 0, \qquad \alpha(t) > 0 \quad t \in [0,1).$$

Therefore:

$$x_0^{os} = z \sim \mathcal{N}(0, I_d), \qquad x_1^{os} = x_1 \sim \rho_1.$$

ONE-SIDED AS A SPECIAL CASE

Start from the general stochastic interpolant:

$$x_t = I(t, x_0, x_1) + \gamma(t)z.$$

Set:

$$I(t, x_0, x_1) = J(t, x_1) + \delta(t)x_0,$$

with:

$$x_0, z \overset{ind}{\sim} \mathcal{N}(0, I_d).$$

Then:

$$\delta(t)x_0 + \gamma(t)z \sim \mathcal{N}(0, (\delta^2(t) + \gamma^2(t))I_d).$$

Choose:

$$\alpha^2(t) = \delta^2(t) + \gamma^2(t).$$

Thus:

$$x_t \overset{d}{=} J(t, x_1) + \alpha(t)z.$$

ONE-SIDED VELOCITY

Differentiate:

$$x_t^{os} = \alpha(t)z + J(t, x_1).$$

Then:

$$\dot{x}_t^{os} = \dot{\alpha}(t)z + \partial_t J(t, x_1).$$

Therefore:

$$\boxed{b(t, x) = \mathbb{E}[\dot{\alpha}(t)z + \partial_t J(t, x_1) \mid x_t^{os} = x].}$$

The velocity objective is:

$$L_b[\hat{b}] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{b}(t, x_t^{os})|^2 - (\dot{\alpha}z + \partial_t J) \cdot \hat{b}(t, x_t^{os})\right] \mathrm{d}t.$$

ONE-SIDED SCORE AND DENOISER

The score is:

$$\boxed{s(t,x) = \nabla \log \rho(t,x) = -\alpha^{-1}(t)\eta_z(t,x),}$$

where:

$$\eta_z(t,x) = \mathbb{E}[z \mid x_t^{os} = x].$$

The denoiser objective is:

$$L_{\eta_z}[\hat{\eta}_z] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_z(t,x_t^{os})|^2 - z \cdot \hat{\eta}_z(t,x_t^{os})\right] \mathrm{d}t.$$

The proof is the same conditional-expectation square completion used in Document II.

ONE-SIDED ASSUMPTIONS

The endpoint assumption reduces to a condition only on ρ_1:

$$\rho_1 > 0, \qquad \rho_1 \in C^2(\mathbb{R}^d), \qquad \int |\nabla \log \rho_1|^2 \rho_1 < \infty.$$

For J:

$$|\partial_t J(t, x_1)| \leq C_1 |x_1|,$$

and:

$$\mathbb{E}|\partial_t J(t, x_1)|^4 \leq M_1, \qquad \mathbb{E}|\partial_t^2 J(t, x_1)|^2 \leq M_2.$$

These replace the two-endpoint assumptions of the general framework.

GENERAL GAUSSIAN BASE

If:

$$\rho_0 = \mathcal{N}(0, C_0),$$

write:

$$C_0 = \sigma_0 \sigma_0^\top.$$

Then:

$$x_t^{os} = \alpha(t)\sigma_0 z + J(t, x_1), \qquad z \sim \mathcal{N}(0, I_d).$$

The score identity becomes:

$$s(t,x) = -\alpha^{-1}(t)\sigma_0^{-\top}\mathbb{E}[z \mid x_t^{os} = x],$$

when σ_0 is invertible.

The standard case $C_0 = I_d$ is recovered by $\sigma_0 = I_d$.

ONE-SIDED SAMPLERS

Once b and s are learned:

$$\dot{X}_t = b(t, X_t), \qquad X_0 \sim \mathcal{N}(0, I_d),$$

or:

$$\mathrm{d}X_t^F = (b + \varepsilon s)(t, X_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t, \qquad X_0^F \sim \mathcal{N}(0, I_d).$$

Both yield:

$$X_1 \sim \rho_1$$

with exact fields.

For learned fields, the SDE version inherits the KL control from Document III.

MIRROR INTERPOLANTS: MOTIVATION

A mirror interpolant bridges a distribution to itself:

$$\rho_0 = \rho_1.$$

This is useful for:

resampling, local data perturbation, data-preserving stochastic transformations.

The interpolant begins and ends at the same distribution, but can move through a noisy intermediate law.

DEFINITION 35: MIRROR INTERPOLANT

Let:

$$x_1 \sim \rho_1, \qquad z \sim \mathcal{N}(0, I_d), \qquad z \perp x_1.$$

A mirror interpolant is:

$$x_t^{mir} = K(t, x_1) + \gamma(t) z.$$

Boundary conditions:

$$K(0, x_1) = x_1, \qquad K(1, x_1) = x_1,$$

$$\gamma(0) = \gamma(1) = 0, \qquad \gamma(t) > 0 \quad 0 < t < 1.$$

Thus:

$$x_0^{mir} = x_1 \sim \rho_1, \qquad x_1^{mir} = x_1 \sim \rho_1.$$

MIRROR VELOCITY

Differentiate:

$$x_t^{mir} = K(t, x_1) + \gamma(t) z.$$

Then:

$$\dot{x}_t^{mir} = \partial_t K(t, x_1) + \dot{\gamma}(t) z.$$

Hence:

$$\boxed{b(t, x) = \mathbb{E}[\partial_t K(t, x_1) + \dot{\gamma}(t) z \mid x_t^{mir} = x].}$$

Objective:

$$L_b[\hat{b}] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{b}(t, x_t^{mir})|^2 - (\partial_t K + \dot{\gamma} z) \cdot \hat{b}(t, x_t^{mir})\right] \mathrm{d}t.$$

MIRROR SCORE AND DENOISER

The score is:

$$s(t,x) = -\gamma^{-1}(t)\eta_z(t,x),$$

where:

$$\eta_z(t,x) = \mathbb{E}[z \mid x_t^{mir} = x].$$

Denoiser objective:

$$L_{\eta_z}[\hat{\eta}_z] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_z(t,x_t^{mir})|^2 - z \cdot \hat{\eta}_z(t,x_t^{mir})\right] \mathrm{d}t.$$

This is identical in structure to the general stochastic-interpolant denoiser objective.

MIRROR ASSUMPTIONS

The density assumption is only on ρ_1:

$$\rho_1 > 0, \qquad \rho_1 \in C^2(\mathbb{R}^d), \qquad \int |\nabla \log \rho_1|^2 \rho_1 < \infty.$$

For K:

$$|\partial_t K(t, x_1)| \leq C_1 |x_1|,$$

and:

$$\mathbb{E}|\partial_t K(t, x_1)|^4 \leq M_1, \qquad \mathbb{E}|\partial_t^2 K(t, x_1)|^2 \leq M_2.$$

SPECIAL MIRROR CASE $K(t, x_1) = x_1$

If:

$$K(t, x_1) = x_1,$$

then:

$$\partial_t K = 0.$$

Therefore:

$$b(t,x) = \dot{\gamma}(t)\mathbb{E}[z \mid x_t^{mir} = x] = \dot{\gamma}(t)\eta_z(t,x).$$

Since:

$$s(t,x) = -\gamma^{-1}(t)\eta_z(t,x),$$

one learned denoiser determines:

$$b, \qquad s, \qquad b^F = b + \varepsilon s, \qquad b^B = b - \varepsilon s.$$

MIRROR INTERPOLANTS AS RESAMPLERS

With exact ODE dynamics:

$$\dot{X}_t = b(t, X_t),$$

and:

$$X_0 \sim \rho_1,$$

we get:

$$X_1 \sim \rho_1.$$

If the ODE map is the identity-like mirror map, one may recover the same sample structure.

With SDE sampling:

$$\mathrm{d}X_t = (b + \varepsilon s)(t, X_t)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t,$$

the sample can move stochastically while preserving the terminal law.

MIRROR AS TWO GLUED ONE-SIDED INTERPOLANTS

If:

$$K(t, x_1) = 0 \qquad t \in [t_1, t_2],$$

then:

$$x_t^{mir} = \gamma(t) z \qquad t \in [t_1, t_2].$$

Thus the process:

$$\rho_1 \longrightarrow \mathcal{N}(0, \gamma^2(t) I_d) \longrightarrow \rho_1$$

looks like two one-sided interpolants glued together.

This creates a reference density in the middle of the bridge.

SCHRÖDINGER BRIDGE: MOTIVATION

The previous constructions fixed the interpolant. Now we optimize over the interpolant itself.

The goal is to solve a stochastic optimal transport problem:

$$\rho_0 \quad \rightsquigarrow \quad \rho_1$$

while minimizing kinetic energy under a diffusive constraint.

This is the Schrödinger bridge problem, an entropy-regularized version of optimal transport.

HYDRODYNAMIC SCHRÖDINGER BRIDGE

For fixed $\varepsilon > 0$, solve:

$$\min_{\hat{u},\hat{\rho}} \int_0^1 \int_{\mathbb{R}^d} |\hat{u}(t,x)|^2 \hat{\rho}(t,x)\,\mathrm{d}x\,\mathrm{d}t$$

subject to:

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{u}\hat{\rho}) = \varepsilon \Delta \hat{\rho},$$

$$\hat{\rho}(0) = \rho_0, \qquad \hat{\rho}(1) = \rho_1\,.$$

The unknowns are:

$$\hat{\rho}(t,x), \qquad \hat{u}(t,x).$$

EULER–LAGRANGE SYSTEM

The optimizer is:

$$u = \nabla \lambda,$$

and (ρ, λ) solve:

$$\boxed{\partial_t \rho + \nabla \cdot (\nabla \lambda \, \rho) = \varepsilon \Delta \rho, \qquad \rho(0) = \rho_0, \quad \rho(1) = \rho_1,}$$

$$\boxed{\partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 = -\varepsilon \Delta \lambda.}$$

The first equation transports density. The second equation is a viscous Hamilton–Jacobi equation.

ASSUMPTION 39: GAUSSIANIZING MAP

Assume there exists a reversible map:

$$T : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d,$$

such that:

$$z \sim \mathcal{N}(0, I_d) \quad \Rightarrow \quad T(t, z) \sim \rho(t),$$

and:

$$x_t \sim \rho(t) \quad \Rightarrow \quad T^{-1}(t, x_t) \sim \mathcal{N}(0, I_d).$$

The exact form of T is not needed; only its existence is used.

LEMMA 40: GAUSSIANIZED INTERPOLANT

Let:

$$\xi_0 = T^{-1}(0, x_0), \qquad \xi_1 = T^{-1}(1, x_1),$$

where:

$$x_0 \sim \rho_0, \qquad x_1 \sim \rho_1.$$

Then:

$$\xi_0, \xi_1, z \sim \mathcal{N}(0, I_d)$$

independently.

If:

$$\alpha^2(t) + \beta^2(t) + \gamma^2(t) = 1,$$

then:

$$\boxed{x_t = T(t, \alpha(t)\xi_0 + \beta(t)\xi_1 + \gamma(t)z)}$$

has density $\rho(t)$.

PROOF OF LEMMA 40

Because:

$$\xi_0, \xi_1, z \overset{ind}{\sim} \mathcal{N}(0, I_d),$$

the linear combination:

$$Y_t = \alpha(t)\xi_0 + \beta(t)\xi_1 + \gamma(t)z$$

is Gaussian with:

$$\mathbb{E}Y_t = 0,$$

$$\mathrm{Cov}(Y_t) = (\alpha^2 + \beta^2 + \gamma^2)I_d = I_d.$$

Therefore:

$$Y_t \sim \mathcal{N}(0, I_d).$$

By the definition of T:

$$T(t, Y_t) \sim \rho(t).$$

ENDPOINT CHECK FOR LEMMA 40

At $t = 0$, choose:

$$\alpha(0) = 1, \qquad \beta(0) = \gamma(0) = 0.$$

Then:

$$x_0 = T(0, \xi_0) = T(0, T^{-1}(0, x_0)) = x_0.$$

At $t = 1$, choose:

$$\beta(1) = 1, \qquad \alpha(1) = \gamma(1) = 0.$$

Then:

$$x_1 = T(1, \xi_1) = T(1, T^{-1}(1, x_1)) = x_1.$$

Thus the construction realizes the desired bridge density.

MAX-MIN PRINCIPLE: IDEA

The Schrödinger bridge asks for the best diffusive current:

$$\partial_t \rho + \nabla \cdot (u\rho) = \varepsilon \Delta \rho.$$

A stochastic interpolant with:

$$\hat{x}_t = \hat{I}(t, x_0, x_1) + \gamma(t) z$$

has forward drift:

$$b^F = \mathbb{E}[\partial_t \hat{I} + (\dot{\gamma} - \varepsilon \gamma^{-1}) z \mid \hat{x}_t = x].$$

Thus minimizing over u forces:

$$u = b^F,$$

while maximizing over $\hat{I}$ searches over density paths.

THEOREM 41: MAX-MIN PROBLEM

Fix γ with:

$$\gamma(0) = \gamma(1) = 0, \qquad \gamma(t) > 0 \quad 0 < t < 1.$$

Consider:

$$\boxed{\max_{\hat{I}} \min_{\hat{u}} \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{u}(t,\hat{x}_t)|^2 - \left(\partial_t \hat{I} + (\dot{\gamma} - \varepsilon\gamma^{-1})z\right) \cdot \hat{u}(t,\hat{x}_t)\right] \mathrm{d}t.}$$

If Assumption 39 holds, optimizers (I, u) have density ρ solving the Schrödinger bridge, and:

$$u = \nabla\lambda.$$

PROOF OF THEOREM 41: CONDITIONAL CURRENT

For:

$$\hat{x}_t = \hat{I}(t, x_0, x_1) + \gamma(t) z,$$

define:

$$\hat{\rho}(t) = \mathcal{L}(\hat{x}_t).$$

The forward drift is:

$$\hat{u}_F(t, x) = \mathbb{E}[\partial_t \hat{I} + (\dot{\gamma} - \varepsilon \gamma^{-1}) z \mid \hat{x}_t = x].$$

Define the current:

$$\hat{j}(t, x) = \hat{u}_F(t, x) \hat{\rho}(t, x).$$

Then:

$$\partial_t \hat{\rho} + \nabla \cdot \hat{j} = \varepsilon \Delta \hat{\rho}.$$

PROOF: OBJECTIVE IN (ρ, j, u)

Using conditional expectation, the max-min objective becomes:

$$\max_{\hat{\rho},\hat{j}} \min_{\hat{u}} \int_0^1 \int \left[\frac{1}{2} |\hat{u}|^2 \hat{\rho} - \hat{u} \cdot \hat{j} \right] \mathrm{d}x \, \mathrm{d}t$$

subject to:

$$\partial_t \hat{\rho} + \nabla \cdot \hat{j} = \varepsilon \Delta \hat{\rho},$$

$$\hat{\rho}(0) = \rho_0, \qquad \hat{\rho}(1) = \rho_1 .$$

This is now a constrained variational problem over density and current.

PROOF: INTRODUCE LAGRANGE MULTIPLIER

Let $\lambda(t, x)$ enforce:

$$\partial_t \rho + \nabla \cdot j - \varepsilon \Delta \rho = 0.$$

The augmented functional is:

$$\mathcal{A} = \int_0^1 \int \left[\frac{1}{2} |u|^2 \rho - u \cdot j - \lambda(\partial_t \rho + \nabla \cdot j - \varepsilon \Delta \rho) \right] \mathrm{d}x \, \mathrm{d}t.$$

Endpoint constraints are enforced separately by endpoint multipliers.

PROOF: VARIATION IN j

The j-dependent terms are:

$$\int [-u \cdot j - \lambda \nabla \cdot j] \, \mathrm{d}x.$$

Integrate by parts:

$$-\int \lambda \nabla \cdot j \, \mathrm{d}x = \int \nabla \lambda \cdot j \, \mathrm{d}x.$$

Thus variation in j gives:

$$-u + \nabla \lambda = 0.$$

Therefore:

$$\boxed{u = \nabla \lambda.}$$

PROOF: VARIATION IN u

The u-dependent part is:

$$\int \left[\frac{1}{2}|u|^2\rho - u \cdot j\right] \mathrm{d}x.$$

Variation gives:

$$u\rho - j = 0.$$

Thus:

$$\boxed{j = u\rho.}$$

Combining with the previous slide:

$$j = \rho\nabla\lambda.$$

PROOF: VARIATION IN ρ

Using:

$$j = u\rho, \qquad u = \nabla\lambda,$$

the first equation becomes:

$$\partial_t \rho + \nabla \cdot (\rho\nabla\lambda) = \varepsilon\Delta\rho.$$

Variation in ρ gives:

$$\partial_t \lambda + \frac{1}{2}|u|^2 = -\varepsilon\Delta\lambda.$$

Since $u = \nabla\lambda$, this becomes:

$$\partial_t \lambda + \frac{1}{2}|\nabla\lambda|^2 = -\varepsilon\Delta\lambda.$$

PROOF: CONCLUDE THEOREM 41

The Euler–Lagrange equations are:

$$\partial_t \rho + \nabla \cdot (\nabla \lambda\, \rho) = \varepsilon \Delta \rho,$$
$$\partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 = -\varepsilon \Delta \lambda,$$

with:

$$\rho(0) = \rho_0, \qquad \rho(1) = \rho_1.$$

These are exactly the Schrödinger bridge equations.

Assumption 39 ensures that an interpolant exists whose density equals this optimizer.

Hence optimizers of the max-min problem recover the Schrödinger bridge.

OPTIMAL TRANSPORT LIMIT

Let:

$$\varepsilon \to 0.$$

Then the Schrödinger bridge constraint:

$$\partial_t \rho + \nabla \cdot (u\rho) = \varepsilon \Delta \rho$$

formally becomes:

$$\partial_t \rho + \nabla \cdot (u\rho) = 0.$$

The objective:

$$\int_0^1 \int |u|^2 \rho$$

becomes the Benamou–Brenier dynamic optimal transport cost.

Thus:

$$\text{Schrödinger bridge} \longrightarrow \text{optimal transport} \qquad \varepsilon \downarrow 0.$$

COMPARISON OF FOUR INSTANTIATIONS

construction	interpolant	main role
diffusive	$I + \sqrt{2a}B_t$	bridge-process interpretation
one-sided	$\alpha z + J(t, x_1)$	Gaussian base generation
mirror	$K(t, x_1) + \gamma z$	self-transport and resampling
Schrödinger	optimized I	entropy-regularized transport

All remain instances of:

probability path first, sampler second.

PROOF EXERCISE 1

Show that:

$$x_t^d = I(t, x_0, x_1) + \sqrt{2a} B_t$$

has the same one-time marginal as:

$$x_t = I(t, x_0, x_1) + \sqrt{2at(1-t)} z.$$

Then prove that:

$$\partial_t \rho + \nabla \cdot (u\rho) = a\Delta\rho$$

with:

$$u = \mathbb{E}\left[\partial_t I - \sqrt{\frac{2at}{1-t}} z \mid x_t = x\right].$$

PROOF EXERCISE 2

For a one-sided interpolant:

$$x_t^{os} = \alpha(t)z + J(t, x_1),$$

prove:

$$s(t, x) = -\alpha^{-1}(t)\mathbb{E}[z \mid x_t^{os} = x].$$

Hint:

$$\mathbb{E}[ze^{ik\cdot x_t^{os}}] = i\alpha(t)k\, g(t, k).$$

Then use Fourier inversion to identify:

$$\mathbb{E}[z \mid x_t = x]\rho(t, x) = -\alpha(t)\nabla\rho(t, x).$$

PROOF EXERCISE 3

For the mirror interpolant:

$$x_t^{mir} = x_1 + \gamma(t)z,$$

prove:

$$b(t,x) = \dot{\gamma}(t)\eta_z(t,x),$$

and:

$$s(t,x) = -\gamma^{-1}(t)\eta_z(t,x).$$

Then write the forward SDE drift:

$$b^F(t,x) = \left(\dot{\gamma}(t) - \varepsilon(t)\gamma^{-1}(t)\right)\eta_z(t,x).$$

PROOF EXERCISE 4

Derive the Schrödinger bridge Euler–Lagrange equations.

Start from:

$$\min_{\rho,u} \int_0^1 \int |u|^2 \rho$$

subject to:

$$\partial_t \rho + \nabla \cdot (u\rho) = \varepsilon \Delta \rho, \qquad \rho(0) = \rho_0, \quad \rho(1) = \rho_1.$$

Use a Lagrange multiplier λ and prove:

$$u = \nabla \lambda,$$

$$\partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 = -\varepsilon \Delta \lambda.$$

DOCUMENT IV SUMMARY

We extended the framework in four directions:

Brownian bridge paths: $x_t^d = I + \sqrt{2a}B_t,$
Gaussian-base one-sided paths: $x_t^{os} = \alpha z + J(t, x_1),$
mirror paths: $x_t^{mir} = K(t, x_1) + \gamma z,$
Schrödinger bridges by optimizing I.

The key lesson:

different bridges change the training targets and geometry, but not the core proof pattern.

Next document:

Spatially linear interpolants and their factorized structure.