STOCHASTIC INTERPOLANTS II

Definitions, Assumptions, Transport Equations, Quadratic Objectives, and Proofs

Ziseok Lee

April 30, 2026

DOCUMENT II: MAIN THEOREM CHAIN

This document proves the core structural results:

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

implies:

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

is smooth and solves:

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

Then:

$$b(t, x) = \mathbb{E}[\dot{x}_t \mid x_t = x]$$

and

$$s(t, x) = \nabla \log \rho(t, x) = -\gamma^{-1}(t) \mathbb{E}[z \mid x_t = x].$$

Both are unique minimizers of explicit quadratic objectives.

DEFINITION: STOCHASTIC INTERPOLANT

Given two densities ρ_0, ρ_1 on $\mathbb{R}^d$, a stochastic interpolant is:

$$x_t = I(t, x_0, x_1) + \gamma(t)z, \qquad t \in [0,1],$$

where:

$$(x_0, x_1) \sim \nu, \qquad z \sim \mathcal{N}(0, I_d), \qquad z \perp (x_0, x_1).$$

The coupling ν satisfies:

$$\nu(\,\mathrm{d}x_0, \mathbb{R}^d) = \rho_0(x_0)\,\mathrm{d}x_0, \qquad \nu(\mathbb{R}^d,\,\mathrm{d}x_1) = \rho_1(x_1)\,\mathrm{d}x_1.$$

ENDPOINT CONDITIONS

The interpolation map satisfies:

$$I(0, x_0, x_1) = x_0, \qquad I(1, x_0, x_1) = x_1.$$

The latent amplitude satisfies:

$$\gamma(0) = \gamma(1) = 0.$$

Therefore:

$$x_{t=0} = x_0, \qquad x_{t=1} = x_1.$$

Hence:

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

REGULARITY ASSUMPTIONS ON I

The paper assumes:

$$I \in C^2([0,1], C^2(\mathbb{R}^d \times \mathbb{R}^d)^d).$$

Growth condition:

$$\exists C_1 < \infty : \quad |\partial_t I(t, x_0, x_1)| \leq C_1 |x_0 - x_1|.$$

This prevents the path from moving too fast relative to its endpoints.

Later moment assumptions ensure:

$$\mathbb{E}|\partial_t I|^4 < \infty, \qquad \mathbb{E}|\partial_t^2 I|^2 < \infty.$$

REGULARITY ASSUMPTIONS ON γ

The latent amplitude satisfies:

$$\gamma : [0,1] \to \mathbb{R},$$

$$\gamma(0) = \gamma(1) = 0, \qquad \gamma(t) > 0 \quad \forall t \in (0,1),$$

and:

$$\gamma^2 \in C^2([0,1]).$$

Important point:

γ itself need not be C^1 at the endpoints.

Example:

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

has $\gamma^2 \in C^2$, while $\dot{\gamma}$ is singular at endpoints.

ASSUMPTION ON ENDPOINT DENSITIES

The endpoint densities satisfy:

$$\rho_i > 0, \qquad \rho_i \in C^2(\mathbb{R}^d), \qquad i = 0, 1.$$

Endpoint Fisher information is finite:

$$\int_{\mathbb{R}^d} |\nabla \log \rho_0(x)|^2 \rho_0(x) \, \mathrm{d}x < \infty,$$
$$\int_{\mathbb{R}^d} |\nabla \log \rho_1(x)|^2 \rho_1(x) \, \mathrm{d}x < \infty.$$

These assumptions are needed for endpoint score control.

MOMENT ASSUMPTIONS

The coupling ν and interpolation map I satisfy:

$$\exists M_1, M_2 < \infty :$$

$$\mathbb{E}|\partial_t I(t, x_0, x_1)|^4 \leq M_1,$$

$$\mathbb{E}|\partial_t^2 I(t, x_0, x_1)|^2 \leq M_2,$$

for all $t \in [0, 1]$.

For the linear interpolant:

$$I(t, x_0, x_1) = (1 - t)x_0 + tx_1,$$

$$\partial_t I = x_1 - x_0.$$

Thus finite fourth moments of ρ_0, ρ_1 suffice.

CONDITIONAL EXPECTATION CONVENTION

For a function $f(t, x_0, x_1, z)$, define:

$$\mathbb{E}[f(t, x_0, x_1, z) \mid x_t = x]$$

by:

$$\int_{\mathbb{R}^d} \phi(x)\mathbb{E}[f \mid x_t = x]\rho(t, x)\,\mathrm{d}x = \mathbb{E}[\phi(x_t)f]$$

for all $\phi \in C_0^\infty(\mathbb{R}^d)$.

This is the rigorous meaning of conditioning on $x_t = x$.

THE DENSITY $\rho(t, x)$

The density $\rho(t, x)$ is defined through:

$$\int_{\mathbb{R}^d} \phi(x)\rho(t, x)\,\mathrm{d}x = \mathbb{E}[\phi(x_t)].$$

The proof will show:

$$\rho(t, \cdot) \text{ exists for all } t,$$

$$\rho(t, x) > 0,$$

and

$$\rho \in C^1([0, 1]; C^p(\mathbb{R}^d)) \qquad \forall p \in \mathbb{N}.$$

The latent Gaussian factor is the reason for interior smoothness.

THEOREM 6: STATEMENT

Let $x_t = I(t, x_0, x_1) + \gamma(t)z$. Then its law is absolutely continuous with density $\rho(t, x)$.

Moreover:

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

The density solves:

$$\boxed{\partial_t \rho + \nabla \cdot (b\rho) = 0}$$

where:

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

Also:

$$\int |b(t, x)|^2 \rho(t, x) \, \mathrm{d}x < \infty.$$

PROOF OF THEOREM 6: CHARACTERISTIC FUNCTION

Define:

$$g(t,k) = \mathbb{E}e^{ik\cdot x_t}.$$

Since:

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

with $z \perp (x_0, x_1)$,

$$g(t,k) = \mathbb{E}e^{ik\cdot I(t,x_0,x_1)}\mathbb{E}e^{i\gamma(t)k\cdot z}.$$

Because $z \sim \mathcal{N}(0, I_d)$:

$$\mathbb{E}e^{i\gamma k\cdot z} = e^{-\frac{1}{2}\gamma^2|k|^2}.$$

Thus:

$$\boxed{g(t,k) = g_I(t,k)e^{-\frac{1}{2}\gamma^2(t)|k|^2}.}$$

GAUSSIAN DECAY

Because:

$$|g_I(t,k)| \leq 1,$$

we have:

$$|g(t,k)| \leq e^{-\frac{1}{2}\gamma^2(t)|k|^2}.$$

For $0 < t < 1, \gamma(t) > 0$, hence:

$$\int_{\mathbb{R}^d} |k|^p |g(t,k)| \, \mathrm{d}k < \infty \qquad \forall p \in \mathbb{N}.$$

Therefore Fourier inversion gives:

$$\rho(t,x) = (2\pi)^{-d} \int e^{-ik\cdot x} g(t,k) \, \mathrm{d}k,$$

and:

$$\rho(t,\cdot) \in C^\infty(\mathbb{R}^d).$$

POSITIVITY OF ρ

Let $\mu_I(t) = \mathcal{L}(I(t, x_0, x_1))$. Then:

$$\rho(t,x) = \int_{\mathbb{R}^d} \frac{1}{(2\pi\gamma^2(t))^{d/2}} \exp\left(-\frac{|x-y|^2}{2\gamma^2(t)}\right) \mu_I(t, \mathrm{d}y).$$

The Gaussian kernel is strictly positive:

$$\varphi_{\gamma(t)}(x-y) > 0.$$

Therefore:

$$\rho(t,x) > 0 \qquad \forall x, \quad 0 < t < 1.$$

At endpoints, strict positivity follows from the assumptions on ρ_0, ρ_1.

DIFFERENTIATING THE CHARACTERISTIC FUNCTION

Differentiate:

$$g(t,k) = \mathbb{E}e^{ik\cdot x_t}.$$

Since:

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

we obtain:

$$\partial_t g(t,k) = \mathbb{E}\left[ik \cdot \dot{x}_t \, e^{ik\cdot x_t}\right].$$

Define:

$$m(t,k) = \mathbb{E}[\dot{x}_t e^{ik\cdot x_t}].$$

Then:

$$\boxed{\partial_t g(t,k) = ik \cdot m(t,k).}$$

IDENTIFYING $m(t, k)$

Using conditional expectation:

$$
\begin{aligned}
m(t, k) &= \mathbb{E}[\dot{x}_t e^{ik \cdot x_t}] \\
&= \mathbb{E}\left[\mathbb{E}[\dot{x}_t \mid x_t] e^{ik \cdot x_t}\right].
\end{aligned}
$$

Since:

$$
b(t, x) = \mathbb{E}[\dot{x}_t \mid x_t = x],
$$

we get:

$$
m(t, k) = \int_{\mathbb{R}^d} e^{ik \cdot x} b(t, x) \rho(t, x) \, \mathrm{d}x.
$$

Thus $m(t, k)$ is the Fourier transform of $b\rho$.

TRANSPORT EQUATION IN FOURIER VARIABLES

We have:

$$\partial_t g(t,k) = ik \cdot m(t,k).$$

But:

$$g(t,k) = \int e^{ik\cdot x} \rho(t,x)\,\mathrm{d}x,$$

so:

$$\partial_t g(t,k) = \int e^{ik\cdot x} \partial_t \rho(t,x)\,\mathrm{d}x.$$

Also:

$$ik \cdot m(t,k) = \int e^{ik\cdot x} ik \cdot b(t,x)\rho(t,x)\,\mathrm{d}x.$$

Since:

$$\mathcal{F}[\nabla \cdot (b\rho)](k) = -ik \cdot m(t,k)$$

under the $e^{ik\cdot x}$ convention, we get:

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

WEAK PROOF OF TRANSPORT EQUATION

Equivalently, for $\phi \in C_0^\infty$:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \phi\rho = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\phi(x_t)] = \mathbb{E}[\nabla\phi(x_t) \cdot \dot{x}_t].$$

Use conditional expectation:

$$\mathbb{E}[\nabla\phi(x_t) \cdot \dot{x}_t] = \mathbb{E}[\nabla\phi(x_t) \cdot b(t, x_t)].$$

Therefore:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \phi\rho = \int \nabla\phi \cdot b\rho.$$

This is:

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

REGULARITY OF b: GAUSSIAN INTEGRATION BY PARTS

Recall:

$$m(t,k) = \mathbb{E}[(\partial_t I + \dot{\gamma} z) e^{ik\cdot(I+\gamma z)}].$$

Separate:

$$m(t,k) = \mathbb{E}[\partial_t I \, e^{ik\cdot I}] e^{-\frac{1}{2}\gamma^2 |k|^2} + \dot{\gamma}\mathbb{E}[e^{ik\cdot I}]\mathbb{E}[z e^{i\gamma k\cdot z}].$$

Gaussian identity:

$$\mathbb{E}[z e^{i\gamma k\cdot z}] = i\gamma k \, e^{-\frac{1}{2}\gamma^2 |k|^2}.$$

Thus:

$$m(t,k) = \mathbb{E}[(\partial_t I + i\gamma\dot{\gamma} k) e^{ik\cdot I}] e^{-\frac{1}{2}\gamma^2 |k|^2}.$$

REGULARITY OF $b\rho$

The Gaussian factor gives:

$$\int |k|^p |m(t,k)| \, \mathrm{d}k < \infty.$$

By Fourier inversion:

$$b(t,\cdot)\rho(t,\cdot) \in C^p(\mathbb{R}^d).$$

Since:

$$\rho(t,x) > 0,$$

we can divide:

$$b(t,x) = \frac{b(t,x)\rho(t,x)}{\rho(t,x)}.$$

Thus:

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

under the stated moment assumptions.

$L^2(\rho)$ BOUND FOR b

By Jensen:

$$|b(t,x)|^2 = |\mathbb{E}[\dot{x}_t \mid x_t = x]|^2 \le \mathbb{E}[|\dot{x}_t|^2 \mid x_t = x].$$

Integrate against $\rho(t,x)\,\mathrm{d}x$:

$$\int |b(t,x)|^2 \rho(t,x)\,\mathrm{d}x \le \mathbb{E}|\dot{x}_t|^2.$$

Since:

$$\dot{x}_t = \partial_t I + \dot{\gamma} z,$$

the moment assumptions and Gaussian moments imply:

$$\mathbb{E}|\dot{x}_t|^2 < \infty.$$

ENDPOINT VELOCITY FORMULA

At $t = 0, x_t \to x_0$. The endpoint velocity can contain score terms:

$$b(0, x) = \mathbb{E}_1[\partial_t I(0, x, x_1)] - \lim_{t \to 0} \dot{\gamma}(t)\gamma(t)\nabla \log \rho_0(x).$$

At $t = 1$:

$$b(1, x) = \mathbb{E}_0[\partial_t I(1, x_0, x)] - \lim_{t \to 1} \dot{\gamma}(t)\gamma(t)\nabla \log \rho_1(x).$$

Here $\mathbb{E}_1$ and $\mathbb{E}_0$ denote conditional expectations over the other endpoint under the coupling.

THEOREM 6: PROOF SUMMARY

The proof consists of:

1. $g(t,k) = g_I(t,k)e^{-\gamma^2|k|^2/2}$;
2. Gaussian decay gives smooth density by Fourier inversion;
3. Gaussian convolution gives positivity;
4. $\partial_t g = ik \cdot m$;
5. $m = \mathcal{F}[b\rho]$;
6. Fourier inversion gives $\partial_t \rho + \nabla \cdot (b\rho) = 0$;
7. Jensen gives $b \in L^2(\rho)$.

THEOREM 7: VELOCITY OBJECTIVE

Define:

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

The theorem states:

$$\boxed{b = \arg\min_{\hat{b}} L_b[\hat{b}].}$$

The minimizer is unique because:

$$\rho(t, x) > 0$$

for all t, x, and the objective is strictly quadratic pointwise in $\hat{b}$.

PROOF OF THEOREM 7: CONDITIONING

Let:

$$\dot{x}_t = \partial_t I + \dot{\gamma} z.$$

Then:

$$\mathbb{E}[\dot{x}_t \cdot \hat{b}(t, x_t)] = \mathbb{E}[\mathbb{E}[\dot{x}_t \mid x_t] \cdot \hat{b}(t, x_t)].$$

Since:

$$\mathbb{E}[\dot{x}_t \mid x_t = x] = b(t, x),$$

we get:

$$L_b[\hat{b}] = \int_0^1 \int \left[\frac{1}{2} |\hat{b}(t, x)|^2 - b(t, x) \cdot \hat{b}(t, x) \right] \rho(t, x) \, \mathrm{d}x \, \mathrm{d}t.$$

PROOF OF THEOREM 7: COMPLETING THE SQUARE

Compute:

$$\frac{1}{2}|\hat{b}|^2 - b \cdot \hat{b} = \frac{1}{2}|\hat{b} - b|^2 - \frac{1}{2}|b|^2.$$

Therefore:

$$L_b[\hat{b}] = \frac{1}{2}\int_0^1 \int |\hat{b} - b|^2 \rho \, \mathrm{d}x \, \mathrm{d}t - \frac{1}{2}\int_0^1 \int |b|^2 \rho \, \mathrm{d}x \, \mathrm{d}t.$$

The second term is independent of $\hat{b}$. Thus:

$$\hat{b} = b$$

is the unique minimizer.

EXCESS VELOCITY LOSS

The proof also gives the exact excess-risk identity:

$$L_b[\hat{b}] - L_b[b] = \frac{1}{2}\int_0^1 \int |\hat{b}(t,x) - b(t,x)|^2 \rho(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

This identity is later used in the likelihood bound:

$$\Delta L_b = \frac{1}{2}\|\hat{b} - b\|^2_{L^2(\rho\ \mathrm{d}t)}.$$

VELOCITY v WITHOUT LATENT DERIVATIVE

Define:

$$v(t,x) = \mathbb{E}[\partial_t I(t,x_0,x_1) \mid x_t = x].$$

Since:

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

and:

$$\mathbb{E}[z \mid x_t = x] = -\gamma s(t,x),$$

we obtain:

$$\boxed{b(t,x) = v(t,x) - \gamma(t)\dot{\gamma}(t)s(t,x).}$$

The corresponding objective:

$$L_v[\hat{v}] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{v}(t,x_t)|^2 - \partial_t I \cdot \hat{v}(t,x_t)\right] \mathrm{d}t.$$

THEOREM 8: SCORE IDENTITY

Theorem 8 states:

$$\boxed{s(t,x) = \nabla \log \rho(t,x) = -\gamma^{-1}(t)\mathbb{E}[z \mid x_t = x],}$$

for $0 < t < 1$.

Equivalently:

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

This is a denoising identity: the conditional mean of the injected noise gives the score.

PROOF OF THEOREM 8: FOURIER TRANSFORM OF CONDITIONAL NOISE

Compute:

$$\mathbb{E}[ze^{ik\cdot x_t}] = \mathbb{E}[e^{ik\cdot I}]\mathbb{E}[ze^{i\gamma k\cdot z}].$$

For $z \sim \mathcal{N}(0, I_d)$:

$$\mathbb{E}[ze^{i\gamma k\cdot z}] = i\gamma k e^{-\frac{1}{2}\gamma^2|k|^2}.$$

Thus:

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

PROOF OF THEOREM 8: COMPARE WITH $\nabla\rho$

By definition of conditional expectation:

$$\mathbb{E}[ze^{ik\cdot x_t}] = \int e^{ik\cdot x}\mathbb{E}[z \mid x_t = x]\rho(t,x)\,\mathrm{d}x.$$

Also:

$$\mathcal{F}[\nabla\rho](k) = \int e^{ik\cdot x}\nabla\rho(x)\,\mathrm{d}x = -ik\,g(k).$$

Since:

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

we identify:

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

PROOF OF SCORE IDENTITY: DIVIDE BY ρ

Because:

$$\rho(t,x) > 0,$$

we may divide:

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

Therefore:

$$\frac{\nabla\rho(t,x)}{\rho(t,x)} = -\gamma^{-1}(t)\mathbb{E}[z \mid x_t = x].$$

That is:

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

where:

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

$L^2(\rho)$ BOUND FOR THE SCORE

By Jensen:

$$|s(t,x)|^2 = \gamma^{-2}(t)|\mathbb{E}[z \mid x_t = x]|^2 \leq \gamma^{-2}(t)\mathbb{E}[|z|^2 \mid x_t = x].$$

Integrating:

$$\int |s(t,x)|^2 \rho(t,x)\,\mathrm{d}x \leq \gamma^{-2}(t)\mathbb{E}|z|^2.$$

Since:

$$\mathbb{E}|z|^2 = d,$$

we get:

$$\int |s|^2 \rho \leq d\,\gamma^{-2}(t).$$

ENDPOINT SCORE INTEGRABILITY

Although:

$$d\,\gamma^{-2}(t) \to \infty$$

near $t = 0, 1$, the endpoint assumptions ensure:

$$\int |\nabla \log \rho_0|^2 \rho_0 < \infty,$$

$$\int |\nabla \log \rho_1|^2 \rho_1 < \infty.$$

Thus the limiting endpoint score behavior is controlled by the finite Fisher information assumptions.

THE SCORE OBJECTIVE

Define:

$$L_s[\hat{s}] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{s}(t, x_t)|^2 + \gamma^{-1}(t) z \cdot \hat{s}(t, x_t)\right] \mathrm{d}t.$$

Using:

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

we rewrite:

$$L_s[\hat{s}] = \int_0^1 \int \left[\frac{1}{2}|\hat{s}|^2 - s \cdot \hat{s}\right] \rho \, \mathrm{d}x \, \mathrm{d}t.$$

Therefore:

$$s = \arg\min_{\hat{s}} L_s[\hat{s}].$$

PROOF OF SCORE OBJECTIVE

Complete the square:

$$\frac{1}{2}|\hat{s}|^2 - s \cdot \hat{s} = \frac{1}{2}|\hat{s} - s|^2 - \frac{1}{2}|s|^2.$$

Therefore:

$$L_s[\hat{s}] = \frac{1}{2}\int_0^1 \int |\hat{s} - s|^2 \rho - \frac{1}{2}\int_0^1 \int |s|^2 \rho.$$

Hence:

$$L_s[\hat{s}] - L_s[s] = \frac{1}{2}\int_0^1 \int |\hat{s} - s|^2 \rho.$$

The unique minimizer is:

$$\hat{s} = s.$$

DENOISER

Define:

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

Then:

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

The denoiser objective is:

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

Its minimizer is:

$$\hat{\eta}_z = \eta_z.$$

PROOF OF DENOISER OBJECTIVE

Condition on x_t:

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

Thus:

$$L_{\eta_z}[\hat{\eta}_z] = \frac{1}{2} \int |\hat{\eta}_z - \eta_z|^2 \rho - \frac{1}{2} \int |\eta_z|^2 \rho.$$

Hence the unique minimizer is:

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

WHY LEARN THE DENOISER INSTEAD OF THE SCORE?

The score objective contains:

$$\gamma^{-1}(t)z.$$

This can be singular near endpoints because:

$$\gamma(0) = \gamma(1) = 0.$$

The denoiser objective:

$$L_{\eta_z} = \int \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_z|^2 - z \cdot \hat{\eta}_z\right] \mathrm{d}t$$

has no γ^{-1} factor.

Then recover:

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

ALTERNATIVE SCORE MATCHING OBJECTIVE

Since $s = \nabla \log \rho$, one may also use:

$$\int_0^1 \mathbb{E}\left[|\hat{s}(t, x_t)|^2 + 2\nabla \cdot \hat{s}(t, x_t) \right] \, \mathrm{d}t.$$

Integration by parts gives:

$$\mathbb{E}[\nabla \cdot \hat{s}(t, x_t)] = \int \nabla \cdot \hat{s} \, \rho = -\int \hat{s} \cdot \nabla \rho = -\int \hat{s} \cdot s \, \rho.$$

Hence this objective differs from:

$$\int |\hat{s} - s|^2 \rho$$

by a constant.

ENERGY-BASED INTERPRETATION

Because the score is a gradient:

$$s(t, x) = \nabla \log \rho(t, x),$$

we may parameterize:

$$\hat{s}(t, x) = -\nabla \hat{E}(t, x).$$

The score objective becomes:

$$L_E[\hat{E}] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\nabla \hat{E}(t, x_t)|^2 + \gamma^{-1} z \cdot \nabla \hat{E}(t, x_t)\right] \mathrm{d}t.$$

The energy is defined up to additive constants.

COROLLARY 10: FORWARD FOKKER–PLANCK EQUATION

For any:

$$\varepsilon \in C^0([0,1]), \qquad \varepsilon(t) \geq 0,$$

define:

$$b^F(t,x) = b(t,x) + \varepsilon(t) s(t,x).$$

Then:

$$\boxed{\partial_t \rho + \nabla \cdot (b^F \rho) = \varepsilon(t) \Delta \rho.}$$

Initial condition:

$$\rho(0) = \rho_0.$$

Solution at the terminal time:

$$\rho(1) = \rho_1.$$

PROOF OF FORWARD FPE

Start with the transport equation:

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

Since:

$$s = \nabla \log \rho,$$

we have:

$$s\rho = \nabla \rho.$$

Thus:

$$\Delta \rho = \nabla \cdot \nabla \rho = \nabla \cdot (s\rho).$$

Now:

$$\nabla \cdot (b^F \rho) = \nabla \cdot (b\rho) + \varepsilon \nabla \cdot (s\rho) = \nabla \cdot (b\rho) + \varepsilon \Delta \rho.$$

Therefore:

$$\partial_t \rho + \nabla \cdot (b^F \rho) = \varepsilon \Delta \rho.$$

COROLLARY 10: BACKWARD FOKKER–PLANCK EQUATION

Define:

$$b^B(t,x) = b(t,x) - \varepsilon(t)s(t,x).$$

Then:

$$\boxed{\partial_t \rho + \nabla \cdot (b^B \rho) = -\varepsilon(t)\Delta\rho.}$$

Final condition:

$$\rho(1) = \rho_1.$$

When solved backward in time:

$$\rho(0) = \rho_0.$$

PROOF OF BACKWARD FPE

Again:

$$\Delta \rho = \nabla \cdot (s\rho).$$

Using:

$$b^B = b - \varepsilon s,$$

we obtain:

$$\nabla \cdot (b^B \rho) = \nabla \cdot (b\rho) - \varepsilon \nabla \cdot (s\rho) = \nabla \cdot (b\rho) - \varepsilon \Delta \rho.$$

Since:

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

we get:

$$\partial_t \rho + \nabla \cdot (b^B \rho) = -\varepsilon \Delta \rho.$$

ONE DENSITY, INFINITELY MANY DYNAMICS

The same $\rho(t)$ satisfies:

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

and for every $\varepsilon(t) \geq 0$:

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

Thus:

The diffusion coefficient changes the sampler, not the density path.

This is one of the central conceptual points of stochastic interpolants.

FORWARD AND BACKWARD DRIFTS

The two stochastic drifts are:

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

Their average gives the probability-flow velocity:

$$b = \frac{1}{2}(b^F + b^B).$$

Their difference gives the score:

$$s = \frac{1}{2\varepsilon}(b^F - b^B).$$

Thus learning b and s learns both forward and backward stochastic samplers.

EXACT EXCESS-RISK IDENTITIES

The objectives give exact $L^2(\rho)$ errors:

$$L_b[\hat{b}] - L_b[b] = \frac{1}{2} \int_0^1 \int |\hat{b} - b|^2 \rho \, \mathrm{d}x \, \mathrm{d}t.$$

$$L_s[\hat{s}] - L_s[s] = \frac{1}{2} \int_0^1 \int |\hat{s} - s|^2 \rho \, \mathrm{d}x \, \mathrm{d}t.$$

$$L_{\eta_z}[\hat{\eta}_z] - L_{\eta_z}[\eta_z] = \frac{1}{2} \int_0^1 \int |\hat{\eta}_z - \eta_z|^2 \rho \, \mathrm{d}x \, \mathrm{d}t.$$

These identities are the bridge from regression to likelihood control.

TRAINING SAMPLES

To estimate the objectives, sample:

$$t_i \sim \mathrm{Unif}[0,1], \qquad (x_0^i, x_1^i) \sim \nu, \qquad z_i \sim \mathcal{N}(0, I_d).$$

Construct:

$$x_{t_i}^i = I(t_i, x_0^i, x_1^i) + \gamma(t_i) z_i.$$

Empirical velocity loss:

$$\widehat{L}_b = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{2} |\hat{b}(t_i, x_{t_i}^i)|^2 - (\partial_t I_i + \dot{\gamma}_i z_i) \cdot \hat{b}(t_i, x_{t_i}^i) \right].$$

EMPIRICAL DENOISER LOSS

The empirical denoiser loss is:

$$\widehat{L}_{\eta_z} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{2}|\hat{\eta}_z(t_i, x^i_{t_i})|^2 - z_i \cdot \hat{\eta}_z(t_i, x^i_{t_i})\right].$$

Then:

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

For stochastic sampling:

$$\hat{b}^F = \hat{b} + \varepsilon\hat{s}.$$

CANONICAL EXAMPLE

Let:

$$I(t, x_0, x_1) = (1-t)x_0 + tx_1,$$
$$\gamma(t) = \sqrt{2t(1-t)}.$$

Then:

$$x_t = (1-t)x_0 + tx_1 + \sqrt{2t(1-t)}z.$$

Velocity target:

$$\dot{x}_t = x_1 - x_0 + \frac{1-2t}{\sqrt{2t(1-t)}}z.$$

The learned velocity is:

$$b(t, x) = \mathbb{E}[\dot{x}_t \mid x_t = x].$$

ENDPOINT BEHAVIOR IN THE CANONICAL EXAMPLE

For:

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

we have:

$$\gamma(t)\dot{\gamma}(t) = 1 - 2t.$$

Thus:

$$\lim_{t\to 0} \gamma\dot{\gamma} = 1, \qquad \lim_{t\to 1} \gamma\dot{\gamma} = -1.$$

Endpoint velocities contain score corrections:

$$b(0,x) = \mathbb{E}[x_1 - x \mid x_0 = x] - \nabla \log \rho_0(x),$$

$$b(1,x) = \mathbb{E}[x - x_0 \mid x_1 = x] + \nabla \log \rho_1(x).$$

COMMON PITFALL: x_t IS NOT THE SAMPLER

The interpolant:

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

uses both x_0 and x_1. It is used for training because samples from both endpoint distributions are available.

But at generation time, x_1 is unknown.

Instead, after learning b, s, generate with:

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

or:

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

COMMON PITFALL: ε AND γ

$\gamma(t)$ appears in the interpolant:

$$x_t = I + \gamma z.$$

It changes the density path:

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

$\varepsilon(t)$ appears in the sampler:

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

It does not change $\rho(t)$ when b, s are exact.

Thus:

$$\boxed{\gamma = \text{path design}; \qquad \varepsilon = \text{sampler design.}}$$

PROOF EXERCISE 1

Prove the least-squares conditional expectation identity:

For $Y \in L^2$, show:

$$g^* = \mathbb{E}[Y \mid X = \cdot]$$

minimizes:

$$J[g] = \mathbb{E}\left[\frac{1}{2}|g(X)|^2 - Y \cdot g(X)\right].$$

Required proof:

$$J[g] = \mathbb{E}\left[\frac{1}{2}|g(X)|^2 - \mathbb{E}[Y \mid X] \cdot g(X)\right],$$

then complete the square.

PROOF EXERCISE 2

Starting from:

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

prove:

$$\mathbb{E}[z e^{ik \cdot x_t}] = i\gamma(t) k \, g(t, k).$$

Then use Fourier inversion to prove:

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

Conclude:

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

PROOF EXERCISE 3

Prove Corollary 10.

Given:

$$\partial_t \rho + \nabla \cdot (b\rho) = 0, \qquad s = \nabla \log \rho.$$

Show:

$$\Delta \rho = \nabla \cdot (s\rho).$$

Then prove:

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

and:

$$\partial_t \rho + \nabla \cdot ((b - \varepsilon s)\rho) = -\varepsilon \Delta \rho.$$

DOCUMENT II SUMMARY

We proved:

$$x_t = I + \gamma z \Rightarrow \rho(t) = \mathcal{L}(x_t)$$

is smooth, positive, and satisfies:

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

The key identities:

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

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

The key objectives:

$$b = \arg\min L_b, \qquad s = \arg\min L_s, \qquad \eta_z = \arg\min L_{\eta_z}.$$

The key PDE family:

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