STOCHASTIC INTERPOLANTS V

Spatially Linear Interpolants: Factorization, Design Choices, Latent Noise, Diffusion Coefficients, and Proofs

Ziseok Lee

April 30, 2026

This document specializes the general interpolant:

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

to the spatially linear form:

$$\boxed{x_t^{lin} = \alpha(t)x_0 + \beta(t)x_1 + \gamma(t)z.}$$

This case is central because:

the velocity and score reduce to conditional means.

The main proof theme:

$$\eta_0, \eta_1, \eta_z \quad \Longrightarrow \quad b, s, b^F, b^B.$$

Let:

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

Define:

$$\boxed{x_t^{lin} = \alpha(t)x_0 + \beta(t)x_1 + \gamma(t)z.}$$

The coefficient functions satisfy:

$$\alpha, \beta, \gamma^2 \in C^2([0,1]),$$
$$\alpha(0) = 1, \qquad \beta(1) = 1,$$
$$\alpha(1) = \beta(0) = \gamma(0) = \gamma(1) = 0,$$
$$\gamma(t) > 0 \qquad 0 < t < 1.$$

At $t = 0$:

$$x_0^{lin} = \alpha(0)x_0 + \beta(0)x_1 + \gamma(0)z = x_0.$$

At $t = 1$:

$$x_1^{lin} = \alpha(1)x_0 + \beta(1)x_1 + \gamma(1)z = x_1.$$

Thus:

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

For $0 < t < 1$, the Gaussian latent term smooths the density.

Define:

$$\eta_0(t,x) = \mathbb{E}[x_0 \mid x_t^{lin} = x],$$
$$\eta_1(t,x) = \mathbb{E}[x_1 \mid x_t^{lin} = x],$$
$$\eta_z(t,x) = \mathbb{E}[z \mid x_t^{lin} = x].$$

These are the basic building blocks.

They are all regression functions:

$$\eta_\bullet(t,x) = \arg\min_{\hat{\eta}} \mathbb{E}\left[\frac{1}{2}|\hat{\eta}(t,x_t)|^2 - y_\bullet \cdot \hat{\eta}(t,x_t)\right].$$

VELOCITY FACTORIZATION

Differentiate:

$$x_t^{lin} = \alpha x_0 + \beta x_1 + \gamma z.$$

Then:

$$\dot{x}_t^{lin} = \dot{\alpha} x_0 + \dot{\beta} x_1 + \dot{\gamma} z.$$

By the definition of the interpolant velocity:

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

Therefore:

$$\boxed{b(t, x) = \dot{\alpha}(t)\eta_0(t, x) + \dot{\beta}(t)\eta_1(t, x) + \dot{\gamma}(t)\eta_z(t, x).}$$

SCORE FACTORIZATION

The general score identity gives:

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

For the spatially linear interpolant:

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

This holds for:

$$0 < t < 1.$$

Thus η_z controls:

score, forward/backward SDE drifts, part of the ODE velocity.

Because:

$$\mathbb{E}[x_t^{lin} \mid x_t^{lin} = x] = x,$$

we have:

$$\mathbb{E}[\alpha x_0 + \beta x_1 + \gamma z \mid x_t^{lin} = x] = x.$$

Therefore:

$$\boxed{\alpha(t)\eta_0(t,x) + \beta(t)\eta_1(t,x) + \gamma(t)\eta_z(t,x) = x.}$$

This means:

any two of η_0, η_1, η_z determine the third.

SOLVING FOR EACH CONDITIONAL MEAN

If $\alpha(t) \neq 0$:

$$\eta_0 = \alpha^{-1}(x - \beta\eta_1 - \gamma\eta_z).$$

If $\beta(t) \neq 0$:

$$\eta_1 = \beta^{-1}(x - \alpha\eta_0 - \gamma\eta_z).$$

If $\gamma(t) \neq 0$:

$$\eta_z = \gamma^{-1}(x - \alpha\eta_0 - \beta\eta_1).$$

Because $\gamma(0) = \gamma(1) = 0$, the last formula is only safe away from endpoints.

Define:

$$L_{\eta_0}[\hat{\eta}_0] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_0(t, x_t^{lin})|^2 - x_0 \cdot \hat{\eta}_0(t, x_t^{lin})\right] \mathrm{d}t.$$

Conditioning:

$$L_{\eta_0} = \int_0^1 \int \left[\frac{1}{2}|\hat{\eta}_0(t, x)|^2 - \eta_0(t, x) \cdot \hat{\eta}_0(t, x)\right] \rho(t, x)\, \mathrm{d}x\, \mathrm{d}t.$$

Complete the square:

$$L_{\eta_0}[\hat{\eta}_0] - L_{\eta_0}[\eta_0] = \frac{1}{2} \int |\hat{\eta}_0 - \eta_0|^2 \rho.$$

REGRESSION OBJECTIVES FOR η_1, η_z

Similarly:

$$L_{\eta_1}[\hat{\eta}_1] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_1(t, x_t^{lin})|^2 - x_1 \cdot \hat{\eta}_1(t, x_t^{lin})\right] \mathrm{d}t,$$

and:

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

Their unique minimizers are:

$$\eta_1(t, x) = \mathbb{E}[x_1 \mid x_t = x], \qquad \eta_z(t, x) = \mathbb{E}[z \mid x_t = x].$$

For $Y \in \{x_0, x_1, z\}$, define:

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

The objective:

$$L_Y[\hat{\eta}] = \int \mathbb{E}\left[\frac{1}{2}|\hat{\eta}(t, x_t)|^2 - Y \cdot \hat{\eta}(t, x_t)\right] \mathrm{d}t$$

becomes:

$$\int\int \left[\frac{1}{2}|\hat{\eta}|^2 - \eta_Y \cdot \hat{\eta}\right] \rho.$$

Then:

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

Hence η_Y is the unique minimizer.

Since:

$$\alpha\eta_0 + \beta\eta_1 + \gamma\eta_z = x,$$

we can learn only two conditional means.

Examples:

$$(\eta_0, \eta_1) \Rightarrow \eta_z \Rightarrow s,$$

$$(\eta_1, \eta_z) \Rightarrow \eta_0 \Rightarrow b,$$

$$(\eta_0, \eta_z) \Rightarrow \eta_1 \Rightarrow b.$$

For one-sided Gaussian-base models, often one denoiser is enough.

Assume:

$$\mathbb{E}x_0 = \mathbb{E}x_1 = 0,$$

$$\mathrm{Cov}(x_0) = \mathrm{Cov}(x_1) = I_d,$$

and:

$$x_0 \perp x_1 \perp z.$$

Then:

$$\mathbb{E}x_t^{lin} = 0.$$

The covariance is:

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

COVARIANCE-PRESERVING CONSTRAINT

A natural design condition is:

$$\boxed{\alpha^2(t) + \beta^2(t) + \gamma^2(t) = 1.}$$

Then:

$$\mathrm{Cov}(x_t^{lin}) = I_d$$

whenever:

$$\mathrm{Cov}(x_0) = \mathrm{Cov}(x_1) = I_d.$$

This avoids large changes in scale along the bridge.

It is especially useful when the endpoints are standardized datasets.

name	$\alpha(t)$	$\beta(t)$	$\gamma(t)$
linear	$1-t$	t	$\sqrt{a\,t(1-t)}$
trigonometric	$\cos(\pi t/2)$	$\sin(\pi t/2)$	$\sqrt{a\,t(1-t)}$
encoding–decoding	$\cos^2(\pi t)\mathbf{1}_{[0,1/2)}$	$\cos^2(\pi t)\mathbf{1}_{(1/2,1]}$	$\sin^2(\pi t)$
one-sided linear	$1-t$	t	0
one-sided trig	$\cos(\pi t/2)$	$\sin(\pi t/2)$	0
SBDM VP	$\sqrt{1-t^2}$	t	0
mirror	0	1	$\sqrt{a\,t(1-t)}$

The linear design is:

$$\alpha(t) = 1 - t, \qquad \beta(t) = t, \qquad \gamma(t) = \sqrt{a\, t(1-t)}.$$

Then:

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

Without the latent term:

$$\gamma = 0,$$

this reduces to straight interpolation:

$$x_t = (1-t)x_0 + tx_1.$$

With $\gamma > 0$, intermediate densities are smoothed.

The trigonometric design is:

$$\alpha(t) = \cos(\pi t/2), \qquad \beta(t) = \sin(\pi t/2).$$

If:

$$\gamma(t) = 0,$$

then:

$$\alpha^2 + \beta^2 = 1.$$

With a latent term, one can instead set:

$$\alpha(t) = \sqrt{1 - \gamma^2(t)}\cos(\pi t/2),$$
$$\beta(t) = \sqrt{1 - \gamma^2(t)}\sin(\pi t/2).$$

This preserves:

$$\alpha^2 + \beta^2 + \gamma^2 = 1.$$

A bridge can deliberately pass through pure Gaussian noise at $t = 1/2$.

Choose:

$$\alpha(t) = \cos^2(\pi t)\mathbf{1}_{[0,1/2)}(t),$$
$$\beta(t) = \cos^2(\pi t)\mathbf{1}_{(1/2,1]}(t),$$
$$\gamma(t) = \sin^2(\pi t).$$

At $t = 1/2$:

$$\alpha(1/2) = \beta(1/2) = 0, \qquad \gamma(1/2) = 1.$$

Therefore:

$$x_{1/2} = z \sim \mathcal{N}(0, I_d).$$

The path is:

$$\rho_0 \longrightarrow \mathcal{N}(0, I_d) \longrightarrow \rho_1.$$

But it is still a single interpolant density path:

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

The probability-flow ODE:

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

still defines a map:

$$X_0 \sim \rho_0 \quad \rightsquigarrow \quad X_1 \sim \rho_1.$$

The midpoint Gaussian does not destroy the endpoint coupling encoded by the learned dynamics.

LATENT VARIABLE $\gamma(t)z$: DENSITY SMOOTHING

Let:

$$x_t^0 = \alpha x_0 + \beta x_1$$

denote the no-latent interpolant.

Then:

$$x_t = x_t^0 + \gamma(t)z.$$

Therefore:

$$\boxed{\rho_\gamma(t,\cdot) = \rho_0^{lin}(t,\cdot) * \mathcal{N}(0, \gamma^2(t) I_d).}$$

This convolution smooths:

$$\rho(t,\cdot), \qquad b(t,\cdot), \qquad s(t,\cdot).$$

PROOF OF CONVOLUTION IDENTITY

Let:

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

Since:

$$x_t = x_t^0 + \gamma z, \qquad z \perp x_t^0,$$

we have:

$$g_\gamma(t,k) = \mathbb{E}e^{ik\cdot x_t^0}\mathbb{E}e^{i\gamma k\cdot z}.$$

The Gaussian factor is:

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

Thus:

$$g_\gamma(t,k) = g_0(t,k)e^{-\frac{1}{2}\gamma^2|k|^2},$$

which is the Fourier transform of convolution with $\mathcal{N}(0,\gamma^2 I_d)$.

WHY SMOOTHING HELPS

If ρ_0 and ρ_1 are multimodal, then:

$$\alpha x_0 + \beta x_1$$

can create many spurious intermediate modes.

Adding:

$$\gamma(t) z$$

blurs these modes.

For learning:

$$\text{smoother } \rho(t) \Rightarrow \text{smoother } b(t,\cdot), s(t,\cdot) \Rightarrow \text{easier regression.}$$

This is a statistical role of the latent variable, separate from sampling stochasticity.

For the spatially linear interpolant:

$$x_t = \alpha x_0 + \beta x_1 + \gamma z,$$

the endpoint velocities satisfy:

$$\boxed{b(0, x) = \dot{\alpha}(0) x + \dot{\beta}(0) \mathbb{E}[x_1 \mid x_0 = x] - \lim_{t \to 0} \gamma(t) \dot{\gamma}(t) s_0(x),}$$

$$\boxed{b(1, x) = \dot{\alpha}(1) \mathbb{E}[x_0 \mid x_1 = x] + \dot{\beta}(1) x - \lim_{t \to 1} \gamma(t) \dot{\gamma}(t) s_1(x),}$$

where:

$$s_i = \nabla \log \rho_i.$$

Recall:

$$b = v - \gamma\dot{\gamma}s,$$

where:

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

For spatially linear I:

$$\partial_t I = \dot{\alpha}x_0 + \dot{\beta}x_1.$$

As $t \to 0$:

$$x_t \to x_0,$$

so:

$$v(0,x) = \dot{\alpha}(0)x + \dot{\beta}(0)\mathbb{E}[x_1 \mid x_0 = x].$$

The remaining term is:

$$-\lim_{t\to 0} \gamma\dot{\gamma}s(t,x) = -\lim_{t\to 0} \gamma\dot{\gamma}s_0(x).$$

If:

$$\gamma \in C^1([0,1]), \qquad \gamma(0) = \gamma(1) = 0,$$

then:

$$\gamma(0)\dot{\gamma}(0) = 0, \qquad \gamma(1)\dot{\gamma}(1) = 0.$$

Therefore endpoint score terms vanish:

$$b(0,x) = \dot{\alpha}(0)x + \dot{\beta}(0)\mathbb{E}[x_1 \mid x_0 = x],$$

$$b(1,x) = \dot{\alpha}(1)\mathbb{E}[x_0 \mid x_1 = x] + \dot{\beta}(1)x.$$

Take:

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

Then:

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

Hence:

$$\lim_{t\to 0} \gamma\dot{\gamma} = \frac{a}{2}, \qquad \lim_{t\to 1} \gamma\dot{\gamma} = -\frac{a}{2}.$$

Thus the velocity contains endpoint score information:

$$-\frac{a}{2}s_0(x) \quad \text{at } t=0, \qquad +\frac{a}{2}s_1(x) \quad \text{at } t=1.$$

ROLE OF $\varepsilon(t)$

The diffusion coefficient $\varepsilon(t)$ appears in the sampler:

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

It does not appear in the interpolant:

$$x_t = \alpha x_0 + \beta x_1 + \gamma z.$$

Thus:

$$\boxed{\gamma(t) \text{ changes the density path,} \qquad \varepsilon(t) \text{ changes only the pathwise sampler.}}$$

PROOF THAT ε DOES NOT CHANGE $\rho(t)$

Start with:

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

Set:

$$b^F = b + \varepsilon s.$$

Then:

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

Since:

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

we have:

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

Therefore:

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

The marginal path is the same.

ODE MAP VS SDE ENSEMBLE

ODE sampler:

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

A single initial point maps to a single terminal point:

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

SDE sampler:

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

A single initial point maps to an ensemble:

$$X_0 = x_0 \quad \Rightarrow \quad X_1 \text{ random.}$$

With exact fields and $X_0 \sim \rho_0$, both give:

$$X_1 \sim \rho_1.$$

quantity	where it appears	effect
$\gamma(t)$	$x_t = \alpha x_0 + \beta x_1 + \gamma z$	changes $\rho(t)$
$\varepsilon(t)$	$\mathrm{d}X_t = (b + \varepsilon s)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t$	changes sampler paths
ν	$(x_0, x_1) \sim \nu$	changes endpoint coupling
α, β	$\alpha x_0 + \beta x_1$	change bridge geometry

The path design variables are:

$$(\alpha, \beta, \gamma, \nu).$$

The sampler design variable is:

$$\varepsilon.$$

Let:

$$\rho_0(x) = \sum_{i=1}^{N_0} p_i^0 \mathcal{N}(x \mid m_i^0, C_i^0),$$

$$\rho_1(x) = \sum_{j=1}^{N_1} p_j^1 \mathcal{N}(x \mid m_j^1, C_j^1).$$

For independent coupling:

$$x_0 \perp x_1,$$

and spatially linear:

$$x_t = \alpha x_0 + \beta x_1 + \gamma z.$$

Conditioned on mixture components (i, j):

$$x_t \mid (i, j) \sim \mathcal{N}(m_{ij}(t), C_{ij}(t)),$$

where:

$$m_{ij} = \alpha m_i^0 + \beta m_j^1,$$

$$C_{ij} = \alpha^2 C_i^0 + \beta^2 C_j^1 + \gamma^2 I_d.$$

The interpolant density is:

$$\rho(t,x) = \sum_{i=1}^{N_0}\sum_{j=1}^{N_1} p_i^0 p_j^1 \mathcal{N}(x \mid m_{ij}(t), C_{ij}(t)).$$

Define posterior mixture weights:

$$w_{ij}(t,x) = \frac{p_i^0 p_j^1 \mathcal{N}(x \mid m_{ij}, C_{ij})}{\sum_{\ell,r} p_\ell^0 p_r^1 \mathcal{N}(x \mid m_{\ell r}, C_{\ell r})}.$$

Then:

$$\sum_{i,j} w_{ij}(t,x) = 1.$$

For fixed components (i, j):

$$\mathbb{E}[x_0 \mid x_t = x, i, j] = m_i^0 + \alpha C_i^0 C_{ij}^{-1}(x - m_{ij}),$$
$$\mathbb{E}[x_1 \mid x_t = x, i, j] = m_j^1 + \beta C_j^1 C_{ij}^{-1}(x - m_{ij}),$$
$$\mathbb{E}[z \mid x_t = x, i, j] = \gamma C_{ij}^{-1}(x - m_{ij}).$$

Therefore:

$$\eta_0(t, x) = \sum_{i,j} w_{ij}(t, x) \left[m_i^0 + \alpha C_i^0 C_{ij}^{-1}(x - m_{ij})\right],$$

and analogously for η_1, η_z.

GAUSSIAN MIXTURE SCORE AND VELOCITY

The mixture score is:

$$s(t,x) = -\sum_{i,j} w_{ij}(t,x) C_{ij}^{-1}(x - m_{ij}).$$

Since:

$$\eta_z = \gamma \sum_{i,j} w_{ij} C_{ij}^{-1}(x - m_{ij}),$$

we recover:

$$s = -\gamma^{-1}\eta_z.$$

The velocity is:

$$b = \dot{\alpha}\,\eta_0 + \dot{\beta}\,\eta_1 + \dot{\gamma}\,\eta_z.$$

Thus Gaussian mixtures provide closed-form ground truth for numerical experiments.

Let $\rho_0 = \mathcal{N}(0, I_d)$. Absorb the Gaussian base into the latent variable:

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

Boundary conditions:

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

This is the main finite-time formulation of diffusion-style models in this framework.

Define:

$$\eta_z^{os}(t,x) = \mathbb{E}[z \mid x_t^{os,lin} = x],$$
$$\eta_1^{os}(t,x) = \mathbb{E}[x_1 \mid x_t^{os,lin} = x].$$

Velocity:

$$\boxed{b(t,x) = \dot{\alpha}(t)\eta_z^{os}(t,x) + \dot{\beta}(t)\eta_1^{os}(t,x).}$$

Score:

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

ONE-SIDED ALGEBRAIC CONSTRAINT

Because:

$$x_t^{os,lin} = \alpha z + \beta x_1,$$

conditioning on $x_t = x$ gives:

$$\boxed{\alpha(t)\eta_z^{os}(t,x) + \beta(t)\eta_1^{os}(t,x) = x.}$$

If:

$$\beta(t) \neq 0,$$

then:

$$\eta_1^{os}(t,x) = \beta^{-1}(t)(x - \alpha(t)\eta_z^{os}(t,x)).$$

Therefore one denoiser η_z^{os} determines η_1^{os}.

Substitute:

$$\eta_1^{os} = \beta^{-1}(x - \alpha\eta_z^{os})$$

into:

$$b = \dot{\alpha}\eta_z^{os} + \dot{\beta}\eta_1^{os}.$$

Then:

$$\boxed{b(t,x) = \frac{\dot{\beta}(t)}{\beta(t)}x + \left(\dot{\alpha}(t) - \frac{\alpha(t)\dot{\beta}(t)}{\beta(t)}\right)\eta_z^{os}(t,x).}$$

This is the key computational simplification.

ENDPOINT $t = 0$ FOR ONE-SIDED LINEAR CASE

At $t = 0$:

$$x_0^{os,lin} = z.$$

Thus:

$$\eta_z^{os}(0, x) = x.$$

Since $z \perp x_1$:

$$\eta_1^{os}(0, x) = \mathbb{E}[x_1].$$

Therefore:

$$\boxed{b(0, x) = \dot{\alpha}(0)x + \dot{\beta}(0)\mathbb{E}[x_1].}$$

The denoiser objective:

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

The data-prediction objective:

$$L_{\eta_1}^{os}[\hat{\eta}_1] = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_1(t, x_t^{os,lin})|^2 - x_1 \cdot \hat{\eta}_1(t, x_t^{os,lin})\right] \mathrm{d}t.$$

Both are conditional-expectation regression objectives.

ONE-SIDED ODE SAMPLER

Given $\hat{\eta}_z^{os}$, define:

$$\hat{b}(t,x) = \frac{\dot{\beta}}{\beta}x + \left(\dot{\alpha} - \frac{\alpha\dot{\beta}}{\beta}\right)\hat{\eta}_z^{os}(t,x).$$

Then sample with:

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

With exact η_z^{os}:

$$X_1 \sim \rho_1.$$

The score estimate is:

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

The forward drift is:

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

Thus:

$$\mathrm{d}X_t^F = \left[\frac{\dot{\beta}}{\beta}X_t^F + \left(\dot{\alpha} - \frac{\alpha\dot{\beta}}{\beta} - \frac{\varepsilon(t)}{\alpha(t)}\right)\hat{\eta}_z^{os}(t, X_t^F)\right]\mathrm{d}t + \sqrt{2\varepsilon(t)}\,\mathrm{d}W_t.$$

HANDLING ENDPOINT SINGULARITIES

The denoiser objective is well-defined for:

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

But:

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

or:

$$s(t, x) = -\frac{\eta_z^{os}(t, x)}{\alpha(t)}$$

can be singular near endpoints.

Practical remedies:

$$t_0 > 0, \qquad t_f < 1,$$

or use a final denoising step.

Another option:

$$\varepsilon(t) = 0$$

near singular endpoints.

MIRROR AS SPATIALLY LINEAR

A linear mirror interpolant has:

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

with:

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

where:

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

In the table notation:

$$\alpha(t) = 0, \qquad \beta(t) = 1, \qquad \gamma(t) = \sqrt{a\, t(1-t)}$$

for the simplest mirror:

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

Then:

$$b = \dot{\gamma}\, \eta_z, \qquad s = -\gamma^{-1} \eta_z.$$

DESIGN CHOICE: COUPLING ν

The law of:

$$x_t = \alpha x_0 + \beta x_1 + \gamma z$$

depends on the joint law:

$$(x_0, x_1) \sim \nu.$$

Independent coupling:

$$\nu = \rho_0 \otimes \rho_1$$

is easy to sample.

Data-dependent coupling can reduce path complexity:

$$x_0 \leftrightarrow x_1$$

according to semantic or transport structure.

The proof theory remains the same as long as the assumptions hold.

Recall:

$$b = v - \gamma\dot{\gamma}s, \qquad v = \mathbb{E}[\partial_t I \mid x_t = x].$$

For spatially linear:

$$v = \dot{\alpha}\eta_0 + \dot{\beta}\eta_1.$$

Learning choices:

$$\hat{b} \text{ directly},$$

or:

$$\hat{v}, \hat{s} \quad \Rightarrow \quad \hat{b} = \hat{v} - \gamma\dot{\gamma}\hat{s}.$$

The likelihood bound from Document III applies to either parameterization, with different constants.

LEARNING η_z VS LEARNING s

The score objective contains:

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

The denoiser objective contains only:

$$z.$$

Therefore:

$$L_{\eta_z}$$

is usually better behaved numerically near endpoints.

But sampling still requires:

$$\hat{s} = -\gamma^{-1}\hat{\eta}_z.$$

Thus denoiser learning separates stable training from careful sampling.

$$x_t = \alpha x_0 + \beta x_1 + \gamma z$$

$$\Downarrow$$

$$(\eta_0, \eta_1, \eta_z) = (\mathbb{E}[x_0 \mid x_t], \mathbb{E}[x_1 \mid x_t], \mathbb{E}[z \mid x_t])$$

$$\Downarrow$$

$$b = \dot{\alpha}\eta_0 + \dot{\beta}\eta_1 + \dot{\gamma}\eta_z, \qquad s = -\gamma^{-1}\eta_z$$

$$\Downarrow$$

$$\dot{X}_t = b(t, X_t) \qquad \text{or} \qquad \mathrm{d}X_t = (b + \varepsilon s)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t.$$

PROOF EXERCISE 1

Prove:

$$b = \dot{\alpha}\eta_0 + \dot{\beta}\eta_1 + \dot{\gamma}\eta_z.$$

Start from:

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

and:

$$\dot{x}_t = \dot{\alpha}x_0 + \dot{\beta}x_1 + \dot{\gamma}z.$$

Use linearity of conditional expectation.

PROOF EXERCISE 2

Prove the algebraic constraint:

$$\alpha\eta_0 + \beta\eta_1 + \gamma\eta_z = x.$$

Hint:

$$\mathbb{E}[x_t \mid x_t = x] = x.$$

Then apply:

$$x_t = \alpha x_0 + \beta x_1 + \gamma z.$$

Derive the one-sided velocity formula:

$$b(t,x) = \frac{\dot{\beta}}{\beta} x + \left(\dot{\alpha} - \frac{\alpha \dot{\beta}}{\beta} \right) \eta_z^{os}(t,x).$$

Use:

$$b = \dot{\alpha} \eta_z^{os} + \dot{\beta} \eta_1^{os},$$

and:

$$\alpha \eta_z^{os} + \beta \eta_1^{os} = x.$$

Assume:

$$x_0, x_1, z$$

are independent and centered with identity covariance.

Prove:

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

and:

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

Then explain why:

$$\alpha^2 + \beta^2 + \gamma^2 = 1$$

is a natural design constraint.

Spatially linear interpolants give a tractable core model:

$$x_t = \alpha x_0 + \beta x_1 + \gamma z.$$

Main identities:

$$b = \dot{\alpha}\eta_0 + \dot{\beta}\eta_1 + \dot{\gamma}\eta_z,$$
$$s = -\gamma^{-1}\eta_z,$$
$$\alpha\eta_0 + \beta\eta_1 + \gamma\eta_z = x.$$

Design variables:

$$\alpha, \beta, \gamma, \nu, \varepsilon.$$

Next document:

$\boxed{\text{connections with SBDM, denoising, rectified flows, algorithms, experiments, and semester summary.}}$