STOCHASTIC INTERPOLANTS VII

Composing Generative Paths: Feynman–Kac Formula, Correctors, and SMC

Ziseok Lee

April 30, 2026

Documents I–VI studied a single stochastic-interpolant path

$$\rho_0 \rightsquigarrow \rho_1 .$$

This document studies how to compose multiple already-trained paths:

$$q_t^{(1)}, q_t^{(2)}, \ldots, q_t^{(m)} .$$

The central object is a time-indexed ratio-of-densities path:

$$p_t^\star(x) = \frac{h_t(x)}{Z_t}, \qquad h_t(x) = \prod_{i=1}^{m} q_t^{(i)}(x)^{\gamma_i} .$$

The central question is:

If each $q_t^{(i)}$ is sampleable, how do we sample $p_t^\star$?

Answer in VII:

derive a Feynman–Kac PDE and simulate it with weighted particles.

WHY COMPOSE GENERATIVE PATHS?

Single generative models learn one data distribution.

Inference-time composition asks for new distributions without retraining:

conditional generation, guidance, annealing, product of experts, reward tilting.

Examples:

$$p(x \mid c) \quad \text{from} \quad p(x), p(x \mid c),$$

$$p(x) \exp(\beta r(x)),$$

$$q_1(x) q_2(x),$$

$$q(x)^\beta.$$

Composition is attractive because it reuses pretrained models:

modular experts instead of monolithic retraining.

Let $q_t^{(i)}$ be time-indexed marginal densities from trained diffusion or stochastic-interpolant models.

Define:

$$h_t(x) = \prod_{i=1}^{m} q_t^{(i)}(x)^{\gamma_i}, \qquad Z_t = \int h_t(x)\,\mathrm{d}x.$$

If:

$$0 < Z_t < \infty,$$

then:

$$p_t^\star(x) = \frac{h_t(x)}{Z_t}$$

is a valid probability path.

The score is:

$$s_t^\star(x) = \nabla \log p_t^\star(x) = \sum_{i=1}^{m} \gamma_i \nabla \log q_t^{(i)}(x),$$

because Z_t is independent of x.

Annealing / tempering:

$$p^{\mathrm{ann}}_{t,\beta}(x) = \frac{q_t(x)^\beta}{Z_t(\beta)}.$$

Product of experts:

$$p^{\mathrm{prod}}_{t}(x) = \frac{q^{(1)}_t(x)q^{(2)}_t(x)}{Z_t}.$$

Geometric average:

$$p^{\mathrm{geo}}_{t,\beta}(x) = \frac{q^{(1)}_t(x)^{1-\beta}q^{(2)}_t(x)^\beta}{Z_t(\beta)}.$$

Classifier-free guidance as a ratio path:

$$p^{\mathrm{CFG}}_{t,\beta}(x) \propto q_t(x \mid \varnothing)^{1-\beta} q_t(x \mid c)^\beta.$$

When $\beta > 1$, the unconditional term has a negative exponent:

$$q_t(x \mid \varnothing)^{1-\beta} = \frac{1}{q_t(x \mid \varnothing)^{\beta-1}}.$$

A common heuristic is:

replace the score by $s_t^\star$.

For diffusion-style reverse SDEs:

$$\mathrm{d}X_t = \left[-f_t(X_t) + \sigma_t^2 s_t^\star(X_t)\right]\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t.$$

For geometric averaging:

$$s_t^\star = (1-\beta)s_t^{(1)} + \beta s_t^{(2)}.$$

This is exactly the score used in classifier-free guidance.

But the problem is:

the SDE with the mixed score need not have marginals $p_t^\star$.

The missing terms are time-evolution terms in the PDE.

We want:

$$p_t^\star = \frac{\prod_i q_t^{(i)\gamma_i}}{Z_t}.$$

We can choose a particle dynamics:

$$\mathrm{d}X_t = v_t(X_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t.$$

Its unweighted density solves:

$$\partial_t r_t = -\nabla \cdot (v_t r_t) + \frac{\sigma_t^2}{2}\Delta r_t.$$

If $p_t^\star$ does not solve this PDE, add weights:

$$\mathrm{d}A_t = g_t(X_t)\,\mathrm{d}t.$$

Then weighted particles solve:

$$\boxed{\partial_t p_t = -\nabla \cdot (v_t p_t) + \frac{\sigma_t^2}{2}\Delta p_t + \bar{g}_t p_t.}$$

Let p_t be a density.

Transport / continuity:

$$\partial_t p_t = -\nabla \cdot (p_t v_t).$$

Diffusion:

$$\partial_t p_t = \frac{\sigma_t^2}{2} \Delta p_t.$$

Reweighting:

$$\partial_t p_t = \bar{g}_t p_t, \qquad \bar{g}_t(x) = g_t(x) - \int g_t(y) p_t(y) \, \mathrm{d}y.$$

The centering condition gives:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int p_t(x) \, \mathrm{d}x = \int \bar{g}_t(x) p_t(x) \, \mathrm{d}x = 0.$$

Feynman–Kac combines all three.

Let:

$$\mathcal{L}_t f = v_t \cdot \nabla f + \frac{\sigma_t^2}{2} \Delta f,$$

and:

$$\mathcal{L}_t^* p = -\nabla \cdot (v_t p) + \frac{\sigma_t^2}{2} \Delta p.$$

The normalized Feynman–Kac PDE is:

$$\boxed{\partial_t p_t = \mathcal{L}_t^* p_t + (g_t - \langle p_t, g_t \rangle)\, p_t.}$$

Here:

$$\langle p_t, g_t \rangle = \int g_t(x) p_t(x)\, \mathrm{d}x.$$

This PDE is nonlinear only because of the normalization term.

UNNORMALIZED FEYNMAN–KAC PDE

Define an unnormalized density $\tilde{p}_t$ by:

$$\partial_t \tilde{p}_t = \mathcal{L}_t^* \tilde{p}_t + g_t \tilde{p}_t, \qquad \tilde{p}_0 = p_0.$$

Let:

$$Z_t = \int \tilde{p}_t(x)\,\mathrm{d}x.$$

Then:

$$p_t = \frac{\tilde{p}_t}{Z_t}.$$

Differentiate:

$$\partial_t p_t = \frac{\partial_t \tilde{p}_t}{Z_t} - \frac{\tilde{p}_t}{Z_t^2} \dot{Z}_t.$$

Since:

$$\dot{Z}_t = \int g_t \tilde{p}_t = \mathcal{Z}_t \left\langle p_t, g_t \right\rangle,$$

we recover:

$$\partial_t p_t = \mathcal{L}_t^* p_t + (g_t - \langle p_t, g_t \rangle) p_t.$$

Simulate:

$$\boxed{\mathrm{d}X_t = v_t(X_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t, \qquad \mathrm{d}A_t = g_t(X_t)\,\mathrm{d}t.}$$

Then:

$$A_T = \int_0^T g_s(X_s)\,\mathrm{d}s.$$

The Feynman–Kac identity says:

$$\boxed{\mathbb{E}_{p_T}[\varphi] = \frac{\mathbb{E}\left[e^{A_T}\varphi(X_T)\right]}{\mathbb{E}\left[e^{A_T}\right]}.}$$

Thus samples from p_T can be approximated by weighted particles:

$$\sum_{k=1}^{K} \frac{e^{A_T^{(k)}}}{\sum_j e^{A_T^{(j)}}}\varphi(X_T^{(k)}).$$

Assume p_t solves:

$$\partial_t p_t = \mathcal{L}_t^* p_t + (g_t - \langle p_t, g_t \rangle) p_t.$$

Let:

$$\mathrm{d}X_t = v_t(X_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t, \qquad X_0 \sim p_0.$$

For bounded φ:

$$\boxed{\mathbb{E}_{p_T}[\varphi(X)] = \frac{\mathbb{E}\left[e^{\int_0^T g_s(X_s)\,\mathrm{d}s} \varphi(X_T)\right]}{\mathbb{E}\left[e^{\int_0^T g_s(X_s)\,\mathrm{d}s}\right]}.}$$

The denominator is:

$$Z_T = \mathbb{E}\left[e^{\int_0^T g_s(X_s)\,\mathrm{d}s}\right].$$

Define:

$$\Phi_T(x,t) = \mathbb{E}\left[e^{\int_t^T g_s(X_s)\,\mathrm{d}s}\varphi(X_T) \mid X_t = x\right].$$

Terminal condition:

$$\Phi_T(x,T) = \varphi(x).$$

Dynamic programming gives, for $t < \tau < T$:

$$\Phi_T(x,t) = \mathbb{E}\left[e^{\int_t^\tau g_s(X_s)\,\mathrm{d}s}\Phi_T(X_\tau,\tau) \mid X_t = x\right].$$

Apply Itô to:

$$e^{\int_t^\tau g_s(X_s)\,\mathrm{d}s}\Phi_T(X_\tau,\tau).$$

The drift must vanish, hence:

$$\boxed{\partial_t \Phi_T + \mathcal{L}_t \Phi_T + g_t \Phi_T = 0.}$$

Let:

$$\partial_t \tilde{p}_t = \mathcal{L}_t^* \tilde{p}_t + g_t \tilde{p}_t.$$

Compute:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \Phi_T(x,t) \tilde{p}_t(x)\,\mathrm{d}x = \int (\partial_t \Phi_T) \tilde{p}_t + \int \Phi_T \partial_t \tilde{p}_t.$$

Insert equations:

$$= \int (-\mathcal{L}_t \Phi_T - g_t \Phi_T) \tilde{p}_t + \int \Phi_T (\mathcal{L}_t^* \tilde{p}_t + g_t \tilde{p}_t).$$

Use adjointness:

$$\int (\mathcal{L}_t \Phi_T) \tilde{p}_t = \int \Phi_T (\mathcal{L}_t^* \tilde{p}_t).$$

All terms cancel:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \Phi_T \tilde{p}_t = 0.$$

Since:

$$\int \Phi_T(x,t)\tilde{p}_t(x)\,\mathrm{d}x$$

is constant in t, evaluate at $t = T$:

$$\int \Phi_T(x,T)\tilde{p}_T(x)\,\mathrm{d}x = \int \varphi(x)\tilde{p}_T(x)\,\mathrm{d}x.$$

Evaluate at $t = 0$:

$$\int \Phi_T(x,0)p_0(x)\,\mathrm{d}x = \mathbb{E}\left[e^{\int_0^T g_s(X_s)\,\mathrm{d}s}\varphi(X_T)\right].$$

Therefore:

$$\int \varphi(x)\tilde{p}_T(x)\,\mathrm{d}x = \mathbb{E}\left[e^{\int_0^T g_s(X_s)\,\mathrm{d}s}\varphi(X_T)\right].$$

Divide by:

$$Z_T = \int \tilde{p}_T.$$

With particles:

$$X_T^{(k)}, \qquad A_T^{(k)} = \int_0^T g_s(X_s^{(k)})\,\mathrm{d}s,$$

define normalized weights:

$$\omega_T^{(k)} = \frac{e^{A_T^{(k)}}}{\sum_{j=1}^K e^{A_T^{(j)}}}.$$

Then:

$$\mathbb{E}_{p_T}[\varphi] \approx \sum_{k=1}^K \omega_T^{(k)} \varphi(X_T^{(k)}).$$

As:

$$K \to \infty,$$

this converges under standard self-normalized importance sampling assumptions.

In practice, full-trajectory weights may degenerate, so FKC uses SMC.

Suppose:

$$\partial_t p_t = -\nabla \cdot (p_t v_t).$$

Rewrite as:

$$\partial_t p_t = g_t^{\mathrm{tr}} p_t.$$

Since:

$$-\nabla \cdot (pv) = -p\nabla \cdot v - v \cdot \nabla p,$$

and:

$$\nabla p = p\nabla \log p,$$

we get:

$$\boxed{g_t^{\mathrm{tr}}(x) = -\nabla \cdot v_t(x) - \nabla \log p_t(x) \cdot v_t(x).}$$

Proof:

$$g_t^{\mathrm{tr}} p_t = [-\nabla \cdot v_t - s_t \cdot v_t]\, p_t = -\nabla \cdot (p_t v_t).$$

Suppose:

$$\partial_t p_t = \frac{\sigma_t^2}{2} \Delta p_t.$$

Use:

$$\Delta p = \nabla \cdot (p \nabla \log p) = p \left(\Delta \log p + |\nabla \log p|^2 \right).$$

Thus:

$$\boxed{g_t^{\text{diff}}(x) = \frac{\sigma_t^2}{2} \left(\Delta \log p_t(x) + |\nabla \log p_t(x)|^2 \right).}$$

Proof:

$$\nabla p = ps,$$

$$\Delta p = \nabla \cdot (ps) = p \nabla \cdot s + s \cdot \nabla p = p(\Delta \log p + |s|^2).$$

CONVERSION RULE: DIFFUSION TO CONTINUITY

Diffusion can also be written as transport:

$$\partial_t p_t = \frac{\sigma_t^2}{2} \Delta p_t.$$

Since:

$$\Delta p_t = \nabla \cdot (p_t s_t), \qquad s_t = \nabla \log p_t,$$

we get:

$$\frac{\sigma_t^2}{2} \Delta p_t = -\nabla \cdot \left[p_t \left(-\frac{\sigma_t^2}{2} s_t \right) \right].$$

Therefore the pure-diffusion marginal evolution can be simulated by:

$$\boxed{\dot{X}_t = -\frac{\sigma_t^2}{2} \nabla \log p_t(X_t).}$$

This is the probability-flow idea in miniature.

Let each pretrained expert $q_t^{(i)}$ satisfy:

$$\partial_t q_t^{(i)} = -\nabla \cdot \left(q_t^{(i)} \left[-f_t + \sigma_t^2 s_t^{(i)} \right] \right) + \frac{\sigma_t^2}{2} \Delta q_t^{(i)},$$

where:

$$s_t^{(i)} = \nabla \log q_t^{(i)}.$$

Equivalently:

$$\partial_t q_t^{(i)} = \nabla \cdot (f_t q_t^{(i)}) - \frac{\sigma_t^2}{2} \Delta q_t^{(i)}.$$

Dividing by $q_t^{(i)}$:

$$\partial_t \log q_t^{(i)} = \nabla \cdot f_t + f_t \cdot s_t^{(i)} - \frac{\sigma_t^2}{2} \left[\nabla \cdot s_t^{(i)} + |s_t^{(i)}|^2 \right].$$

Let:

$$p^{\mathrm{geo}}_{t,\beta}(x) = \frac{q_t^{(1)}(x)^{1-\beta} q_t^{(2)}(x)^{\beta}}{Z_t(\beta)}.$$

Its score is:

$$s_t^\star = (1-\beta) s_t^{(1)} + \beta s_t^{(2)}.$$

The heuristic SDE is:

$$\mathrm{d}X_t = \left[-f_t(X_t) + \sigma_t^2 s_t^\star(X_t)\right] \mathrm{d}t + \sigma_t \, \mathrm{d}W_t.$$

FKC asks:

What weight g_t makes this SDE target $p^{\mathrm{geo}}_{t,\beta}$?

For:

$$p_t^\star \propto q_t^{(1)1-\beta} q_t^{(2)\beta},$$

simulate:

$$\boxed{\mathrm{d}X_t = \left[-f_t(X_t) + \sigma_t^2 \left((1-\beta)s_t^{(1)}(X_t) + \beta s_t^{(2)}(X_t)\right)\right] \mathrm{d}t + \sigma_t \,\mathrm{d}W_t.}$$

Update weights:

$$\boxed{\mathrm{d}A_t = \frac{\sigma_t^2}{2}\beta(\beta-1)\left\|s_t^{(1)}(X_t) - s_t^{(2)}(X_t)\right\|^2 \mathrm{d}t.}$$

If $\beta = 0$ or $\beta = 1$, the correction vanishes.

Let:

$$a = 1 - \beta, \qquad b = \beta, \qquad s^\star = a s_1 + b s_2.$$

Ignoring the scalar normalizer:

$$\partial_t \log h = a \partial_t \log q_1 + b \partial_t \log q_2.$$

Using the expert log-PDE:

$$\partial_t \log q_i = \nabla \cdot f + f \cdot s_i - \frac{\sigma^2}{2}(\nabla \cdot s_i + |s_i|^2).$$

The SDE with score $s^\star$ has log-density evolution:

$$\nabla \cdot f + f \cdot s^\star - \frac{\sigma^2}{2}(\nabla \cdot s^\star + |s^\star|^2).$$

Subtract:

$$g = -\frac{\sigma^2}{2}\left[a|s_1|^2 + b|s_2|^2 - |a s_1 + b s_2|^2\right].$$

Use:

$$a|u|^2 + b|v|^2 - |au + bv|^2 = ab|u - v|^2,$$

with $ab = \beta(1 - \beta)$. Thus:

$$\sigma^2 \quad \beta(\beta - 1)| \quad |^2$$

Let:

$$q_t^{(1)}(x) = q_t(x \mid \varnothing), \qquad q_t^{(2)}(x) = q_t(x \mid c).$$

CFG uses:

$$s_t^{\mathrm{CFG}} = (1-\beta)s_t(x \mid \varnothing) + \beta s_t(x \mid c).$$

FKC interprets this as targeting:

$$p_t^{\mathrm{CFG}} \propto q_t(x \mid \varnothing)^{1-\beta} q_t(x \mid c)^{\beta}.$$

The correction is:

$$\mathrm{d}A_t = \frac{\sigma_t^2}{2}\beta(\beta-1)\left\|s_t(x \mid \varnothing) - s_t(x \mid c)\right\|^2 \, \mathrm{d}t.$$

Thus FKC corrects the intermediate marginals targeted by guidance.

For one expert:

$$p_{t,\beta}^{\mathrm{ann}}(x) = \frac{q_t(x)^\beta}{Z_t(\beta)}.$$

Its target score is:

$$s_t^\star = \beta s_t, \qquad s_t = \nabla \log q_t.$$

But there are many SDEs one may simulate.

FKC uses a one-parameter family:

$$\mathrm{d}X_t = \left[-f_t(X_t) + \eta\sigma_t^2 s_t(X_t)\right]\,\mathrm{d}t + \zeta\sigma_t\,\mathrm{d}W_t,$$

where:

$$\eta = \beta + (1-\beta)a,$$

$$\zeta = \sqrt{\frac{\beta + (1-\beta)2a}{\beta}}.$$

For:

$$p_t^\star \propto q_t^\beta, \qquad \beta > 0,$$

simulate:

$$\boxed{\mathrm{d}X_t = \left[-f_t(X_t) + \eta\sigma_t^2 s_t(X_t)\right]\,\mathrm{d}t + \zeta\sigma_t\,\mathrm{d}W_t,}$$

where:

$$\eta = \beta + (1-\beta)a, \qquad \zeta = \sqrt{\frac{\beta + (1-\beta)2a}{\beta}}.$$

Then update:

$$\boxed{\mathrm{d}A_t = (\beta - 1)\left[\nabla \cdot f_t(X_t) + \frac{\sigma_t^2}{2}\beta\|s_t(X_t)\|^2\right]\,\mathrm{d}t.}$$

For linear $f_t(x)$, the divergence term is constant and cancels after self-normalization.

The unnormalized target is:

$$h_t = q_t^{\beta}.$$

Thus:

$$\partial_t \log h_t = \beta \partial_t \log q_t.$$

From the expert PDE:

$$\partial_t \log q_t = \nabla \cdot f_t + f_t \cdot s_t - \frac{\sigma_t^2}{2}(\nabla \cdot s_t + \|s_t\|^2).$$

The simulated SDE has drift:

$$-f_t + \eta \sigma_t^2 s_t,$$

and diffusion:

$$\zeta \sigma_t.$$

Its Fokker–Planck log-evolution differs from $\partial_t \log h_t$ by the displayed g_t.

The choice:

$$\zeta^2 = \frac{\beta + (1-\beta)2a}{\beta}$$

is exactly the condition that cancels the $\nabla \cdot s_t$ terms.

Target-score simulation:

$$a = 0, \qquad \eta = \beta, \qquad \zeta = 1.$$

$$\mathrm{d}X_t = [-f_t + \beta\sigma_t^2 s_t]\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t.$$

Tempered-noise simulation:

$$a = \frac{1}{2}, \qquad \eta = \frac{1+\beta}{2}, \qquad \zeta = \frac{1}{\sqrt{\beta}}.$$

$$\mathrm{d}X_t = \left[-f_t + \frac{1+\beta}{2}\sigma_t^2 s_t\right]\mathrm{d}t + \frac{\sigma_t}{\sqrt{\beta}}\,\mathrm{d}W_t.$$

Both target:

$$p_t^\star \propto q_t^\beta$$

after FKC reweighting.

For two experts:

$$p_{t,\beta}^{\text{prod}}(x) \propto \left(q_t^{(1)}(x) q_t^{(2)}(x)\right)^{\beta}.$$

The target score is:

$$s_t^{\star} = \beta(s_t^{(1)} + s_t^{(2)}).$$

FKC simulates:

$$\mathrm{d}X_t = \left[-f_t + \eta\sigma_t^2(s_t^{(1)} + s_t^{(2)})\right] \mathrm{d}t + \zeta\sigma_t \, \mathrm{d}W_t,$$

with the same:

$$\eta = \beta + (1-\beta)a, \qquad \zeta = \sqrt{\frac{\beta + (1-\beta)2a}{\beta}}.$$

For:

$$p_{t,\beta}^{\mathrm{prod}} \propto (q_t^{(1)} q_t^{(2)})^{\beta},$$

the corrector is:

$$\boxed{\begin{aligned} \mathrm{d}A_t = \beta(\beta-1)\frac{\sigma_t^2}{2}\left\| s_t^{(1)}(X_t) + s_t^{(2)}(X_t) \right\|^2 \mathrm{d}t \\ + \beta\sigma_t^2 \left\langle s_t^{(1)}(X_t), s_t^{(2)}(X_t) \right\rangle \mathrm{d}t \\ + (2\beta-1)\nabla\cdot f_t(X_t)\,\mathrm{d}t. \end{aligned}}$$

When f_t is linear, $\nabla \cdot f_t$ is constant and cancels under normalized weights.

PROOF IDEA: PRODUCT-OF-EXPERTS

Let:

$$h_t = (q_1 q_2)^\beta.$$

Then:

$$\partial_t \log h_t = \beta \partial_t \log q_1 + \beta \partial_t \log q_2.$$

The target score is:

$$s^\star = \beta(s_1 + s_2).$$

The Fokker–Planck log-evolution of the chosen SDE contains quadratic terms in:

$$s_1 + s_2.$$

The difference between:

$$\beta(|s_1|^2 + |s_2|^2)$$

and:

$$\beta^2 |s_1 + s_2|^2$$

produces:

$$\beta(\beta - 1)|s_1 + s_2|^2 + 2\beta\, s_1 \cdot s_2.$$

Multiplying by $\sigma^2/2$ yields the stated correction.

Let one model q_t be tilted by a reward $r(x)$:

$$p_t^{\text{rew}}(x) \propto q_t(x)\exp(\beta_t r(x)).$$

Score:

$$s_t^\star = s_t + \beta_t \nabla r.$$

One FKC sampler is:

$$\mathrm{d}X_t = \left[-f_t(X_t) + \sigma_t^2\left(s_t(X_t) + \frac{\beta_t}{2}\nabla r(X_t)\right)\right]\mathrm{d}t + \sigma_t\,\mathrm{d}W_t.$$

Corrector:

$$\boxed{\begin{aligned}\mathrm{d}A_t = \dot{\beta}_t r(X_t)\,\mathrm{d}t - \langle \beta_t \nabla r(X_t), f_t(X_t)\rangle\ \mathrm{d}t \\ + \left\langle \beta_t \nabla r(X_t), \frac{\sigma_t^2}{2} s_t(X_t)\right\rangle \mathrm{d}t.\end{aligned}}$$

Full-trajectory weights:

$$e^{A_T}$$

can have high variance.

Weight degeneracy means:

$$\omega_T^{(k)} \approx 1 \quad \text{for one particle,} \qquad \omega_T^{(j)} \approx 0 \quad \text{for most others.}$$

Sequential Monte Carlo combats this by alternating:

$$\text{propagate} \quad \longrightarrow \quad \text{incremental reweight} \quad \longrightarrow \quad \text{resample if necessary.}$$

The philosophy:

$$\boxed{\text{correct the population along the path, not only at the end.}}$$

At time t_n, represent:

$$p_{t_n}$$

by weighted particles:

$$\left\{X_{t_n}^{(k)}, \omega_{t_n}^{(k)}\right\}_{k=1}^{K}.$$

Propagation:

$$X_{t_{n+1}}^{(k)} = X_{t_n}^{(k)} + v_{t_n}(X_{t_n}^{(k)})\Delta t + \sigma_{t_n}\sqrt{\Delta t}\xi_n^{(k)}.$$

Weight update:

$$A_{t_{n+1}}^{(k)} = A_{t_n}^{(k)} + g_{t_n}(X_{t_n}^{(k)})\Delta t.$$

Normalize:

$$\omega_{t_{n+1}}^{(k)} = \frac{e^{A_{t_{n+1}}^{(k)}}}{\sum_j e^{A_{t_{n+1}}^{(j)}}}.$$

EFFECTIVE SAMPLE SIZE

The effective sample size is:

$$\boxed{\mathrm{ESS} = \frac{\left(\sum_{k=1}^{K} w_k\right)^2}{\sum_{k=1}^{K} w_k^2}.}$$

If weights are uniform:

$$w_1 = \cdots = w_K,$$

then:

$$\mathrm{ESS} = K.$$

If one weight dominates:

$$\mathrm{ESS} \approx 1.$$

Resampling rule:

$$\text{if } \mathrm{ESS} < \tau K, \quad \text{resample.}$$

Here:

$$\tau \in (0, 1)$$

is a threshold.

SYSTEMATIC RESAMPLING

Given normalized weights:

$$\omega_1, \dots, \omega_K, \qquad \sum_k \omega_k = 1,$$

define cumulative sums:

$$C_k = \sum_{j=1}^{k} \omega_j .$$

Draw:

$$U \sim \mathrm{Unif}(0, 1/K).$$

Use thresholds:

$$U_m = U + \frac{m-1}{K}, \qquad m = 1, \dots, K.$$

Choose ancestor a_m such that:

$$C_{a_m - 1} < U_m \le C_{a_m} .$$

Set:

$$\tilde{X}_m = X_{a_m}, \qquad \tilde{w}_m = 1/K.$$

ACTIVE RESAMPLING INTERVAL

In practice, FKC often resamples only on:

$$t \in [t_{\min}, t_{\max}].$$

Outside the active interval:

$$g_t = 0 \quad \text{or weights are not accumulated.}$$

Motivation:

early resampling may reduce diversity,

late resampling may improve final correctness.

This is an algorithmic design choice:

FK theorem gives the target; SMC design controls variance and diversity.

The reweighting PDE is:

$$\partial_t p_t(x) = p_t(x)\left(g_t(x) - \mathbb{E}_{p_t} g_t\right).$$

This can be realized by a mean-field jump process:

$$\partial_t p_t^{\mathrm{jump}}(x) = \int \lambda_t(y) J_t(x \mid y) p_t(y)\,\mathrm{d}y - \lambda_t(x) p_t(x).$$

The first term is inflow. The second term is outflow.

The goal:

choose λ_t, J_t so that jump evolution equals reweighting.

Let:

$$c_t(x) = g_t(x) - \mathbb{E}_{p_t} g_t.$$

Define:

$$c_t^+(x) = \max(c_t(x), 0), \qquad c_t^-(x) = \max(-c_t(x), 0).$$

Choose:

$$\boxed{\lambda_t(x) = c_t^-(x),}$$

and:

$$\boxed{J_t(y \mid x) = \frac{c_t^+(y) p_t(y)}{\int c_t^+(z) p_t(z)\, \mathrm{d}z}.}$$

Then:

$$\boxed{\partial_t p_t^{\text{jump}}(x) = c_t(x) p_t(x) = \partial_t p_t^{\text{weight}}(x).}$$

PROOF OF JUMP REPRESENTATION

Since:

$$\mathbb{E}_{p_t} c_t = 0,$$

we have:

$$\int c_t^+(z) p_t(z) \, \mathrm{d}z = \int c_t^-(z) p_t(z) \, \mathrm{d}z =: M_t.$$

The inflow term is:

$$\int \lambda_t(y) J_t(x \mid y) p_t(y) \, \mathrm{d}y = \int c_t^-(y) \frac{c_t^+(x) p_t(x)}{M_t} p_t(y) \, \mathrm{d}y.$$

Thus:

$$\text{inflow} = \frac{c_t^+(x) p_t(x)}{M_t} \int c_t^-(y) p_t(y) \, \mathrm{d}y = c_t^+(x) p_t(x).$$

The outflow term is:

$$\lambda_t(x) p_t(x) = c_t^-(x) p_t(x).$$

Therefore:

$$\partial_t p_t^{\text{jump}} = (c_t^+ - c_t^-) p_t = c_t p_t.$$

Given:

$$p_t^\star \propto \prod_i q_t^{(i)\gamma_i}, \qquad s_t^\star = \sum_i \gamma_i s_t^{(i)}.$$

Choose a base SDE:

$$\mathrm{d}X_t = \mu_t(X_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t.$$

Derive:

$$g_t(x) = \frac{\partial_t h_t(x) - \mathcal{L}_t^* h_t(x)}{h_t(x)}$$

up to an additive constant.

Then simulate:

$$\mathrm{d}A_t = g_t(X_t)\,\mathrm{d}t.$$

Approximate expectations:

$$\mathbb{E}_{p_T^\star}[\varphi] \approx \sum_k \frac{e^{A_T^{(k)}}}{\sum_j e^{A_T^{(j)}}} \varphi(X_T^{(k)}).$$

FKC AS CORRECTED PATH COMPOSITION

FKC does not train a new model.

It takes pretrained quantities:

$$s_t^{(i)} = \nabla \log q_t^{(i)}, \qquad f_t, \qquad \sigma_t.$$

It defines:

$$p_t^\star \propto \prod_i q_t^{(i)\gamma_i}.$$

Then it constructs weighted particles whose weighted law matches $p_t^\star$.

The correction lives in:

particle weights, not neural-network parameters.

Each expert path $q_t^{(i)}$ may itself come from a stochastic interpolant:

$$x_t^{(i)} = I_i(t, x_0^{(i)}, x_1^{(i)}) + \gamma_i^{\mathrm{SI}}(t) z_i.$$

Documents I–VI give:

$$q_t^{(i)}, \quad v_t^{(i)}, \quad s_t^{(i)}, \quad \text{ODE/SDE samplers.}$$

FKC adds a new layer:

$$\boxed{q_t^{(1)}, \ldots, q_t^{(m)} \quad \longmapsto \quad p_t^\star \propto \prod_i q_t^{(i)\gamma_i}.}$$

Thus composition is a second-order operation on already constructed probability paths.

A HIDDEN ASSUMPTION

Everything above assumes:

$$Z_t = \int h_t(x)\,\mathrm{d}x < \infty \qquad \forall t.$$

If:

$$Z_t = \infty,$$

then:

$$p_t^\star = \frac{h_t}{Z_t}$$

does not exist.

Then:

$$s_t^\star = \nabla \log h_t$$

is not the score of any probability density.

The SDE and weight updates may still be numerically computable, but:

they no longer target the intended probability path.

This is the starting point of ACE.

The Feynman–Kac PDE can be written:

$$\partial_t p_t = \mathcal{L}_t^* p_t + \bar{g}_t p_t.$$

Here:

$$\mathcal{L}_t$$

is a Markov generator:

$$\mathcal{L}_t f = v_t \cdot \nabla f + \frac{\sigma_t^2}{2} \Delta f.$$

The term:

$$\bar{g}_t p_t$$

is not ordinary diffusion or flow; it is a selection / branching / reweighting mechanism.

SMC implements this through:

weights, resampling, jump processes.

Semester II will generalize this language:

$$\boxed{\text{generative modeling by arbitrary Markov generators.}}$$

Generator matching studies Markov processes through:

$$\partial_t \langle p_t, f \rangle = \langle p_t, \mathcal{L}_t f \rangle.$$

FKC introduces a Feynman–Kac semigroup:

$$\partial_t \tilde{p}_t = \mathcal{L}_t^* \tilde{p}_t + g_t \tilde{p}_t.$$

Equivalently, for observables:

$$\partial_t u_t + \mathcal{L}_t u_t + g_t u_t = 0.$$

The potential g_t is not mass-preserving until normalized.

This suggests a broader view:

composition may require transport, diffusion, jumps, and selection.

Prove the normalized Feynman–Kac PDE.

Given:

$$\partial_t \tilde{p}_t = \mathcal{L}_t^* \tilde{p}_t + g_t \tilde{p}_t, \qquad p_t = \tilde{p}_t / Z_t,$$

show:

$$\partial_t p_t = \mathcal{L}_t^* p_t + (g_t - \mathbb{E}_{p_t} g_t) p_t.$$

Required steps:

$$\dot{Z}_t = \int g_t \tilde{p}_t,$$

$$\partial_t (\tilde{p}_t / Z_t) = (\partial_t \tilde{p}_t)/Z_t - p_t \dot{Z}_t / Z_t.$$

Derive the geometric-average FKC correction:

$$p_t^\star \propto q_t^{(1)1-\beta} q_t^{(2)\beta}.$$

Show:

$$g_t = \frac{\sigma_t^2}{2}\beta(\beta - 1)\|s_t^{(1)} - s_t^{(2)}\|^2.$$

Hint:

$$a|u|^2 + b|v|^2 - |au + bv|^2 = ab|u - v|^2,$$

with:

$$a = 1 - \beta, \qquad b = \beta.$$

PROOF EXERCISE 3

Prove the jump-process representation of reweighting.

Let:

$$c(x) = g(x) - \mathbb{E}_p g.$$

Set:

$$\lambda(x) = c^-(x), \qquad J(y \mid x) = \frac{c^+(y)p(y)}{\int c^+(z)p(z)\,\mathrm{d}z}.$$

Show:

$$\int \lambda(y)J(x \mid y)p(y)\,\mathrm{d}y - \lambda(x)p(x) = c(x)p(x).$$

Key fact:

$$\int c^+ p = \int c^- p.$$

We developed the FKC mechanism:

$$p_t^\star \propto \prod_i q_t^{(i)\gamma_i}$$

$$\Rightarrow \partial_t p_t^\star = \mathcal{L}_t^* p_t^\star + \bar{g}_t p_t^\star$$

$$\Rightarrow \; \mathrm{d}X_t = v_t(X_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t, \qquad \mathrm{d}A_t = g_t(X_t)\,\mathrm{d}t$$

$$\Rightarrow \mathbb{E}_{p_T^\star}[\varphi] = \frac{\mathbb{E}[e^{A_T}\varphi(X_T)]}{\mathbb{E}[e^{A_T}]}$$

$\Rightarrow$ SMC resampling corrects particles along the path.

The unresolved issue:

What if $Z_t = \int h_t = \infty$ at an intermediate time?

FKC is principled only when:

$$p_t^\star = \frac{h_t}{Z_t}$$

exists at every time used by the sampler. But heterogeneous path composition can fail:

$$Z_t = \infty \quad \text{even if} \quad Z_0, Z_1 < \infty.$$

This is:

$$\boxed{\text{Marginal Path Collapse.}}$$

Document VIII develops:

$$\text{Path Existence Criterion} \quad + \quad \text{Adaptive Correction with Exponents.}$$

ACE extends FKC from:

$$\gamma_i = \text{constant}$$

to:

$$\gamma_i = \gamma_i(t),$$

while guaranteeing the composed path exists.