Diffusion Models and SDEs

Lecture 3:

Schrodinger Bridges, IPF/Sinkhorn, Entropic Optimal Transport

Schrodinger Bridges – Intuition

Schrodinger 1931/32

In 1931/32, Erwin Schrodinger proposed the following Gedankenexperiment [52, 53]:

Consider the evolution of a cloud of N independent Brownian particles in R^3 . This cloud of particles has been observed having at the initial time t = 0 an empirical distribution equal to π_0 .

Schrodinger Bridges – Intuition

Schrodinger 1931/32

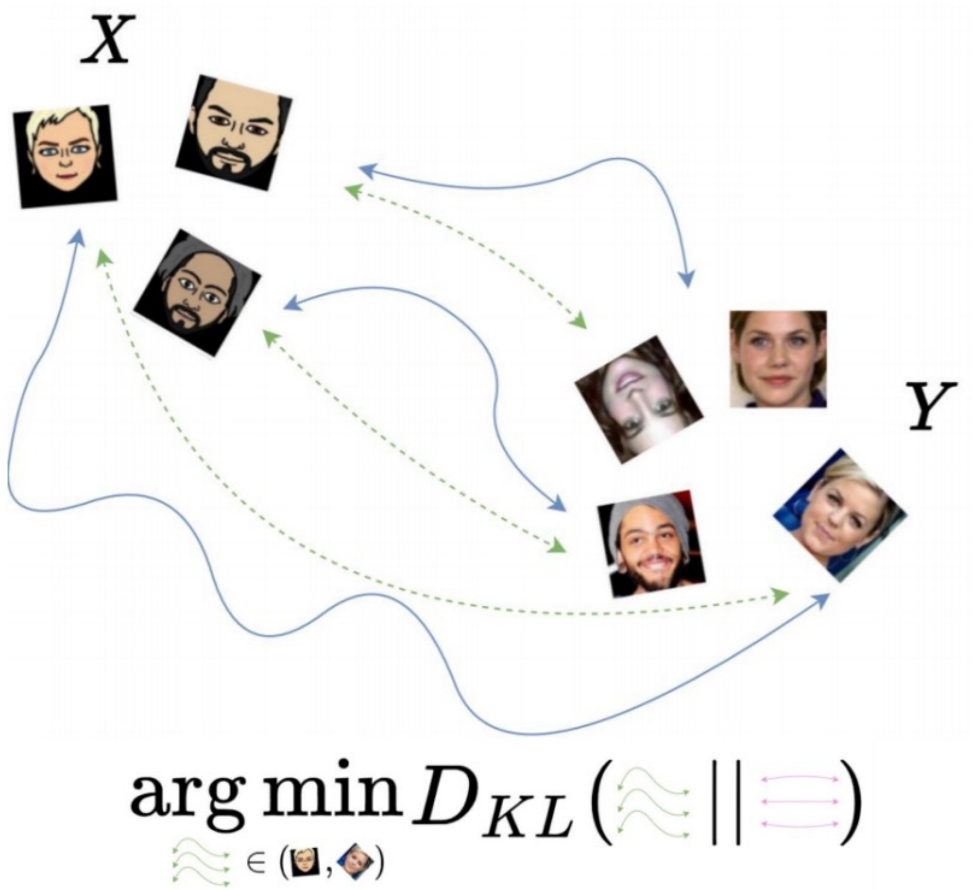
At time t = T, an empirical distribution π_1 is observed which considerably differs from what it should be according to the law of large numbers (N is large, typically of the order of Avogadro's number), namely

$$\pi_1(y) \neq \int_{\mathbb{R}^3} \mathcal{N}(y; x, T)\pi_0(x)dx$$

It seems that the particles have been transported in an unlikely way. But of the many unlikely ways in which this could have happened, which one is the most likely?

Schrodinger Bridges – Motivation

Schrodinger 1931/32



Schrodinger Bridges – Constrained KL minimisation

Constrained Optimisation

$$\mathbb{P}^* = \underset{\mathbb{P}\,:\,\text{s.t. } \mathbb{P}_0=\pi_0, \mathbb{P}_T=\pi_1}{\arg\min} D_{KL}(\mathbb{P}||\mathbb{P}^\rho)$$

Much harder problem than half bridges. Does not admit such a simple unconstrained formulation. Lets disintegrate:

$$\underset{\mathbb{P}\,:\,\text{s.t. } \mathbb{P}_0=\pi_0, \mathbb{P}_T=\pi_1}{\arg\min} D_{KL}(\mathbb{P}_{0,T}||\mathbb{P}^\rho_{0,T}) + \mathbb{E}_{\mathbb{P}_{0,T}} D_{KL}(\mathbb{P}_{|0,T}||\mathbb{P}^\rho_{|0,T})$$

Schrodinger Bridges – Entropic Optimal Transport

From Dynamic SBP to Static Entropic OT

$$\underset{\mathbb{P}\,:\,\text{s.t. } \mathbb{P}_0=\pi_0, \mathbb{P}_T=\pi_1}{\arg\min} \quad D_{KL}(\mathbb{P}_{0,T}||\mathbb{P}^{\rho}_{0,T}) + \mathbb{E}_{\mathbb{P}_{0,T}} \cancel{D_{KL}(\mathbb{P}_{|0,T}||\mathbb{P}^{\rho}_{|0,T})}$$

$$\underset{\mathbb{P}\,:\,\text{s.t. } \mathbb{P}_0=\pi_0, \mathbb{P}_T=\pi_1}{\arg\min} \quad D_{KL}(\mathbb{P}_{0,T}||\mathbb{P}^{\rho}_{0,T})$$

$$\underset{p(x,y):\,\text{s.t.} p(x)=\pi_0, p(y)=\pi_1}{\arg\min} \quad \mathbb{E}[\sigma^2 \ln p^{\rho}_{T|0}(y|x)] + \sigma^2 H(p)$$

Already looking like Entropic OT simply let p(x|y) = exp(-c(x,y)/sigma^2) and we arrive at your usual entropic OT objective.

Schrodinger Bridges – Entropic Optimal Transport

From Dynamic SBP to Static Entropic OT

$$\min_{p(x,y):\ \mathrm{s.t.}\, p(x)=\pi_0,\, p(y)=\pi_1} \mathbb{E}[\sigma^2 \ln p^{\rho}_{T|0}(y|x)] + \sigma^2 H(p)$$

Let \rho=0 then we have :

$$\min_{p(x,y):\ \mathrm{s.t.}\, p(x)=\pi_0,\, p(y)=\pi_1} \mathbb{E}[||y-x||^2] + \sigma^2 H(p) = \mathcal{W}^2_{2,\sigma^2}(\pi_0, \pi_1)$$

Aka the entropy regularized Wasserstein distance between the boundary distributions.

Schrodinger Bridges – IPF/Sinkhorn Algorithm

Solution - Alternating Subproblems (Coordinate Ascent - Sinkhorn Algorithm)

$$\mathbb{P}_0^* = \mathbb{P}^\rho$$

$$\mathbb{Q}_i^* = \underset{\mathbb{Q} \,:\, \text{s.t. } \mathbb{Q}_T = \pi_1}{\arg\min} \quad D_{KL}(\mathbb{Q}||\mathbb{P}_i^*)$$

$$\mathbb{P}_{i+1}^* = \underset{\mathbb{P} \,:\, \text{s.t. } \mathbb{P}_0 = \pi_0}{\arg\min} \quad D_{KL}(\mathbb{P}||\mathbb{Q}_i^*)$$

The above IPF (Iterative Proportional Fitting) iterates also known as sinkhorn have been proved to converge to the Schrodinger bridge solution. This approach dates back to Kullback.

Schrodinger Bridges – IPF/Sinkhorn Algorithm

Solution - Alternating Subproblems (Coordinate Ascent - Sinkhorn Algorithm)

These should look familiar

$$\mathbb{Q}_i^* = \underset{\mathbb{Q}\,:\,\text{s.t. }\mathbb{Q}_T = \pi_1}{\arg\min} \; D_{KL}(\mathbb{Q}||\mathbb{P}_i^*)$$

$$\mathbb{P}_{i+1}^* = \underset{\mathbb{P}\,:\,\text{s.t. }\mathbb{P}_0 = \pi_0}{\arg\min} \; D_{KL}(\mathbb{P}||\mathbb{Q}_i^*)$$

They are half bridges, and we know how to solve via score matching or stochastic control (i.e., via minimizing forward or reverse KL iteratively more later in the paper presentations).

Schrodinger Bridges – Schrodinger System

Solution – Functional System of Potentials

Another way to formulate the solution (and construct iterations) is based on the Schrodinger system:

$$\hat{\phi}_0(x)\phi_0(x) = \pi_0(x), \quad \hat{\phi}_1(y)\phi_1(y) = \pi_1(y)$$

$$\phi_0(x) = \int p_{T|0}(x|y)\phi_1(y)dy, \quad \hat{\phi}_1(y) = \int p_{T|0}(y|x)\hat{\phi}_0(x)dx$$

Result can be arrived at via Disintegration Theorem -> Lagrange Multipliers -> Calc of Variations. (The potentials are the Lagrange multipliers).

Schrodinger Bridges – Schrodinger System

Solution – Functional System of Potentials

Then given the potentials we have that

$$X_0 \sim \pi_0$$

$$dX_t = \left(\rho + \sigma^2\left(\nabla_{X_t} \ln \int \phi_1(z) p^{\rho}_{T|t}(z|X_t)\mathrm{d}x\right)\right)\mathrm{d}t + \sigma \mathrm{d}W_t$$

$$Y_0 \sim \pi_1$$

$$dY_t = \left(\rho - \sigma^2\left(\nabla_{Y_t} \ln \int \hat{\phi}_0(z) p^{\rho}_{t|0}(Y_t|z)\mathrm{d}z\right)\right)\mathrm{d}t + \sigma \mathrm{d}W_t^-$$

Solve The Schrodinger Bridge when the path measures represent SDE solutions.

Schrodinger Bridges – Schrodinger System

Solution – PDE Formulation

Furthermore, the potentials

$$\phi_t(x) = \int \phi_1(z) p^{\rho}_{T|t}(z|x) \mathrm{d}x \qquad \hat{\phi}_t(y) = \int \hat{\phi}_0(z) p^{\rho}_{t|0}(y|z) \mathrm{d}z$$

Solve the Following PDEs (remember space-time regularity from Doobs transform):

$$-\partial_t \phi_t = \nabla \phi_t \cdot \rho + \sigma^2 \Delta \phi_t, \qquad \hat{\phi}_0(x)\phi_0(x) = \pi_0(x)$$

$$\partial_t \hat{\phi}_t = -\nabla \cdot (\hat{\phi}_t \rho) + \sigma^2 \Delta \hat{\phi}_t, \qquad \hat{\phi}_1(y)\phi_1(y) = \pi_1(y)$$

These are just the FPK and the backward Kolmogorov equations. With funky boundary conditions.

Schrodinger Bridges – HJB/Hopf-Cole/Flemming

Solution – PDE Formulation

Via reversing Flemings/Hopf-Cole transform that is:

$$\psi_t(x) = \exp(\phi_t(x)), \quad \hat{\psi}_t(y) = \exp(\hat{\phi}_t(y))$$

Then through some standard calculus we arrive at the following HJB-PDEs:

$$-\partial_t \psi_t = ||\sigma \nabla \psi_t||^2 + \nabla \psi_t \cdot \rho + \sigma^2 \Delta \psi_t, \qquad \hat{\psi}_0(x) + \psi_0(x) = \ln \pi_0(x)$$

$$\partial_t \hat{\psi}_t = ||\sigma \nabla \hat{\psi}_t||^2 - \nabla \hat{\psi}_t \cdot (\rho - \ln p_t) + \sigma^2 \Delta \hat{\psi}_t, \quad \hat{\psi}_1(y) + \psi_1(y) = \ln \pi_1(y)$$

And thus, connecting to stochastic control / verification results etc.

Recap and Take Aways

OU and Pinned Brownian Motion

We studied two SDEs which transform complex distributions into simple distributions:

$$X_0 \sim \pi$$

$$\mathrm{d}X_t = \alpha(\mu - X_t)\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}W_t$$

$$X_0 \sim \pi$$

$$\mathrm{d}X_t = \frac{\mu - X_t}{T - t}\mathrm{d}t + \sqrt{\sigma}\mathrm{d}W_t$$

The OU process which rapidly mixes into a Gaussian, and the Pinned Brownian motion which instantaneously maps any distribution into a point mass.

$$Z_0 \sim \mathrm{law} X_T \approx \mathcal{N}(\mu, 1)$$

$$\mathrm{d}Z_t = (\alpha(Z_t - \mu) + 2\alpha\nabla \ln p_{T-t}(Z_t))\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}B_t$$

$$Z_0 = \mu$$

$$\mathrm{d}Z_t = \left(\frac{Z_t - \mu}{t} + \sigma^2\nabla \ln p_{T-t}(Z_t)\right)\mathrm{d}t + \sigma\mathrm{d}B_t$$

Their respective time reversals provide us with tractable generative models!

Recap and Take Aways

OU and Pinned Brownian Motion

In both settings we can learn the score and thus the time reversal via solving simple MSE/Regression objectives where we sample from the original noising processes to generate the “data” for the objectives.

$$Z_0 \sim \mathrm{law} X_T \approx \mathcal{N}(\mu, 1)$$

$$\mathrm{d}Z_t = (\alpha(Z_t - \mu) + 2\alpha \nabla \ln p_{T-t}(Z_t))\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}B_t$$

$$Z_0 = \mu$$

$$\mathrm{d}Z_t = \left(\frac{Z_t - \mu}{t} + \sigma^2 \nabla \ln p_{T-t}(Z_t)\right)\mathrm{d}t + \sigma \mathrm{d}B_t$$

In both cases learning the score / time reversal has an equivalent variational formulation in terms of half/full bridges:

$$\arg\min_{\mathbb{P}\,:\,\text{s.t. } \mathbb{P}_T = \pi} D_{KL}(\mathbb{P} || \mathbb{P}^{\alpha(\mu - x)})$$

$$\arg\min_{\mathbb{P}\,:\,\text{s.t. } \mathbb{P}_0 = \delta_0, \mathbb{P}_T = \pi} D_{KL}(\mathbb{P} || \mathbb{P}^0)$$

Which can be applied to gen modelling, sampling, path simulation, etc.

Sampling – Quick Change of Gears !

The Sampling Problem

Imagine I have access to a probability density function of the form

$$p(x) = \frac{e^{-U(x)}}{\int e^{-U(x)} dx}$$

Where U(x) is a “regular” function that can be computed pointwise however the denominator (partition function) we do not have access to. We would like to be able to sample random variables such that

$$X \sim p(x)$$

Sampling – Quick Change of Gears !

The Sampling Problem – Motivation Bayesian Inference

Consider a posterior we wish to draw samples from (in order to predict via the posterior predictive):

$$p(\theta|X) = \frac{e^{\ln p(X|\theta)p(\theta)}}{p(X)}$$

Here

$$U(x) = -\ln p(X|\theta)p(\theta)$$

And P(X) is typically intractable.

Sampling – Quick Change of Gears !

The Sampling Problem – SDEs to the rescue (Unadapted Langevin Algorithm – ULA)

The following SDE (looks a bit like OU ?)

$$\mathrm{d}X_t = -\alpha \nabla U(X_t)\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}W_t$$

Has the property:

$$\lim_{t \to \infty} \mathrm{law} X_t = \frac{e^{-U(x)}}{\int e^{-U(x)} dx}$$

However, unlike OU it mixes slowly (sqrt convergence).

Could we use half bridges to learn a sampler ? Yes ! More Later.

Sampling – Quick Change of Gears !

The Sampling Problem – Bayesian Inference with ULA – Test of time award ICML 2021

Consider the posterior example:

$$\mathrm{d}\Theta_t = \alpha \nabla \ln p(X|\Theta_t)p(\Theta_t)\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}W_t$$

Then:

$$\lim_{t \to \infty} \mathrm{law}\Theta_t = p(\theta|X)$$

Could we use half bridges to learn a sampler ? Yes ! More Later.