Reversed diffusion processes

Samy Tindel

Purdue University

Stochastic calculus and Machine Learning - MA 595

PURDUE UNIVERSITY

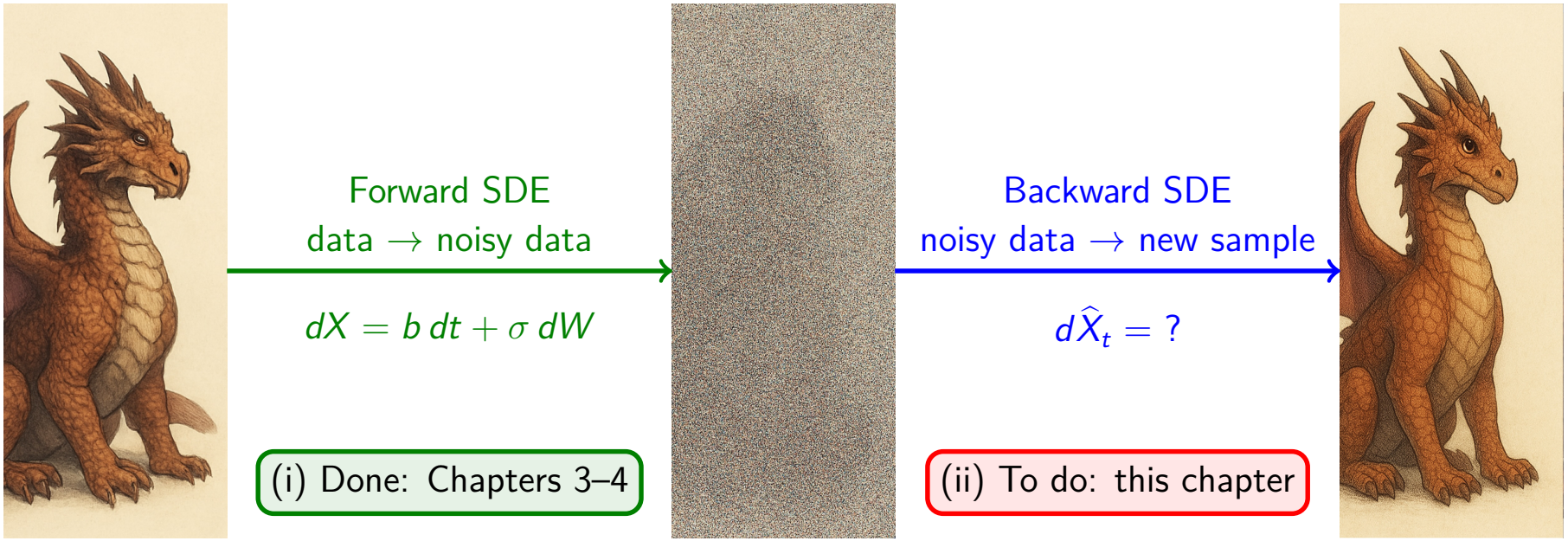
Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Where we stand



(i) Forward step: Markov property, generator, Kolmogorov equations

(ii) Backward step: Is $X_{\tau-t}$ a diffusion, and with which coefficients?

Generative modeling by reversal

Aim: Sample from a law μ on $\mathbb{R}^n$, only known through a data set

Forward step:

- Diffusion X started from $X_0 \sim \mu$, run on $[0, \tau]$
- Law of X_τ close to a simple law (e.g. $\mathcal{N}(0, \mathrm{Id})$)

Backward step: Run the process backward from a sample of the simple law

$$\widehat{X}_t = X_{\tau - t}, \qquad t \in [0, \tau]$$

$\hookrightarrow$ Lands on a sample of μ at $t = \tau$

Questions for this chapter

Direction of time in Chapters 3–4:

- Itô integral: fresh noise in the future
- Markov property: future given the present

Questions:

1. Is $\widehat{X}$ again a diffusion?
2. If so, what are its coefficients?
3. Can we simulate it?

Answer (preview): Same noise, drift corrected by the score $\nabla \log p_t$

Setting (recall)

Equation with additive noise:

$$X_t = X_0 + \int_0^t b(r, X_r)\, dr + \int_0^t \sigma(r)\, dW_r, \qquad X_0 \sim \mu$$

Smoothness assumption:

- b smooth, bounded, with bounded derivatives
- σ continuously differentiable, $\sigma(t)$ invertible, $a(t) = \sigma(t)\sigma(t)^*$

Densities: Transition density $\widetilde{q}_{s,t}(x, y)$, and density of X_t

$$p_t(x) = \int_{\mathbb{R}^n} \widetilde{q}_{0,t}(z, x)\, \mu(dz)$$

Generator and adjoint (recall)

Generator: For $\varphi \in \mathcal{C}^2(\mathbb{R}^n)$,

$$\mathcal{L}_s\varphi(x) = \sum_{j=1}^{n} b^j(s,x)\,\partial_{x_j}\varphi(x) + \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(s)\,\partial^2_{x_i x_j}\varphi(x)$$

Adjoint: For $\psi \in \mathcal{C}^2(\mathbb{R}^n)$,

$$\mathcal{L}_s^*\psi(x) = -\sum_{j=1}^{n} \partial_{x_j}\left[b^j(s,x)\,\psi(x)\right] + \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(s)\,\partial^2_{x_i x_j}\psi(x)$$

$\hookrightarrow$ Formal adjoint of $\mathcal{L}_s$ in $L^2(\mathbb{R}^n)$

Kolmogorov equations (recall)

Backward equation: (t, y) frozen, $\mathcal{L}_s$ acts on the starting point x

$$\partial_s \widetilde{q}_{s,t}(x, y) = -\mathcal{L}_s \widetilde{q}_{s,t}(\cdot, y)(x)$$

Forward equation: Density of X_t, $\mathcal{L}_t^*$ acts on x

$$\partial_t p_t(x) = \mathcal{L}_t^* p_t(x)$$

Moreover: Under the smoothness assumption on b and σ
$\hookrightarrow$ p smooth and $p_t(x) > 0$ on $(0, \tau] \times \mathbb{R}^n$
$\hookrightarrow$ $\nabla \log p_t$ well defined for $t > 0$

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Conditional densities (recall)

Setting: Couple (X, Z) in $\mathbb{R}^n \times \mathbb{R}^n$

- With joint density $f(x, z)$
- Marginal density of X: $f_X(x) = \int_{\mathbb{R}^n} f(x, z)\, dz$

Bayes rule: For x such that $f_X(x) > 0$
$\hookrightarrow$ conditional density of Z given $X = x$

$$f_{Z|X}(z \,|\, x) = \frac{f(x, z)}{f_X(x)}$$

Conditional expectation: For φ bounded measurable,

$$\mathbf{E}\left[\varphi(Z) \,|\, X = x\right] = \frac{1}{f_X(x)} \int_{\mathbb{R}^n} \varphi(z)\, f(x, z)\, dz$$

Looking backward: Bayes rule

Joint density: For $0 < s - h < s$, by the Markov property

$$(X_{s-h}, X_s) \text{ has density } (z, x) \mapsto p_{s-h}(z)\, \widetilde{q}_{s-h,s}(z, x)$$

Bayes rule: With $X = X_s$ and $Z = X_{s-h}$, we get

$$f_{X_{s-h}|X_s}(z \,|\, x) = \frac{p_{s-h}(z)\, \widetilde{q}_{s-h,s}(z, x)}{p_s(x)}$$

Reading:

- Forward transition $\widetilde{q}_{s-h,s}(\cdot, x)$, reweighted by p_{s-h}
- Transition of $\widehat{X}$, explicit in terms of p and $\widetilde{q}$

Reading coefficients from moments (recall)

Infinitesimal drift and covariance: For the forward diffusion,

$$\lim_{h \to 0^+} \frac{1}{h} \mathbf{E}\left[X_{s+h} - X_s \,|\, \mathcal{F}_s\right] = b(s, X_s)$$
$$\lim_{h \to 0^+} \frac{1}{h} \mathbf{Cov}\left(X_{s+h} - X_s \,|\, \mathcal{F}_s\right) = a(s)$$

Strategy for the time reversal equation: Take backward limits

- Backward mean: $\lim_{h \to 0^+} \frac{1}{h} \mathbf{E}\left[X_{s-h} - X_s \,|\, X_s = x\right]$
- Backward second moment: $\lim_{h \to 0^+} \frac{1}{h} \mathbf{E}\left[(X_{s-h} - X_s)^{\otimes 2} \,|\, X_s = x\right]$
- These limits: drift and diffusion matrix of $\widehat{X}$, if it is a diffusion

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Brownian case

Setting: $b \equiv 0$ and $\sigma \equiv \mathrm{Id}$, that is $X = X_0 + W$ with $X_0 \sim \mu$

Heat kernel: Transition density

$$\begin{aligned} \widetilde{q}_{s,t}(x,y) &= g_{t-s}(y-x) \\ g_u(w) &= \frac{1}{(2\pi u)^{n/2}} \exp\left(-\frac{|w|^2}{2u}\right) \end{aligned}$$

Marginal density:

$$p_s(x) = \int_{\mathbb{R}^n} g_s(x-z)\,\mu(dz) = (\mu * g_s)(x)$$

$\hookrightarrow$ Smooth and positive for $s > 0$, whatever μ

Backward drift of Brownian motion

Proposition 1

Consider:

- $X = X_0 + W$ *with* $X_0 \sim \mu$
- $p_s = \mu * g_s$ *the density of* X_s

Then for $s \in (0, \tau]$ *and* $x \in \mathbb{R}^n$*, we have*

$$\lim_{h \to 0^+} \frac{1}{h} \mathbf{E}\left[X_{s-h} - X_s \,|\, X_s = x\right] = \nabla \log p_s(x)$$

Remark: Forward, Brownian motion has no drift

$\hookrightarrow$ Backward, it acquires the drift $\nabla \log p_s$

Proof sketch for Proposition 1 (1)

Formal computation: Taylor expansions, limits under the integral sign

Bayes rule: Conditional density of X_{s-h} given $X_s = x$
$\hookrightarrow$ with $\widetilde{q}_{s-h,s}(z,x) = g_h(x-z)$

$$\frac{1}{h}\,\mathbf{E}\left[X_{s-h} - X_s \,|\, X_s = x\right] = \frac{1}{h}\int_{\mathbb{R}^n} (z-x)\,\frac{p_{s-h}(z)\,g_h(x-z)}{p_s(x)}\,dz$$

Proof sketch for Proposition 1 (2)

Gaussian scaling:

- Change of variables $z = x + \sqrt{h}\,\xi$
  $\hookrightarrow$ so that $dz = h^{n/2}\,d\xi$ and $z - x = \sqrt{h}\,\xi$
- Scaling: $h^{n/2} g_h(\sqrt{h}\,\xi) = \phi(\xi)$, standard Gaussian density
- g_h even, hence $g_h(x - z)\,dz = \phi(\xi)\,d\xi$

Then we have:

$$\frac{1}{h}\,\mathbf{E}\left[X_{s-h} - X_s \,|\, X_s = x\right] = \frac{1}{\sqrt{h}} \int_{\mathbb{R}^n} \xi\, \frac{p_{s-h}(x + \sqrt{h}\,\xi)}{p_s(x)}\, \phi(\xi)\,d\xi$$

$\hookrightarrow$ Prefactor $h^{-1/2}$ diverges: a cancellation is needed

Proof sketch for Proposition 1 (3)

Taylor expansion of the tilt: Around (s, x),

$$\frac{p_{s-h}(x + \sqrt{h}\,\xi)}{p_s(x)} = 1 + \sqrt{h}\,\xi \cdot \nabla \log p_s(x) + O(h)$$

Plugging into the previous slide: Formally,

$$\frac{1}{h}\,\mathbf{E}\left[X_{s-h} - X_s \,|\, X_s = x\right]$$
$$= \frac{1}{\sqrt{h}} \int_{\mathbb{R}^n} \xi\,\phi(\xi)\,d\xi + \int_{\mathbb{R}^n} \xi\,(\xi \cdot \nabla \log p_s(x))\,\phi(\xi)\,d\xi + O(\sqrt{h})$$

Proof sketch for Proposition 1 (4)

Three contributions:

- Order $h^{-1/2}$: vanishes, since $\int \xi \, \phi(\xi) \, d\xi = 0$
- Order 1: survives, since $\int \xi \xi^* \, \phi(\xi) \, d\xi = \mathrm{Id}$
- Order $\sqrt{h}$: vanishes in the limit

Conclusion: Only the order 1 term survives

$$\lim_{h \to 0^+} \frac{1}{h} \mathbf{E} \left[X_{s-h} - X_s \,|\, X_s = x \right] = \mathrm{Id} \, \nabla \log p_s(x) = \nabla \log p_s(x)$$

Backward covariance

Proposition 2

Let $\Delta_h = X_{s-h} - X_s$. *For* $s \in (0, \tau]$ *and* $x \in \mathbb{R}^n$,

$$\lim_{h \to 0^+} \frac{1}{h} \mathbf{E}\left[\Delta_h^i \Delta_h^j \,|\, X_s = x\right] = \delta_{ij}$$

Formal proof: Same Bayes rule and same scaling

- Quadratic integrand: $(z-x)^i(z-x)^j = h\,\xi_i\xi_j$ absorbs h^{-1}
- Leading term 1 of the tilt: $\int \xi_i\xi_j\,\phi(\xi)\,d\xi = \delta_{ij}$
- Other terms of order $\sqrt{h}$: vanish

Reading the computation

Where the drift comes from:

- Gaussian $g_h(x - \cdot)$: past positions symmetric around x
- Tilt by p_{s-h}: more likely origins are preferred
- Tilt of order $\sqrt{h} \times$ displacement of order $\sqrt{h} =$ drift of order h

Direction: $\nabla \log p_s = \nabla p_s / p_s$
$\hookrightarrow$ Steepest relative increase of the density of X_s

Diffusion matrix unchanged: Even (second-order) quantity
$\hookrightarrow$ Blind to the direction of time

Identifying coefficients in reversed time

Reversed clock:

- Time t for $\widehat{X}$ is time $\tau - t$ for X
- Drift of $\widehat{X}$ at time t: $\nabla \log p_{\tau - t}$. Diffusion matrix: Id

Guess:

$$d\widehat{X}_t = \nabla \log p_{\tau - t}(\widehat{X}_t)\, dt + d\widehat{W}_t$$

Comments:

- Forward: no drift. Backward: time dependent drift
- $\widehat{X}$ drifts toward regions where $X_{\tau - t}$ is likely

The score

Definition 3

Let p_s be a positive $\mathcal{C}^1$ density on $\mathbb{R}^n$. The score of p_s is

$$x \longmapsto \nabla \log p_s(x) = \frac{\nabla p_s(x)}{p_s(x)}$$

Properties:

- Insensitive to normalization: $\nabla \log(c\, p_s) = \nabla \log p_s$
- Depends on μ through the whole evolution
- Only unknown term of the reversed drift: learned in Chapter 6

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Two clusters (1)

Setting: $n = 1$, Brownian motion started from

$$\mu = \tfrac{1}{2}\,\mathcal{N}(-m, v_0) + \tfrac{1}{2}\,\mathcal{N}(m, v_0), \qquad m, v_0 > 0$$

Law of X_s: $X_s = X_0 + W_s$, with $W_s \sim \mathcal{N}(0, s)$ independent of X_0
$\hookrightarrow$ Each component $\mathcal{N}(\pm m, v_0)$ becomes $\mathcal{N}(\pm m, v)$, with $v = v_0 + s$

Density: With g_v the $\mathcal{N}(0, v)$ density,

$$p_s(x) = \tfrac{1}{2}\,g_v(x + m) + \tfrac{1}{2}\,g_v(x - m)$$

Two clusters (2)

Derivative: Since $g_v'(w) = -\frac{w}{v}\, g_v(w)$, with $g_\pm = g_v(x \pm m)$,

$$\partial_x \log p_s(x) = -\frac{(x+m)\, g_+ + (x-m)\, g_-}{v\,(g_+ + g_-)} = \frac{1}{v}\left(-x + m\,\frac{g_- - g_+}{g_- + g_+}\right)$$

Weights: $g_-/g_+ = e^{2mx/v}$, and $\frac{e^{2a}-1}{e^{2a}+1} = \tanh(a)$ with $a = mx/v$, so

$$\partial_x \log p_s(x) = \frac{m\tanh(mx/v) - x}{v}$$

Two clusters (3)

Slope at 0: Since $\tanh'(0) = 1$,

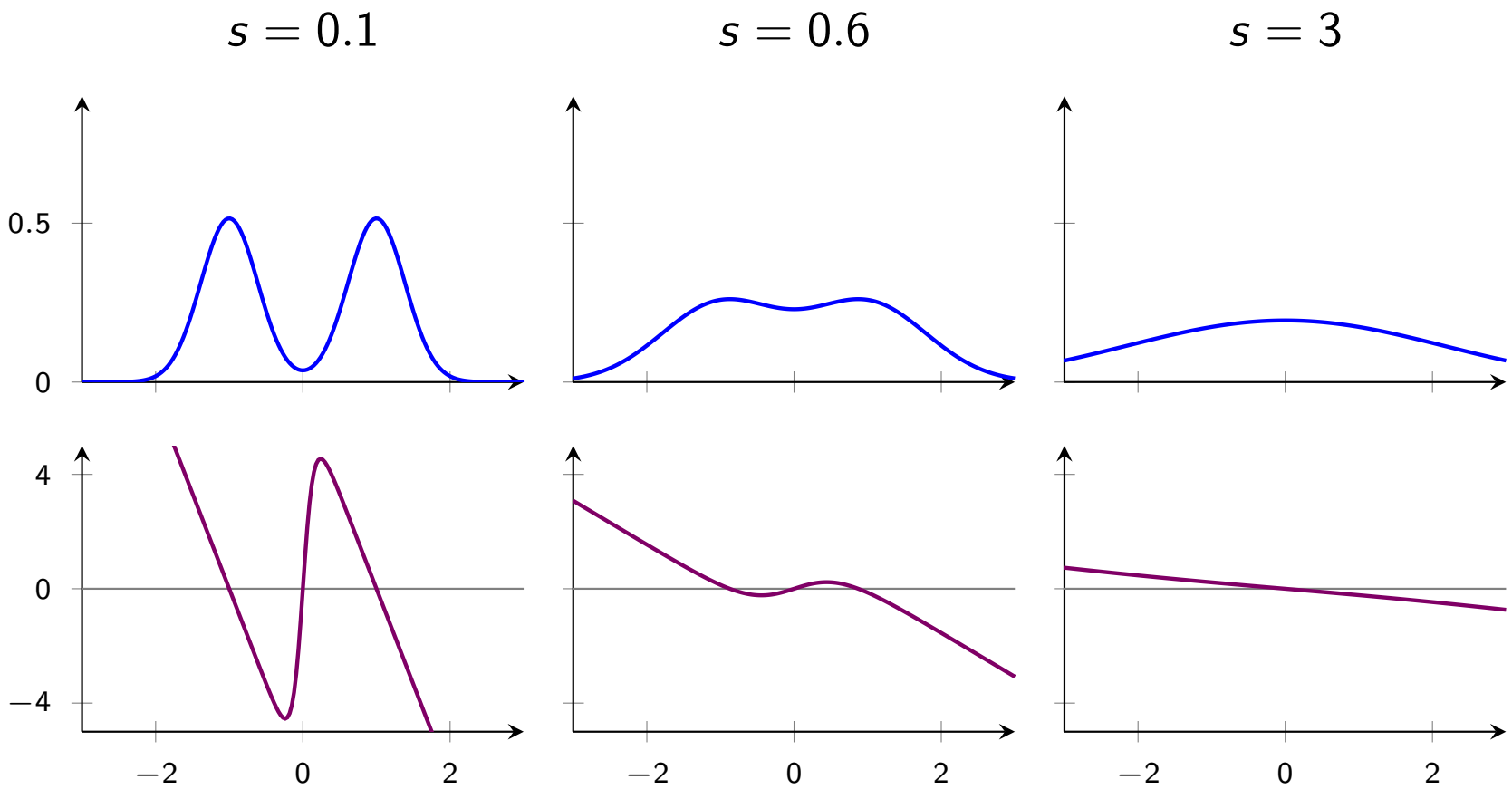
$$\kappa := \partial_x \left(\partial_x \log p_s\right)(0) = \frac{1}{v}\left(\frac{m^2}{v} - 1\right)$$

Reversed dynamics near 0: $d\widehat{X}_t \approx \kappa\, \widehat{X}_t\, dt + d\widehat{W}_t$

- $\kappa < 0$ ($v > m^2$): drift pulls back to 0, Ornstein–Uhlenbeck like
- $\kappa > 0$ ($v < m^2$): drift pushes away from 0, toward $\pm m$

Conclusion: 0 attracting for $v > m^2$, repelling for $v < m^2$

Two clusters: picture



Top: density p_s. Bottom: score $\partial_x \log p_s$ ($m = 1$, $v_0 = 0.05$)

Two clusters: the reversed dynamics

Early times ($\tau - t$ large):

- Score close to $-x/v$: pull toward the center
- p_τ close to a single Gaussian, two clusters invisible

Late times ($v_0 + \tau - t < m^2$):

- 0 becomes repelling, attractors near $\pm m$
- Each path is pushed into one of the two clusters

Conclusion: Bimodal structure recovered through the score alone

Tweedie's identity

Proposition 4

Let $X = X_0 + W$ with $X_0 \sim \mu$ arbitrary. Then for $s > 0$ and $x \in \mathbb{R}^n$,

$$\nabla \log p_s(x) = \frac{\mathbf{E}[X_0 \mid X_s = x] - x}{s}$$

Interpretation:

- $X_s = x$: noisy observation of the clean sample X_0
- $\mathbf{E}[X_0 \mid X_s = x]$: best guess of X_0, i.e. denoised version of x
- Score at x: points from x toward its denoised version

Proof of Tweedie's identity

Step 1, differentiation: Since $\nabla g_s(w) = -\frac{w}{s}\, g_s(w)$,

$$\nabla p_s(x) = \int_{\mathbb{R}^n} \nabla g_s(x-z)\, \mu(dz) = \frac{1}{s} \int_{\mathbb{R}^n} (z-x)\, g_s(x-z)\, \mu(dz)$$

Step 2, Bayes rule: (X_0, X_s) has law $\mu(dz)\, g_s(x-z)\, dx$, thus

$$f_{X_0|X_s}(z|x) = \frac{g_s(x-z)\, \mu(dz)}{p_s(x)}$$

Step 3, division by $p_s(x)$:

$$\nabla \log p_s(x) = \frac{1}{s} \int_{\mathbb{R}^n} (z-x)\, \frac{g_s(x-z)\, \mu(dz)}{p_s(x)} = \frac{\mathbf{E}\left[X_0 \,|\, X_s = x\right] - x}{s}$$

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

From Brownian motion to a general drift

Brownian case: Drift $\nabla \log p$, diffusion matrix unchanged

General guess: In the forward clock

- $s = \tau - t$, drift $-b + a\nabla \log p$
- $-b$: drift run backward
- $a\nabla \log p$: tilt by the density

Strategy (Anderson, 1982):

1. Bayes rule gives the transition density of $\widehat{X}$
2. Check that it solves a Fokker–Planck equation
3. Read off the coefficients

Candidate coefficients

Definition 5

For $t \in [0, \tau)$ and $x \in \mathbb{R}^n$, the reversed drift is

$$\widehat{b}(t,x) = -b(\tau - t, x) + a(\tau - t)\,\nabla \log p_{\tau - t}(x)$$

For $s \in (0, \tau]$, $\widehat{\mathcal{L}}_s^*$ is $\mathcal{L}_s^*$ with $b(s,\cdot)$ replaced by $\widehat{b}(\tau - s, \cdot)$

$$\widehat{\mathcal{L}}_s^* \psi = -\sum_{j=1}^{n} \partial_{x_j} \left[\widehat{b}^j(\tau - s, \cdot)\,\psi\right] + \frac{1}{2} \sum_{i,j=1}^{n} a^{ij}(s)\,\partial^2_{x_i x_j} \psi$$

Remark: Diffusion matrix a unchanged, as in the Brownian case

Consistency check on the marginals

Setting: We consider the case

- $n = 1$, $' = \partial_x$
- $a = a(s)$
- Since $\partial_x \log p_s = p_s'/p_s$, we have
  $\hookrightarrow \widehat{b}(\tau - s, \cdot)\, p_s = -b(s, \cdot)\, p_s + a\, p_s'$

Fokker-Planck type equation: We compute

$$\begin{aligned}
\widehat{\mathcal{L}}_s^* p_s &= -\left(\widehat{b}(\tau - s, \cdot)\, p_s\right)' + \tfrac{1}{2} a\, p_s'' \qquad \text{(definition of } \widehat{\mathcal{L}}_s^*\text{)} \\
&= -\left(-b\, p_s + a\, p_s'\right)' + \tfrac{1}{2} a\, p_s'' \qquad \text{(identity above)} \\
&= (b\, p_s)' - \tfrac{1}{2} a\, p_s'' \qquad (a \text{ free of } x) \\
&= -\mathcal{L}_s^* p_s \;=\; -\partial_s p_s \qquad \text{(forward equation)}
\end{aligned}$$

Why the marginals are not enough

Reversed clock: $\partial_t [p_{\tau - t}] = -(\partial_s p_s)_{s=\tau-t}$, hence

$$\partial_t p_{\tau - t} = \widehat{\mathcal{L}}^*_{\tau - t} \, p_{\tau - t}$$

$\hookrightarrow$ $(p_{\tau-t})_t$ solves the Fokker–Planck equation with drift $\widehat{b}$

But: Marginals do not determine a process

- W_t and $\sqrt{t}\, Z$, $Z \sim \mathcal{N}(0,1)$: same marginals
- Only the first one is Brownian motion

Right object: Transition density of $\widehat{X}$

Reversed transition density (1)

Definition 6

For $0 < s < t \leq \tau$ and $x, y \in \mathbb{R}^n$, we set

$$\widehat{q}_{s,t}(y,x) = \frac{p_s(x)\,\widetilde{q}_{s,t}(x,y)}{p_t(y)}$$

Notation:

- $p_s(x)$: density of X_s at x
- $\widetilde{q}_{s,t}(x,y)$: forward transition, from $X_s = x$ to $X_t = y$
- $\widehat{q}_{s,t}(y,x)$: backward transition, from $X_t = y$ to $X_s = x$

Reversed transition density (2)

Bayes rule: $\widehat{q}_{s,t}(y,\cdot)$ is the density of X_s given $X_t = y$
$\hookrightarrow$ Joint density $p_s(x)\,\widetilde{q}_{s,t}(x,y)$, divided by the density of X_t

Reversed clock: Transition density of $\widehat{X}$

- From time $\tau - t$, where $\widehat{X}$ sits at y
- To the later time $\tau - s$

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Reverse-time Fokker–Planck equation

Theorem 7

Consider:

- b, σ *satisfying the smoothness assumption,* $X_0 \sim \mu$
- $t \in (0, \tau]$ *and* $y \in \mathbb{R}^n$ *fixed*

Then for $s \in (0, t)$ *and* $x \in \mathbb{R}^n$,

$$-\partial_s \widehat{q}_{s,t}(y, x) = \widehat{\mathcal{L}}_s^* \, \widehat{q}_{s,t}(y, \cdot)(x)$$

Proof: Below for $n = 1$. General n: same steps, with sums

Proof for $n = 1$ (1)

Reduction: $p_t(y)$ constant in (s, x), so we work with the joint density

$$P = p_s(x), \qquad Q = \widetilde{q}_{s,t}(x, y), \qquad \pi = P\,Q$$

$\hookrightarrow$ Notation: $' = \partial_x$, and $a = a(s) = \sigma(s)^2$ does not depend on x

Kolmogorov equations: Forward for P, backward for Q

$$\begin{aligned} \partial_s P &= -(bP)' + \tfrac{1}{2} a\, P'' \\ \partial_s Q &= -b\, Q' - \tfrac{1}{2} a\, Q'' \end{aligned}$$

Proof for $n = 1$ (2)

Leibniz rule:

$$\begin{aligned}\partial_s \pi &= Q\,\partial_s P + P\,\partial_s Q \\ &= -\left[Q\,(bP)' + P\,b\,Q'\right] + \tfrac{1}{2}a\left[Q\,P'' - P\,Q''\right]\end{aligned}$$

First-order terms: Leibniz rule for the product $b\,P\,Q$

$$Q\,(bP)' + P\,b\,Q' = (b\,\pi)'$$

$\hookrightarrow$ The drift terms recombine into a single derivative

Proof for $n = 1$ (3)

Second-order terms: Two elementary identities

$$Q\,P'' - P\,Q'' = (Q\,P' - P\,Q')', \qquad Q\,P' - P\,Q' = 2\,Q\,P' - \pi'$$

Hence:

$$\tfrac{1}{2}a\,[Q\,P'' - P\,Q''] = (a\,Q\,P')' - \tfrac{1}{2}a\,\pi''$$

Observation: Derivatives of Q have disappeared
$\hookrightarrow$ Only π and P' are left

Proof for $n = 1$ (4)

Gathering: From steps (2) and (3)

$$\partial_s \pi = (-b\,\pi + a\,Q\,P')' - \tfrac{1}{2} a\,\pi''$$

Score appears: $Q\,P' = \frac{P'}{P}\,\pi = \pi\,\partial_x \log p_s$, so

$$-b\,\pi + a\,Q\,P' = (-b + a\,\partial_x \log p_s)\,\pi = \widehat{b}(\tau - s, \cdot)\,\pi$$

Conclusion: Changing signs,

$$-\partial_s \pi = -\left(\widehat{b}(\tau - s, \cdot)\,\pi\right)' + \tfrac{1}{2} a\,\pi'' = \widehat{\mathcal{L}}_s^* \pi$$

Reading the coefficients

Reversed clock: $\theta = \tau - s$, and $v_\theta(x) = \widehat{q}_{\tau-\theta,t}(y,x)$
$\hookrightarrow$ Density of $\widehat{X}_\theta$ given $\widehat{X}_{\tau-t} = y$

Chain rule: $\partial_\theta v_\theta = -\partial_s \widehat{q}_{s,t}(y,\cdot)$ at $s = \tau - \theta$, hence

$$\partial_\theta v_\theta = \widehat{\mathcal{L}}^*_{\tau-\theta} v_\theta$$

$\hookrightarrow$ Fokker–Planck equation, drift $\widehat{b}(\theta,\cdot)$, diffusion matrix $a(\tau-\theta)$

Caution: Correspondence, not a theorem
$\hookrightarrow$ Identifies the candidate. The SDE for $\widehat{X}$ needs more work

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Time reversal theorem

Theorem 8

Consider:

- b, σ *satisfying the smoothness assumption,* $X_0 \sim \mu$, $\widehat{X}_t = X_{\tau - t}$
- *Assumption, finite Fisher information: for all* $\varepsilon > 0$

$$\int_\varepsilon^\tau \mathbf{E}\left[|\nabla \log p_v(X_v)|^2\right] dv < \infty$$

Then for $\varepsilon \in (0, \tau)$ *there is a* $\mathcal{F}^{\widehat{X}}$*-Wiener process* $\widehat{W}$ *with*

$$\widehat{X}_t = \widehat{X}_0 + \int_0^t \widehat{b}(r, \widehat{X}_r)\, dr + \int_0^t \sigma(\tau - r)\, d\widehat{W}_r, \quad t \in [0, \tau - \varepsilon]$$

Comments on the theorem

Coefficients:

- Same diffusion coefficient, read in the reversed clock
- Drift $-b$, corrected by $a\nabla \log p$

Hypotheses:

- Fisher information: automatic under the smoothness assumption
- Cutoff ε: score may blow up as $s \to 0$ (e.g. μ Dirac mass)

In practice: Reversed equation run down to a small positive time
$\hookrightarrow$ Cutoff error, Chapter 7

Idea of proof

From densities to an SDE: Three stages (admitted)

1. Backward equation for $\widehat{q}$ (second Kolmogorov identity)
2. $\widehat{X}$ Markov with transition $\widehat{q}$, martingale problem for $\widehat{\mathcal{L}}$
3. Lévy's characterization: martingale part of $\widehat{X}$ gives $\widehat{W}$

Explicit reversed noise:

$$\widehat{W}_t = -\left(W_\tau - W_{\tau-t}\right) - \int_{\tau-t}^{\tau} \sigma(v)^* \nabla \log p_v(X_v)\, dv$$

$\hookrightarrow$ Forward noise run backward, compensated by the score

Outline

1. Setting
2. Bayes rule and the Brownian score
   - Looking backward
   - The Brownian computation
   - Example: two clusters
3. The Kolmogorov route
   - Reversed transition density
   - Anderson's computation
   - The time reversal theorem
4. The variance preserving equation

Variance preserving equation

Noise schedule: $\beta : [0, \tau] \to (0, \infty)$ smooth, $\bar{\beta}_t = \int_0^t \beta(r)\, dr$

Forward equation:

$$dX_t = -\tfrac{1}{2}\, \beta(t)\, X_t\, dt + \sqrt{\beta(t)}\, dW_t$$

Law of X_t given $X_0 = x$:

$$\mathcal{N}\left(e^{-\bar{\beta}_t/2}\, x,\, (1 - e^{-\bar{\beta}_t})\, \mathrm{Id}\right)$$

$\hookrightarrow$ Close to $\mathcal{N}(0, \mathrm{Id})$ when $\bar{\beta}_\tau$ is large, whatever μ

Reversed variance preserving equation

Time reversal: Theorem extended to this affine setting (Chapter 6)

$$d\widehat{X}_t = \left[\tfrac{1}{2}\,\beta(\tau - t)\,\widehat{X}_t + \beta(\tau - t)\,\nabla \log p_{\tau - t}(\widehat{X}_t)\right] dt + \sqrt{\beta(\tau - t)}\, d\widehat{W}_t$$

Toward an algorithm:

- Start from $\mathcal{N}(0, \mathrm{Id})$ instead of the law of X_τ
- Replace $\nabla \log p$ by a learned score s_θ
- Discretize in time (Euler–Maruyama)

Sampling algorithm (simple form)

Data: Learned score $s_\theta(t, x)$, cutoff $\varepsilon > 0$, K steps
$\hookrightarrow$ Step $h = (\tau - \varepsilon)/K$, grid $t_k = kh$

Algorithm:

1. Draw $Y_0 \sim \mathcal{N}(0, \mathrm{Id})$
2. For $0 \leq k < K$, draw $Z_{k+1} \sim \mathcal{N}(0, \mathrm{Id})$ independent and set

$$Y_{k+1} = Y_k + \beta(\tau - t_k)\left[\tfrac{1}{2} Y_k + s_\theta(\tau - t_k, Y_k)\right] h + \sqrt{\beta(\tau - t_k)\, h}\, Z_{k+1}$$

3. Return Y_K, approximate sample of μ

Summary

Reversal of a diffusion:

- Bayes rule: transition of $\widehat{X}$ in terms of p and $\widetilde{q}$
- Brownian case: backward drift $\nabla \log p$, same diffusion matrix
- Kolmogorov route: drift $-b + a\nabla \log p$, read off a Fokker–Planck equation

Next:

- Chapter 6: learning the score from a sample
- Chapter 7: sampling, and the error of the whole procedure