Advanced computational methods-Lecture 12

1 Detailed balance in SDEs and a splitting strategy

Let us consider the general SDE:

$$dX = b(X)\,dt + \sigma(X)dW,$$

where σ is a $d \times m$ matrix while W is an m-dimensional standard Wiener process.

1.1 The detailed balance

Proposition 1. *If $b = -\nabla V$ and $\sigma = \sqrt{2/\beta}I$, then the SDE has invariant measure*

$$\pi(z) \propto \exp\left(-\beta V\right). \tag{1}$$

Moreover, if $p(z,t;x_i)$ denotes the transition density from x_i, then the detailed balance condition holds

$$\pi(x_i)p(x_i^*,t;x_i) = \pi(x_i^*)p(x_i,t;x_i^*). \tag{2}$$

Proof. The law of X (or the density of the law of X), $p(x,t)$, satisfies the Fokker-Planck equation:

$$\partial_t p = -\nabla\cdot(bp) + \frac{1}{2}\sum_{i,j=1}^{d}\partial_{ij}(\Lambda_{ij}p) =: \mathcal{L}^* p, \tag{3}$$

where $\Lambda = \sigma\sigma^T$ and the operator

$$\mathcal{L}^* = -\nabla\cdot(b\cdot) + \frac{1}{2}\sum_{i,j=1}^{d}\partial_{ij}(\Lambda_{ij}\cdot)$$

is the dual operator of the generator of the SDE [**?**, Theorem 7.3.3] given by

$$\mathcal{L} := -b\cdot\nabla + \frac{1}{2}\Lambda : \nabla^2.$$

Let $p(\cdot,t;x_i)$ be the Green's function which is the solution of Eq. (3) with initial condition $p(x,0) = \delta(x - x_i)$ and gives the transition density starting from x_i:

$$p(\cdot,t;x_i) = e^{t\mathcal{L}^*}\delta(\cdot - x_i).$$

The detailed balance condition is therefore the distributional identity

$$\pi(x_i)\mathcal{L}_z^*\delta(z-x_i) = \pi(z)\mathcal{L}_{x_i}^*\delta(x_i-z)$$

We pick a test function φ. It can be shown easily that

$$\langle \pi(x_i)\mathcal{L}^*\delta(\cdot - x_i), \varphi(\cdot)\rangle = \pi(x_i)\mathcal{L}\varphi(x_i),$$

and that

$$\langle \pi(z)\mathcal{L}_{x_i}^*\delta(x_i - \cdot), \varphi(\cdot)\rangle = \mathcal{L}_{x_i}^*(\varphi(x_i)\pi(x_i)).$$

Hence, to verify the detailed balance condition, one needs the following

$$\pi(x_i)\mathcal{L}\varphi(x_i) = \mathcal{L}_{x_i}^*(\varphi(x_i)\pi(x_i)), \tag{4}$$

which is reduced to

$$-\pi b + \frac{1}{2}\nabla\cdot\left(\sigma\sigma^T\pi\right) = 0.$$

This holds for b being a gradient field and σ being square with $\sigma \propto I$.

$\square$

1.2 A splitting strategy

Consider that we want to sample from some density ρ. Then, we can define

$$U = -\frac{1}{\beta}\log\rho.$$

This potential U may have some singularity.

Consider decomposing the potential as

$$U = U_1 + U_2.$$

Given some X_{n-1}, we run the following SDE

$$dY = -\nabla U_1\,dt + \sqrt{\frac{2}{\beta}}\,dW$$

for some time $T > 0$, with initial data

$$Y(0) = X_{n-1}.$$

Then, we compute

$$\alpha = \min\Big(1, \exp[-\beta(U_2(Y(T)) - U_2(X_{n-1}))]\Big)$$

With probability, we accept $Y(T)$ as the new sample $X_n = Y(T)$, and otherwise, $X_n = X_{n-1}$.

This can be viewed as special case of the Metropolis-Hastings algorithm. In fact, for the SDE step:

$$e^{-\beta U_1(x)} p(x, y; T) = e^{-\beta U_1(y)} p(y, x; T)$$

Then, by MH algorithm:

$$\frac{e^{-\beta U(y)} p(y, x; T)}{e^{-\beta U(x)} p(x, y; T)} = \exp[-\beta(U_2(y) - U_2(x))]$$

The SDE step can be finished by typical SDE schemes and maybe applying minibatch technique, like in SGLD.

Example: sampling from the Gibbs distribution for many body systems.