Notes on Detailed Balance

Xuda Ye, 1600010616

May 27, 2020

Contents

1 Brownian dynamics

Consider the simple Brownian dynamics:

$$\mathrm{d}X = b(X)\mathrm{d}t + \sigma(X)\mathrm{d}W$$

where $\sigma \in \mathbb{R}^{d\times m}$ and W is the m-dimensional Wiener process.

1.1 Fokker-Planck equation

Define $a = \sigma\sigma^{\mathrm{T}} \in \mathbb{R}^{d\times d}$, then the generator of the SDE is

$$\mathcal{L}\varphi = b\cdot\nabla\varphi + \frac{1}{2}a : \nabla^2\varphi$$

and the forward Kolmogorov operator is

$$\mathcal{L}^*\varphi = -\nabla \cdot (b\varphi) + \frac{1}{2}\nabla^2 : (a\varphi)$$

which leads to the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p(x,t), \quad x \in \mathbb{R}^d,\ t \geqslant 0$$

The invariant distribution $\pi(x)$ of the SDE is defined as

$$\mathcal{L}^*\pi(x) = 0$$

1.2 Detailed balance

1.2.1 Formal definition

The detailed balance can be defined in several ways.

Definition 1.1 *Suppose the Fokker-Planck equation can be written as*

$$\frac{\partial p}{\partial t} + \nabla \cdot j(p) = 0$$

where $j(p) \in \mathbb{R}^d$ is a functional of p. The detailed balance is given by the relation

$$j(\pi) \equiv 0$$

where $\pi(x)$ is the invariant distribution.

For the Brownian dynamics, the flux is given by

$$j(p) = bp - \frac{1}{2}\nabla \cdot (ap)$$

so the condition for detailed balance becomes

$$b(x)\pi(x) = \frac{1}{2}\nabla \cdot (a(x)\pi(x)), \quad x \in \mathbb{R}^d \tag{$*$}$$

1.2.2 Derivation from the definition

We can also derive the relation $(*)$ from the definition of detailed balance.

Let $T_t(x,y)$ be the transition probability density of the SDE from $x \in \mathbb{R}^d$ to $y \in \mathbb{R}^d$ in time t. Using the Fokker-Planck equation, $T_t(x,y)$ can be represented as

$$T_t(x,y) = e^{t\mathcal{L}^*}\delta(y-x)$$

The detailed balance of the SDE is

$$\pi(x)T_t(x,y) = \pi(y)T_t(y,x) \tag{1}$$

or

$$\pi(x)\mathcal{L}_y^*\delta(y-x) = \pi(y)\mathcal{L}_x^*\delta(x-y) \tag{2}$$

where the subscript x means that the operator acts on a function of x.

To simply (2), pick a test function $\varphi \in C^2(\mathbb{R}^d)$ and integrate (2) with $\varphi(y)$:

$$\pi(x)\int_{\mathbb{R}^d} \varphi(y)\mathcal{L}_y^*\delta(y-x)\mathrm{d}y = \mathcal{L}_x^*\left(\int_{\mathbb{R}^d} \pi(y)\varphi(y)\delta(x-y)\mathrm{d}y\right)$$

or

$$\pi(x)\langle\varphi(y), \mathcal{L}_y^*\delta(y-x)\rangle = \mathcal{L}_x^*(\varphi(x)\pi(x))$$

or

$$\pi(x)\langle\mathcal{L}_y\varphi(y), \delta(y-x)\rangle = \mathcal{L}_x^*(\varphi(x)\pi(x))$$

or

$$\pi(x)\mathcal{L}\varphi(x) = \mathcal{L}^*(\varphi(x)\pi(x)) \tag{3}$$

Now using the formulation of $\mathcal{L}, \mathcal{L}^*$, (3) becomes

$$\pi b \cdot \nabla\varphi + \frac{1}{2}\pi a : \nabla^2\varphi = -\nabla\cdot(b\varphi\pi) + \frac{1}{2}\nabla^2 : (a\varphi\pi)$$

or

$$2\pi b\cdot\nabla\varphi = -\varphi\nabla\cdot(\pi b) + \frac{1}{2}\varphi\nabla^2 : (\pi a) + \nabla\varphi\cdot(\nabla\cdot(\pi a)) \tag{4}$$

Since $\pi(x)$ is the invariant distribution, $\mathcal{L}^*\pi(x) = 0$, i.e.,

$$-\nabla\cdot(\pi b) + \frac{1}{2}\nabla^2 : (\pi a) = 0$$

hence (4) becomes

$$2\pi b = \nabla\cdot(\pi a)$$

This is just $(*)$.

1.3 Examples

In some cases the detailed balance is automatically satisfied. If $b(x) = -\nabla V(x)$ and $\sigma(x) = \sqrt{2}$, then the invariant distribution is

$$\pi(x) = e^{-V(x)}$$

In this case the detailed balance becomes

$$-\nabla V(x)e^{-V(x)} = \nabla(e^{-V(x)})$$

2 Langevin dynamics

2.1 Theorem of detailed balance

Consider the second order Langevin dynamics:

$$\begin{aligned} \mathrm{d}q &= v\mathrm{d}t \\ \mathrm{d}v &= f(q)\mathrm{d}t - \gamma v\mathrm{d}t + \sigma\mathrm{d}W \end{aligned}$$

where $q, v \in \mathbb{R}^d$ are the position and velocity of the particle, $f(q)$ is the force, $\gamma > 0$ is the friction constant, $\sigma \in \mathbb{R}^{d\times m}$ is a constant matrix and W is the m-dimensional Wiener process. The detailed balance can also be established for Langevin dynamics.

Theorem 2.1 *Consider the general Langevin dynamics*

$$\begin{aligned} \mathrm{d}q &= v\mathrm{d}t \\ \mathrm{d}v &= f(q)\mathrm{d}t - \gamma v\mathrm{d}t + \sigma\mathrm{d}W \end{aligned}$$

for $q, v \in \mathbb{R}^d$*. Assume* $\pi(q, v)$ *is the invariant distribution with symmetry in* v

$$\pi(q, v) = \pi(q, -v)$$

Let $T_t((q, v), (q', v'))$ *be the transition probability in time* t*, then the detailed balance*

$$\pi(q, v)T_t((q, v), (q', v')) = \pi(q', -v')T_t((q', -v'), (q, -v))$$

holds if and only if

$$-2\gamma\pi v = \sigma\sigma^{\mathrm{T}}\frac{\partial\pi}{\partial v}$$

Note: The detailed balance presented in the Langevin dynamics is closely related to the reversibility in the Hamiltonian dynamics.

Proof: To make the notations more compact, introduce the phase $x = (q, v) \in \mathbb{R}^{2d}$ and its conjugate $x^* = (q, -v)$. Then the symmetry of π becomes

$$\pi(x) = \pi(x^*)$$

We can compactly write the SDE as

$$\mathrm{d}x = b(x)\mathrm{d}t + \Sigma\mathrm{d}W \tag{1}$$

where

$$b(x) = \begin{bmatrix} v \\ f(q) - \gamma v \end{bmatrix} \in \mathbb{R}^{2d}, \qquad \Sigma = \begin{bmatrix} 0 \\ \sigma \end{bmatrix} \in \mathbb{R}^{2d\times m}$$

Define the diffusion matrix

$$A = \Sigma\Sigma^{\mathrm{T}} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma\sigma^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{2d\times 2d}$$

then the Kolmogorov operators $\mathcal{L}, \mathcal{L}^*$ of SDE (1) are

$$\begin{aligned} \mathcal{L}\varphi(x) &= b(x)\cdot\nabla\varphi(x) + \frac{1}{2}A : \nabla^2\varphi(x) \\ \mathcal{L}^*\varphi(x) &= -\nabla\cdot(b(x)\varphi(x)) + \frac{1}{2}\nabla^2 : (A\varphi(x)) \end{aligned} \tag{2}$$

and the Fokker-Planck equation of SDE (1) is

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p(x,t) \tag{3}$$

Now we consider the detailed balance

$$\pi(x)T_t(x,y) = \pi(y^*)T_t(y^*,x^*) \tag{4}$$

Observing that $T_t(x,y)$ can be represented as

$$T_t(x,y) = \exp\Big(t\mathcal{L}_y^*\delta(y-x)\Big)$$

we can equivalently write the detailed balance (4) as

$$\pi(x)\mathcal{L}_y^*\delta(y-x) = \pi(y^*)\mathcal{L}_{x^*}^*\delta(x^*-y^*) \tag{5}$$

Multiply (5) with a test function $\varphi(y)\in C^2(\mathbb{R}^{2d})$ and integrate in $\mathbb{R}^{2d}$:

$$\pi(x)\int_{\mathbb{R}^{2d}}\varphi(y)\mathcal{L}_y^*\delta(y-x)\mathrm{d}y = \mathcal{L}_{x^*}^*\bigg(\int_{\mathbb{R}^{2d}}\varphi(y)\pi(y^*)\delta(x^*-y^*)\mathrm{d}y\bigg)$$

which immediately reduces to

$$\pi(x)\mathcal{L}_x\varphi(x) = \mathcal{L}_{x^*}^*(\varphi(x)\pi(x^*)) \tag{6}$$

Substituting the formulation of $\mathcal{L}, \mathcal{L}^*$, we obtain

$$\begin{aligned} \pi(x)\mathcal{L}_x\varphi(x) &= \pi(x)\Big(b(x)\cdot\nabla\varphi(x) + \frac{1}{2}A : \nabla^2\varphi(x)\Big) \\ &= \pi(x)b(x)\cdot\nabla\varphi(x) + \frac{1}{2}\pi(x)A : \nabla^2\varphi(x) \end{aligned} \tag{7}$$

Let ∇_* be the gradient with respect to x^*, then

$$\mathcal{L}_{x^*}^*(\varphi(x)\pi(x^*)) = -\nabla_*\cdot(\pi(x^*)b(x^*)\varphi(x)) + \frac{1}{2}\nabla_*^2 : (\pi(x^*)A\varphi(x))$$

Using the equations

$$-\nabla_* \cdot (\pi(x^*)b(x^*)\varphi(x)) = -\pi(x^*)b(x^*) \cdot \nabla_*\varphi(x) - \nabla_* \cdot (\pi(x^*)b(x^*))\varphi(x)$$
$$\frac{1}{2}\nabla_*^2 : (\pi(x^*)A\varphi(x)) = \frac{1}{2}\nabla_*^2 : (\pi(x^*)A)\varphi(x) + \frac{1}{2}\pi(x^*)A : \nabla_*^2\varphi(x) + \nabla_* \cdot (A\pi(x^*)) \cdot \nabla_*\varphi(x)$$

We obtain

$$\begin{aligned}\mathcal{L}_{x^*}^*(\varphi(x)\pi(x^*)) &= \Big(-\nabla_* \cdot (\pi(x^*)b(x^*)) + \frac{1}{2}\nabla_*^2 : (\pi(x^*)A)\Big)\varphi(x) \\ &\quad + \Big(-\pi(x^*)b(x^*) + \nabla_* \cdot (A\pi(x^*))\Big) \cdot \nabla_*\varphi(x) \\ &\quad + \frac{1}{2}\pi(x^*)A : \nabla_*^2\varphi(x) \qquad (8)\end{aligned}$$

Terms in (8) correspond to different order of gradients of $\varphi(x)$. Note that $\pi(x)$ is the invariant distribution of (1), we have

$$\mathcal{L}_{x^*}^*\pi(x^*) = 0 \Longrightarrow -\nabla_* \cdot (\pi(x^*)b(x^*)) + \frac{1}{2}\nabla_*^2 : (\pi(x^*)A) = 0$$

hence (8) reduces to

$$\mathcal{L}_{x^*}^*(\varphi(x)\pi(x^*)) = \Big(-\pi(x^*)b(x^*) + \nabla_* \cdot (A\pi(x^*))\Big) \cdot \nabla_*\varphi(x) + \frac{1}{2}\pi(x^*)A : \nabla_*^2\varphi(x) \qquad (9)$$

Combine (6)(7)(9), the detailed balance becomes

$$\begin{aligned}&\pi(x)b(x) \cdot \nabla\varphi(x) + \frac{1}{2}\pi(x)A : \nabla^2\varphi(x) = \\ &\qquad \Big(-\pi(x^*)b(x^*) + \nabla_* \cdot (A\pi(x^*))\Big) \cdot \nabla_*\varphi(x) + \frac{1}{2}\pi(x^*)A : \nabla_*^2\varphi(x) \qquad (10)\end{aligned}$$

Note that in (10) we have $\pi(x) = \pi(x^*)$ and

$$A : \nabla^2\varphi = \begin{bmatrix} 0 & 0 \\ 0 & \sigma\sigma^{\mathrm{T}} \end{bmatrix} : \begin{bmatrix} \varphi_{qq} & \varphi_{qv} \\ \varphi_{vq} & \varphi_{vv} \end{bmatrix} = \sigma\sigma^{\mathrm{T}} : \varphi_{vv} = \begin{bmatrix} \varphi_{qq} & -\varphi_{qv} \\ -\varphi_{vq} & \varphi_{vv} \end{bmatrix} = A : \nabla_*^2\varphi(x)$$

hence

$$\frac{1}{2}\pi(x)A : \nabla^2\varphi(x) = \frac{1}{2}\pi(x^*)A : \nabla_*^2\varphi(x)$$

Using $\pi(x) = \pi(x^*)$ again, (10) reduces to

$$\pi(x)\Big(b(x) \cdot \nabla\varphi(x) + b(x^*) \cdot \nabla_*\varphi(x)\Big) = \nabla_* \cdot (A\pi(x^*)) \cdot \nabla_*\varphi(x) \qquad (11)$$

On the one hand,

$$b(x) = \begin{bmatrix} v \\ f(q) - \gamma v \end{bmatrix}, \nabla\varphi(x) = \begin{bmatrix} \varphi_q \\ \varphi_v \end{bmatrix} \Longrightarrow b(x) \cdot \nabla\varphi(x) = v \cdot \varphi_q + f(q) \cdot \varphi_v - \gamma v \cdot \varphi_v$$

$$b(x^*) = \begin{bmatrix} -v \\ f(q) + \gamma v \end{bmatrix}, \nabla_* \varphi(x) = \begin{bmatrix} \varphi_q \\ -\varphi_v \end{bmatrix} \Longrightarrow b(x^*) \cdot \nabla_* \varphi(x) = -v \cdot \varphi_q - f(q) \cdot \varphi_v - \gamma v \cdot \varphi_v$$

Hence

$$b(x) \cdot \nabla \varphi(x) + b(x^*) \cdot \nabla_* \varphi(x) = -2\gamma v \cdot \varphi_v \tag{12}$$

On the other hand,

$$\begin{aligned} \nabla_* \cdot (A\pi(x^*)) \cdot \nabla_* \varphi(x) &= (A \nabla_* \pi(x)) \cdot \nabla_* \varphi(x) \\ &= \left(\begin{bmatrix} 0 & 0 \\ 0 & \sigma\sigma^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \pi_q \\ -\pi_v \end{bmatrix} \right) \cdot \begin{bmatrix} \varphi_q \\ -\varphi_v \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -\sigma\sigma^{\mathrm{T}} \pi_v \end{bmatrix} \cdot \begin{bmatrix} \varphi_q \\ -\varphi_v \end{bmatrix} \\ &= (\sigma\sigma^{\mathrm{T}} \pi_v) \cdot \varphi_v \end{aligned} \tag{13}$$

Combine (11)(12)(13) and the detailed balance becomes

$$-2\gamma \pi v \cdot \varphi_v = (\sigma\sigma^{\mathrm{T}} \pi_v) \cdot \varphi_v \tag{14}$$

Since φ_v is arbitrary, (14) is just

$$-2\gamma \pi v = \sigma\sigma^{\mathrm{T}} \frac{\partial \pi}{\partial v}$$

Theorem 2.2 *If the detailed balance*

$$\pi(q, v) T_t((q, v), (q', v')) = \pi(q', -v') T_t((q', -v'), (q, -v))$$

holds for some symmetric $\pi(q, v) = \pi(q, -v)$*, then* $\pi(q, v)$ *is the invariant distribution.*

Integrating the detailed balance in $q \in \mathbb{R}^d$, we obtain

$$\int_{\mathbb{R}^{2d}} \pi(q, v) T_t((q, v), (q', v')) \mathrm{d}q \mathrm{d}v = \pi(q', -v') = \pi(q', v')$$

hence $\pi(q, v)$ is the invariant distribution.

2.2 Applications in thermal equilibrium

Let $U(q)$ be the potential function and assume the mass of the particle is 1. Consider the Langevin dynamics

$$\begin{aligned} \mathrm{d}q &= v \mathrm{d}t \\ \mathrm{d}v &= -\nabla U(q) \mathrm{d}t - \gamma v \mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}} \mathrm{d}W \end{aligned}$$

and the invariant distribution is

$$\pi(q,v) = \exp\left(-\beta\left(\frac{|v|^2}{2} + U(q)\right)\right)$$

It's easy to verify

$$\sigma\sigma^{\mathrm{T}}\pi_v = \frac{2\gamma}{\beta}\pi_v = -2\gamma\pi v$$

hence the Langevin dynamics of a classical system satisfies the detailed balance.

Now consider the pmmLang:

$$\begin{aligned} \mathrm{d}\boldsymbol{q} &= \boldsymbol{v}\mathrm{d}t \\ \mathrm{d}\boldsymbol{v} &= -\boldsymbol{q}\mathrm{d}t - (L^\alpha)^{-1}\nabla U^\alpha(\boldsymbol{q})\mathrm{d}t - \gamma\boldsymbol{v}\mathrm{d}t + \sqrt{\frac{2\gamma(L^\alpha)^{-1}}{\beta_N}}\mathrm{d}\boldsymbol{W} \end{aligned}$$

The invariant distribution is

$$\pi(\boldsymbol{q},\boldsymbol{v}) = \exp\left(-\beta_N\left(\frac{1}{2}\boldsymbol{v}^{\mathrm{T}}L^\alpha\boldsymbol{v} + U(\boldsymbol{q})\right)\right)$$

It's easy to verify the detailed balance condition.

3 Potential splitting method

3.1 Methodology

Suppose we want to sample the Boltzmann distribution

$$\pi(q,v) = \exp\left(-\beta\left(\frac{|v|^2}{2} + U(q)\right)\right)$$

for some potential function $U(q)$. Decompose

$$U(q) = U_1(q) + U_2(q)$$

and consider the following update scheme in timestep Δt:

Algorithm 1: Potential splitting method

1. Given the state (q, v), evolve the following SDE

$$\begin{aligned} \mathrm{d}q &= v\mathrm{d}t \\ \mathrm{d}v &= -\nabla U_1(q)\mathrm{d}t - \gamma v\mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}}\mathrm{d}W \end{aligned}$$

   with initial value (q, v) in time Δt to obtain (q^*, v^*).

2. Set $(q, v) = (q^*, v^*)$ with probability

$$a(q, q^*) = \min\{1, e^{-\beta(U_2(q^*)-U_2(q))}\}$$

   otherwise $(q, v) = (q, -v)$.

Theorem 3.1 *Let $T((q, v),(q', v'))$ be the transition probability of Algorithm 1, then the detailed balance holds:*

$$\pi(q, v)T((q, v),(q', v')) = \pi(q', -v')T((q', -v'),(q, -v))$$

A direct consequence is that

$$\int_{\mathbb{R}^{2d}} \pi(q, v)T((q, v),(q', v'))\mathrm{d}q\mathrm{d}v = \pi(q', v')$$

i.e., the scheme preserves $\pi(q, v)$.

Proof Define the distributions

$$\pi_1(q, v) = \exp\Big(-\beta\Big(\frac{|v|^2}{2} + U_1(q)\Big)\Big), \qquad \pi_2(q) = \exp\Big(-\beta U_2(q)\Big)$$

Let $T_1((q, v),(q', v'))$ be the transition probability of Step 1, and $a(q, q')$ be the acceptance probability in Step 2, then

$$\pi_1(q, v)T_1((q, v),(q', v')) = \pi_1(q', -v')T_1((q', -v'),(q, -v)) \tag{1}$$

$$\pi_2(q)a(q, q') = \pi_2(q')a(q', q) \tag{2}$$

Notice that $\pi(q, v) = \pi_1(q, v)\pi_2(q)$, and the transition probability of Algorithm 1 is

$$T((q, v),(q', v')) = T_1((q, v),(q', v'))a(q, q') + \delta(q' - q)\delta(v' + v)(1 - A(q, v)) \tag{3}$$

where $A(q,v)$ is the total acceptance probability

$$A(q,v) = \int T_1((q,v),(q',v'))a(q,q')\mathrm{d}q'\mathrm{d}v'$$

To verify the detailed balance, we need

$$\pi(q,v)T((q,v),(q',v')) = \pi(q',-v')T((q',-v'),(q,-v))$$

or

$$\pi(q,v)T_1((q,v),(q',v'))a(q,q') = \pi(q',-v')T_1((q',-v'),(q,-v))a(q',q) \tag{4}$$

$$\pi(q,v)\delta(q'-q)\delta(v'+v)(1-A(q,v)) = \pi(q',-v')\delta(q'-q)\delta(v'+v)(1-A(q',-v')) \tag{5}$$

It's easy to see (4) is the product of (1)(2), and (5) holds from the Dirac function. In conclusion, the detailed balance holds.

3.2 Application in pmmLang

For pmmLang, assume the effective potential $U(\boldsymbol{q})$ is decomposed into

$$U(\boldsymbol{q}) = U_1(\boldsymbol{q}) + U_2(\boldsymbol{q})$$

and define $U_1^\alpha(\boldsymbol{q}) = U(\boldsymbol{q}) - \frac{\alpha}{2}|\boldsymbol{q}|^2$. Similarly we have

Algorithm 2:

1. Given the initial state $(\boldsymbol{q},\boldsymbol{v})$, evolve the following SDE in Δt time

$$\mathrm{d}\boldsymbol{q} = \boldsymbol{v}\mathrm{d}t$$

$$\mathrm{d}\boldsymbol{v} = -\boldsymbol{q}\mathrm{d}t - (L^\alpha)^{-1}\nabla U_1^\alpha(\boldsymbol{q})\mathrm{d}t - \gamma\boldsymbol{v}\mathrm{d}t + \sqrt{\frac{2\gamma(L^\alpha)^{-1}}{\beta_N}}\mathrm{d}\boldsymbol{W}$$

   to obtain $(\boldsymbol{q}^*,\boldsymbol{v}^*)$.

2. Set $(\boldsymbol{q},\boldsymbol{v}) = (\boldsymbol{q}^*,\boldsymbol{v}^*)$ with probability

$$a(\boldsymbol{q},\boldsymbol{q}^*) = \min\{1, e^{-\beta(U_2(\boldsymbol{q}^*)-U_2(\boldsymbol{q}))}\}$$

   otherwise $(\boldsymbol{q},\boldsymbol{v}) = (\boldsymbol{q},-\boldsymbol{v})$.

Theorem 3.2 *Let $T((\boldsymbol{q},\boldsymbol{v}),(\boldsymbol{q}',\boldsymbol{v}'))$ be the transition probability of the scheme, then the detailed balance holds:*

$$\pi(\boldsymbol{q},\boldsymbol{v})T((\boldsymbol{q},\boldsymbol{v}),(\boldsymbol{q}',\boldsymbol{v}')) = \pi(\boldsymbol{q}',-\boldsymbol{v}')T((\boldsymbol{q}',-\boldsymbol{v}'),(\boldsymbol{q},-\boldsymbol{v}))$$

A direct consequence is that

$$\int_{\mathbb{R}^{2dN}} \pi(\boldsymbol{q},\boldsymbol{v})T((\boldsymbol{q},\boldsymbol{v}),(\boldsymbol{q}',\boldsymbol{v}'))\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{v} = \pi(\boldsymbol{q}',\boldsymbol{v}')$$

i.e., the update scheme preserves $\pi(\boldsymbol{q},\boldsymbol{v})$.