On score-based generative model: a thermodynamic perspective

Shi Feng

January 15, 2023

In Song (2021) [1] the authors proposed an ingenious trick by which the stochastic diffusion process can be effectively replaced by a deterministic drift process, thus enable the retrieve of the original data distribution from random noise. In this draft I would like to show that the spirit behind Song (2021) is precisely a simulation of quasi-static equilibrium process free of entropy production, and can be perceived simply as an equilibrated Ohm's law.

1 Thermodynamic equilibrium inside a conductor

The current of particles, say electrons, reads

$$J(x,t) = u(x,t)\rho(x,t) \tag{1}$$

where u is the drift velocity of electrons induced by an electric field $u = \mu E$, with μ the mobility of electrons, and ρ the local density of electrons. Indeed this is equivalent to Ohm's law of the vector form $J = \sigma E$. WLOG, I assumed a 1D system so that all variables are scalers.

However, this is only true inside a conductor of infinite size, or a circular loop, when the particles are free to move into certain direction without ending up a static equilibrium state. Eq.(1) needs some dynamical revision to account for the finial equilibrium state. Figure.1 is an illustration of the

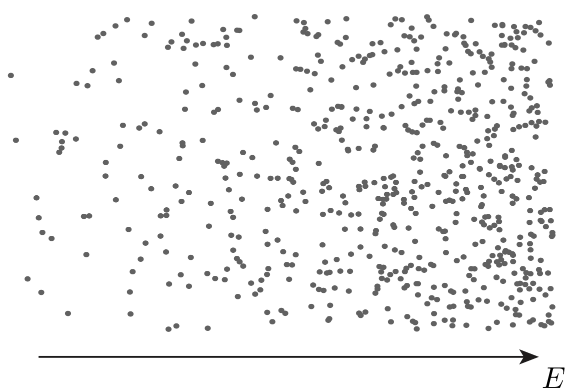

Figure 1: Illustration of the finial density profile in a finite system subjected to an external field.

electron density distribution inside a finite conductor that is subjected to an external electric field. Electron density ρ stops its drift evolution once such thermodynamic equilibrium is established; and from this moment forth, the thermal diffusion completely counteracts the drift velocity so the

ρ remains static. This process can be described by a diffusive velocity opposite to the drift velocity, such that the dynamical current becomes

$$J(x,t) = u(x,t)\rho(x,t) - D\partial_x \rho(x,t) \tag{2}$$

where D is a phenomenological diffusion coefficient that is ultimately related to the Gaussian variance of Brownian motion. At the moment t_e when equilibrum is established, ρ stops evolving, hence $J(x,t_e)$ must vanish. This means the drift contribution must cancel the diffusive contribution:

$$u(x,t_e)\rho(x,t_e) = D\partial_x \rho(x,t_e) \tag{3}$$

so that

$$u = D\frac{\partial_x \rho}{\rho} = D\,\partial_x(\log\rho) \tag{4}$$

This is exactly the same as in Song (2021), where the drift velocity is

$$u(X,t) = -\frac{\sigma^2}{2}\left(\partial_X \log P_t(X)\right) \tag{5}$$

with the phenomenological diffusion coefficient $D = \sigma^2/2$; and the negative sign is because in Song (2021) diffusion is set as the positive direction.

Therefore, it is clear that the method proposed by Song *et al* is a simulation of quasi-static equilibrium, reversible process with zero-entropy production. It can be perceived as starting from the distribution profile shown in Fig.1 which is held there by some fictitious field, and quasi-statically weakening the field so that the density profile diffuses quasi-statically while retaining the equilibrium.

2 From Stochastic Differential Equation[1]

The Fokker–Planck equation (FPE) plays a role in stochastic systems analogous to that of the Liouville equation in deterministic mechanical systems. Namely, the FPE describes in a statistical sense how a collection of initial data evolves in time. Just as Liouville equation can be inferred from the statistical ensumble of macroscopically large number of microscopic particles, FPE can be inferred from the microscopic stochastic differential equation (SDE). The generic SDE reads

$$X_{t+\Delta t} = X_t + \mu(X_t,t)\Delta t + \sigma(X_t,t)\sqrt{\Delta t}\,e_t \tag{6}$$

where X_t is a random variable that denotes the position of a Brownian particle, μ and σ correspond to the drift and diffusive dynamics respectively, and e_t is a stochastic Gaussian force that drives the diffusive process. We are interested in a deterministic description of this random process. We start with the simplest case, i.e. in absence of the drift process. The SDE then reads

$$X_{t+\Delta t} = X_t + \sigma\sqrt{\Delta t}\,e_t \tag{7}$$

Formally we can define an ancillary test function $h(X)$ which would assist us in deriving the probability evolution $X_t \sim P_t(X)$. The expectation of $h(X_t)$ evolves according to

$$\begin{aligned} E[h(X_{t+\Delta t})] &= E[h(X_t + \sigma\sqrt{\Delta t}\,e_t)] \\ &= E\left[h(X_t) + h'(X_t)\sigma\sqrt{\Delta t}\,e_t + \frac{1}{2}h''(X_t)\sigma^2\Delta t\,e_t^2\right] \\ &= E[h(X_t)] + \frac{1}{2}\sigma^2\Delta t\,E[h''(X_t)] \end{aligned} \tag{8}$$

[1]This section is an annotation of Prof. Y.N. Wu's online lecture on Jan.21 2022, UCLA

where we used $E(e_t) = 0$ and $E(e_t^2) = Var(e_t) = 1$. Hence we have

$$E[h(X_{t+\Delta t})] - E[h(X_t)] = \frac{1}{2}\sigma^2 \Delta t \; E[h''(X_t)] \tag{9}$$

On the other hand, the above equation can be written in probablistic language as

$$\begin{aligned} \int h(X) P_{t+\Delta t}(X) dX - \int h(x) P_t(X) dX &= \frac{1}{2}\sigma^2 \Delta t \int h''(X) P_t(X) dX \\ &= \frac{1}{2}\sigma^2 \Delta t \; \int h(X) \partial_X^2 P_t(X) dX \end{aligned} \tag{10}$$

Noting that the LHS is actually

$$L.H.S. = \int h(X) \frac{\partial P_t(X)}{\partial t} dX \tag{11}$$

we have

$$\boxed{\frac{\partial P_t(X)}{\partial t} = \frac{\sigma^2}{2} \partial_X^2 P_t(X)} \tag{12}$$

which is the diffusion process that's responsible for the local heat flow.

Now we turn to the other extreme, where the dynamics is due only to the drift process. The SDE reads

$$X_{t+\Delta t} = X_t + \mu(X_t, t)\Delta t \tag{13}$$

By the same token we have

$$E[h(X_{t+\Delta t})] = E[h(X_t + \mu(X_t, t)\Delta t)] = E[h(X_t)] + E[h'(X_t)\mu(X_t, t)\Delta t] \tag{14}$$

This is equivalent to

$$\int h(X) \frac{\partial P_t(X)}{\partial t} dX = \int h'(X) \mu(X, t) P_t(X) dX = -\int h(X) d[\mu(X, t) P_t(X)] \tag{15}$$

so we have

$$\boxed{\frac{\partial P_t(X)}{\partial t} = -\frac{\partial}{\partial X}[\mu(X, t) P_t(X)]} \tag{16}$$

Now we've derived the Fokker-Planck processes driving by drift and diffusion. By equating the two we can derive the condition by which the diffusion can be effectively described by drift:

$$-\frac{\partial}{\partial X}\left(\mu(X, t) P_t(X)\right) = \frac{\sigma^2}{2} \frac{\partial^2}{\partial X^2} P_t(X) \tag{17}$$

$$-\mu(X, t) P_t(X) = \frac{\sigma^2}{2} \partial_X P_t(X) = \frac{\sigma^2}{2} \left(\frac{\partial}{\partial X} \log P_t(X)\right) P_t(X) \tag{18}$$

$$\boxed{\mu(X, t) = -\frac{\sigma^2}{2} \left(\frac{\partial}{\partial X} \log P_t(X)\right)} \tag{19}$$

that is, diffusion can be effectively captured by a drift process, if the drift velocity satisfies the above equation.

References

[1] Song, Y. *et al.* Score-based generative modeling through stochastic differential equations (2021). URL https://openreview.net/forum?id=PxTIG12RRHS.