

# 1 Lyapunov Function

**Notation 1.** Overall notations in this section are:

- $\mathcal{M}$  a manifold, and  $\mu$  its measure, e.g.  $\mu(x) = \sqrt{g(x)}$  if  $\mathcal{M}$  is Riemannian with metric  $g_{ab}$ ;
- if  $p(x)$  the distribution of random variable  $X$ , then

$$\langle f \rangle_p = \langle f \rangle_X := \int_{\mathcal{M}} d\mu(x) p(x) f(x);$$

- if  $D$  is a set of samples, then

$$\langle f \rangle_D := \frac{1}{|D|} \sum_{x \in D} f(x);$$

- let  $\mathcal{N}(\mu, \Sigma)$  denotes normal distribution with mean  $\mu$  and covariance  $\Sigma$ ;
- given function  $g$ , let  $f\{g\}$ , or  $f_{\{g\}}$ , denote a function constructed out of  $g$ , that is,

$$f\{\cdot\}: (\mathcal{M} \rightarrow A) \rightarrow (\mathcal{M} \rightarrow B);$$

- for conditional maps  $f$ , let  $f(x|y)$  denotes the map of  $x$  with  $y$  given and fixed, and  $f(x; y)$  denotes the map of  $x$  with  $y$  given but mutable;
- r.v. is short for random variable, i.i.d. for independent identically distributed, s.t. for such that, and a.e. for almost every.

## 1 Relaxation

Next, we illustrate how, during a non-equilibrium process, a distribution  $p$  relaxes to its stationary distribution  $q$ , and how this process relates to the variational inference. Further, we try to find the most generic dynamics that underlies the non-equilibrium to equilibrium process, on both macroscopic (distribution) and microscopic (“particle”) viewpoints.

First, we shall define what relaxation is, via free energy.

**Definition 2.** [Free Energy]

Let  $E(x): \mathcal{M} \rightarrow \mathbb{R}$ . Define stationary distribution

$$q_E(x) := \frac{\exp(-E(x)/T)}{Z},$$

where  $T > 0$  and  $Z_E := \int_{\mathcal{M}} d\mu(x) \exp(-E(x)/T)$ . Given  $E$ , for any time-dependent distribution  $p(x, t)$ , define free energy as

$$F_E[p(\cdot, t)] := TD_{\text{KL}}(p \| q_E) - T \ln Z_E = T \int_{\mathcal{M}} d\mu(x) p(x, t) \ln \frac{p(x, t)}{q_E(x)} - T \ln Z_E.$$

Or, equivalently,

$$F_E[p(\cdot, t)] := \langle E \rangle_{p(\cdot, t)} - TH[p(\cdot, t)],$$

where entropy functional  $H[p(\cdot, t)] := \langle -\ln p(\cdot, t) \rangle_p$ .

**Definition 3.** [Relaxation]

For a time-dependent distribution  $p(x, t)$  on  $\mathcal{M}$ , we say  $p$  relaxes to  $q_E$  if and only if the free energy  $F_E[p(\cdot, t)]$  monotonically decreases to its minimum, where  $p(\cdot, t) = q_E$ .

We can visualize this relaxation process by an imaginary ensemble of juggling “particles” (or “bees”). Initially, they are arbitrarily positioned. This forms a distribution of “particles”  $p$ . With some underlying dynamics, these “particles” moves and finally the distribution relaxes, if it can, to a stationary distribution  $q_E$ . Apparently, the underlying dynamics and the  $E$  are correlated. We first provide a way of peeping the underlying dynamics, that is, the “flux”.

**Lemma 4.** [Conservation of “Mass”]

For any time-dependent distribution  $p(x, t)$ , there exists a “flux”  $f^a\{p\}(x, t)$  s.t.

$$\frac{\partial p}{\partial t}(x, t) + \nabla_a(f^a\{p\}(x, t) p(x, t)) = 0.$$

What is the dynamics of  $p$  by which any initial  $p$  will finally relax to  $q_E$ ? That is, what is the sufficient (and essential) condition of relaxing to  $q_E$  for any  $p$ ? Because of the conservation of “mass”, the dynamics of  $p$ , i.e.  $\partial p / \partial t$ , is determined by a “flux”,  $f^a$ . Thus, this sufficient (and essential) condition must be about the  $f^a$ .

**Lemma 5.** *Given  $p$  and  $(x, t)$ , for any  $f^a\{p\}(x, t)$ , we can always construct a  $K^{ab}\{p\}(x, t)$  s.t.*

$$f^a\{p\}(x, t) = -K^{ab}\{p\}(x, t) \nabla_b \{T \ln p(x, t) + E(x)\}.$$

**Proof.** For any vector  $f^a$  and  $v_a$ , we can always construct a tensor  $K^{ab}$  s.t.  $f^a = K^{ab} v_b$ . Indeed, we can rotate  $v_a$  to the direction of  $f^a$  and then dimension-wise recalc to  $f^a$ . This rotation and dimension-wise rescaling compose the linear transform  $K^{ab}$ . Now, letting

$$v_a = -\nabla_a \{T \ln p(x, t) + E(x)\},$$

we arrive at the conclusion.  $\square$

Now, we claim a sufficient condition of relaxing to  $q_E$  for any  $p$ .

**Theorem 6.** *[Fokker-Planck Equation]*

*If, for any  $p$  and  $t$ , the symmetric part of  $K^{ab}\{p\}(x, t)$  is a.e. positive definite on  $\mathcal{M}$ , then any  $p$  evolves by this “flux” will relax to  $q_E$ .*

**Proof.** Directly

$$\begin{aligned} \frac{dF_E}{dt}[p(\cdot, t)] &= T \int_{\mathcal{M}} d\mu(x) \frac{\partial p}{\partial t}(x, t) \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \\ \{\text{Conservation of mass}\} &= -T \int_{\mathcal{M}} d\mu(x) \nabla_a [f^a\{p\}(x, t) p(x, t)] \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right]. \end{aligned}$$

Since

$$\nabla_a [f^a\{p\}(x, t) p(x, t)] \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] = \nabla_a \left\{ [f^a\{p\}(x, t) p(x, t)] \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \right\} - [f^a\{p\}(x, t) p(x, t)] \nabla_a \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right],$$

we have

$$\begin{aligned} \frac{dF_E}{dt}[p(\cdot, t)] &= -T \int_{\mathcal{M}} d\mu(x) \nabla_a [f^a\{p\}(x, t) p(x, t)] \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \\ &= -T \int_{\mathcal{M}} d\mu(x) \nabla_a \left\{ [f^a\{p\}(x, t) p(x, t)] \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \right\} \\ &\quad + T \int_{\mathcal{M}} d\mu(x) [f^a\{p\}(x, t) p(x, t)] \nabla_a \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \\ [\text{Divergence theorem}] &= -T \int_{\partial \mathcal{M}} dS_a p(x, t) f^a\{p\}(x, t) \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \\ &\quad + T \int_{\mathcal{M}} d\mu(x) p(x, t) f^a\{p\}(x, t) \nabla_a \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \end{aligned}$$

The first term vanishes.<sup>1</sup> Then, direct calculus shows

$$\begin{aligned} \frac{dF_E}{dt}[p(\cdot, t)] &= T \int_{\mathcal{M}} d\mu(x) p(x, t) f^a\{p\}(x, t) \nabla_a \left[ \ln \frac{p(x, t)}{q(x)} + 1 \right] \\ &= T \int_{\mathcal{M}} d\mu(x) p(x, t) f^a\{p\}(x, t) [\nabla_a \ln p(x, t) - \nabla_a \ln q(x)] \\ \{q(x) := \dots\} &= \int_{\mathcal{M}} d\mu(x) p(x, t) f^a\{p\}(x, t) [T \nabla_a \ln p(x, t) + \nabla_a E(x)] \\ &= \int_{\mathcal{M}} d\mu(x) p(x, t) f^a\{p\}(x, t) \nabla_a \{T \ln p(x, t) + E(x)\}. \end{aligned}$$

By the previous lemma, we have

$$\begin{aligned} \frac{dF_E}{dt}[p(\cdot, t)] &= \int_{\mathcal{M}} d\mu(x) p(x, t) f^a\{p\}(x, t) \nabla_a \{T \ln p(x, t) + E(x)\} \\ \{f^a = \dots\} &= - \int_{\mathcal{M}} d\mu(x) p(x, t) K^{ab}\{p\}(x, t) \nabla_a \{T \ln p(x, t) + E(x)\} \nabla_b \{T \ln p(x, t) + E(x)\}. \end{aligned}$$

Letting  $S^{ab} := (K^{ab} + K^{ba})/2$  and  $A^{ab} := (K^{ab} - K^{ba})/2$ , we have  $K^{ab} = S^{ab} + A^{ab}$ , where  $S^{ab}$  is symmetric and  $A^{ab}$  anti-symmetric. Then,

$$\begin{aligned} \frac{dF_E}{dt}[p(\cdot, t)] &= - \int_{\mathcal{M}} d\mu(x) p(x, t) [S^{ab}\{p\}(x, t) + A^{ab}\{p\}(x, t)] \nabla_a \{T \ln p(x, t) + E(x)\} \nabla_b \{T \ln p(x, t) + E(x)\} \\ \{A^{ab} = A^{ba}\} &= - \int_{\mathcal{M}} d\mu(x) p(x, t) S^{ab}\{p\}(x, t) \nabla_a \{T \ln p(x, t) + E(x)\} \nabla_b \{T \ln p(x, t) + E(x)\}. \end{aligned}$$

<sup>1</sup>. To-do: Explain the reason explicitly.

The condition claims that  $S^{ab}\{p\}(x, t)$  is positive definite for any  $p$  and  $(x, t)$ . Then, the integrand is a positive definite quadratic form, being positive if and only if  $\nabla_a\{T \ln p(x, t) + E(x)\} \neq 0$ . Then, we find  $(dF_E/dt)[p(\cdot, t)] < 0$  as long as  $\nabla_a\{T \ln p(x, t) + E(x)\} \neq 0$  at some  $x$ , i.e.  $p \neq q$ , and  $(dF_E/dt)[p(\cdot, t)] = 0$  if and only if  $\nabla_a\{T \ln p(x, t) + E(x)\} = 0$  for  $\forall x$ , i.e.  $p = q$ . Thus proof ends.  $\square$

**Remark 7.** [Sufficient but Not Essential]

However, this is not an essential condition of relaxing to  $q_E$  for any  $p$ . Indeed, we proved the integrand of  $(dF_E/dt)[p(\cdot, t)]$  is negative everywhere, which implies the integral, i.e.  $(dF_E/dt)[p(\cdot, t)]$ , is negative. But, we cannot exclude the case where the integrand is not negative everywhere, whereas the integral is still negative. During the proof, this is the only place that leads to the non-essential-ness, which is hard to overcome.

As the dynamics of distribution is a macroscopic viewpoint, the microscopic viewpoint, i.e. the stochastic dynamics of single “particle”<sup>2</sup>, is as follow.

**Theorem 8.** [Stochastic Dynamics]

If  $K^{ab}$  is symmetric, independent of  $p$  and almost everywhere smooth on  $\mathcal{M}^3$ , then Fokker-Planck equation is equivalent to the stochastic dynamics

$$dx^a = [T \nabla_b K^{ab}(x, t) - K^{ab}(x, t) \nabla_b E(x)] dt + \sqrt{2T} dW^a(x, t),$$

where

$$dW \sim \mathcal{N}(0, K(x, t) dt).$$

**Proof.** By the lemma 22, we find

$$\mu^a(x, t) = T \nabla_b K^{ab}(x, t) - K^{ab}(x, t) \nabla_b E(x)$$

and

$$\Sigma^{ab}(x, t) = 2TK^{ab}(x, t).$$

Then, directly,

$$\begin{aligned} \frac{\partial p}{\partial t}(x, t) &= -\nabla_a\{p(x, t) \mu^a(x, t)\} + \frac{1}{2} \nabla_a \nabla_b (p(x, t) \Sigma^{ab}(x, t)) \\ &= \nabla_a\{p(x, t) [K^{ab}(x, t) \nabla_b E(x) - T \nabla_b K^{ab}(x, t)]\} + \nabla_a \nabla_b \{Tp(x, t) K^{ab}(x, t)\} \\ \{\text{Expand}\} &= \nabla_a\{K^{ab}(x, t) \nabla_b E(x) p(x, t)\} - \nabla_a\{T \nabla_b K^{ab}(x, t) p(x, t)\} \\ &\quad + \nabla_a\{TK^{ab}(x, t) \nabla_b p(x, t)\} + \nabla_a\{T \nabla_b K^{ab}(x, t) p(x, t)\} \\ &= \nabla_a\{K^{ab}(x, t) \nabla_b E(x) p(x, t)\} + \nabla_a\{TK^{ab}(x, t) \nabla_b p(x, t)\}, \end{aligned}$$

which is just the Fokker-Planck equation. Indeed, the Fokker-Planck equation 6 is

$$\begin{aligned} \frac{\partial p}{\partial t}(x, t) &= -\nabla_a(f^a\{p\}(x, t) p(x, t)) \\ \{f^a = \dots\} &= \nabla_a(K^{ab}\{p\}(x, t) \nabla_b\{T \ln p(x, t) + E(x)\} p(x, t)) \\ \{K^{ab} \text{ independent of } p\} &= \nabla_a(K^{ab}(x, t) \nabla_b\{T \ln p(x, t) + E(x)\} p(x, t)) \\ \{\text{Expand}\} &= \nabla_a\{K^{ab}(x, t) \nabla_b E(x) p(x, t)\} + \nabla_a\{TK^{ab}(x, t) \nabla_b p(x, t)\}, \end{aligned}$$

exactly the same. Thus proof ends.  $\square$

**Question 1.** Given a stochastic dynamics, how can we determine if there exists the  $E$ , or the stationary distribution  $q_E$ ?

**Question 2.** Further, if it exists, then how can we reveal it? Precisely, in the case  $T \rightarrow 0$ , given  $(dx^a/dt) = h^a(x, t)$ , how can we reconstruct the  $E$  and find a positive definite  $K^{ab}$ , s.t.  $h^a(x) = K^{ab}(x, t) \nabla_b E(x)$ ?

The scalar function  $E$  is called *Lyapunov function*. Given an autonomous dynamical system<sup>4</sup>, along the phase trajectory, Lyapunov function will monotonically decrease. So, it reflects the stability of the dynamical system. Next, we investigate, for any autonomous dynamical system, if there's a Lyapunov function, s.t.

$$\frac{dx^a}{dt} = f^a(x) = K^{ab}(x) \nabla_b E(x),$$

Notice that, this is an autonomous stocastical dynamics with  $T \rightarrow 0$ .

2. For the conception of stochastic dynamics, c.f. B.2.

3. **TODO: Check this.**

4. That is, ordinary differential equations that do not explicitly depend on time. The word autonomous means independent of time.

**Definition 9.** *[Phase Tape]*

Given an autonomous dynamical system  $dx^a/dt = f^a(x)$  on  $\mathcal{M}$  and a hypersurface  $\mathcal{S}$  with  $\dim(\mathcal{S}) < \dim(\mathcal{M})$ . A phase tape  $\text{PT}(\mathcal{S}, t) := \{x^a(s) : s \in [0, t], x^a(0) \in \mathcal{S}\}$ .

**Theorem 10.** *[Existence of Lyapunov Function]*

If there's no limit cycle, then there exists a Lyapunov function, constructed as follow. First

## 2 Ambient & Latent Variables

In the real world, there can be two types of variables: ambient and latent. The ambient variables are those observed directly, like sensory inputs or experimental observations. While the latent are usually more simple and basic aspects, like wave-function in QM.

We formulate the  $E$  as a function of  $(v, h) \in \mathcal{V} \times \mathcal{H}$ , where  $v$ , for visible, represents the ambient and  $h$ , for hidden, represents the latent. Then, we extend the free energy to

**Definition 11.** *[Conditional Free Energy]*

Given  $v$ , if define

$$Z_E(v) := \int_{\mathcal{H}} d\mu(h) \exp(-E(v, h)/T),$$

then we have a conditional free energy of distribution  $p(h)$  defined as

$$F_E[p|v] := TD_{\text{KL}}(p||q_E(\cdot|v)) - T \ln Z_E(v).$$

Directly, we have

**Lemma 12.**

$$q_E(h|v) = \frac{\exp(-E(v, h)/T)}{\int_{\mathcal{H}} d\mu(h) \exp(-E(v, h)/T)},$$

which is simply the  $q_E$  with the  $v$  in the  $E(v, h)$  fixed.

Thus,

**Theorem 13.**

$$F_E[p|v] = \langle E(v, \cdot) \rangle_p - TH[p].$$

## 3 Minimize Free Energy Principle

If  $E$  is in a function family parameterized by  $\theta \in \mathbb{R}^N$ , denoted as  $E(x; \theta)$ , then we want to find the most generic distribution  $q_E$  in the function family of  $E$  s.t. the expectation  $\langle E(\cdot; \theta) \rangle_{q_E(\cdot; \theta)}$  is minimized. For instance, given ambient  $v$ , we want to locate  $v$  on the minimum of  $E$ , that is  $\langle E(v, \cdot; \theta) \rangle_{q_E(\cdot|v; \theta)}$  (c.f. lemma 12).

On one hand, we want to minimize  $\langle E(\cdot; \theta) \rangle_{q_E(\cdot; \theta)}$ ; on the other hand, we shall keep the minimal prior knowledge on  $q_E(\cdot; \theta)$ , that is, maximize  $H[q_E(\cdot; \theta)]$ . So, we find the  $\theta$  that minimizes  $\langle E(\cdot; \theta) \rangle_{q_E(\cdot; \theta)} - TH[q_E(\cdot; \theta)]$ , where the positive constant  $T$  balances the two aspects. This happens to be the free energy.

Next, we propose an algorithm that establishes the free energy minimization. First, notice the relation

**Lemma 14.**

$$\frac{\partial}{\partial \theta^\alpha} \{-T \ln Z_E(\cdot; \theta)\} = \left\langle \frac{\partial E}{\partial \theta^\alpha}(\cdot; \theta) \right\rangle_{q_E(\cdot; \theta)}.$$

So, we have an EM-like algorithm, as

**Algorithm 15.** *[Recall and Learn (RL)]*

To minimize free energy  $F_E[p]$ , we have two steps:

1. minimize  $\langle E(\cdot; \theta) \rangle_p - TH[p]$  by the stochastic dynamics until relaxation, where  $p = q_E(\cdot; \theta)$ ; then

2. minimize  $-T \ln Z_{E(\cdot; \theta)}$  by gradient descent and replacing  $\left\langle \frac{\partial E}{\partial \theta^\alpha}(\cdot; \theta) \right\rangle_{q_{E(\cdot; \theta)}} \rightarrow \left\langle \frac{\partial E}{\partial \theta^\alpha}(\cdot; \theta) \right\rangle_p$ .

By repeating these two steps, we get smaller and smaller free energy.

For instance, in a brain, the first step can be illustrated as recalling, and the second as learning (searching for a more proper memory, or code of information). So we call this algorithm *recall and learn*.

During the optimization, the first term minimizes the expectation of  $E(\cdot; \theta)$ , while the second term smoothes  $E(\cdot; \theta)$ . Since the  $q_E(\cdot; \theta)$  is invariant for  $E(x; \theta) \rightarrow E(x; \theta) + \text{Const}$ , we shall eliminate this symmetry by re-defining

$$E(x; \theta) \rightarrow E(x; \theta) - E(x_\star; \theta),$$

for any  $x_\star \in \mathcal{M}$  given.

## 4 Example: Continuous Hopfield Network

Here, we provide a biological inspired example, for illustrating both the stochastic dynamics 8 and the RL algorithm 15.

**Definition 16.** [Continuous Hopfield Network]

Let  $U^{\alpha\beta}$  and  $I^\alpha$  constants, and  $L_v(v)$  and  $L_h(h)$  scalar functions. Define  $f_\alpha := \partial_\alpha L_h$ ,  $g_\alpha := \partial_\alpha L_v$ . Then the dynamics of continuous Hopfield network is defined as<sup>5</sup>

$$\begin{aligned} \frac{dv^\alpha}{dt} &= U^{\alpha\beta} f_\beta(h) - v^\alpha + I^\alpha; \\ \frac{dh^\alpha}{dt} &= U^{\beta\alpha} g_\beta(v) - h^\alpha, \end{aligned}$$

where  $U$  describes the strength of connection between neurons, and  $f, g$  the activation functions of latent and ambient, respectively. Further, we have the  $E$  constructed as

$$E(v, h) = [(v^\alpha - I^\alpha) g_\alpha(v) - L_v(v)] + [h^\alpha f_\alpha(h) - L_h(h)] - U_{\alpha\beta} g^\alpha(v) f^\beta(h).$$

Next, we convert this deterministic dynamics to its stochastic version.

**Theorem 17.** If  $f = \partial L_h$  and  $g = \partial L_v$  are linear functions, and the Hessian matrix of  $L_v$  and  $L_h$  are positive definite, then the stochastic dynamics of the continuous Hopfield network is

$$\begin{aligned} \frac{dv^\alpha}{dt} &= K_v^{\alpha\beta} [U_{\beta\gamma} f^\gamma(h) - v_\beta + I_\beta] + \sqrt{2T} dW_v^\alpha; \\ \frac{dh^\alpha}{dt} &= K_h^{\alpha\beta} [U_{\gamma\beta} g^\gamma(v) - h_\beta] + \sqrt{2T} dW_h^\alpha, \end{aligned}$$

where  $K_v := [\partial^2 L_v(v)]^{-1}$  and  $K_h := [\partial^2 L_h(h)]^{-1}$  are constant matrices.<sup>6</sup>

**Proof.** Directly, we have

$$\begin{aligned} \frac{\partial E}{\partial v^\alpha}(v, h) &= g_\alpha(v) + (v^\beta - I^\beta) \frac{\partial g_\beta(v)}{\partial v^\alpha} - \frac{\partial L_v}{\partial v^\alpha}(v) - U^{\beta\gamma} f_\gamma(h) \frac{\partial g_\beta(v)}{\partial v^\alpha} \\ \left\{ g_\alpha = \frac{\partial L_v}{\partial v^\alpha} \right\} &= -[U^{\beta\gamma} f_\gamma(h) + v^\beta - I^\beta] \frac{\partial g_\beta(v)}{\partial v^\alpha}; \end{aligned}$$

and

$$\begin{aligned} \frac{\partial E}{\partial h^\alpha}(v, h) &= f_\alpha(h) + h^\beta \frac{\partial f_\beta(h)}{\partial h^\alpha} - \frac{\partial L_h}{\partial h^\alpha}(h) - U^{\gamma\beta} g_\beta(v) \frac{\partial f_\beta(h)}{\partial h^\alpha} \\ \left\{ f_\alpha = \frac{\partial L_h}{\partial h^\alpha} \right\} &= -[U^{\gamma\beta} g_\beta(v) + h^\beta] \frac{\partial f_\beta(h)}{\partial h^\alpha}. \end{aligned}$$

If  $f$  and  $g$  are linear functions, then  $\partial^2 f$  and  $\partial^2 g$  vanish. Thus, comparing with 8, we find  $K_v = \partial^2 L_v(v)^{-1}$ ,  $K_h = \partial^2 L_h(h)^{-1}$ , and  $\nabla K = 0$ . That is,

$$\begin{aligned} \frac{dv^\alpha}{dt} &= K_v^{\alpha\beta} [U_{\beta\gamma} f^\gamma(h) - v_\beta + I_\beta] + \sqrt{2T} dW_v^\alpha; \\ \frac{dh^\alpha}{dt} &= K_h^{\alpha\beta} [U_{\gamma\beta} g^\gamma(v) - h_\beta] + \sqrt{2T} dW_h^\alpha, \end{aligned}$$

5. Originally illustrated in Large Associative Memory Problem in Neurobiology and Machine Learning, Dmitry Krotov and John Hopfield, 2020.

6. Here the  $\partial^2 L$  is the Hessian matrix, and  $[\partial^2 L]^{-1}$  the inverse matrix.

Thus proof ends.  $\square$

**Remark 18.** [Hebbian Rule]

In addition, we find, along the gradient descent trajectory of  $U$ , the difference is

$$\Delta U^{\alpha\beta} \propto \left\langle -\frac{\partial E}{\partial U_{\alpha\beta}}(v, h; U) \right\rangle_{q_E(\cdot|v)} = \langle g^\alpha(v) f^\alpha(h) \rangle_{q_E(\cdot|v)}.$$

Since  $f$  and  $g$  are activation functions, we recover the Hebbian rule, that is, neurons that fire together wire together.

**Remark 19.** [Simplified Brain]

This model can be viewed as a simplified model of brain. Indeed, in the equation (1) of Dehaene et al. (2003)<sup>7</sup>, when the  $V$  are limited to a small region, and the  $\tau$ s are large, then the coefficients, i.e. the  $m$ s and  $h$ s, can be regarded as constants. The equation (1), thus, reduces to the continuous Hopfield network.

## Appendix A Useful Lemmas

**Lemma 20.** [Kramers–Moyal Expansion]

Given random variable  $X$  and time parameter  $t$ , consider random variable  $\epsilon$  whose distribution is  $(x, t)$ -dependent. After  $\Delta t$ , particles in position  $x$  jump to  $x + \epsilon$ . Then, we have

$$p(x, t + \Delta t) - p(x, t) = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n!} \nabla_{a_1} \cdots \nabla_{a_n} [p(x, t) M^{a_1 \cdots a_n}(x, t)],$$

where  $M^{a_1 \cdots a_n}(x, t)$  represents the  $n$ -order moments of  $\epsilon$

$$M^{a_1 \cdots a_n}(x, t) := \langle \epsilon^{a_1} \cdots \epsilon^{a_n} \rangle_\epsilon.$$

**Proof.** The trick is introducing a smooth test function,  $h(x)$ . Denote

$$I_{\Delta t}[h] := \int d\mu(x) p(x, t + \Delta t) h(x). \quad \square$$

The transition probability from  $x$  at  $t$  to  $y$  at  $t + \Delta t$  is  $\int d\mu(\epsilon) p_\epsilon(\epsilon; x, t) \delta(x + \epsilon - y)$ . This implies

$$p(y, t + \Delta t) = \int d\mu(x) p(x, t) \left[ \int d\mu(\epsilon) p_\epsilon(\epsilon; x, t) \delta(x + \epsilon - y) \right].$$

With this,

$$\begin{aligned} I_{\Delta t}[h] &:= \int d\mu(x) p(x, t + \Delta t) h(x) \\ \{x \rightarrow y\} &= \int d\mu(y) p(y, t + \Delta t) h(y) \\ [p(y, t + \Delta t) = \cdots] &= \int d\mu(x) p(x, t) \int d\mu(y) \int d\mu(\epsilon) p_\epsilon(\epsilon; x, t) \delta(x + \epsilon - y) h(y) \\ \{\text{Integrate over } y\} &= \int d\mu(x) p(x, t) \int d\mu(\epsilon) p_\epsilon(\epsilon; x, t) h(x + \epsilon). \end{aligned}$$

Taylor expansion  $h(x + \epsilon)$  on  $\epsilon$  gives

$$I_{\Delta t}[h] = \int d\mu(x) p(x, t) h(x) + \sum_{n=1}^{+\infty} \frac{1}{n!} \int d\mu(x) p(x, t) [\nabla_{a_1} \cdots \nabla_{a_n} h(x)] \int d\mu(\epsilon) p_\epsilon(\epsilon; x, t) \epsilon^{a_1} \cdots \epsilon^{a_n}.$$

Integrating by part on  $x$  for the second term, we find

$$I_{\Delta t}[h] = \int d\mu(x) p(x, t) h(x) + \sum_{n=1}^{+\infty} \frac{(-1)^n}{n!} \int d\mu(x) h(x) \nabla_{a_1} \cdots \nabla_{a_n} \left[ p(x, t) \int d\mu(\epsilon) p_\epsilon(\epsilon; x, t) \epsilon^{a_1} \cdots \epsilon^{a_n} \right].$$

Denote  $n$ -order moments of  $\epsilon$  as  $M^{a_1 \cdots a_n}(x, t) := \langle \epsilon^{a_1} \cdots \epsilon^{a_n} \rangle_\epsilon$  and recall the definition of  $I_{\Delta t}[h]$ , then we arrive at

$$\int d\mu(x) [p(x, t + \Delta t) - p(x, t)] h(x) = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n!} \int d\mu(x) h(x) \nabla_{a_1} \cdots \nabla_{a_n} [p(x, t) M^{a_1 \cdots a_n}(x, t)].$$

<sup>7</sup> A neuronal network model linking subjective reports and objective physiological data during conscious perception, Stanislas Dehaene, Claire Sergent, and Jean-Pierre Changeux, 2003.

Since  $h(x)$  is arbitrary, we conclude that

$$p(x, t + \Delta t) - p(x, t) = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n!} \nabla_{a_1} \cdots \nabla_{a_n} [p(x, t) M^{a_1 \cdots a_n}(x, t)].$$

## Appendix B Stochastic Dynamics

### B.1 Random Walk

Given  $\forall x \in \mathcal{M}$  and any time  $t$ , consider a series of i.i.d. random variables (random walks),

$$\{\varepsilon_i^a : i = 1 \dots n(t)\},$$

where, for  $\forall i$ ,  $\varepsilon_i^a \sim P$  for some distribution  $P$ , with the mean 0 and covariance  $\Sigma(x, t)$ , and the walk steps

$$n(t) = \int_0^t d\tau \frac{dn}{d\tau}(x(\tau), \tau).$$

For any time interval  $\Delta t$ , this series of random walks leads to a difference

$$\Delta x^a := \sum_{i=n(t)}^{n(t+\Delta t)} \varepsilon_i^a.$$

Let

$$\tilde{W}^a(x, t) := \frac{1}{\sqrt{n(t+\Delta t) - n(t)}} \sum_{i=n(t)}^{n(t+\Delta t)} \varepsilon_i^a,$$

we have  $\Delta x^a = \sqrt{n(t+\Delta t) - n(t)} \tilde{W}^a(x, t)$ . Since  $n(t+\Delta t) - n(t) = \frac{dn}{dt}(x, t) \Delta t + o(\Delta t)$ , we have

$$\Delta x^a = \sqrt{n(t+\Delta t) - n(t)} \tilde{W}^a(x, t) = \sqrt{\frac{dn}{dt}(x, t) \Delta t} \tilde{W}^a(x, t) + o(\sqrt{\Delta t}).$$

If

$$\frac{dn}{dt}(x, t) \Sigma^{ab}(x, t) = \mathcal{O}(1)$$

as  $dn/dt \rightarrow +\infty$ , that is, more steps per unit time, then, by central limit theorem (for multi-dimension),

$$\Delta x^a = \Delta W^a + o\left(\frac{dn}{dt}(x, t)\right),$$

where

$$\Delta W^a \sim \mathcal{N}(0, \Delta t \Sigma^{ab}(x, t)).$$

### B.2 Stochastic Dynamics

A stochastic dynamics is defined by two parts. The first is deterministic, and the second is a random walk. Precisely,

**Definition 21.** [*Stochastic Dynamics*]

Given  $\mu^a(x, t)$  and  $\Sigma^{ab}(x, t)$  on  $\mathcal{M} \times \mathbb{R}$ ,

$$dx^a = \mu^a(x, t) dt + dW^a(x, t),$$

where  $dW^a(x, t)$  is a random walk with covariance  $\Sigma^{ab}(x, t)$ .

Consider an ensemble of particles, each obeys this stochastic dynamics. This ensemble will form a distribution, evolving with time  $t$ , say  $p(t)$ . The equation of this evolution is

**Lemma 22.** [*Macroscopic Landscape*]

$$\frac{\partial p}{\partial t}(t) = -\nabla_a \{p(x, t) \mu^a(x, t)\} + \frac{1}{2} \nabla_a \nabla_b \{p(x, t) \Sigma^{ab}(x, t)\}.$$

**Proof.** From the difference of the stochastic dynamics,

$$\Delta x^a = \mu^a(x, t) \Delta t + \Delta W^a(x, t),$$

by Kramers–Moyal expansion 20, we have

$$p(x, t + \Delta t) - p(x, t) = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n!} \nabla_{a_1} \cdots \nabla_{a_n} [p(x, t) \langle \Delta x^{a_1} \cdots \Delta x^{a_n} \rangle_{\Delta x}].$$

For  $n=1$ , since  $dW^a(x, t)$  is a random walk,  $\langle \Delta W^a(x, t) \rangle_{\Delta W(x, t)} = 0$ . Then the term is  $-\nabla_a [p(x, t) \langle \Delta x^a \rangle_{\Delta x}] = -\nabla_a \{p(x, t) \mu^a(x, t)\} \Delta t$ . And for  $n=2$ , by noticing that, as a random walk,  $\langle \Delta W^a(x, t) \Delta W^b(x, t) \rangle_{\Delta W(x, t)} = \mathcal{O}(\Delta t)$ , we have, up to  $o(\Delta t)$ , only  $(1/2) \nabla_a \nabla_b [p(x, t) \Sigma^{ab}(x, t)] \Delta t$  left. For  $n \geq 3$ , all are  $o(\Delta t)$ . So, we have

$$\frac{p(x, t + \Delta t) - p(x, t)}{\Delta t} = -\nabla_a \{p(x, t) \mu^a(x, t)\} + \frac{1}{2} \nabla_a \nabla_b \{p(x, t) \Sigma^{ab}(x, t)\} + o(1).$$

Letting  $\Delta t \rightarrow 0$ , we find

$$\frac{\partial p}{\partial t}(x, t) = -\nabla_a \{p(x, t) \mu^a(x, t)\} + \frac{1}{2} \nabla_a \nabla_b \{p(x, t) \Sigma^{ab}(x, t)\}.$$

Thus proof ends. □