1 Lyapunov Function

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

- $\mathcal{M}$ *a manifold, and* μ *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*
$$\langle f \rangle_p = \langle f \rangle_X := \int_{\mathcal{M}} \mathrm{d}\mu(x)\, p(x)\, f(x);$$
- *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* μ *and covariance* Σ*;*
- *given function* g*, let* $f\{g\}$*, or* $f_{\{g\}}$*, denote a function constructed out of* g*, that is,*
$$f\{\cdot\}: (\mathcal{M} \to A) \to (\mathcal{M} \to B);$$
- *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 $E(x): \mathcal{M} \to \mathbb{R}$*. Define stationary distribution*

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

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

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

Or, equivalently,

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

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. TODO □

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

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

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

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

The first term vanishes.[1] Then, direct calculus shows

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

By the previous lemma, we have

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

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,

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

1. 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 integrad 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 $(\mathrm{d}F_E/\mathrm{d}t)[p(\cdot,t)]<0$ as long as $\nabla_a\{T\ln p(x,t)+E(x)\}\neq 0$ at some x, i.e. $p\neq q$, and $(\mathrm{d}F_E/\mathrm{d}t)[p(\cdot,t)]=0$ if and only if $\nabla_a\{T\ln p(x,t)+E(x)\}=0$ for $\forall x$, i.e. $p=q$. Thus proof ends. □

Remark 7. [Sufficent but Not Essential]

However, this is not an essntial condition of relaxing to q_E for any p. Indeed, we proved the integrand of $(\mathrm{d}F_E/\mathrm{d}t)[p(\cdot,t)]$ is negative everywhere, which implies the integral, i.e. $(\mathrm{d}F_E/\mathrm{d}t)[p(\cdot,t)]$, 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"[2], 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

$$\mathrm{d}x^a=[T\,\nabla_b K^{ab}(x,t)-K^{ab}(x,t)\,\nabla_b E(x)]\,\mathrm{d}t+\sqrt{2T}\,\mathrm{d}W^a(x,t),$$

where

$$\mathrm{d}W\sim\mathcal{N}(0,K(x,t)\,\mathrm{d}t).$$

Proof. By the lemma 23, 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

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

exactly the same. Thus proof ends. □

Question 1. Given an autonomous dynamical system[4],

$$\frac{\mathrm{d}x^a}{\mathrm{d}t}=f^a(x),$$

are there $E(x)$ and $K^{ab}(x)$ s.t. $f^a(x)=-K^{ab}(x)\,\nabla_b E(x)$? And if exist, how can we construct them explicitly?

This question calls for a generic method of construction of Lyapunov function.

2 Lyapunov Function

Definition 9. *[Lyapunov Function]*

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.

Given an autonomous dynamical system[5]*,*

$$\frac{\mathrm{d}x^a}{\mathrm{d}t} = f^a(x),$$

a Lyapunov function of this dynamical system, $V(x)$, is a scalar function s.t. $\nabla_a V(x)\, f^a(x) \leqslant 0$ and the equality holds if and only if $f^a(x) = 0$.

Along the phase trajectory, a Lyapunov function will monomotically decrease. So, it reflects the stability of the dynamical system. Again, by lemma 20, for any x out of the neighbourhood of singular points of $\nabla_a V(x)$, i.e. $x \notin \mathcal{U}_\delta(\nabla_a V(x))$, there exists a tensor fields $K^{ab}(x)$ with the same order of smooth as $f^a(x)$ or $\nabla_a V(x)$, s.t. $f^a(x) = -K^{ab}(x)\, \nabla_b V(x)$. Then we have, for any $x \notin \mathcal{U}_\delta(\nabla_a V(x))$,

$$\frac{\mathrm{d}V}{\mathrm{d}t}(x(t)) = \nabla_a V(x)\, f^a(x) = -K^{ab}(x)\, \nabla_a V(x)\, \nabla_b V(x) < 0,$$

meaning that the symmetric part of $K^{ab}(x)$ is positive definite. Comparing with the Fokker-Planck equation 6, in the case $T \to 0$, we recognize that $V(x)$ here is the $E(x)$ there.

3 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 funciton 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 10. *[Conditional Free Energy]*

Given v, if define

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

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

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

Directly, we have

Lemma 11.

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

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

Thus,

Theorem 12.

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

4 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 expection $\langle E(\cdot; \theta) \rangle_{q_{E(\cdot;\theta)}}$ is minimized. For instance, given ambient v, we want to locates v on the minimum of E, that is $\langle E(v, \cdot; \theta) \rangle_{q_E(\cdot|v;\theta)}$ (c.f. lemma 11).

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 θ that minimizes $\langle E(\cdot; \theta) \rangle_{q_{E(\cdot;\theta)}} - T H[q_{E(\cdot;\theta)}]$, where the positive constant T balances the two aspects. This happens to be the free energy.

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

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

Lemma 13.
$$\frac{\partial}{\partial \theta^\alpha}\{-T \ln Z_{E(\cdot;\theta)}\} = \left\langle \frac{\partial E}{\partial \theta^\alpha}(\cdot;\theta) \right\rangle_{q_{E(\cdot;\theta)}}.$$

So, we have an EM-like algorithm, as

Algorithm 14. *[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.

5 Example: Continuous Hopfield Network

Here, we provide a biological inspired example, for illustrating both the stochastic dynamics 8 and the RL algorithm 14.

Definition 15. *[Continuous Hopfield Network]*

Let $U^{\alpha\beta}$ and I^α 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[6]

$$\frac{\mathrm{d}v^\alpha}{\mathrm{d}t} = U^{\alpha\beta} f_\beta(h) - v^\alpha + I^\alpha;$$
$$\frac{\mathrm{d}h^\alpha}{\mathrm{d}t} = U^{\beta\alpha} g_\beta(v) - h^\alpha,$$

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 16. *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*

$$\frac{\mathrm{d}v^\alpha}{\mathrm{d}t} = K_v^{\alpha\beta}\, [U_{\beta\gamma} f^\gamma(h) - v_\beta + I_\beta] + \sqrt{2T}\, \mathrm{d}W_v^\alpha;$$
$$\frac{\mathrm{d}h^\alpha}{\mathrm{d}t} = K_h^{\alpha\beta}\, [U_{\gamma\beta}\, g^\gamma(v) - h^\beta] + \sqrt{2T}\, \mathrm{d}W_h^\alpha,$$

where $K_v := [\partial^2 L_v(v)]^{-1}$ and $K_h := [\partial^2 L_h(h)]^{-1}$ are constant matrices.[7]

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

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

Proof. Directly, we have

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

and

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

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,

$$\begin{aligned}\frac{\mathrm{d}v^{\alpha}}{\mathrm{d}t} &= K_v^{\alpha\beta}\,[U_{\beta\gamma} f^{\gamma}(h) - v_{\beta} + I_{\beta}] + \sqrt{2T}\,\mathrm{d}W_v^{\alpha};\\ \frac{\mathrm{d}h^{\alpha}}{\mathrm{d}t} &= K_h^{\alpha\beta}\,[U_{\gamma\beta} g^{\gamma}(v) - h^{\beta}] + \sqrt{2T}\,\mathrm{d}W_h^{\alpha},\end{aligned}$$

Thus proof ends. □

Remark 17. [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 18. [Simplified Brain]

This model can be viewed as a simplified model of brain. Indeed, in the equation (1) of Dehaene et al. (2003)[8], when the V are limited to a small region, and the τs are large, then the coefficients, i.e. the ms and hs, can be regarded as constants. The equation (1), thus, reduces to the continuous Hopfield network.

Appendix A Useful Lemmas

A.1 Vector Fields

Lemma 19. *Given any vector f^a and g_b, if $g_b \neq 0$, then there exists a tensor K^{ab}, s.t. $f^a = K^{ab} g_b$.*

Proof. We can rotate g_b to the direction of f^a and then dimension-wise recale to f^a. This rotation and dimension-wise rescaling compose the linear transform K^{ab}. □

Now we extends this lemma to vector fields.

Lemma 20. *[Vector Fields]*

Given any vector fields $f^a(x)$ and $g_b(x)$, define $\mathcal{U}_{\delta}(g)$ as the union of δ-neighbourhoods of singular point of $g_b(x)$. then there exists a smooth tensor field $K^{ab}(x)$, s.t. $f^a(x) = K^{ab}(x)\, g_b(x)$ for $\forall x \notin \mathcal{U}_{\delta}(g)$.

Proof. Only smoothness of $K^{ab}(x)$ is to be proved. Since f^a and g_b are smooth, after varying them a little, that is, $f^a(x+\delta x) = K^{ab}(x+\delta x)\, g_b(x+\delta x)$. Taylor expanding f^a and g_b, we find $f^a(x) + \delta x^b\, \nabla_b f^a(x) = K^{ab}(x+\delta x)\, g_b(x) + K^{ab}(x+\delta x)\, \delta x^c\, \nabla_c g_b(x)$. Thus $\delta x^b\, [\nabla_b f^a(x) - K^{ac}(x+\delta x)\, \nabla_b g_c(x)] = [K^{ab}(x+\delta x) - K^{ab}(x)]\, g_b(x)$. Since $g_b(x)$ is not singular, we have $K_b^a(x+\delta x) - K_b^a(x) = \mathcal{O}(\delta x)$, thus the first order derivatives exist. The same process holds for higher order derivatives, untill the order where either f^a or g_b becomes non-smooth. □

When x approaches a singular point of $g_b(x)$, then the $K^{ab}(x)$ may be divergent, since $f^a(x)$ may not vanish at this point. Even excluding the singular points is not enough. For instance, TODO: add a plot. In this example, the $K^{ab}(x)$ cannot be smooth. This is why we have to exclude the neighbours of the singular points, instead of the singular points themselves.

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

A.2 Kramers–Moyal Expansion

Kramers–Moyal Expansion relates the microscopic landscape, i.e. the dynamics of Brownian particles, and the macroscopic landscape, i.e. the evolution of distribution.

Lemma 21. *[Kramers–Moyal Expansion]*

Given random variable X and time parameter t, consider random variable ϵ whose distribution is (x,t)-dependent. After Δt, 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 ϵ

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

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

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

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

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

With this,

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

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

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

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

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

Denote n-order moments of ϵ 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

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

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),

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

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

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

For any time interval Δt, 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 $n(t+\Delta t) - n(t) = \frac{\mathrm{d}n}{\mathrm{d}t}(x,t) \, \Delta t + o(\Delta t)$, we have

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

If

$$\frac{\mathrm{d}n}{\mathrm{d}t}(x,t) \, \Sigma^{ab}(x,t) = \mathcal{O}(1)$$

as $\mathrm{d}n / \mathrm{d}t \to +\infty$, that is, more steps per unit time, then, by central limit theorem (for multi-dimension),

$$\Delta x^a = \Delta W^a + o\left(\frac{\mathrm{d}n}{\mathrm{d}t}(x,t) \right),$$

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 22. *[Stochastic Dynamics]*
Given $\mu^a(x,t)$ and $\Sigma^{ab}(x,t)$ on $\mathcal{M} \times \mathbb{R}$,

$$\mathrm{d}x^a = \mu^a(x,t) \, \mathrm{d}t + \mathrm{d}W^a(x,t),$$

where $\mathrm{d}W^a(x,t)$ 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 23. *[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 21, 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 $\mathrm{d}W^a(x,t)$ 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 $n \geqslant 3$, 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 $\Delta t \to 0$, 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. $\square$