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*

$$\langle f \rangle_p = \langle f \rangle_X = \mathbb{E}_{x\sim p}[f(x)] := \int_{\mathcal{M}} \mathrm{d}\mu(x)\, p(x)\, f(x);$$

- *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*;*
- *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, and i.i.d. for independent identically distributed;*
- *ODE for ordinary differential equation(s), SDE for stochastic differential equation(s);*
- *Laplacian* $\Delta := \nabla_a \nabla^a$*;*
- *let* $\mathcal{D}_v$ *denotes the Lie derivative along direction* v*;*
- $\mathcal{C}^p(A, B)$ *denotes* p*-order smooth function from* A *to* B*, and* $\mathcal{C}^p(A)$ *shorten for* $\mathcal{C}^p(A, \mathbb{R})$*.*

1 Lyapunov Function

1.1 Definition

Definition 2. *[Lyapunov Function]*

Given an autonomous[1] *ODE,*

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

a Lyapunov function, $V(x)$*, of it is a scalar function such that* $\nabla_a V(x)\, f^a(x) \leqslant 0$ *and the equality holds if and only if* $f^a(x) = 0$*.*

Along the phase trajectory, a Lyapunov function monomotically decreases. So, it reflects the stability of the ODE.

1.2 Construction of Lyapunov Function

Question 1. Given an autonomous ODE, whether a Lyapunov function of it exists or not?

Question 2. And how to construct one if there is any?

Here we propose a simulation based method that furnishes a criterion on whether a Lyapunov function exists or not, and then reveals an analytic approximation to the Lyapunov function if it exists.

We first extend the autonomous ODE to a SDE[2], as

$$\mathrm{d}X^a = f^a(X)\,\mathrm{d}t + \sqrt{2T}\,\mathrm{d}W^a,$$

1. That is, ordinary differential equations that do not explicitly depend on time. The word autonomous means independent of time.
2. SDE is defined in 8.

where $\mathrm{d}W^a \sim \mathcal{N}(0, \delta^{ab}\mathrm{d}t)$ and parameter $T > 0$. Then, we sample an essemble of "particles" independently evolving along this SDE. As a set of Markov chains, this simulation will arrive at a stationary distribution. This is true if the Markov chain is irreducible and recurrent. These conditions are hard to check. But, in practice, there is criterion on the convergence of a chain at a finite time.[3] If it has converged, we get an empirical distribution, denoted as p_D, that approximates to the true stationary distribution.

Next, we are to find an analytic approximation to the empirical distribution p_D. This can be taken by any universal approximator, such as neural network. Say, an universal approximator $E(\cdot;\theta)$ parameterized by θ, and define q_E as

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

where $Z_E(\theta) := \int_{\mathcal{M}} \mathrm{d}\mu(x) \exp(-E(x;\theta)/T)$. Then, we construct the loss as

$$\begin{aligned} L(\theta) &:= T D_{\mathrm{KL}}(p_D \| q_E(\cdot;\theta)) \\ &= T\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p_D(x) \ln p_D(x) - T\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p_D(x) \ln q_E(x;\theta). \end{aligned}$$

The first term is independent of θ, thus omitable. Thus, the loss becomes

$$\begin{aligned} L(\theta) &= -T\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p_D(x) \ln q_E(x;\theta) \\ &= \int_{\mathcal{M}} \mathrm{d}\mu(x)\, p_D(x)\, E(x;\theta) + T\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p_D(x) \ln Z_E(\theta) \\ &= \langle E(\cdot;\theta)\rangle_{p_D} \\ \left[\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p_D(x) = 1\right] &+ T \ln Z_E(\theta). \end{aligned}$$

We find the best fit $\theta_\star := \operatorname{argmin}_\theta L(\theta)$ by using gradient descent. Notice the relation

Lemma 3.

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

Proof. Directly,

$$\begin{aligned} T\frac{\partial}{\partial\theta^\alpha} \ln Z_{E(\cdot;\theta)} &= T\frac{1}{Z_{E(\cdot;\theta)}} \frac{\partial}{\partial\theta^\alpha} Z_{E(\cdot;\theta)} \\ \{Z_E := \cdots\} &= T\frac{1}{Z_{E(\cdot;\theta)}} \frac{\partial}{\partial\theta^\alpha} \int_{\mathcal{M}} \mathrm{d}\mu(x)\, \mathrm{e}^{-E(x;\theta)/T} \\ &= -\int_{\mathcal{M}} \mathrm{d}\mu(x)\, \frac{\mathrm{e}^{-E(x;\theta)/T}}{Z_{E(\cdot;\theta)}} \frac{\partial E}{\partial\theta^\alpha}(x;\theta) \\ \{q_E := \cdots\} &= -\int_{\mathcal{M}} \mathrm{d}\mu(x)\, q_E(x;\theta) \frac{\partial E}{\partial\theta^\alpha}(x;\theta). \\ &= -\left\langle \frac{\partial E}{\partial\theta^\alpha}(\cdot;\theta) \right\rangle_{q_E(\cdot;\theta)}. \end{aligned}$$

Thus, proof ends. □

This implies

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

Both of the two terms can be computed by Monte Carlo integral. Since p_D has been an empirical distribution, the computation of the first Monte Carlo integral is straight forward. The second can be computed in the same way of generating the empirical distribution p_D, by noticing

Lemma 4. *Markov chains by SDE*

$$\mathrm{d}X^a = -\nabla^a E(x)\,\mathrm{d}t + \sqrt{2T}\,\mathrm{d}V^a,$$

where $\mathrm{d}V^a \sim \mathcal{N}(0, \delta^{ab}\,\mathrm{d}t)$ and $T > 0$, will converge to q_E.

Proof. By theorem 9, the distribution $p(x,t)$ of the Markov chains generated by the SDE obeys

$$\frac{\partial p}{\partial t}(x,t) = \nabla_a[p(x,t)\,\nabla^a E(x)] + T\,\Delta p(x,t).$$

3. E.g., visualization by animation of histograms. Practically, we find that the Gelman-Rubin-Brooks plot won't work in some cases.

It's straight forward to check that q_E is a stationary solution to this equation. And for any initial value of $p(x,t)$, it always relax to q_E. Indeed,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} T D_{\mathrm{KL}}(p\|q_E) &= \frac{\mathrm{d}}{\mathrm{d}t} T \int_{\mathcal{M}} \mathrm{d}\mu(x)\, p(x,t)\, [\ln p(x,t) - \ln q_E(x)] \\
&= T \int_{\mathcal{M}} \mathrm{d}\mu(x)\, \frac{\partial p}{\partial t}(x,t)\, [\ln p(x,t) - \ln q_E(x) + 1] \\
\left\{\frac{\partial p}{\partial t}(x,t) = \cdots\right\} &= T \int_{\mathcal{M}} \mathrm{d}\mu(x)\, \nabla_a [p(x,t)\, \nabla^a E(x) + T\, \nabla^a p(x,t)]\, [\ln p(x,t) - \ln q_E(x) + 1] \\
\{\text{Integral by part}\} &= -T \int_{\mathcal{M}} \mathrm{d}\mu(x) [p(x,t)\, \nabla^a E(x) + T\, \nabla^a p(x,t)]\, \nabla_a [\ln p(x,t) - \ln q_E(x) + 1] \\
&= -\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p(x,t)\, \nabla_a [E(x) + T \ln p(x,t)]\, \nabla^a [E(x) + T \ln p(x,t)] \\
&\leqslant 0,
\end{aligned}$$

and the equility holds if and only if $\nabla_a [E(x) + T \ln p(x,t)] = 0$ for $\forall x$, that is, $p(x,t) \equiv q_E(x)$. □

Notice that the best fit $\theta_\star$ depends on T implicitly. We take a series of T, $\{T_1, T_2, \dots\}$, that tends to 0_+. And for each T_i, we repeat the previous process to find the best fit $\theta_\star$ for this T_i. When T_i tends to 0, we claim that the best fit E, denoted by $E_\star$, satisfies $\nabla_a E_\star(x)\, f^a(x) \leqslant 0$ for $\forall x \in \mathcal{M}$, and the equality holds if and only if $E_\star(x) = 0$.

Proof. For simplification, we denote $E_\star$ by E within this proof.

By theorem 9, the distribution $p(x,t)$ of the Markov chains generated by the SDE to p_D obeys

$$\frac{\partial p}{\partial t}(x,t) = -\nabla_a [p(x,t)\, f^a(x)] + T\, \Delta p(x,t).$$

In the end, $p \to p_D \approx q_E$ where $\partial p / \partial t \to 0$, we have

$$\begin{aligned}
0 &= [-\nabla_a [q_E(x)\, f^a(x)] + T\, \Delta q_E(x)] \\
\{q_E = \cdots\} &= -\frac{1}{Z_E} \nabla_a [\mathrm{e}^{-E(x)/T} f^a(x)] + T\, \Delta \mathrm{e}^{-E(x)/T} \\
[\text{1st term}] &= \frac{1}{T Z_E} \mathrm{e}^{-E(x)/T}\, \nabla_a E(x) f^a(x) - \mathrm{e}^{-E(x)/T} \nabla_a f^a(x) \\
[\text{2nd term}] &+ \frac{1}{T Z_E}\, \mathrm{e}^{-E(x)/T}\, \nabla_a E(x) \nabla^a E(x) - \mathrm{e}^{-E(x)/T}\, \Delta E(x) \\
&= \frac{1}{T Z_E} \mathrm{e}^{-E(x)/T} \times \{\nabla_a E(x) f^a(x) + \nabla_a E(x) \nabla^a E(x) - T\, [\nabla_a f^a(x) + \Delta E(x)]\}.
\end{aligned}$$

As $T \to 0$, we arrive at

$$\nabla_a E(x) f^a(x) = -\nabla_a E(x) \nabla^a E(x) \leqslant 0,$$

and the equality holds if and only if $\nabla_a E(x) = 0$. □

TODO: Proof that $\nabla_a E(x) = 0 \Rightarrow f^a(x) = 0$.

1.3 Parameterized ODE

Next, we consider parameterized ODE, that is, $f(x) \to f(x;\varphi)$, where φ denotes collection of parameters. For instance, in the case of oscillator, φ can be the stiffness factor. In this situation, we want to construct a Lyapunov function, not only as a function of x, but also of φ, that is, $E(x,\varphi)$, such that for $\forall (x,\varphi)$ in the domain, we have

$$\frac{\partial E}{\partial x^a}(x,\varphi)\, f^a(x;\varphi) \leqslant 0,$$

where the equality holds if and only $f^a(x;\varphi) = 0$.

Even though a completely new question, with some trick, it can be reduced to the one we have solved.

Example 5. Consider the one-dimensional ODE with a parameter r,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = r - x.$$

It is equivalent to an augmented ODE

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= r - x; \\
\frac{\mathrm{d}r}{\mathrm{d}t} &= 0.
\end{aligned}$$

And, it has a Lyapunov function

$$E(x,r)=\frac{1}{2}(r-x)^2.$$

Indeed, since $(\partial E/\partial x)(x,r)=x-r$, $(\partial E/\partial x)(x,r)\,f(x)=-(r-x)^2\leqslant 0$. Given r, it's a valley centered at r along x-aixs. By the way, $(\partial E/\partial r)(x,r)=r-x$. We find that the place where $(\partial E/\partial x)(x,r)=0$ has $(\partial E/\partial r)(x,r)=0$.

Obviously, as a function of (x,r), the Markov chain constructed via $E(x,r)$ cannot relax. However, since the SDE along r-axis is a pure Brownian motion, particles obey a normal distribution with the standard derivative $\sigma_\varphi\sim\sqrt{t}$. Thus $\mathrm{d}\sigma_\varphi/\mathrm{d}t\sim 1/\sqrt{t}\to 0$, as $t\to+\infty$, indicating that the $p_D(x,t)$ becomes slow varying as t large enough. In other word, the p_D will approximately relaxes along r-axis, and finally this approximation becomes "good enough". With this consideration, we can say $q_E(x)=\lim_{t\to+\infty}p_D(x,t)$ again.

The parameterized ODE

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

is equivalent to the augmented ODE without parameter $\frac{\mathrm{d}y^a}{\mathrm{d}t}=F^a(y)$, where $y=(x,\varphi)$ and $F:=(f,0)$. Now, to solve the problem with the parameterized ODE, we simply solve the equivalent one with the augmented ODE, using the same method we have proposed. Following the previous process, again, we find the distribution $p(x,t)$ of the Markov chains generated by the SDE to p_D obeys

$$\frac{\partial p}{\partial t}(y,t)=-\nabla_a[p(y,t)\,F^a(y)]+T\,\Delta p(y,t).$$

Since the φ components of F vanish, it becomes

$$\frac{\partial p}{\partial t}(x,\varphi,t)=-\frac{\partial}{\partial x^\alpha}[p(x,\varphi,t)\,f^\alpha(x;\varphi)]+T\,\Delta_x p(x,\varphi,t)+T\,\Delta_\varphi p(x,\varphi,t).$$

Thus, as $\partial p/\partial t\to 0$ and $p\to q_E$, and as $T\to 0$,

$$\frac{\partial E}{\partial x^\alpha}(x,\varphi)\,f^\alpha(x;\varphi)+\delta^{\alpha\beta}\frac{\partial E}{\partial x^\alpha}(x,\varphi)\frac{\partial E}{\partial x^\beta}(x,\varphi)+\delta^{\alpha\beta}\frac{\partial E}{\partial \varphi^\alpha}(x,\varphi)\frac{\partial E}{\partial \varphi^\beta}(x,\varphi)=0.$$

We arrive at

$$\frac{\partial E}{\partial x^\alpha}(x,\varphi)\,f^\alpha(x;\varphi)=-\left[\delta^{\alpha\beta}\frac{\partial E}{\partial x^\alpha}(x,\varphi)\frac{\partial E}{\partial x^\beta}(x,\varphi)+\delta^{\alpha\beta}\frac{\partial E}{\partial \varphi^\alpha}(x,\varphi)\frac{\partial E}{\partial \varphi^\beta}(x,\varphi)\right]\leqslant 0,$$

and equality holds if and only $(\partial E/\partial x^a)(x,\varphi)=0$ and $(\partial E/\partial \varphi^a)(x,\varphi)=0$.

1.4 Implementation

Implementation can be found in `src/Lyapunov.jl` in `julia`. And tested, found in `test/TestLyapunov.jl`, on damped oscillators with and without fixed stiffness factor. Both underdamped and overdamped cases are considered. This method, as the results manifestly show, is surprisingly effective.

Appendix A Useful Lemmas

A.1 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 6. *[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

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

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 $\Sigma^{ab}(x, t)$, 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

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

Then, we have

Theorem 7. *[Brownian Motion]*

As $\mathrm{d}n/\mathrm{d}t \to +\infty$,

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

where

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

Proof. 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, or stochastic differential equations (SDE), is defined by two parts. The first is deterministic, and the second is a random walk. Precisely,

Definition 8. *[SDE] Given $f^a(x,t)$, $g^{ab}(x,t)$, and $\Sigma^{ab}(x,t)$ on $\mathcal{M} \times \mathbb{R}$, with g^{ab} and Σ^{ab} positive definite, stochastic differential equations is defined as*

$$\mathrm{d}x^a = f^a(x,t)\, \mathrm{d}t + g^{ab}(x,t)\, \mathrm{d}W_b(x,t),$$

where $\mathrm{d}W_a(x,t)$ is a Brownian motion 7 with covariance $\Sigma_{ab}(x,t)\, \mathrm{d}t$.

Consider an ensemble of particles, randomly sampled at an initial time, evolving along a SDE 8. By saying "ensemble", we mean that the number of particles has the order of Avogadro's constant, s.t. the distribution of the particles can be viewed as smooth. Let $p(x,t)$ denotes the distribution. Then we have

Theorem 9. *[Fokker-Planck Equation]*

$$\frac{\partial p}{\partial t}(x,t) = -\nabla_a[f^a(x,t)\, p(x,t)] + \nabla_a \nabla_b[K^{ab}(x,t)\, p(x,t)],$$

where $K^{ab} := (1/2) g^{ac}(x,t)\, g^{bd}(x,t)\, \Sigma_{cd}(x,t)$.

Proof. From the difference of the SDE,

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

by Kramers–Moyal expansion 6, 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

$$-\nabla_a[p(x,t) \langle \Delta x^a \rangle_{\Delta x}] = -\nabla_a[p(x,t)\, f^a(x,t)] \Delta t.$$

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,

$$\frac{1}{2} \nabla_a \nabla_b[p(x,t)\, \langle \Delta x^a\, \Delta x^b \rangle_{\Delta x}] = \frac{1}{2} \nabla_a \nabla_b[p(x,t)\, g_c^a(x,t)\, g_d^b(x,t)\, \Sigma^{cd}(x,t)]\, \Delta t + o(\Delta t).$$

For $n \geqslant 3$, all are $o(\Delta t)$. So, we have

$$p(x,t+\Delta t) - p(x,t) = -\nabla_a[p(x,t)\, f^a(x,t)] + \frac{1}{2}\nabla_a\nabla_b[p(x,t)\, g_c^a(x,t)\, g_d^b(x,t)\, \Sigma^{cd}(x,t)]\, \Delta t + o(\Delta t).$$

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

$$\frac{\partial p}{\partial t}(x,t) = -\nabla_a[p(x,t)\, f^a(x,t)] + \frac{1}{2}\nabla_a\nabla_b[p(x,t)\, g_c^a(x,t)\, g_d^b(x,t)\, \Sigma^{cd}(x,t)].$$

Thus proof ends. □

B.3 Relaxation

Lemma 10. *Given the Fokker-Planck equation 9 and any initial distribution $p(x,t)$, as t increases, the support of $p(x,t)$ will be $\mathcal{M}$.*

Proof. TODO □

First, we notice that Fokker-Planck 9 is linear, thus it can be re-written by linear operator

$$\frac{\partial}{\partial t}p(x,t) = \hat{L}_{\mathrm{FP}}(x,t)\, p(x,t),$$

where $\hat{L}_{\mathrm{FP}}(x,t) := -\nabla_a f^a(x,t) + \nabla_a\nabla_b K^{ab}(x,t)$. Then, we have

Lemma 11. *[Adjoint Operator] The adjoint operator of $\hat{L}_{\mathrm{FP}}(x,t)$ is*

$$\hat{L}_{\mathrm{FP}}^{\dagger}(x,t) = f^a(x,t)\nabla_a + K^{ab}(x,t)\nabla_a\nabla_b.$$

Proof. Directly, for $\forall t$,

$$\begin{aligned}
\langle h|\hat{L}_{\mathrm{FP}}|g\rangle &= \int_{\mathcal{M}} \mathrm{d}\mu(x)\, h(x)\, \hat{L}_{\mathrm{FP}}(x,t)\, g(x) \\
\{\hat{L}_{\mathrm{FP}} = \cdots\} &= -\int_{\mathcal{M}} \mathrm{d}\mu(x)\, h(x)\, \nabla_\alpha[f^a(x,t) g(x)] + \int_{\mathcal{M}} \mathrm{d}\mu(x)\, h(x) \nabla_a\nabla_b[K^{ab}(x,t) g(x)] \\
\{\text{Integral by part}\} &= \int_{\mathcal{M}} \mathrm{d}\mu(x)\, g(x)\, f^a(x,t)\nabla_a h(x) + \int_{\mathcal{M}} \mathrm{d}\mu(x) g(x)\, K^{ab}(x,t)\, \nabla_a\nabla_b h(x) \\
\{\text{By definition}\} &= \langle g|\hat{L}_{\mathrm{FP}}^{\dagger}|h\rangle.
\end{aligned}$$

This is held for any h and g, we thus find

$$\hat{L}_{\mathrm{FP}}^{\dagger}(x,t) = f^a(x,t)\nabla_a + K^{ab}(x,t)\nabla_a\nabla_b.$$ □

Given the Fokker-Planck equation, we claim that any two initial distributions will evolve into one final distribution as time $t \to +\infty$. To prove this, we simply have to show that, for any two initial distributions, the distance between them decreases as time increasing, and finally vanishes in the limit. That is, relaxation happens.

Theorem 12. *[Relaxation]*
For $\forall p(x,t), q(x,t) \in \mathcal{C}^2(\mathcal{M} \times \mathbb{R})$ as time-dependent distributions obeying the Fokker-Planck equation 9, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{KL}}(p(\cdot,t)|q(\cdot,t)) \leqslant 0,$$

and the equality holds if and only if $p(\cdot,t) = q(\cdot,t)$ for some t.

Proof. Let $r(x,t) := p(x,t)\,/\,q(x,t)$. Then,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{KL}}(p(\cdot,t)|q(\cdot,t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p(x,t) \ln r(x,t) \\
&= \int_{\mathcal{M}} \mathrm{d}\mu(x)\, \frac{\partial p}{\partial t}(x,t) \ln r(x,t) + \int_{\mathcal{M}} \mathrm{d}\mu(x)\, \frac{\partial p}{\partial t}(x,t) - \int_{\mathcal{M}} \mathrm{d}\mu(x) \frac{\partial q}{\partial t}(x,t)\, r(x,t).
\end{aligned}$$

The second term vanishes since

$$\begin{aligned}
\int_{\mathcal{M}} \mathrm{d}\mu(x)\, \frac{\partial p}{\partial t}(x,t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p(x,t) \\
\left\{\int_{\mathcal{M}} \mathrm{d}\mu(x)\, p(x,t) = 1\right\} &= 0.
\end{aligned}$$

And the first term (we omit the (x,t) for simplification)

$$\begin{aligned}\int_{\mathcal{M}} \frac{\partial p}{\partial t} \ln r &= \int_{\mathcal{M}} \ln r\, \hat{L}_{\mathrm{FP}}\, p \\ &= \int_{\mathcal{M}} p\, \hat{L}_{\mathrm{FP}}^{\dagger} \ln r \\ &= \int_{\mathcal{M}} p[f^a \nabla_a \ln r + K^{ab} \nabla_a \nabla_b \ln r] \\ &= \int_{\mathcal{M}} p\left[f^a \frac{\nabla_a r}{r} + \frac{K^{ab} \nabla_a \nabla_b r}{r} - \frac{K^{ab} \nabla_a r \nabla_b r}{r^2}\right] \\ &= \int_{\mathcal{M}} q[f^a \nabla_a r + K^{ab} \nabla_a \nabla_b r] - \int_{\mathcal{M}} q \frac{K^{ab} \nabla_a r \nabla_b r}{r} \\ &= \int_{\mathcal{M}} q \hat{L}_{\mathrm{FP}}^{\dagger} r - \int_{\mathcal{M}} \frac{q^2}{p} K^{ab} \nabla_a r \nabla_b r \\ &= \int_{\mathcal{M}} r \hat{L}_{\mathrm{FP}} q - \int_{\mathcal{M}} \frac{q^2}{p} K^{ab} \nabla_a r \nabla_b r \\ &= \int_{\mathcal{M}} r \frac{\partial q}{\partial t} - \int_{\mathcal{M}} \frac{q^2}{p} K^{ab} \nabla_a r \nabla_b r.\end{aligned}$$

Herein, the first term cancels the third term in $(\mathrm{d}D_{\mathrm{KL}}/\mathrm{d}t)(p(\cdot,t)|q(\cdot,t))$. Thus, we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{KL}}(p(\cdot,t)|q(\cdot,t)) = -\int_{\mathcal{M}} \frac{q^2(x,t)}{p(x,t)} K^{ab}(x,t) \nabla_a r(x,t) \nabla_b r(x,t) \leqslant 0,$$

since $K^{ab}(x,t)$ is positive definite, and the equality holds if and only if $\nabla_a r(\cdot,t) = 0$ on $\mathcal{M}$, that is $p(\cdot,t) \propto q(\cdot,t)$. Since both p and q are normalized distributions, we find $p(\cdot,t) = q(\cdot,t)$. □

B.4 Stationary Solution

When the f^a and K^{ab} of the Fokker-Planck equation 9 are independent of t, we can consider the stationary solution $p(x,t)$ such that $\partial p/\partial t = 0$ for $\forall (x,t)$. In this case, the Fokker-Planck equation reduces to

$$\hat{L}_{\mathrm{FP}}(x)\, p(x) = -\nabla_a [f^a(x)\, p(x)] + \nabla_a \nabla_b [K^{ab}(x)\, p(x)] = 0.$$

Notice that, by theorem 12, if there exists such a stationary distribution (i.e. a normalized stationary solution), then any initial distribution will finally relax to it.

First, we claim an useful lemma.

Lemma 13. *[Decompostion] Given $\varphi \in \mathcal{C}^1(\mathcal{M} \times \mathbb{R})$, and*

$$h^a(x,t) := f^a(x,t) + K^{ab}(x,t)\, \nabla_b \varphi(x,t) - \nabla_b K^{ab}(x,t),$$

then

$$\nabla_a (h^a(x,t)\, \mathrm{e}^{-\varphi(x,t)}) = -\hat{L}_{\mathrm{FP}}(x,t)\, \mathrm{e}^{-\varphi(x,t)}.$$

Proof. Because of

$$\begin{aligned} h^a(x,t) &= f^a(x,t) + K^{ab}(x,t)\, \nabla_b \varphi(x,t) - \nabla_b K^{ab}(x,t) \\ &= f^a(x,t) - \mathrm{e}^{\varphi(x,t)}\, \nabla_b (K^{ab}(x,t)\, \mathrm{e}^{-\varphi(x,t)}), \end{aligned}$$

we have

$$\begin{aligned} \nabla_a (h^a(x,t)\, \mathrm{e}^{-\varphi(x,t)}) &= \nabla_a (f^a(x,t)\, \mathrm{e}^{-\varphi(x,t)} - \nabla_b (K^{ab}(x,t)\, \mathrm{e}^{-\varphi(x,t)})) \\ &= -\hat{L}_{\mathrm{FP}}(x,t)\, \mathrm{e}^{-\varphi(x,t)}. \end{aligned}$$

Thus proof ends. □

With the aid of this lemma, we claim the result on stationary solution.

Theorem 14. *[Stationary Solution]*

For $\forall \varphi \in \mathcal{C}^1(\mathcal{M})$, $\exp(-\varphi(x))$ is a stationary solution of Fokker-Planck if and only if the

$$h^a(x) := f^a(x) + K^{ab}(x)\, \nabla_b \varphi(x) - \nabla_b K^{ab}(x)$$

satisfies, for $\forall x \in \mathcal{M}$,

$$\mathcal{D}_h \varphi(x) = \nabla_a h^a(x).$$

Proof. Directly from the lemma 13, $\nabla_a(h^a(x)\,\mathrm{e}^{-\varphi(x)}) = -\hat{L}_{\mathrm{FP}}(x)\,\mathrm{e}^{-\varphi(x)}$. □

In the case $K^{ab}(x) = 2\,g^{ab}(x)$, where g^{ab} is the Riemmann metric, we have $h^a = f^a + \nabla^a\varphi$. Thus, $f^a = -\nabla^a\varphi + h^a$. This is an analogy to the Helmholtz's decompostion, except that the h^a may not be divergence-free. Indeed, the divergence of h^a is compactly related to φ. And when h^a does be divergence-free, then the Lie derivative of φ along direction h^a vanishes, that is, h^a goes along the counterline of φ.

From $\nabla_a(h^a(x)\,\mathrm{e}^{-\varphi(x)}) = 0$, we get

$$\begin{aligned}0 &= \nabla_a(h^a\,\mathrm{e}^{-\varphi})\\ &= \nabla_a(f^a\mathrm{e}^{-\varphi} + K^{ab}\,\nabla_b\varphi\mathrm{e}^{-\varphi} - \nabla_b K^{ab}\,\mathrm{e}^{-\varphi})\\ &= \mathrm{e}^{-\varphi}[\nabla_a f^a - f^a\nabla_a\varphi + \nabla_a K^{ab}\nabla_b\varphi + K^{ab}\nabla_a\nabla_b\varphi - K^{ab}\nabla_a\varphi\nabla_b\varphi - \nabla_a\nabla_b K^{ab} + \nabla_a K^{ab}\nabla_b\varphi]\end{aligned}$$

Thus,

$$\mathcal{D}_f\,\varphi = -K^{ab}\nabla_a\varphi\nabla_b\varphi + \nabla_a K^{ab}\nabla_b\varphi - \nabla_a\nabla_b K^{ab} + \nabla_a K^{ab}\nabla_b\varphi + \nabla_a f^a + K^{ab}\nabla_a\nabla_b\varphi.$$

Let $\varphi = E/T$, $K^{ab} = g^{ab}\,T$, we get

$$\mathcal{D}_f\,E = -\nabla^a E\nabla_a E + T\nabla_a[f^a + \nabla^a E] = -\nabla^a E\nabla_a E + T\nabla_a h^a.$$

Let Helmholtz decomposition $f^a = -\nabla^a\varphi + h^a$. How to prove that the relation $\mathcal{D}_h\varphi = 0$. That is, $h^a\nabla_a\varphi = 0$.

$$\mathcal{D}_h E(x) = T\,\nabla_a h^a(x).$$

As $T \to 0_+$, $\mathcal{D}_h E(x) = 0$. At $\nabla_a E(x_\star) = 0$, $h^a(x)\nabla_a E(x) = h^a(x_\star)\,\nabla_a\nabla_b E(x_\star)\Delta x^b + \mathcal{O}(\Delta x^2)$, held for any tiny Δx. Once $\nabla_a\nabla_b E(x_\star)$ is irreducible, we have $h^a(x_\star) = 0$.