

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

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

- 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 $\Delta := \nabla_a \nabla^a$.

1 Lyapunov Function

1.1 Definition

Definition 2. [Lyapunov Function]

Given an autonomous¹ ODE,

$$\frac{dx^a}{dt} = f^a(x),$$

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

Along the phase trajectory, a Lyapunov function monotonically 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, or approximate to, it 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 SDE², as

$$dX^a = f^a(X) dt + \sqrt{2T} dW^a,$$

where $dW^a \sim \mathcal{N}(0, \delta^{ab} dt)$ and parameter $T > 0$. Then, we sample an ensemble 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

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

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 ?.

³. E.g., Gelman-Rubin-Brooks plot.

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

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

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

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

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,

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

Thus, proof ends. \square

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*

$$dX^a = -\nabla^a E(x) dt + \sqrt{2T} dV^a,$$

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

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

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

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,

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

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

Even though the θ is keep changing during the gradient descent process, as long as it's controlled so as to be slowly varying, we can use the same strategy as the persistent contrastive divergence trick to simplify the computation. So, during the gradient descent steps, we employ two distinct sets of Markov chains that are consistently evolving. The first is generated by the SDE to p_D , and the second by the SDE to q_E . Along the gradient descent steps of θ , on the chains to p_D E is suppressed, while on the chains to q_E E is elevated. Gradient descent stops when the two parts balance, where q_E fits p_D best.

Finally, we claim that the E we find at the best fit θ_* is a Lyapunov function of the original autonomous ODE. By lemma 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, $p \rightarrow p_D \approx q_E$ where $\partial p / \partial t \rightarrow 0$, we have

$$\begin{aligned} 0 &= -\nabla_a[q_E(x) f^a(x)] + T \Delta q_E(x) \\ \{q_E = \dots\} &= -\nabla_a[e^{-E(x)/T} f^a(x)] + T \Delta e^{-E(x)/T} \\ &= \frac{1}{T} e^{-E(x)/T} \nabla_a E(x) f^a(x) - e^{-E(x)/T} \nabla_a f^a(x) \\ &\quad + \frac{1}{T} e^{-E(x)/T} \delta^{ab} \nabla_b E(x) \nabla_a E(x) - e^{-E(x)/T} \delta^{ab} \nabla_a \nabla_b E(x) \\ &= \frac{1}{T} e^{-E(x)/T} \times \{ \nabla_a E(x) f^a(x) + \delta^{ab} \nabla_b E(x) \nabla_a E(x) - T [\nabla_a f^a(x) + \delta^{ab} \nabla_a \nabla_b E(x)] \} \end{aligned}$$

Thus,

$$\nabla_a E(x) f^a(x) + \delta^{ab} \nabla_a E(x) \nabla_b E(x) = T [\nabla_a f^a(x) + \Delta E(x)].$$

As $T \rightarrow 0$, the SDE reduces to the original autonomous ODE, and we arrive at

$$\nabla_a E(x) f^a(x) = -\delta^{ab} \nabla_a E(x) \nabla_b E(x) \leq 0,$$

where equality holds if and only if $\nabla_a E(x) = 0$. Thus, $E(x)$ is a Lyapunov function of $dx^a/dt = f^a(x)$.

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

We summarize this method as three steps:

1. generate stationary empirical distribution p_D by the SDE of $f^a(x)$;
2. fit the empirical distribution p_D by an analytic distribution q_E induced by a parameterized function E , until $q_E \approx p_D$;
3. claim: let $T \rightarrow 0$ and go back to the original ODE, we find $\nabla_a E(x) f^a(x) = -\delta^{ab} \nabla_a E(x) \nabla_b E(x) \leq 0$.

1.3 Parameterized ODE

Next, we consider parameterized ODE, that is, $f(x) \rightarrow f(x; \varphi)$, 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

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

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

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 ,

$$\frac{dx}{dt} = r - x.$$

It is equivalent to an augmented ODE

$$\begin{aligned} \frac{dx}{dt} &= r - x; \\ \frac{dr}{dt} &= 0. \end{aligned}$$

And, it has a Lyapunov function

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

Indeed, since $(\partial E / \partial x)(x, r) = x - r$, $(\partial E / \partial x)(x, r) f(x) = -(r - x)^2 \leq 0$. Given r , it's a valley centered at r along x -axis. 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 $d\sigma_\varphi / dt \sim 1 / \sqrt{t} \rightarrow 0$, as $t \rightarrow +\infty$, 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 $q_E(x) = \lim_{t \rightarrow +\infty} p_D(x, t)$ again.

The parameterized ODE

$$\frac{dx^a}{dt} = f^a(x; \varphi)$$

is equivalent to the augmented ODE without parameter $\frac{dy^a}{dt} = F^a(y)$, 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 $\partial p / \partial t \rightarrow 0$ and $p \rightarrow q_E$, and as $T \rightarrow 0$,

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

We arrive at

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

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

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

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

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

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

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

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

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

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

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

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

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

□

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 $\Sigma^{ab}(x, t)$, and the walk steps

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

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 $dn/dt \rightarrow +\infty$,

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

where

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

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 $n(t + \Delta t) - n(t) = \frac{dn}{dt}(x, t) \Delta t + o(\Delta t)$, 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, or stochastic differential equations (SDE), is defined by two parts. The first is deterministic, and the second is a random walk. Precisely,

Definition 8. Given $f^a(x, t)$, $g_b^a(x, t)$, and $\Sigma^{ab}(x, t)$ on $\mathcal{M} \times \mathbb{R}$, stochastic differential equations is defined as

$$dx^a = f^a(x, t) dt + g_b^a(x, t) dW^b(x, t),$$

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

Lemma 9. [Macroscopic Landscape]

Consider an ensemble of particles, randomly sampled at an initial time, evolving along a SDE. 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

$$\frac{\partial p}{\partial t}(x, t) = -\nabla_a[p(x, t) f^a(x, t)] + \frac{1}{2} \nabla_a \nabla_b[p(x, t) K^{ab}(x, t)],$$

where $K^{ab} := g_c^a(x, t) g_d^b(x, t) \Sigma^{cd}(x, t)$.

Proof. From the difference of the SDE,

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

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 $dW^a(x, t)$ is a random walk, $\langle \Delta W^a(x, t) \rangle_{\Delta W(x, t)} = 0$. Then the term is

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

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 $n \geq 3$, 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 $\Delta t \rightarrow 0$, 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. □