Math 612: Single Cell Analysis **2019W Term 1**

Lecture 16: October 29

Lecturer: Geoffrey Schiebinger *Scribe: Qiong Zhang*

Gradient Flow in Wasserstein Space

Last time, we talked about the Wasserstein geodesic on a line and the curve of probability measures in the Wasserstein space associated with the vector field in the state space. The curve of the probability measures models single cell movement. However, how could we model the movement of cells when there are cell-cell interactions such as the cell touch in the real 3D space? The gradient flow in Wasserstein space, which we are going to cover in today's lecture, can be used to model the movement of cells when there are cell-cell interactions.

16.1 Gradient Flow in Euclidean Space

The **gradient descent** is a numerical algorithm for finding the minimum of a differentiable function $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$. It produces a sequence of points x_k that converges to a local minimum of f. At kth step, let

$$x_k \leftarrow x_{k-1} - \eta_k \underbrace{\nabla f(x_{k-1})}_{\substack{\text{direction of}\\ \text{steepest descent}}} .$$

where η_k is the step size. The kth step can be rewritten as $\frac{x_k - x_{k-1}}{\eta_k} = -\nabla f(x_{k-1})$. For sufficiently small step size η_k, we then have

$$x'(t) = -\nabla f(x(t)),\ x(0) = x_0 \tag{16.1}$$

Definition 16.1 (Gradient Flow) *A curve $x(t) : [0, T] \to \mathbb{R}^d$ is the gradient flow according to f if it is a solution to the differential equation* (**??**).

By definition, we can see that gradient flow is a curve that depicts the direction of steepest descent of f. For example, in Figure **??**, the solid line is the gradient descent step with large step size. As the step size becomes extremely small, the dahsed line is the gradient flow according to f.

In the above definition for gradient flow, the function f is defined over the Euclidean space, if we have some functional F defined over a probability measure space, how to generalize the definition of gradient flow? The current definition via the differential equation is not applicable as ∇F is not well-defined. We view the gradient descent from a different point of view, which can be generalized for non-differentiable f, hopefully, it can help us to generalize the definition of gradient flow according to F.

The gradient descent algorithm is only applicable for f that is differentiable. For non-differentiable functions, a generalization of the gradient descent is called the proximal method.

Definition 16.2 (Proximal Method) *The procedure of proximal method is as follows. At step k,*

$$\begin{aligned} &x_k = \arg\min_x f(x) \\ &\textit{subject to } \|x - x_{k-1}\| \le \epsilon. \end{aligned} \tag{16.2}$$

We show that for differentiable f, the procedure of proximal method reduces to the procedure of gradient descent. The Lagrangian of (**??**) is

$$L(x, \lambda) = f(x) + \lambda \|x - x_{k-1}\|^2$$

At optimality, we have

$$\nabla f(x) + 2\lambda(x - x_{k-1}) = 0.$$

For sufficiently small ϵ, we have $\nabla f(x) \approx \nabla f(x_{k-1})$. Therefore, at optimality we approximately $\nabla f(x_{k-1}) + 2\lambda(x - x_{k-1}) = 0$ which implies

$$x_k = x_{k-1} - \frac{1}{2\lambda}\nabla f(x_{k-1}).$$

Therefore, for differentiable f, the update based on the proximal method is a gradient descent step with a particular choice of step size. Hence, the **take away message** is that proximal methods generalize gradient descent to non-differential optimization problems. As we shall in see in Example **??**, the proximal method allows us to generalize the gradient flow over functional without a need to define the gradient.

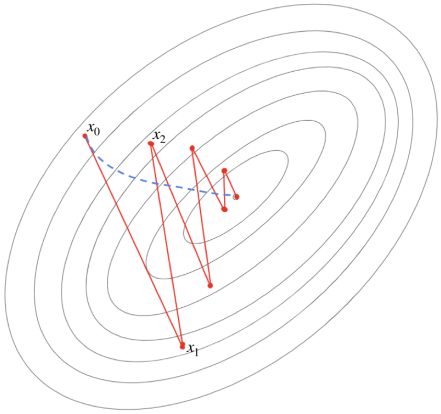

Figure 16.1: An illustration of gradient descent and gradient flow according to some function.

In 1998, [JKO1998] defined gradient flow in $W_2(\mathcal{X})$. In their work, the gradient flow for functional $F : W_2(\mathcal{X}) \to \mathbb{R}$ is defined through the Fokker-Planck equation. The Fokker-Plack equation plays a central rule in statistical physics as a solution to the Fokker–Planck equation represents the probability density for the position of a particle whose motion is described by a corresponding Ito stochastic differential equation. Below is an example of the gradient flow according to negative entropy in $W_2(\mathcal{X})$ and a description of the Fokker-Planck equation is given in the next section.

Example 16.3 (Gradient Flow according to Negative Entropy) *Let ρ be a probability density function, consider the classical diffusion equation*

$$\frac{\partial \rho}{\partial t} = \Delta \rho$$

which is a Fokker-Planck equation associated with a standard Brownian motion. [JKO1998] shows that

$$\rho^{(k)} = \arg\min_{\rho} \frac{1}{2} W_2^2(\rho^{(k-1)}, \rho) + h \int \rho(x) \log \rho(x) dx$$

is a discretization of the diffucsion equation for some step size h. As a comparison to the proximal method we described above, the discretization is very similar to the procedure of proximal method except that the distance is changed from Euclidean distance to Wasserstein distance. Therefore, this allows us to regard the diffusion equation as a steepest descent of the functional $\int \rho(x) \log \rho(x) dx$ with respect to the Wasserstein metric.

16.2 Fokker-Planck Equation

[JKO1998] considers the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \nabla \psi) + \beta^{-1} \Delta \rho, \qquad \rho(x, 0) = \rho^0(x) \tag{16.3}$$

where β^{-1} is a "temperature parameter", ψ is a poential energy function. It is well known that the Fokker-Planck equation is inherently related to the stochastic differential equation

$$dX(t) = -\nabla\psi(X(t))dt + \sqrt{2\beta^{-1}}d\mathbb{B}_t, \qquad X(0) = X^0 \tag{16.4}$$

where X^0 is a random vector with probability density ρ^0. This models a particle doing Brownian motion with drift. The solution $\{X_t\}_{t=0}^{\infty}$ starting from X^0 is a continuous time Markov process.

Example 16.4 *If $\psi \equiv 0$, then $X_t = \mathbb{B}_t$ is a Brownian motion where Law$(X_{t_2}|X_{t_1} = x) = N(x, \sigma(t_2 - t_1))$ for some function σ.*

Example 16.5 *When $\beta^{-1} = 0$, X_t is deterministic.*

The $\rho(x,t)$ in Fokker-Planck equation is the density describing where the particle is at time t. One fact is that if ψ is "smooth enough" and "grows quickly", then there is a unique stationary solution $\rho_s(x)$ of the Fokker-Planck equation

$$\rho_s(x) = Z^{-1}\exp(-\beta\psi(x))$$

where $Z = \int \exp(-\beta\psi(x))dx$. ψ must grow rapidly enough to ensure that Z is finite. Then $\rho_s(x)dx$ is called the Gibbs measure and it ρ_s minimizes the **free energy functional**

$$F(\rho) = E(\rho) + \beta^{-1}S(\rho)$$

where

$$E(\rho) = \int \psi(x)\rho(x)dx$$
$$S(\rho) = \int \rho(x)\log\rho(x)dx$$

Then the solution $\rho(x,t)$ to the Fokker-Planck equation is the gradient flow of the free energy functional F. The free energy can only decrease in time for any solution $\rho(x,t)$ to Fokker-Planck equation.

16.3 Gradient Flow in Wasserstein Space

What is a gradient flow on $W_2(\mathcal{X})$? Define the discrete iterative scheme:

$$(\star)\begin{cases}\text{Start with } \rho^0 \\ \rho^{k+1} = \arg\min_\rho\{\frac{1}{2}W_2^2(\rho,\rho^k) + \eta F(\rho)\}\end{cases}$$

Theorem 16.6 *Let ρ^0 satisfy $F(\rho^0) < \infty$, ψ grows quickly, and for $\eta > 0$. Let $\{\rho_\eta^k\}_{k=0}^{\infty}$ be the iterates from $(\star)$. Define a curve*

$$\rho_\eta(t) = \rho_\eta^{(k)} \quad \text{for} \quad t \in [k\eta, (k+1)\eta)$$

*Then as $\eta \downarrow 0$, $\rho_\eta(t) \to \rho(t)$ weakly where $\rho(t)$ is the unique solution to Fokker-Planck equation (**??**).*

How does the gradient flow in the Wasserstein space related with the course? For example, ψ can be the potential of the Waddington's landscape. We could model the process with cell-cell interactions via

$$dX_t = -\nabla \underbrace{\psi(X_t, \mathbb{P}_t)}_{\substack{\text{Interaction}\\ \text{potential}}} + \epsilon d\mathbb{B}_t$$

The distribution $\mathbb{P}_t$ follows a gradient flow for a more general "free energy" $F(\mathbb{P})$. For example, for the pairwise interactions, $F(\mathbb{P}) = \int \psi(x,y)d\mathbb{P}(x)d\mathbb{P}(y)$.

References

[1] Richard Jordan, David Kinderlehrer, and Felix Otto. ”The variational formulation of the Fokker–Planck equation.” SIAM journal on mathematical analysis 29.1 (1998): 1-17.