Discussing Gradient Flows

Renaud

Informal reading group on Optimal Transport
May 25, 2020

Plan

I will informally discuss:

Plan

I will informally discuss:

- Gradient flows in $\mathbb{R}^N$;

Plan

I will informally discuss:

- Gradient flows in $\mathbb{R}^N$;
- 3 approaches to gradient flows in more general spaces;

Plan

I will informally discuss:

- Gradient flows in $\mathbb{R}^N$;
- 3 approaches to gradient flows in more general spaces;
- Gradient flows in $\mathbb{W}_2$.

In $\mathbb{R}^N$

General first order ODE in $\mathbb{R}^N$:

$$\dot{x}(t) = F(x(t))$$

where $t \mapsto x(t)$ is a curve, $t \mapsto \dot{x}(t)$ the tangent, $x \mapsto F(x)$ a vector field.

In $\mathbb{R}^N$

General first order ODE in $\mathbb{R}^N$:

$$\dot{x}(t) = F(x(t))$$

where $t \mapsto x(t)$ is a curve, $t \mapsto \dot{x}(t)$ the tangent, $x \mapsto F(x)$ a vector field.

Gradient flows in $\mathbb{R}^N$:

$$\dot{x}(t) = -\nabla V(x(t))$$

where $t \mapsto x(t)$ is a curve, $t \mapsto \dot{x}(t)$ the tangent, $x \mapsto V(x)$ a scalar function, called a "potential".

In $\mathbb{R}^N$

General first order ODE in $\mathbb{R}^N$:

$$\dot{x}(t) = F(x(t))$$

where $t \mapsto x(t)$ is a curve, $t \mapsto \dot{x}(t)$ the tangent, $x \mapsto F(x)$ a vector field.

Gradient flows in $\mathbb{R}^N$:

$$\dot{x}(t) = -\nabla V(x(t))$$

where $t \mapsto x(t)$ is a curve, $t \mapsto \dot{x}(t)$ the tangent, $x \mapsto V(x)$ a scalar function, called a "potential".

Newton's second law for conservative force + friction:

$$m\ddot{x}(t) = -\nabla V(x(t)) - \gamma \dot{x}(t).$$

Taking $m \to 0$, we (formally) recover a gradient flow.

In more general spaces

There are at least 3 approaches to generalizing gradient flows to more general spaces:

In more general spaces

There are at least 3 approaches to generalizing gradient flows to more general spaces:

- Riemannian geometry;

In more general spaces

There are at least 3 approaches to generalizing gradient flows to more general spaces:

- Riemannian geometry;
- Metric derivatives and upper gradients;

In more general spaces

There are at least 3 approaches to generalizing gradient flows to more general spaces:

- Riemannian geometry;
- Metric derivatives and upper gradients;
- Discretization in time.

Riemannian geometry

To consider

$$\dot{x}(t) = -\nabla V(x(t)),$$

we would want to:

Riemannian geometry

To consider

$$\dot{x}(t) = -\nabla V(x(t)),$$

we would want to:

- give a meaning to $\dot{x}(t)$ (cannot simply substract points on $\mathcal{M}$);

Riemannian geometry

To consider

$$\dot{x}(t) = -\nabla V(x(t)),$$

we would want to:

- give a meaning to $\dot{x}(t)$ (cannot simply substract points on $\mathcal{M}$);
- give a meaning to $\nabla V(x(t))$ (which is of the same nature as $\dot{x}(t)$).

Riemannian geometry

To consider

$$\dot{x}(t) = -\nabla V(x(t)),$$

we would want to:

- give a meaning to $\dot{x}(t)$ (cannot simply substract points on $\mathcal{M}$);
- give a meaning to $\nabla V(x(t))$ (which is of the same nature as $\dot{x}(t)$).

In Riemannian geometry $(\mathcal{M}, g)$,

Riemannian geometry

To consider

$$\dot{x}(t) = -\nabla V(x(t)),$$

we would want to:

- give a meaning to $\dot{x}(t)$ (cannot simply substract points on $\mathcal{M}$);
- give a meaning to $\nabla V(x(t))$ (which is of the same nature as $\dot{x}(t)$).

In Riemannian geometry $(\mathcal{M}, g)$,

- $\dot{x}(t)$ is a tangent vector to $x(t)$ and is defined — like any tangent vector — by its action on functions $f : \mathcal{M} \to \mathbb{R}$

$$(\dot{x}(t))(f) := \frac{\mathrm{d}}{\mathrm{d}s} f(x(s))\Big|_{s=t};$$

Riemannian geometry

To consider

$$\dot{x}(t) = -\nabla V(x(t)),$$

we would want to:

- give a meaning to $\dot{x}(t)$ (cannot simply substract points on $\mathcal{M}$);
- give a meaning to $\nabla V(x(t))$ (which is of the same nature as $\dot{x}(t)$).

In Riemannian geometry $(\mathcal{M}, g)$,

- $\dot{x}(t)$ is a tangent vector to $x(t)$ and is defined — like any tangent vector — by its action on functions $f \colon \mathcal{M} \to \mathbb{R}$

$$(\dot{x}(t))(f) := \frac{\mathrm{d}}{\mathrm{d}s} f(x(s))\Big|_{s=t};$$

- an inner product g defined on pairs of tangent vectors provides a nice way of defining a vector field from V by asking that

$$g(\nabla V, Z) := Z(V) \qquad \forall Z.$$

More general metric spaces

Sadly, we can't always define those objects (or maybe we can, but it's messy)... In many metric spaces though, we can still

More general metric spaces

Sadly, we can't always define those objects (or maybe we can, but it's messy)... In many metric spaces though, we can still

- define $|\dot{x}(t)|$ as a scalar function (the metric derivative in Gabriel's presentation);

More general metric spaces

Sadly, we can't always define those objects (or maybe we can, but it's messy)... In many metric spaces though, we can still

- define $|\dot{x}(t)|$ as a scalar function (the metric derivative in Gabriel's presentation);
- define $|\nabla V(x)|$ as a scalar function (upper gradient $\sim$ local Lipschitz constant);

More general metric spaces

Sadly, we can't always define those objects (or maybe we can, but it's messy)... In many metric spaces though, we can still

- define $|\dot{x}(t)|$ as a scalar function (the metric derivative in Gabriel's presentation);
- define $|\nabla V(x)|$ as a scalar function (upper gradient $\sim$ local Lipschitz constant);
- interpret
$$\dot{x}(t) = -\nabla V(x(t))$$
as
$$(V \dot{\circ} x)(t) \leq -\frac{1}{2}|\dot{x}(t)| - \frac{1}{2}|\nabla V(x(t))|^2$$
(check that this is equivalent in $\mathbb{R}^N$).

Discretization in time

Again in rather general metric spaces, we can

Discretization in time

Again in rather general metric spaces, we can

1. Divide time into intervals of small length τ

Discretization in time

Again in rather general metric spaces, we can

1. Divide time into intervals of small length τ
2. Ask that the "discrete trajectory" $(x_k^\tau)_{k\in\mathbb{N}}$ satisfies

$$x_{k+1}^\tau \in \operatorname*{argmin}_m \left(V(x) + \frac{1}{2\tau} d(x, x_k^\tau)^2 \right);$$

Discretization in time

Again in rather general metric spaces, we can

1. Divide time into intervals of small length τ
2. Ask that the "discrete trajectory" $(x_k^\tau)_{k\in\mathbb{N}}$ satisfies

$$x_{k+1}^\tau \in \operatorname*{argmin}_m \left(V(x) + \frac{1}{2\tau} d(x, x_k^\tau)^2 \right);$$

3. Hope for convergence as $\tau \to 0$.

Discretization in time

Again in rather general metric spaces, we can

1. Divide time into intervals of small length τ
2. Ask that the "discrete trajectory" $(x_k^\tau)_{k\in\mathbb{N}}$ satisfies

$$x_{k+1}^\tau \in \operatorname*{argmin}_m \left(V(x) + \tfrac{1}{2\tau} d(x, x_k^\tau)^2 \right);$$

3. Hope for convergence as $\tau \to 0$.

To see why this makes sense, note that, in $\mathbb{R}^N$,

$$x_{k+1}^\tau \in \operatorname*{argmin}_x \left(V(x) + \tfrac{1}{2\tau} d(x, x_k^\tau)^2 \right) \quad \leftrightsquigarrow \quad 0 = \nabla V(y) + \tfrac{1}{\tau}(x_{k+1}^\tau - x_k^\tau)$$

is an implicit Euler scheme for $\dot{x} = -\nabla V(x)$.

Gradient flows in $\mathbb{W}_2$

Santambrogio chooses to go through discretization in time, but one can use either of the other two!

Gradient flows in $\mathbb{W}_2$

Santambrogio chooses to go through discretization in time, but one can use either of the other two!

- See [Otto. The geometry of dissipative evolution equations: the porous medium equation, Commun. Partial Differ. Equ., 26 (2001)] or [Lott. Some geometric calculations on Wasserstein space, arXiv:math/0612562v2 (2007)] for the Riemannian approach.

Gradient flows in $\mathbb{W}_2$

Santambrogio chooses to go through discretization in time, but one can use either of the other two!

- See [Otto. The geometry of dissipative evolution equations: the porous medium equation, Commun. Partial Differ. Equ., 26 (2001)] or [Lott. Some geometric calculations on Wasserstein space, arXiv:math/0612562v2 (2007)] for the Riemannian approach.
- See [Ambrosio, Gigli, Savaré. *Gradient Flows*, Birkhäuser (2005)] for the approach with metric derivatives.

Gradient flows in $\mathbb{W}_2$

Our metric space is the space $\mathbb{W}_2$ of prob. meas. with finite variance, equipped with the 2-Wasserstein distance

$$W_2(\mu,\nu)^2 := \inf_{\gamma\in\Gamma(\mu,\nu)} \int d(x,y)^2 \gamma(\mathrm{d}x,\mathrm{d}y)$$

and we want to perform study the gradient flow for a function(al) $\mathcal{F} : \mathbb{W}_2 \to \mathbb{R} \cup \{+\infty\}$.

Gradient flows in $\mathbb{W}_2$

Our metric space is the space $\mathbb{W}_2$ of prob. meas. with finite variance, equipped with the 2-Wasserstein distance

$$W_2(\mu,\nu)^2 := \inf_{\gamma\in\Gamma(\mu,\nu)} \int d(x,y)^2 \gamma(\mathrm{d}x,\mathrm{d}y)$$

and we want to perform study the gradient flow for a function(al) $\mathcal{F} : \mathbb{W}_2 \to \mathbb{R} \cup \{+\infty\}$.

I will completely ignore issues of absolute continuity, smoothness and boundary conditions. The presentation will be full of half-truths and you should really read the book :) .

Gradient flows in $\mathbb{W}_2$

When considering

$$\mu_{k+1}^\tau \in \operatorname{argmin}_\mu \left(\mathcal{F}(\mu) + \frac{1}{2\tau} W_2(\mu,\nu)^2\right),$$

the first variations (Gabriel's presentation) of $\mathcal{F}$ and W_2^2 appear

$$\rightsquigarrow \qquad \frac{\delta F}{\delta \mu} + \frac{\phi}{\tau} = \text{constant}$$

where ϕ is the Kantorovich potential for the optimal transport from μ to μ_k^τ.

Gradient flows in $\mathbb{W}_2$

When considering

$$\mu_{k+1}^\tau \in \operatorname{argmin}_\mu \left(\mathcal{F}(\mu) + \tfrac{1}{2\tau} W_2(\mu,\nu)^2\right),$$

the first variations (Gabriel's presentation) of $\mathcal{F}$ and W_2^2 appear

$$\rightsquigarrow \qquad \frac{\delta F}{\delta \mu} + \frac{\phi}{\tau} = \text{constant}$$

where ϕ is the Kantorovich potential for the optimal transport from μ to μ_k^τ.
Differentiating,

$$\nabla \frac{\delta F}{\delta \mu} = -\frac{\nabla \phi}{\tau} = \frac{T - \mathrm{id}}{\tau}.$$

Gradient flows in $\mathbb{W}_2$

When considering

$$\mu_{k+1}^\tau \in \text{argmin}_\mu \left(\mathcal{F}(\mu) + \tfrac{1}{2\tau} W_2(\mu,\nu)^2 \right),$$

the first variations (Gabriel's presentation) of $\mathcal{F}$ and W_2^2 appear

$$\rightsquigarrow \qquad \frac{\delta F}{\delta \mu} + \frac{\phi}{\tau} = \text{constant}$$

where ϕ is the Kantorovich potential for the optimal transport from μ to μ_k^τ.
Differentiating,

$$\nabla \frac{\delta F}{\delta \mu} = -\frac{\nabla \phi}{\tau} = \frac{T - \text{id}}{\tau}.$$

As $\tau \to 0$, this can hopefully be thought of as coming from a continuity equation on $\mathbb{R}^N$

$$\rightsquigarrow \qquad \dot{\mu} = \text{div}\left(\mu \nabla \frac{\delta F}{\delta \mu}\right).$$

An Important example

Consider

$$\mathcal{F}(\mu) = -\Theta\,\mathsf{Ent}(\mu) + \int V(x)\mu(\mathrm{d}x).$$

for some constant $\Theta > 0$.

An Important example

Consider

$$\mathcal{F}(\mu) = -\Theta\,\mathsf{Ent}(\mu) + \int V(x)\mu(\mathrm{d}x).$$

for some constant $\Theta > 0$. We saw last time that

$$\frac{\delta \mathcal{F}}{\delta \mu} = \Theta + \Theta \log \mu + V$$

An Important example

Consider

$$\mathcal{F}(\mu) = -\Theta \, \mathrm{Ent}(\mu) + \int V(x)\mu(\mathrm{d}x).$$

for some constant $\Theta > 0$. We saw last time that

$$\frac{\delta \mathcal{F}}{\delta \mu} = \Theta + \Theta \log \mu + V$$

so that

$$\mu \nabla \frac{\delta F}{\delta \mu} = \Theta \nabla \mu + \mu \nabla V$$

An Important example

Consider

$$\mathcal{F}(\mu) = -\Theta\,\mathrm{Ent}(\mu) + \int V(x)\mu(\mathrm{d}x).$$

for some constant $\Theta > 0$. We saw last time that

$$\frac{\delta \mathcal{F}}{\delta \mu} = \Theta + \Theta \log \mu + V$$

so that

$$\mu \nabla \frac{\delta F}{\delta \mu} = \Theta \nabla \mu + \mu \nabla V$$

and

$$\dot{\mu} = \mathrm{div}(\Theta \nabla \mu + \mu \nabla V) = \Theta \Delta \mu - \mathrm{div}\left(\mu(-\nabla V)\right).$$

This is the evolution of the probability density for the SDE

$$\mathrm{d}x = -\nabla V(x)\,\mathrm{d}t + \sqrt{2\Theta}\,\mathrm{d}w.$$