Gradient Flow in the Wasserstein Metric

Katy Craig
University of California, Santa Barbara

JMM, AMS Metric Geometry & Topology
January 11th, 2018
### Gradient Flow and PDE

#### Examples:

<table>
<thead>
<tr>
<th>Metric</th>
<th>Energy</th>
<th>Gradient Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(L^2(\mathbb{R}^d), | \cdot |_{L^2})$</td>
<td>$E(f) = \frac{1}{2} \int</td>
<td>\nabla f</td>
</tr>
<tr>
<td>$(\mathcal{P}_2(\mathbb{R}^d), W_2)$</td>
<td>$E(\rho) = \frac{1}{2} \int K * \rho d\rho + \int V d\rho + \frac{1}{m-1} \int \rho^m$</td>
<td>$\frac{d}{dt} \rho = \nabla \cdot ((\nabla K * \rho) \rho) + \nabla \cdot (\nabla V \rho) + \Delta \rho^m$</td>
</tr>
</tbody>
</table>
Wasserstein metric

• Given two probability measures \( \mu \) and \( \nu \) on \( \mathbb{R}^d \), \( t : \mathbb{R}^d \to \mathbb{R}^d \) transports \( \mu \) onto \( \nu \) if \( \nu(B) = \mu(t^{-1}(B)) \). Write this as \( t \# \mu = \nu \).

• The Wasserstein distance between \( \mu \) and \( \nu \in P_{2,\text{ac}}(\mathbb{R}^d) \) is

\[
W_2(\mu, \nu) := \inf \left\{ \left( \int |t(x) - x|^2 d\mu(x) \right)^{1/2} : t \# \mu = \nu \right\}
\]

[Diagram showing \( \mu \) being transported to \( \nu \) by \( t(x) \)]

- Effort to rearrange \( \mu \) to look like \( \nu \), using \( t(x) \)
- \( t \) sends \( \mu \) to \( \nu \)
geodesics

Not just a metric space... a **geodesic metric space**: there is a constant speed geodesic \( \sigma : [0, 1] \to P_2(\mathbb{R}^d) \) connecting any \( \mu \) and \( \nu \).

\[
\sigma(0) = \mu, \quad \sigma(1) = \nu, \quad W_2(\sigma(t), \sigma(s)) = |t - s| W_2(\mu, \nu)
\]

**Monge**

\( \mu \)

Wasserstein geodesic \( \sigma(t) \)

**Kantorovich**

\( \nu \)

\( L^2 \) geodesic \( (1 - t)\mu + t\nu \)
convexity

Since the Wasserstein metric has geodesics, it has a notion of convexity.

Recall: \( E: L^2(\mathbb{R}^d) \to \mathbb{R} \) is \( \lambda \)-convex if

\[
E((1 - t)f + tg) \leq (1 - t)E(f) + tE(g)
\]

For any \( g \in L^2(\mathbb{R}^d) \), \( E(f) = \|f - g\|_2^2 \) is 2-convex \( \implies L^2 \) is NPC.

Likewise, in the Wasserstein metric, \( E: P_2(\mathbb{R}^d) \to \mathbb{R} \) is \( \lambda \)-convex if

\[
E(\sigma(t)) \leq (1 - t)E(\mu) + tE(\nu) - t(1 - t)\frac{\lambda}{2}W_2^2(\mu, \nu)
\]

For any \( \nu \in P_2(\mathbb{R}^d) \), \( E(\mu) = W_2^2(\mu, \nu) \) is 2-concave \( \implies W_2 \) is PC.
gradient flow

We want to define the gradient flow as \[ \frac{d}{dt} \rho(t) = -\nabla_{W_2} E(\rho(t)) \], but without a Riemannian structure, we don’t have a notion of gradient.

- Given \( E: P_2(\mathbb{R}^d) \to \mathbb{R} \), its \textbf{local slope} is:
  \[ |\partial E|(\mu) := \limsup_{\nu \to \mu} \frac{(E(\mu) - E(\nu))^+}{W_2(\mu, \nu)} \]

- Given \( \rho: [0,T] \to P_2(\mathbb{R}^d) \), its \textbf{metric derivative} is:
  \[ |\rho'| (t) = \lim_{s \to t} \frac{W_2(\rho(s), \rho(t))}{|s - t|} \]

\textbf{DEF:} \( \rho(t): \mathbb{R} \to P_2(\mathbb{R}^d) \) is the \textbf{Wasserstein gradient flow} of \( E:P_2(\mathbb{R}^d) \to \mathbb{R} \) if
\[ \frac{d}{dt} E(\rho(t)) \leq -\frac{1}{2} |\partial E(\rho(t))| - \frac{1}{2} |\rho'| (t) \]
**Wasserstein gradient flow**

**DEF:** \( \rho(t) : \mathbb{R} \rightarrow P_2(\mathbb{R}^d) \) is the Wasserstein gradient flow of \( E : P_2(\mathbb{R}^d) \rightarrow \mathbb{R} \) if

\[
\frac{d}{dt} E(\rho(t)) \leq -\frac{1}{2} |\partial E(\rho(t))| - \frac{1}{2} |\rho'| (t)
\]

Analogy with \( L^2 \) gradient flow:

Abbreviating \( \nabla_{L^2} \) by \( \nabla \),

\[
\frac{d}{dt} f(t) = -\nabla E(f(t)) \iff \begin{cases} 
\left| \frac{d}{dt} f(t) \right| = |\nabla E(f(t))| \\
\frac{d}{dt} E(f(t)) = -|\nabla E(f(t))| \left| \frac{d}{dt} f(t) \right| 
\end{cases}
\]

\[
\frac{d}{dt} E(f(t)) \leq -\frac{1}{2} |\nabla E(f(t))| - \frac{1}{2} \left| \frac{d}{dt} f(t) \right|
\]
**Gradient flow and PDE**

\[ \frac{d}{dt} x(t) = -\nabla_x E(x(t)) \]

**Good news:** gradient flows structure is very useful in PDE

- existence
- uniqueness
- approximation
- stability

**Bad news:** Wasserstein metric has more complicated geometry

<table>
<thead>
<tr>
<th>(L^2)</th>
<th>Wasserstein metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riemannian manifold</td>
<td>metric space</td>
</tr>
<tr>
<td>non-positively curved</td>
<td>positively curved</td>
</tr>
</tbody>
</table>
time discretization: $L^2$

Analogous results hold in any NPC metric space [Mayer, ’98], [CL ’71].

What about when the metric space isn’t NPC?

Assume: $E$ is $\lambda$-convex. Since $L^2(\mathbb{R}^d)$ is NPC, $\Phi$ is $\frac{1}{\tau} + \lambda$-convex.

Prop: $\|f_n - \tilde{f}_n\|_2 \leq \frac{1}{1 + \lambda \tau} \|f_{n-1} - \tilde{f}_{n-1}\|_2$

Thm: For $\tau = \frac{t}{n}$, $\|f(t) - f_n\|_2 \leq \frac{C}{\sqrt{n}}$, $\|f(t) - \tilde{f}(t)\|_2 \leq e^{-\lambda t} \|f(0) - \tilde{f}(0)\|_2$

---

contraction inequality  
time discretization
time discretization: $W_2$

Gradient flow

\[ \frac{d}{dt} E(\rho(t)) \leq -\frac{1}{2} |\partial E(\rho(t))| - \frac{1}{2} |\rho'|(t) \]

\[ \rho(0) = \mu \]

Time discretization (JKO)

\[ \rho_n = \arg\min_{\nu} \left\{ \frac{1}{2\tau} W_2^2(\nu, \rho_{n-1}) + E(\nu) \right\} \]

\[ \rho_0 = \mu \]

**Assume:** $E$ is bounded below and $\lambda$-convex along generalized geodesics.

Then $\Phi(\nu) = \frac{1}{2\tau} W_2^2(\nu, \rho_{n-1}) + E(\nu)$ is $\frac{1}{\tau} + \lambda$-convex along gen geodesics.

**Thm:** For $\tau = \frac{t}{n}$,

\[ W_2(\rho(t), \rho_n) \leq \frac{C}{\sqrt{n}} \]

\[ W_2(\rho(t), \tilde{\rho}(t)) \leq e^{-\lambda t} W_2(\rho(0), \tilde{\rho}(0)) \]

**Prop:**

\[ W_2(\rho_n, \tilde{\rho}_n) \leq \frac{1}{1 + \lambda \tau} W_2(\rho_{n-1}, \tilde{\rho}_{n-1}) + O(\tau^2) \]

Overcome $W_2$ geometry issues... what about when $E$ isn’t $\lambda$-convex?
**ω-convexity**

**Recall:**

E: \( P_2(\mathbb{R}^d) \to \mathbb{R} \) is \( \lambda \)-convex if

\[
E(\sigma(t)) \leq (1 - t)E(\mu) + tE(\nu) - t(1 - t)\frac{\lambda}{2} W_2^2(\mu, \nu)
\]

**Def:** Given a modulus of convexity \( \omega(x) \) and \( \lambda \in \mathbb{R} \), E is \( \omega \)-convex if

\[
E((\sigma(t))) \leq (1 - t)E(\mu) + tE(\nu) - \frac{\lambda}{2} \left[ (1 - t)\omega(t^2 W_2^2(\mu, \nu)) + t\omega((1 - t)^2 W_2^2(\mu, \nu)) \right]
\]

**Examples:**

- \( \omega(x) = x \), reduces to \( \lambda \)-convexity
- \( \omega(x) = x|\log(x)| \), [Ambrosio Serfaty, 2008] [Carrillo Lisini Mainini, 2014]
- \( \omega(x) = x^p, \ p > 1 \), [Carrillo McCann Villani, 2006]
Assume: $E$ is bounded below and $\omega$-convex along generalized geodesics for $\omega(x)$ satisfying Osgood’s condition: 

$$\int_0^1 \frac{dx}{\omega(x)} = +\infty$$

**Thm:** For $\tau = \frac{t}{n}$, $W_2(\rho(t), \rho_n) \to 0$, $F_{2t}(W_2^2(\rho_1(t), \rho_2(t))) \leq W_2^2(\rho_1(0), \rho_2(0))$

In particular, for $\omega(x) = x|\log(x)|$ and $W_2(\rho(0), \tilde{\rho}(0)) \leq 1$, 

$$W_2(\rho(t), \tilde{\rho}(t)) \leq W_2(\rho(0), \tilde{\rho}(0))e^{2\lambda t}$$
Questions
Thank you!