Gradient Flow in the 2-Wasserstein Metric

a Crandall and Liggett type proof of the exponential formula

Katy Craig

Rutgers University

MSRI: Introductory Workshop on Optimal Transport

August 28, 2013

Gradient flow and discrete gradient flow

- Consider $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, coercive, and convex

Gradient flow and discrete gradient flow

- Consider $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, coercive, and convex
- Assume $\mu << \mathcal{L}^d$, so by [Brenier] $\mathbf{t}_\mu^\nu \# \mu = \nu$ exists and

$$W_2(\mu, \nu) = \left(\int |\mathbf{id} - \mathbf{t}_\mu^\nu|^2 d\mu \right)^{1/2}$$

Assume $E(\mu) < \infty \implies \mu << \mathcal{L}^d$

Gradient flow and discrete gradient flow

- Consider $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, coercive, and convex
- Assume $\mu << \mathcal{L}^d$, so by [Brenier] $\mathbf{t}_\mu^\nu \# \mu = \nu$ exists and

$$W_2(\mu, \nu) = \left(\int |\mathbf{id} - \mathbf{t}_\mu^\nu|^2 d\mu \right)^{1/2}$$

  Assume $E(\mu) < \infty \implies \mu << \mathcal{L}^d$
- Formally, the gradient flow of $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is

$$\frac{d}{ds}\mu(s) = -\nabla E(\mu(s)) \ , \quad \mu(0) = \mu$$

Gradient flow and discrete gradient flow

- Consider $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, coercive, and convex
- Assume $\mu << \mathcal{L}^d$, so by [Brenier] $\mathbf{t}_\mu^\nu \# \mu = \nu$ exists and

$$W_2(\mu, \nu) = \left(\int |\mathbf{id} - \mathbf{t}_\mu^\nu|^2 d\mu \right)^{1/2}$$

  Assume $E(\mu) < \infty \implies \mu << \mathcal{L}^d$
- Formally, the gradient flow of $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is

$$\frac{d}{ds}\mu(s) = -\nabla E(\mu(s)) \ , \quad \mu(0) = \mu$$

- If we were in Euclidean space, we could approximate the gradient flow

$$\frac{d}{ds}x(s) = -\nabla E(x(s)) \ , \quad x(0) = x$$

  using the implicit finite difference scheme

$$\frac{x_n - x_{n-1}}{\tau} = -\nabla E(x_n) \ , \quad x_0 = x$$

Gradient flow and discrete gradient flow

- Consider $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, coercive, and convex
- Assume $\mu << \mathcal{L}^d$, so by [Brenier] $\mathbf{t}_\mu^\nu \# \mu = \nu$ exists and

$$W_2(\mu, \nu) = \left(\int |\mathbf{id} - \mathbf{t}_\mu^\nu|^2 d\mu \right)^{1/2}$$

  Assume $E(\mu) < \infty \implies \mu << \mathcal{L}^d$
- Formally, the gradient flow of $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is

$$\frac{d}{ds}\mu(s) = -\nabla E(\mu(s)) \ , \quad \mu(0) = \mu$$

- If we were in Euclidean space, we could approximate the gradient flow

$$\frac{d}{ds}x(s) = -\nabla E(x(s)) \ , \quad x(0) = x$$

  using the implicit finite difference scheme

$$\frac{x_n - x_{n-1}}{\tau} + \nabla E(x_n) = 0 \ , \quad x_0 = x$$

Gradient flow and discrete gradient flow

- Consider $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, coercive, and convex
- Assume $\mu << \mathcal{L}^d$, so by [Brenier] $\mathbf{t}_\mu^\nu \# \mu = \nu$ exists and

$$W_2(\mu, \nu) = \left(\int |\mathbf{id} - \mathbf{t}_\mu^\nu|^2 d\mu \right)^{1/2}$$

Assume $E(\mu) < \infty \implies \mu << \mathcal{L}^d$

- Formally, the gradient flow of $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is

$$\frac{d}{ds}\mu(s) = -\nabla E(\mu(s)) \ , \quad \mu(0) = \mu$$

- If we were in Euclidean space, we could approximate the gradient flow

$$\frac{d}{ds}x(s) = -\nabla E(x(s)) \ , \quad x(0) = x$$

using the implicit finite difference scheme

$$\frac{x_n - x_{n-1}}{\tau} + \nabla E(x_n) = 0 \ , \quad x_0 = x$$

$\implies x_n$ is a critical point of $y \mapsto \frac{1}{2\tau}|y - x_{n-1}|^2 + E(y)$

Gradient flow and discrete gradient flow

- Consider $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, coercive, and convex
- Assume $\mu << \mathcal{L}^d$, so by [Brenier] $\mathbf{t}_\mu^\nu \# \mu = \nu$ exists and

$$W_2(\mu, \nu) = \left(\int |\mathbf{id} - \mathbf{t}_\mu^\nu|^2 d\mu \right)^{1/2}$$

  Assume $E(\mu) < \infty \implies \mu << \mathcal{L}^d$
- Formally, the gradient flow of $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is

$$\frac{d}{ds}\mu(s) = -\nabla E(\mu(s)) \ , \quad \mu(0) = \mu$$

- If we were in Euclidean space, we could approximate the gradient flow

$$\frac{d}{ds}x(s) = -\nabla E(x(s)) \ , \quad x(0) = x$$

  using the implicit finite difference scheme

$$\frac{x_n - x_{n-1}}{\tau} + \nabla E(x_n) = 0 \ , \quad x_0 = x$$

$\implies x_n$ is a minimum of $y \mapsto \frac{1}{2\tau}|y - x_{n-1}|^2 + E(y)$

Gradient flow and discrete gradient flow

- Analogously, define the discrete gradient flow sequence of E in W_2 as

$$\mu_n \text{ is a minimum of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu_{n-1}, \nu) + E(\nu) \ , \quad \mu_0 = \mu$$

Gradient flow and discrete gradient flow

- Analogously, define the discrete gradient flow sequence of E in W_2 as

$$\mu_n \text{ is a minimum of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu_{n-1}, \nu) + E(\nu) \ , \quad \mu_0 = \mu$$

- Unlike in the Euclidean case, $\nu \mapsto W_2^2(\mu, \nu)$ is not convex [AGS, Example 9.1.5]. This is a recurring problem when trying to extend results from Banach and Hilbert spaces to the Wasserstein metric.

Gradient flow and discrete gradient flow

- Analogously, define the discrete gradient flow sequence of E in W_2 as

$$\mu_n \text{ is a minimum of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu_{n-1}, \nu) + E(\nu) \ , \quad \mu_0 = \mu$$

- Unlike in the Euclidean case, $\nu \mapsto W_2^2(\mu, \nu)$ is not convex [AGS, Example 9.1.5]. This is a recurring problem when trying to extend results from Banach and Hilbert spaces to the Wasserstein metric.
- To compensate for this, I follow [AGS] and require E to be convex along generalized geodesics (a class of curves in $\mathcal{P}_2(\mathbb{R}^d)$—will define soon).

Gradient flow and discrete gradient flow

- With this assumption, [AGS, Theorem 4.1.2] showed

$$E(\mu) < \infty \implies \forall \tau > 0 \,, \exists! \text{ minimizer of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$$

Gradient flow and discrete gradient flow

- With this assumption, [AGS, Theorem 4.1.2] showed

$$E(\mu) < \infty \implies \forall \tau > 0 \text{ , } \exists! \text{ minimizer of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$$

- Write the minimizer as $J_\tau \mu$ and call $J_\tau : \mu \mapsto J_\tau \mu$ the proximal map with time step τ.
  The nth term of the discrete gradient flow sequence with time step τ is $J_\tau^n \mu$.

Gradient flow and discrete gradient flow

- With this assumption, [AGS, Theorem 4.1.2] showed

$$E(\mu) < \infty \implies \forall \tau > 0 \text{ , } \exists! \text{ minimizer of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$$

- Write the minimizer as $J_\tau \mu$ and call $J_\tau : \mu \mapsto J_\tau \mu$ the proximal map with time step τ.
  The nth term of the discrete gradient flow sequence with time step τ is $J_\tau^n \mu$.

Theorem (Exponential formula [AGS])

$$\lim_{n \to \infty} J_{s/n}^n \mu = \mu(s)$$

Gradient flow and discrete gradient flow

- With this assumption, [AGS, Theorem 4.1.2] showed

$$E(\mu) < \infty \implies \forall \tau > 0 \text{ , } \exists! \text{ minimizer of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$$

- Write the minimizer as $J_\tau \mu$ and call $J_\tau : \mu \mapsto J_\tau \mu$ the proximal map with time step τ.
  The nth term of the discrete gradient flow sequence with time step τ is $J_\tau^n \mu$.

Theorem (Exponential formula [AGS])

$$\lim_{n \to \infty} J_{s/n}^n \mu = \mu(s)$$

- *the limit exists*

Gradient flow and discrete gradient flow

- With this assumption, [AGS, Theorem 4.1.2] showed

$$E(\mu) < \infty \implies \forall \tau > 0 \text{ , } \exists! \text{ minimizer of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$$

- Write the minimizer as $J_\tau \mu$ and call $J_\tau : \mu \mapsto J_\tau \mu$ the proximal map with time step τ.
  The nth term of the discrete gradient flow sequence with time step τ is $J_\tau^n \mu$.

Theorem (Exponential formula [AGS])

$$\lim_{n \to \infty} J_{s/n}^n \mu = \mu(s)$$

- *the limit exists*
- *the limit is a solution to the gradient flow*

Gradient flow and discrete gradient flow

- With this assumption, [AGS, Theorem 4.1.2] showed

$$E(\mu) < \infty \implies \forall \tau > 0 \,, \exists! \text{ minimizer of } \nu \mapsto \frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$$

- Write the minimizer as $J_\tau \mu$ and call $J_\tau : \mu \mapsto J_\tau \mu$ the proximal map with time step τ.
  The nth term of the discrete gradient flow sequence with time step τ is $J_\tau^n \mu$.

Theorem (Exponential formula [AGS])

$$\lim_{n \to \infty} J_{s/n}^n \mu = \mu(s)$$

- *the limit exists*
- *the limit is a solution to the gradient flow*

Key elements of Crandall and Liggett's method

Banach space $(X, ||\cdot||)$

1. Contraction inequality: $||J_\tau x - J_\tau y|| \leq ||x - y||$

Key elements of Crandall and Liggett's method

Banach space $(X, ||\cdot||)$

1. Contraction inequality: $||J_\tau x - J_\tau y|| \leq ||x - y||$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h \left[\frac{\tau - h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$

Key elements of Crandall and Liggett's method

Banach space $(X, ||\cdot||)$

1. Contraction inequality: $||J_\tau x - J_\tau y|| \leq ||x-y||$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h \left[\frac{\tau-h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$
3. Combine these to get a recursive inequality:

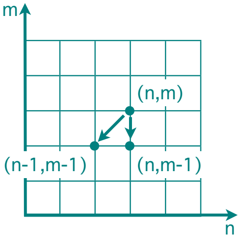
$$
\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left| J_h \left[\frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x\right] - J_h^m x \right|\right| && \text{by } ② \\
&\leq \left|\left| \left[\frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x\right] - J_h^{m-1} x \right|\right| && \text{by } ① \\
&\leq \frac{\tau - h}{\tau} ||J_\tau^n x - J_h^{m-1} x|| + \frac{h}{\tau} ||J_\tau^{n-1} x - J_h^{m-1} x||
\end{aligned}
$$

Key elements of Crandall and Liggett's method

Banach space $(X, ||\cdot||)$

1. Contraction inequality: $||J_\tau x - J_\tau y|| \leq ||x - y||$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h \left[\frac{\tau - h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$
3. Combine these to get a recursive inequality:

$$\begin{aligned} ||J_\tau^n x - J_h^m x|| &= \left|\left| J_h \left[\frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x\right] - J_h^m x \right|\right| && \text{by (2)} \\ &\leq \left|\left| \left[\frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x\right] - J_h^{m-1} x \right|\right| && \text{by (1)} \\ &\leq \frac{\tau - h}{\tau} ||J_\tau^n x - J_h^{m-1} x|| + \frac{h}{\tau} ||J_\tau^{n-1} x - J_h^{m-1} x|| \end{aligned}$$



Goal

1. Contraction inequality: $||J_\tau x - J_\tau y|| \leq ||x - y||$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h \left[\frac{\tau - h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$
3. Combine these to get a recursive inequality:

$$
\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left| J_h \left[\frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x\right] - J_h^m x \right|\right| && \text{by } ❷ \\
&\leq \left|\left| \frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x - J_h^{m-1} x \right|\right| && \text{by } ❶ \\
&\leq \frac{\tau - h}{\tau} ||J_\tau^n x - J_h^{m-1} x|| + \frac{h}{\tau} ||J_\tau^{n-1} x - J_h^{m-1} x||
\end{aligned}
$$

Goal

1. Contraction inequality: $||J_\tau x - J_\tau y|| \leq ||x - y||$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h \left[\frac{\tau-h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$
3. Combine these to get a recursive inequality:

$$
\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left| J_h \left[\frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x \right] - J_h^m x \right|\right| && \text{by } ② \\
&\leq \left|\left| \frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x - J_h^{m-1} x \right|\right| && \text{by } ① \\
&\leq \frac{\tau - h}{\tau} ||J_\tau^n x - J_h^{m-1} x|| + \frac{h}{\tau} ||J_\tau^{n-1} x - J_h^{m-1} x||
\end{aligned}
$$

1) Contraction Inequality

The W_2 analogue of $||J_\tau x - J_\tau y|| \leq ||x - y||$ is

Almost contraction inequality [C.]

$W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$

where $|\partial E|(u) := \limsup_{v\to u} \frac{(E(u)-E(v))^+}{d(u,v)}$ is the metric slope.

1) Contraction Inequality

The W_2 analogue of $||J_\tau x - J_\tau y|| \leq ||x - y||$ is

Almost contraction inequality [C.]

$W_2^2(J_\tau \mu, J_\tau \nu) \leq W_2^2(\mu, \nu) + \tau^2 |\partial E|^2(\mu)$

where $|\partial E|(u) := \limsup_{v \to u} \frac{(E(u) - E(v))^+}{d(u,v)}$ is the metric slope.

This is proved using the discrete variational inequality [AGS, Theorem 4.1.2]:

$$\frac{1}{2\tau}[W_2^2(J_\tau \mu, \nu) - W_2^2(\mu, \nu)] \leq E(\nu) - E(J_\tau \mu) - \frac{1}{2\tau} W_2^2(\mu, J_\tau \mu)$$

Goal

1. Contraction inequality: $||J_\tau x - J_\tau y|| \leq ||x - y||$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h\left[\frac{\tau-h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$
3. Combine these to get a recursive inequality:

$$
\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left| J_h\left[\frac{\tau-h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x\right] - J_h^m x \right|\right| && \text{by } 2 \\
&\leq \left|\left| \frac{\tau-h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x - J_h^{m-1} x \right|\right| && \text{by } 1 \\
&\leq \frac{\tau-h}{\tau} ||J_\tau^n x - J_h^{m-1} x|| + \frac{h}{\tau} ||J_\tau^{n-1} x - J_h^{m-1} x||
\end{aligned}
$$

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h\left[\frac{\tau-h}{\tau}J_\tau x + \frac{h}{\tau}x\right]$
3. Combine these to get a recursive inequality:

$$\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right|\right| && \text{by } ❷ \\
&\leq \left|\left|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right|\right| && \text{by } ❶ \\
&\leq \frac{\tau-h}{\tau}||J_\tau^n x - J_h^{m-1}x|| + \frac{h}{\tau}||J_\tau^{n-1}x - J_h^{m-1}x||
\end{aligned}$$

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h\left[\frac{\tau-h}{\tau}J_\tau x + \frac{h}{\tau}x\right]$
3. Combine these to get a recursive inequality:

$$\begin{aligned}
\|J_\tau^n x - J_h^m x\| &= \left\|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right\| && \text{by } ❷\\
&\leq \left\|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right\| && \text{by } ❶\\
&\leq \frac{\tau-h}{\tau}\|J_\tau^n x - J_h^{m-1}x\| + \frac{h}{\tau}\|J_\tau^{n-1}x - J_h^{m-1}x\|
\end{aligned}$$

2) Proximal maps at different time steps

The W_2 analogue of $J_\tau x = J_h \left[\frac{\tau-h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$ is

Proximal maps at different time steps [Jost, Mayer, C.]

If $0 < h \leq \tau$, $J_\tau \mu = J_h \left[\left(\frac{\tau-h}{\tau} \mathbf{t}_\mu^{J_\tau \mu} + \frac{h}{\tau} \mathbf{id}\right) \# \mu\right] = J_h \left[\mu^{\mu \to J_\tau \mu}_{\frac{\tau-h}{\tau}}\right]$

2) Proximal maps at different time steps

The W_2 analogue of $J_\tau x = J_h\left[\frac{\tau-h}{\tau}J_\tau x + \frac{h}{\tau}x\right]$ is

Proximal maps at different time steps [Jost, Mayer, C.]

If $0 < h \leq \tau$, $J_\tau \mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$

This is analogous to the Banach space result since, for $\alpha \in [0,1]$,

$\alpha J_\tau x + (1-\alpha)x$ is the Banach space geodesic from x to $J_\tau x$ at time α

$(\alpha\mathbf{t}_\mu^{J_\tau\mu} + (1-\alpha)\mathbf{id})\#\mu$ is the Wasserstein geodesic from μ to $J_\tau\mu$ at time α.

2) Proximal maps at different time steps

The W_2 analogue of $J_\tau x = J_h\left[\frac{\tau-h}{\tau} J_\tau x + \frac{h}{\tau} x\right]$ is

Proximal maps at different time steps [Jost, Mayer, C.]

If $0 < h \leq \tau$, $J_\tau \mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$

This is analogous to the Banach space result since, for $\alpha \in [0,1]$,

$\alpha J_\tau x + (1-\alpha)x$ is the Banach space geodesic from x to $J_\tau x$ at time α

$(\alpha \mathbf{t}_\mu^{J_\tau\mu} + (1-\alpha)\mathbf{id})\#\mu$ is the Wasserstein geodesic from μ to $J_\tau\mu$ at time α.

This follows from the Euler-Lagrange equation characterizing the minimizer $J_\tau\mu$.

Euler-Lagrange equation [AGS, C.]

ν is the unique minimizer of $\frac{1}{2\tau}W_2^2(\mu,\nu) + E(\nu)$ if and only if $\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu)$ is a strong subdifferential.

Euler-Lagrange equation

Definition (Subdifferential)

$\boldsymbol{\xi} \in L^2(\mu)$ belongs to $\partial E(\mu)$ in case

$$E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t}_\mu^\nu - \mathbf{id} \rangle d\mu + o(W_2(\mu, \nu)) \quad \text{as } \nu \xrightarrow{W_2} \mu.$$

Euler-Lagrange equation

Definition (Subdifferential)

$\boldsymbol{\xi} \in L^2(\mu)$ belongs to $\partial E(\mu)$ in case

$$E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t}_\mu^\nu - \mathbf{id} \rangle d\mu + o(W_2(\mu, \nu)) \quad \text{as } \nu \xrightarrow{W_2} \mu.$$

If E is convex, we may drop $o(W_2(\mu, \nu))$. Hence, $0 \in \partial E(\mu) \implies \mu$ minimizes E.

Euler-Lagrange equation

Definition (Subdifferential)

$\boldsymbol{\xi} \in L^2(\mu)$ belongs to $\partial E(\mu)$ in case

$$E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t}_\mu^\nu - \mathbf{id} \rangle d\mu + o(W_2(\mu, \nu)) \quad \text{as } \nu \xrightarrow{W_2} \mu.$$

If E is convex, we may drop $o(W_2(\mu, \nu))$. Hence, $0 \in \partial E(\mu) \implies \mu$ minimizes E.

Definition (Strong subdifferential)

If, in addition,

$$E(\mathbf{t}\#\mu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t} - \mathbf{id} \rangle d\mu + o(||\mathbf{t} - \mathbf{id}||_{L^2(\mu)}) \quad \text{as } \mathbf{t} \xrightarrow{L^2(\mu)} \mathbf{id}$$

$\boldsymbol{\xi}$ is a strong subdifferential.

Euler-Lagrange equation

Definition (Subdifferential)

$\boldsymbol{\xi} \in L^2(\mu)$ belongs to $\partial E(\mu)$ in case

$$E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t}_\mu^\nu - \mathbf{id} \rangle d\mu + o(W_2(\mu, \nu)) \quad \text{as } \nu \xrightarrow{W_2} \mu.$$

If E is convex, we may drop $o(W_2(\mu,\nu))$. Hence, $0 \in \partial E(\mu) \implies \mu$ minimizes E.

Definition (Strong subdifferential)

If, in addition,

$$E(\mathbf{t}\#\mu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t} - \mathbf{id} \rangle d\mu + o(||\mathbf{t} - \mathbf{id}||_{L^2(\mu)}) \quad \text{as } \mathbf{t} \xrightarrow{L^2(\mu)} \mathbf{id}$$

$\boldsymbol{\xi}$ is a strong subdifferential.

[AGS, Lemma 10.1.2], following a method introduced by [Otto], prove

ν minimizes $\frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu) \implies \frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu)$ is a strong subdifferential.

Euler-Lagrange equation

To prove the converse,

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential } \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

define the quadratic perturbation $\Phi_{\tau,\mu}(\nu) := \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu)$.

Euler-Lagrange equation

To prove the converse,

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential } \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

define the quadratic perturbation $\Phi_{\tau,\mu}(\nu) := \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu)$.

We'd like to use that $\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu)$ to conclude $0 \in \partial \Phi_{\tau,\mu}(\nu)$, hence ν minimizes $\Phi_{\tau,\mu}$. However, this argument doesn't work since $\Phi_{\tau,\mu}$ is not convex.

Euler-Lagrange equation

To prove the converse,

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential } \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

define the quadratic perturbation $\Phi_{\tau,\mu}(\nu) := \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu)$.

We'd like to use that $\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu)$ to conclude $0 \in \partial \Phi_{\tau,\mu}(\nu)$, hence ν minimizes $\Phi_{\tau,\mu}$. However, this argument doesn't work since $\Phi_{\tau,\mu}$ is not convex.

Again, we confront the problem that $\nu \mapsto W_2^2(\mu,\nu)$ is not convex.

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h\left[\frac{\tau-h}{\tau}J_\tau x + \frac{h}{\tau}x\right]$
   Will come back to this.
3. Combine these to get a recursive inequality:

$$\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right|\right| && \text{by 2} \\
&\leq \left|\left|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right|\right| && \text{by 1} \\
&\leq \frac{\tau-h}{\tau}||J_\tau^n x - J_h^{m-1}x|| + \frac{h}{\tau}||J_\tau^{n-1}x - J_h^{m-1}x||
\end{aligned}$$

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h\left[\frac{\tau-h}{\tau}J_\tau x + \frac{h}{\tau}x\right]$
   Will come back to this.
3. Combine these to get a recursive inequality:

$$\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right|\right| && \text{by 2} \\
&\leq \left|\left|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right|\right| && \text{by 1} \\
&\leq \frac{\tau-h}{\tau}||J_\tau^n x - J_h^{m-1}x|| + \frac{h}{\tau}||J_\tau^{n-1}x - J_h^{m-1}x||
\end{aligned}$$

3) Recursive Inequality

Combining

1. the almost contraction inequality and
2. the relation between proximal maps at different time steps

we get

$$\begin{aligned} W_2^2(J_\tau^n\mu, J_h^m\mu) &= W_2^2(J_h(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu\to J_\tau^n\mu}), J_h^m\mu) && \text{by ②} \\ &\leq W_2^2(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu\to J_\tau^n\mu}, J_h^{m-1}\mu) + h^2|\partial E|^2(J_h^{m-1}\mu) && \text{by ①} \end{aligned}$$

3) Recursive Inequality

Combining

1. the almost contraction inequality and
2. the relation between proximal maps at different time steps

we get

$$\begin{aligned} W_2^2(J_\tau^n\mu, J_h^m\mu) &= W_2^2(J_h(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu\to J_\tau^n\mu}), J_h^m\mu) && \text{by } ❷ \\ &\le W_2^2(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu\to J_\tau^n\mu}, J_h^{m-1}\mu) + h^2|\partial E|^2(J_h^{m-1}\mu) && \text{by } ❶ \end{aligned}$$

To follow the Banach space case, we would like to bound the first term by a convex combination of $W_2^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu)$ and $W_2^2(J_\tau^n\mu, J_h^{m-1}\mu)$, but this again requires the convexity of

$$\nu \mapsto W_2^2(\mu, \nu) \ .$$

Generalized geodesics

[AGS] ran into the same problem. Their solution was to consider convexity along a different class of curves. While $\nu \mapsto W_2^2(\mu, \nu)$ is not convex along all geodesics, it is convex along all generalized geodesics with base μ [AGS, Lemma 9.2.1].

Generalized geodesics

[AGS] ran into the same problem. Their solution was to consider convexity along a different class of curves. While $\nu \mapsto W_2^2(\mu, \nu)$ is not convex along all geodesics, it is convex along all generalized geodesics with base μ [AGS, Lemma 9.2.1].

Definition (Generalized geodesic)

The generalized geodesic from μ_0 to μ_1 with base ω is

$$\mu_\alpha := ((1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha \mathbf{t}_\omega^{\mu_1})\,\#\omega.$$

Note that if $\omega = \mu_0$ or μ_1, this is the standard geodesic from μ_0 to μ_1.

Generalized geodesics

[AGS] ran into the same problem. Their solution was to consider convexity along a different class of curves. While $\nu \mapsto W_2^2(\mu, \nu)$ is not convex along all geodesics, it is convex along all generalized geodesics with base μ [AGS, Lemma 9.2.1].

Definition (Generalized geodesic)

The generalized geodesic from μ_0 to μ_1 with base ω is

$$\mu_\alpha := \left((1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha \mathbf{t}_\omega^{\mu_1}\right) \# \omega.$$

Note that if $\omega = \mu_0$ or μ_1, this is the standard geodesic from μ_0 to μ_1.

The assumption that E is convex along generalized geodesics means that E is convex along all generalized geodesics, with any base.

Generalized geodesics

[AGS] ran into the same problem. Their solution was to consider convexity along a different class of curves. While $\nu \mapsto W_2^2(\mu, \nu)$ is not convex along all geodesics, it is convex along all generalized geodesics with base μ [AGS, Lemma 9.2.1].

Definition (Generalized geodesic)

The generalized geodesic from μ_0 to μ_1 with base ω is

$$\mu_\alpha := ((1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha \mathbf{t}_\omega^{\mu_1}) \# \omega.$$

Note that if $\omega = \mu_0$ or μ_1, this is the standard geodesic from μ_0 to μ_1.

The assumption that E is convex along generalized geodesics means that E is convex along all generalized geodesics, with any base.

On its own, the convexity of $\nu \mapsto W_2^2(\mu, \nu)$ along generalized geodesics with base μ is not enough to control

$$W_2^2(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu \to J_\tau^n \mu}, J_h^{m-1}\mu) ,$$

since $\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu \to J_\tau^n \mu}$ is not a generalized geodesic with base $J_h^{m-1}\mu$. For this reason, consider a related notion, also introduced by [AGS, Equation 7.3.2].

Transport Metric

Definition (Transport Metric)

Given ω, the transport metric from μ to ν with base ω is

$$W_{2,\omega}(\mu,\nu) := \left(\int |\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\nu|^2 d\omega \right)^{1/2}$$

Transport Metric

Definition (Transport Metric)

Given ω, the transport metric from μ to ν with base ω is

$$W_{2,\omega}(\mu,\nu) := \left(\int |\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\nu|^2 d\omega \right)^{1/2}$$

Properties of the Transport Metric [C.]

- $W_{2,\omega}$ is a metric

Transport Metric

Definition (Transport Metric)

Given ω, the transport metric from μ to ν with base ω is

$$W_{2,\omega}(\mu,\nu) := \left(\int |\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\nu|^2 d\omega \right)^{1/2}$$

Properties of the Transport Metric [C.]

- $W_{2,\omega}$ is a metric
- $W_2(\mu,\nu) \leq W_{2,\omega}(\mu,\nu)$, and equality holds if $\omega = \mu$ or $\omega = \nu$

Transport Metric

Definition (Transport Metric)

Given ω, the transport metric from μ to ν with base ω is

$$W_{2,\omega}(\mu,\nu) := \left(\int |\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\nu|^2 d\omega \right)^{1/2}$$

Properties of the Transport Metric [C.]

- $W_{2,\omega}$ is a metric
- $W_2(\mu,\nu) \leq W_{2,\omega}(\mu,\nu)$, and equality holds if $\omega = \mu$ or $\omega = \nu$
- $\nu \mapsto W_{2,\omega}^2(\mu,\nu)$ is convex along generalized geodesics with base ω

Transport Metric

Definition (Transport Metric)

Given ω, the transport metric from μ to ν with base ω is

$$W_{2,\omega}(\mu,\nu) := \left(\int |\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\nu|^2 d\omega \right)^{1/2}$$

Properties of the Transport Metric [C.]

- $W_{2,\omega}$ is a metric
- $W_2(\mu,\nu) \leq W_{2,\omega}(\mu,\nu)$, and equality holds if $\omega = \mu$ or $\omega = \nu$
- $\nu \mapsto W_{2,\omega}^2(\mu,\nu)$ is convex along generalized geodesics with base ω
- The geodesics of $W_{2,\omega}$ are the generalized geodesics with base ω

Transport Metric

Definition (Transport Metric)

Given ω, the transport metric from μ to ν with base ω is

$$W_{2,\omega}(\mu,\nu) := \left(\int |\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\nu|^2 d\omega \right)^{1/2}$$

Properties of the Transport Metric [C.]

- $W_{2,\omega}$ is a metric
- $W_2(\mu,\nu) \leq W_{2,\omega}(\mu,\nu)$, and equality holds if $\omega = \mu$ or $\omega = \nu$
- $\nu \mapsto W_{2,\omega}^2(\mu,\nu)$ is convex along generalized geodesics with base ω
- The geodesics of $W_{2,\omega}$ are the generalized geodesics with base ω
- $\Phi_{\tau,\mu}(\nu) := \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu)$ is convex in $W_{2,\mu}$

Goal

1. Contraction inequality: $W_2^2(J_\tau \mu, J_\tau \nu) \leq W_2^2(\mu, \nu) + \tau^2 |\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$, $J_\tau x = J_h \left[\frac{\tau - h}{\tau} J_\tau x + \frac{h}{\tau} x \right]$
3. Combine these to get a recursive inequality:

$$\begin{aligned} \|J_\tau^n x - J_h^m x\| &= \left\| J_h \left[\frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x \right] - J_h^m x \right\| && \text{by } 2 \\ &\leq \left\| \frac{\tau - h}{\tau} J_\tau^n x + \frac{h}{\tau} J_\tau^{n-1} x - J_h^{m-1} x \right\| && \text{by } 1 \\ &\leq \frac{\tau - h}{\tau} \|J_\tau^n x - J_h^{m-1} x\| + \frac{h}{\tau} \|J_\tau^{n-1} x - J_h^{m-1} x\| \end{aligned}$$

Euler-Lagrange equation

We want to prove

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential } \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

so we will have an Euler-Lagrange equation that characterizes $J_\tau \mu$.

Euler-Lagrange equation

We want to prove

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential} \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

so we will have an Euler-Lagrange equation that characterizes $J_\tau \mu$.

To do this, define what it means for $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ and show

Lemma ([C.])

1. *E convex in $W_{2,\omega}$ and $0 \in \partial_{2,\omega} E(\mu) \implies \mu$ minimizes E.*
2. *$\boldsymbol{\xi} \in \partial E(\rho)$ strong subdifferential $\implies \boldsymbol{\xi} \circ \mathbf{t}_\omega^\rho \in \partial_{2,\omega} E(\rho)$.*

Euler-Lagrange equation

We want to prove

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential } \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

so we will have an Euler-Lagrange equation that characterizes $J_\tau \mu$.

To do this, define what it means for $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ and show

Lemma ([C.])

1. *E convex in $W_{2,\omega}$ and $0 \in \partial_{2,\omega} E(\mu) \implies \mu$ minimizes E.*
2. *$\boldsymbol{\xi} \in \partial E(\rho)$ strong subdifferential $\implies \boldsymbol{\xi} \circ \mathbf{t}_\omega^\rho \in \partial_{2,\omega} E(\rho)$.*

Sketch of Proof.

- By 2, $\frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) \in \partial_{2,\mu} E(\nu)$

□

Euler-Lagrange equation

We want to prove

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential } \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

so we will have an Euler-Lagrange equation that characterizes $J_\tau \mu$.

To do this, define what it means for $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ and show

Lemma ([C.])

1. *E convex in $W_{2,\omega}$ and $0 \in \partial_{2,\omega} E(\mu) \implies \mu$ minimizes E.*
2. *$\boldsymbol{\xi} \in \partial E(\rho)$ strong subdifferential $\implies \boldsymbol{\xi} \circ \mathbf{t}_\omega^\rho \in \partial_{2,\omega} E(\rho)$.*

Sketch of Proof.

- By 2, $\frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) \in \partial_{2,\mu} E(\nu)$
- $2(\mathbf{t}_\mu^\nu - \mathbf{id}) \in \partial_{2,\mu} W_2^2(\mu,\nu)$

□

Euler-Lagrange equation

We want to prove

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential } \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

so we will have an Euler-Lagrange equation that characterizes $J_\tau \mu$.

To do this, define what it means for $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ and show

Lemma ([C.])

1. *E convex in $W_{2,\omega}$ and $0 \in \partial_{2,\omega} E(\mu) \implies \mu$ minimizes E.*
2. *$\boldsymbol{\xi} \in \partial E(\rho)$ strong subdifferential $\implies \boldsymbol{\xi} \circ \mathbf{t}_\omega^\rho \in \partial_{2,\omega} E(\rho)$.*

Sketch of Proof.

- By 2, $\frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) \in \partial_{2,\mu} E(\nu)$
- $2(\mathbf{t}_\mu^\nu - \mathbf{id}) \in \partial_{2,\mu} W_2^2(\mu,\nu)$
- By additivity of $\partial_{2,\mu}$, $\frac{2}{2\tau}(\mathbf{t}_\mu^\nu - \mathbf{id}) + \frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) = 0 \in \partial_{2,\mu} \Phi_{\tau,\mu}(\nu)$

□

Euler-Lagrange equation

We want to prove

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential} \implies \nu \text{ minimizes } \frac{1}{2\tau} W_2^2(\mu,\nu) + E(\nu),$$

so we will have an Euler-Lagrange equation that characterizes $J_\tau \mu$.

To do this, define what it means for $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ and show

Lemma ([C.])

1. *E convex in $W_{2,\omega}$ and $0 \in \partial_{2,\omega} E(\mu) \implies \mu$ minimizes E.*
2. *$\boldsymbol{\xi} \in \partial E(\rho)$ strong subdifferential $\implies \boldsymbol{\xi} \circ \mathbf{t}_\omega^\rho \in \partial_{2,\omega} E(\rho)$.*

Sketch of Proof.

- By 2, $\frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) \in \partial_{2,\mu} E(\nu)$
- $2(\mathbf{t}_\mu^\nu - \mathbf{id}) \in \partial_{2,\mu} W_2^2(\mu,\nu)$
- By additivity of $\partial_{2,\mu}$, $\frac{2}{2\tau}(\mathbf{t}_\mu^\nu - \mathbf{id}) + \frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) = 0 \in \partial_{2,\mu} \Phi_{\tau,\mu}(\nu)$
- $\Phi_{\tau,\mu}$ is convex in $W_{2,\mu} \implies \nu$ must be a minimizer, by 1

□

Proximal maps at different time steps

Thus, we have shown,

Euler-Lagrange equation [AGS, C.]

ν is the unique minimizer of $\frac{1}{2\tau}W_2^2(\mu,\nu) + E(\nu)$ if and only if $\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu)$ is a strong subdifferential.

Proximal maps at different time steps

Thus, we have shown,

Euler-Lagrange equation [AGS, C.]

ν is the unique minimizer of $\frac{1}{2\tau}W_2^2(\mu,\nu)+E(\nu)$ if and only if $\frac{1}{\tau}(\mathbf{t}_\nu^\mu-\mathbf{id})\in\partial E(\nu)$ is a strong subdifferential.

With this, we can now prove

Proximal maps at different time steps [Jost, Mayer, C.]

If $0<h\le\tau$, $J_\tau\mu=J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu}+\frac{h}{\tau}\mathbf{id}\right)\#\mu\right]=J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$

Proximal maps at different time steps

Thus, we have shown,

Euler-Lagrange equation [AGS, C.]

ν is the unique minimizer of $\frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$ if and only if $\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu)$ is a strong subdifferential.

With this, we can now prove

Proximal maps at different time steps [Jost, Mayer, C.]

If $0 < h \leq \tau$, $J_\tau \mu = J_h \left[\left(\frac{\tau - h}{\tau} \mathbf{t}_\mu^{J_\tau \mu} + \frac{h}{\tau} \mathbf{id} \right) \# \mu \right] = J_h \left[\mu_{\frac{\tau-h}{\tau}}^{\mu \to J_\tau \mu} \right]$

Sketch of Proof.

- Define $\nu := (\mathbf{id} + h\boldsymbol{\xi}) \# J_\tau \mu$, where $\boldsymbol{\xi} := \frac{1}{\tau}(\mathbf{t}_{J_\tau \mu}^\mu - \mathbf{id})$

□

Proximal maps at different time steps

Thus, we have shown,

Euler-Lagrange equation [AGS, C.]

ν is the unique minimizer of $\frac{1}{2\tau}W_2^2(\mu,\nu)+E(\nu)$ if and only if $\frac{1}{\tau}(\mathbf{t}_\nu^\mu-\mathbf{id})\in\partial E(\nu)$ is a strong subdifferential.

With this, we can now prove

Proximal maps at different time steps [Jost, Mayer, C.]

If $0<h\leq\tau$, $J_\tau\mu=J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu}+\frac{h}{\tau}\mathbf{id}\right)\#\mu\right]=J_h\left[\mu_{\frac{\tau-h}{\tau}}^{\mu\to J_\tau\mu}\right]$

Sketch of Proof.

- Define $\nu:=(\mathbf{id}+h\boldsymbol{\xi})\#J_\tau\mu$, where $\boldsymbol{\xi}:=\frac{1}{\tau}(\mathbf{t}_{J_\tau\mu}^\mu-\mathbf{id})$
- Use cyclic monotonicity and the EL equation to conclude $J_\tau\mu=J_h\nu$

□

Proximal maps at different time steps

Thus, we have shown,

Euler-Lagrange equation [AGS, C.]

ν is the unique minimizer of $\frac{1}{2\tau}W_2^2(\mu,\nu)+E(\nu)$ if and only if $\frac{1}{\tau}(\mathbf{t}_\nu^\mu-\mathbf{id})\in\partial E(\nu)$ is a strong subdifferential.

With this, we can now prove

Proximal maps at different time steps [Jost, Mayer, C.]

If $0<h\leq\tau$, $J_\tau\mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu}+\frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$

Sketch of Proof.

- Define $\nu := (\mathbf{id}+h\boldsymbol{\xi})\#J_\tau\mu$, where $\boldsymbol{\xi} := \frac{1}{\tau}(\mathbf{t}^\mu_{J_\tau\mu}-\mathbf{id})$
- Use cyclic monotonicity and the EL equation to conclude $J_\tau\mu = J_h\nu$
- Rewrite ν as $\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu}+\frac{h}{\tau}\mathbf{id}\right)\#\mu$ to conclude the result

□

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau, J_\tau x = J_h\left[\frac{\tau-h}{\tau}J_\tau x + \frac{h}{\tau}x\right]$
3. Combine these to get a recursive inequality:

$$\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right|\right| && \text{by ❷}\\
&\leq \left|\left|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right|\right| && \text{by ❶}\\
&\leq \frac{\tau-h}{\tau}||J_\tau^n x - J_h^{m-1}x|| + \frac{h}{\tau}||J_\tau^{n-1}x - J_h^{m-1}x||
\end{aligned}$$

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$,
   $J_\tau\mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$
3. Combine these to get a recursive inequality:

$$\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right|\right| && \text{by } 2\\
&\leq \left|\left|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right|\right| && \text{by } 1\\
&\leq \frac{\tau-h}{\tau}||J_\tau^n x - J_h^{m-1}x|| + \frac{h}{\tau}||J_\tau^{n-1}x - J_h^{m-1}x||
\end{aligned}$$

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$,
   $J_\tau\mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$
3. Combine these to get a recursive inequality:

$$\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right|\right| && \text{by } ❷ \\
&\leq \left|\left|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right|\right| && \text{by } ❶ \\
&\leq \frac{\tau-h}{\tau}||J_\tau^n x - J_h^{m-1}x|| + \frac{h}{\tau}||J_\tau^{n-1}x - J_h^{m-1}x||
\end{aligned}$$

Recursive Inequality

Proof.

Combining the contraction inequality with the relation between proximal maps at different time steps, we get

$$
\begin{aligned}
&W_2^2(J_\tau^n\mu, J_h^m\mu)\\
&= W_2^2(J_h(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu\to J_\tau^n\mu}), J_h^m\mu)\\
&\le W_2^2(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu\to J_\tau^n\mu}, J_h^{m-1}\mu) + h^2|\partial E|^2(J_h^{m-1}\mu)\\
&\le W_{2,J_\tau^{n-1}\mu}^2(\mu_{\frac{\tau-h}{\tau}}^{J_\tau^{n-1}\mu\to J_\tau^n\mu}, J_h^{m-1}\mu) + h^2|\partial E|^2(J_h^{m-1}\mu)\\
&\le \frac{h}{\tau}W_{2,J_\tau^{n-1}\mu}^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu) + \frac{\tau-h}{\tau}W_{2,J_\tau^{n-1}\mu}^2(J_\tau^n\mu, J_h^{m-1}\mu) + h^2|\partial E|^2(J_h^{m-1}\mu)\\
&= \frac{h}{\tau}W_2^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu) + \frac{\tau-h}{\tau}W_{2,J_\tau^{n-1}\mu}^2(J_\tau^n\mu, J_h^{m-1}\mu) + h^2|\partial E|^2(J_h^{m-1}\mu)
\end{aligned}
$$

Now we need to go from the transport metric back to the W_2 metric.

Recursive Inequality

Lemma (Controlling transport metric in terms of W_2 [C.])

$$W^2_{2,J_\tau^{n-1}\mu}(J_h^{m-1}\mu, J_\tau^n\mu) \leq \frac{\tau}{h}\left(W_2^2(J_h^{m-1}\mu, J_\tau^n\mu) - W_2^2(J_h^m\mu, J_\tau^n\mu)\right) + W_2^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu) + \tau h|\partial E|^2(J_h^{m-1}\mu)$$

Recursive Inequality

Lemma (Controlling transport metric in terms of W_2 [C.])

$$W^2_{2,J_\tau^{n-1}\mu}(J_h^{m-1}\mu, J_\tau^n\mu) \leq \frac{\tau}{h}\left(W_2^2(J_h^{m-1}\mu, J_\tau^n\mu) - W_2^2(J_h^m\mu, J_\tau^n\mu)\right) + W_2^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu) + \tau h|\partial E|^2(J_h^{m-1}\mu)$$

We can now finish the proof.

Proof.

Plugging this into the previous inequality, rearranging, and simplifying gives

$$W_2^2(J_\tau^n\mu, J_h^m\mu) \leq \frac{h}{\tau}W_2^2(J_h^{m-1}\mu, J_\tau^{n-1}\mu) + \frac{\tau - h}{\tau}W_2^2(J_h^{m-1}\mu, J_\tau^n\mu) + 2h^2|\partial E|^2(\mu)$$

□

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$,
   $J_\tau\mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$
3. Combine these to get a recursive inequality:

$$
\begin{aligned}
||J_\tau^n x - J_h^m x|| &= \left|\left|J_h\left[\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x\right] - J_h^m x\right|\right| && \text{by } 2 \\
&\leq \left|\left|\frac{\tau-h}{\tau}J_\tau^n x + \frac{h}{\tau}J_\tau^{n-1}x - J_h^{m-1}x\right|\right| && \text{by } 1 \\
&\leq \frac{\tau-h}{\tau}||J_\tau^n x - J_h^{m-1}x|| + \frac{h}{\tau}||J_\tau^{n-1}x - J_h^{m-1}x||
\end{aligned}
$$

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \leq \tau$,
   $J_\tau\mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu^{\mu\to J_\tau\mu}_{\frac{\tau-h}{\tau}}\right]$
3. Combine these to get a recursive inequality:

$$W_2^2(J_\tau^n\mu, J_h^m\mu) \leq \frac{h}{\tau}W_2^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu) + \frac{\tau-h}{\tau}W_2^2(J_\tau^n\mu, J_h^{m-1}\mu) + 2h^2|\partial E|^2(\mu)$$

Goal

1. Contraction inequality: $W_2^2(J_\tau\mu, J_\tau\nu) \le W_2^2(\mu,\nu) + \tau^2|\partial E|^2(\mu)$
2. Proximal maps at different time steps: if $0 < h \le \tau$,
   $J_\tau\mu = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{J_\tau\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu\right] = J_h\left[\mu_{\frac{\tau-h}{\tau}}^{\mu\to J_\tau\mu}\right]$
3. Combine these to get a recursive inequality:

$$W_2^2(J_\tau^n\mu, J_h^m\mu) \le \frac{h}{\tau}W_2^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu) + \frac{\tau-h}{\tau}W_2^2(J_\tau^n\mu, J_h^{m-1}\mu) + 2h^2|\partial E|^2(\mu)$$

Exponential formula

- Iterating the recursive inequality in a manner similar to [Rasmussen] gives

$$W_2^2(J_\tau^n\mu, J_h^m\mu) \leq \left[(n\tau - mh)^2 + \tau hm + 2h^2 m\right] |\partial E|^2(\mu)$$

Exponential formula

- Iterating the recursive inequality in a manner similar to [Rasmussen] gives

$$W_2^2(J_\tau^n \mu, J_h^m \mu) \leq \left[(n\tau - mh)^2 + \tau hm + 2h^2 m\right] |\partial E|^2(\mu)$$

- Therefore, taking $\tau = \frac{t}{n}$ and $h = \frac{t}{m}$ with $m \geq n$, so $h \leq \tau$, we obtain

$$W_2^2(J_{t/n}^n \mu, J_{t/m}^m \mu) \leq \left[\frac{t^2}{n} + 2\frac{t^2}{m}\right] |\partial E|^2(\mu) \ ,$$

  and the sequence $J_{t/n}^n \mu$ is Cauchy

Exponential formula

- Iterating the recursive inequality in a manner similar to [Rasmussen] gives

$$W_2^2(J_\tau^n \mu, J_h^m \mu) \leq \left[(n\tau - mh)^2 + \tau hm + 2h^2 m\right] |\partial E|^2(\mu)$$

- Therefore, taking $\tau = \frac{t}{n}$ and $h = \frac{t}{m}$ with $m \geq n$, so $h \leq \tau$, we obtain

$$W_2^2(J_{t/n}^n \mu, J_{t/m}^m \mu) \leq \left[\frac{t^2}{n} + 2\frac{t^2}{m}\right] |\partial E|^2(\mu) \ ,$$

  and the sequence $J_{t/n}^n \mu$ is Cauchy

- Since W_2 is complete [AGS, Prop 7.1.5], $\lim_{n\to\infty} J_{t/n}^n \mu$ exists

Exponential formula

- Iterating the recursive inequality in a manner similar to [Rasmussen] gives

$$W_2^2(J_\tau^n\mu, J_h^m\mu) \leq \left[(n\tau - mh)^2 + \tau hm + 2h^2 m\right] |\partial E|^2(\mu)$$

- Therefore, taking $\tau = \frac{t}{n}$ and $h = \frac{t}{m}$ with $m \geq n$, so $h \leq \tau$, we obtain

$$W_2^2(J_{t/n}^n\mu, J_{t/m}^m\mu) \leq \left[\frac{t^2}{n} + 2\frac{t^2}{m}\right] |\partial E|^2(\mu) \ ,$$

  and the sequence $J_{t/n}^n\mu$ is Cauchy
- Since W_2 is complete [AGS, Prop 7.1.5], $\lim_{n\to\infty} J_{t/n}^n\mu$ exists
- This gives a rate of convergence of $W_2(J_{t/n}^n\mu, \mu(s)) \leq O\left(\frac{1}{\sqrt{n}}\right) |\partial E|(\mu)$

Exponential formula

- Iterating the recursive inequality in a manner similar to [Rasmussen] gives

$$W_2^2(J_\tau^n\mu, J_h^m\mu) \leq \left[(n\tau - mh)^2 + \tau hm + 2h^2 m\right] |\partial E|^2(\mu)$$

- Therefore, taking $\tau = \frac{t}{n}$ and $h = \frac{t}{m}$ with $m \geq n$, so $h \leq \tau$, we obtain

$$W_2^2(J_{t/n}^n\mu, J_{t/m}^m\mu) \leq \left[\frac{t^2}{n} + 2\frac{t^2}{m}\right] |\partial E|^2(\mu) \ ,$$

  and the sequence $J_{t/n}^n\mu$ is Cauchy
- Since W_2 is complete [AGS, Prop 7.1.5], $\lim_{n\to\infty} J_{t/n}^n\mu$ exists
- This gives a rate of convergence of $W_2(J_{t/n}^n\mu, \mu(s)) \leq O\left(\frac{1}{\sqrt{n}}\right) |\partial E|(\mu)$
- This rate is not as good as the optimal rate showed by [AGS, Theorem 4.0.4] $W_2(J_{t/n}^n\mu, \mu(s)) \leq O\left(\frac{1}{n}\right) |\partial E|(\mu)$

Exponential formula

- Iterating the recursive inequality in a manner similar to [Rasmussen] gives

$$W_2^2(J_\tau^n \mu, J_h^m \mu) \leq \left[(n\tau - mh)^2 + \tau hm + 2h^2 m\right] |\partial E|^2(\mu)$$

- Therefore, taking $\tau = \frac{t}{n}$ and $h = \frac{t}{m}$ with $m \geq n$, so $h \leq \tau$, we obtain

$$W_2^2(J_{t/n}^n \mu, J_{t/m}^m \mu) \leq \left[\frac{t^2}{n} + 2\frac{t^2}{m}\right] |\partial E|^2(\mu) \ ,$$

  and the sequence $J_{t/n}^n \mu$ is Cauchy
- Since W_2 is complete [AGS, Prop 7.1.5], $\lim_{n\to\infty} J_{t/n}^n \mu$ exists
- This gives a rate of convergence of
  $W_2(J_{t/n}^n \mu, \mu(s)) \leq O\left(\frac{1}{\sqrt{n}}\right) |\partial E|(\mu)$
- This rate is not as good as the optimal rate showed by [AGS, Theorem 4.0.4]
  $W_2(J_{t/n}^n \mu, \mu(s)) \leq O\left(\frac{1}{n}\right) |\partial E|(\mu)$
- This rate improves upon the metric space result of [Clément, Desch]
  $W_2(J_{t/n}^n \mu, \mu(s)) \leq O\left(\frac{1}{n^{1/4}}\right) |\partial E|(\mu)$
  (though their result holds in greater generality)

Generalizations and directions for future work

The previous results continue to hold if...

- E λ-convex along generalized geodesics, $\lambda \in \mathbb{R}$

Generalizations and directions for future work

The previous results continue to hold if...

- E λ-convex along generalized geodesics, $\lambda \in \mathbb{R}$
- variable time steps τ_i

Generalizations and directions for future work

The previous results continue to hold if...

- E λ-convex along generalized geodesics, $\lambda \in \mathbb{R}$
- variable time steps τ_i
- $E(\mu) < \infty$ for measures μ that give mass to small sets

Generalizations and directions for future work

The previous results continue to hold if...

- E λ-convex along generalized geodesics, $\lambda \in \mathbb{R}$
- variable time steps τ_i
- $E(\mu) < \infty$ for measures μ that give mass to small sets

Directions for future work...

Generalizations and directions for future work

The previous results continue to hold if...

- E λ-convex along generalized geodesics, $\lambda \in \mathbb{R}$
- variable time steps τ_i
- $E(\mu) < \infty$ for measures μ that give mass to small sets

Directions for future work...

- Gradient flow for irregular functionals

Generalizations and directions for future work

The previous results continue to hold if...

- E λ-convex along generalized geodesics, $\lambda \in \mathbb{R}$
- variable time steps τ_i
- $E(\mu) < \infty$ for measures μ that give mass to small sets

Directions for future work...

- Gradient flow for irregular functionals
- How does the gradient flow behave as a regularization is removed?

Generalizations and directions for future work

The previous results continue to hold if...

- E λ-convex along generalized geodesics, $\lambda \in \mathbb{R}$
- variable time steps τ_i
- $E(\mu) < \infty$ for measures μ that give mass to small sets

Directions for future work...

- Gradient flow for irregular functionals
- How does the gradient flow behave as a regularization is removed?
- How can we tune time steps in the discrete gradient flow to avoid irregularities?

References I

L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second edition, Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 2008. MR2401600 (2009h:49002)

Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417. MR1100809 (92d:46088)

H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland, Amsterdam, 1973. MR0348562 (50 #1060)

E. Carlen and K. Craig, Contraction of the proximal map and generalized convexity of the Moreau-Yosida regularization in the 2-Wasserstein metric, Math. and Mech. of Complex Systems **1** (2013), no. 1, 33–65.

References II

M. G. Crandall, Semigroups of nonlinear transformations in Banach spaces, in *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971)*, 157–179. Publ. Math. Res. Center Univ. Wisconsin, 27, Academic Press, New York. MR0470787 (57 #10532)

M. G. Crandall and T. M. Liggett, Generation of semi-groups of nonlinear transformations on general Banach spaces, Amer. J. Math. **93** (1971), 265–298. MR0287357 (44 #4563)

E. De Giorgi, New problems on minimizing movements, in *Boundary value problems for partial differential equations and applications*, 81–98, RMA Res. Notes Appl. Math., 29 Masson, Paris. MR1260440 (95a:35057)

R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17. MR1617171 (2000b:35258)

References III

J. Jost, Convex functionals and generalized harmonic maps into spaces of nonpositive curvature, Comment. Math. Helv. **70** (1995), no. 4, 659–673. MR1360608 (96j:58043)

Y. Kobayashi, Difference approximation of Cauchy problems for quasi-dissipative operators and generation of nonlinear semigroups, J. Math. Soc. Japan **27** (1975), no. 4, 640–665. MR0399974 (53 #3812)

U. F. Mayer, Gradient flows on nonpositively curved metric spaces and harmonic maps, Comm. Anal. Geom. **6** (1998), no. 2, 199–253. MR1651416 (99m:58067)

R. J. McCann, A convexity principle for interacting gases, Adv. Math. **128** (1997), no. 1, 153–179. MR1451422 (98e:82003)

F. Otto, Doubly degenerate diffusion equations as steepest descent, Manuscript (1996).

References IV

F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174. MR1842429 (2002j:35180)

S. Rasmussen, *Non-linear semi-groups, evolution equations and product integral representations*, Various Publication Series, (1971), no. 2, Aarhus Universitet.

R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28 Princeton Univ. Press, Princeton, NJ, 1970. MR0274683 (43 #445)

C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58, Amer. Math. Soc., Providence, RI, 2003. MR1964483 (2004e:90003)

K. Yosida, *Functional analysis*, reprint of the sixth (1980) edition, Classics in Mathematics, Springer, Berlin, 1995. MR1336382 (96a:46001)