PDEs from Monge-Kantorovich Mass Transportation Theory

Luca Petrelli

Math & Computer Science Dept.
Mount St Mary’s University
Outline

• Monge-Kantorovich mass transportation problem

• Gradient Flow formalism

• Time-step discretization of gradient flows

• Application of theory to nonlinear diffusion problems

• Signed measures
Monge’s original problem

move a pile of soil from a deposit to an excavation with minimum amount of work

from “Memoir sur la theorie des deblais et des remblais” - 1781
Mathematical Model of Monge’s Problem

\[ \mu^+, \mu^- \text{ nonnegative Radon measures on } \mathbb{R}^d \]

\[ \mu^+ (\mathbb{R}^d) = \mu^- (\mathbb{R}^d) < \infty \]

\[ s : \mathbb{R}^d \rightarrow \mathbb{R}^d \text{ one-to-one mapping rearranging } \mu^+ \text{ into } \mu^- \]

\[ s \# \mu^+ = \mu^- \quad (s \#) \]

or

\[ \int_X h(s(x)) \, d\mu^+(x) = \int_Y h(y) \, d\mu^-(y) \quad \forall h \in C(\mathbb{R}^d; \mathbb{R}^d) \]

for \[ X = spt(\mu^+), \quad Y = spt(\mu^-) \]
\( c(x,y) \) cost of moving a unit mass from \( x \in \mathbb{R}^d \) to \( y \in \mathbb{R}^d \)

total cost \( I[s] := \int_{\mathbb{R}^d} c(x, s(x)) \, d\mu^+(x) \)

Monge’s problem is then to find \( s^* \in \mathcal{A} \) (admissable set) such that:

\[ I[s^*] = \min_{s \in \mathcal{A}} I[s] \quad (M) \]

with \( \mathcal{A} = \{ s \mid s#(\mu^+) = \mu^- \} \)
Problem is too hard!

- Constraint is highly nonlinear!
  \[
  \int_X h(s(x)) \, d\mu^+(x) = \int_Y h(y) \, d\mu^-(y) \quad \forall \, h \in C(\mathbb{R}^d; \mathbb{R}^d)
  \]

- Hard to identify minimum!
  \[
  \{s_k\}_{k=1}^\infty \subset A \text{ minimizing sequence such that } I[s_k] \to \inf_{s \in A} I[s]
  \]
  Hard to find \(\{s_{kj}\}\) subsequence such that \(s_{kj} \to s^* \) optimal.

- Classical methods of Calculus of Variation fail!
  - No terms create compactness for \(I[\cdot]\)
  - \(I[\cdot]\) does not involve gradients hence it can not be shown coercive on any Sobolev space
Kantorovich’s relaxation - 1940’s

Kantorovich’s idea: transform \((M)\) into linear problem

Define:

\[
M := \left\{ \text{prob. meas. } \mu \text{ on } \mathbb{R}^d \times \mathbb{R}^d \mid \text{proj}_x \mu = \mu^+, \text{proj}_y \mu = \mu^- \right\}
\]

\[
J[\mu] := \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \, d\mu(x, y)
\]

Find \(\mu^* \in M\) such that \(J[\mu^*] = \min_{\mu \in M} J[\mu]\) \quad (K)
Motivation

given \( s \in A \) we can define \( \mu \in \mathcal{M} \) as

\[
\mu(E) := \mu^+ \left\{ x \in \mathbb{R}^d \mid (x, s(x)) \in E \right\}
\]

\( E \subset \mathbb{R}^d \times \mathbb{R}^d \), \( E \) Borel

Problem

\( \mu^* \) need not be generated by any one-to-one one mapping \( s \in A \)

Solution

only look for “weak” or generalized solutions
Linear programming analogy

(Finite dimensional case)

\[
\begin{align*}
\mu^+(x) & \longrightarrow \mu^+_i & \mu^-(y) & \longrightarrow \mu^-_j \\
\mu(x, y) & \longrightarrow \mu_{i,j} & c(x, y) & \longrightarrow c_{i,j} \\
(i = 1, \ldots, n, & j = 1, \ldots, m)
\end{align*}
\]

Mass Balance Condition
\[
\sum_{i=1}^{n} \mu^+_i = \sum_{j=1}^{m} \mu^-_j < \infty
\]

Constraints
\[
\sum_{j=1}^{m} \mu_{i,j} = \mu^+_i, \quad \sum_{i=1}^{n} \mu_{i,j} = \mu^-_j, \quad \mu_{i,j} > 0
\]

Linear programming problem
\[
\text{minimize} \quad \sum_{i=1}^{n} \sum_{j=1}^{m} c_{i,j} \mu_{i,j}
\]

Then dual problem is
\[
\text{maximize} \quad \sum_{i=1}^{n} u_i \mu^+_i + \sum_{j=1}^{m} v_j \mu^-_j
\]

Subject to \( u_i + v_j \leq c_{i,j} \)
Kantorovich’s Dual Problem

Define:

\[ \mathcal{L} := \left\{ (u, v) \mid u, v : \mathbb{R}^d \to \mathbb{R}^+ \text{ continuous}, \ u(x) + v(y) \leq c(x, y) \ (x, y \in \mathbb{R}^d) \right\} \]

\[ K(u, v) := \int_{\mathbb{R}^d} u(x) \, d\mu^+(x) + \int_{\mathbb{R}^d} v(y) \, d\mu^-(y) \]

Then dual problem to (K) is:

Find \( u^*, v^* \) such that \( K(u^*, v^*) = \max_{(u,v) \in \mathcal{L}} K(u, v) \)
Gradient Flows

To define a gradient flow we need:

- a differentiable manifold \( \mathcal{M} \)
- a metric tensor \( g \) on \( \mathcal{M} \) which makes \( (\mathcal{M}, g) \) a Riemannian manifold
- and a functional \( E \) on \( \mathcal{M} \)

Then \( \frac{du}{dt} = -\text{grad} \ E(u) \) is the gradient flow of \( E \) on \( (\mathcal{M}, g) \).

where \( g(\text{grad}E, s) = \text{diff} E \cdot s \) for all vector fields \( s \) on \( \mathcal{M} \).

Then \( g_u(\frac{du}{dt}, s) + \text{diff} E|_u \cdot s = 0 \) for all vector fields \( s \) along \( u \).

Main property of gradient flows:
- energy of system is decreasing along trajectories, i.e.
  \[ \frac{d}{dt} E(u) = \text{diff} E|_u \cdot \frac{du}{dt} = -g_u(\frac{du}{dt}, \frac{du}{dt}) \]
Partial Differential Equations as gradient flows

Let \( \mathcal{M} := \left\{ u \geq 0, \text{measurable, with } \int u \, dx = 1 \right\} \)

define the tangent space to \( \mathcal{M} \) as

\[ T_u \mathcal{M} := \left\{ s \text{ measurable, with } \int s \, dx = 0 \right\} \]

and identify it with \( \left\{ p \text{ measurable } \right\} \! ~/~ \sim \)

via the elliptic equation

\[ -\nabla \cdot (u \nabla p) = s \, . \]
Define
\[ g_u(s_1, s_2) = \int u \nabla p_1 \cdot \nabla p_2 \, dx \quad \left( \equiv \int s_1 p_2 \, dx \right) \]

and \[ E(u) = \int e(u) \, dx \]

Then
\[ g_u(du/dt, s) + \text{diff } E|u \cdot s = \int \left( \frac{\partial u}{\partial t} p - \nabla \cdot (u \nabla p) e'(u) \right) \, dx = \]
\[ = \int \left( \frac{\partial u}{\partial t} p + \nabla p \cdot (u \nabla e'(u)) \right) \, dx = \int p \left( \frac{\partial u}{\partial t} - \nabla \cdot (u \nabla e'(u)) \right) \, dx = 0 \]

\[ \Rightarrow \quad \frac{\partial u}{\partial t} = \nabla \cdot (u \nabla e'(u)) \]
## Examples of PDE that can be obtained as Gradient Flows

<table>
<thead>
<tr>
<th>$e(u) = u \log u$</th>
<th>$\frac{\partial u}{\partial t} = \Delta u$</th>
<th>Heat Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e(u) = u \log u + u V$</td>
<td>$\frac{\partial u}{\partial t} = \Delta u + \nabla \cdot (u \nabla V)$</td>
<td>Fokker-Planck Equation</td>
</tr>
<tr>
<td>$e(u) = \frac{1}{m-1} u^m$</td>
<td>$\frac{\partial u}{\partial t} = \Delta u^m$</td>
<td>Porous Medium Equation</td>
</tr>
</tbody>
</table>

Note: equations are only solved in a weak or generalized way.
Important fact! Can implement gradient flow without making explicit use of gradient operator through \textit{time-discretization} and then passing to the limit as the time step goes to 0.

  \[
  \frac{\partial u(x, t)}{\partial t} - div(u\nabla \psi(x)) - \Delta u = 0
  \]

  \[
  \frac{\partial u(x, t)}{\partial t} - \Delta u^2 = 0
  \]

- Kinderlehrer and Walkington (1999)
  \[
  \frac{\partial u(x, t)}{\partial t} - \frac{\partial}{\partial x}(u\nabla \psi(x) + K(u)_x) = g(x, t)
  \]

- Agueh (2002)
  \[
  \frac{\partial u(x, t)}{\partial t} - div \left\{ u\nabla c^* \left[ \nabla (F'(u) + V(x)) \right] \right\} = 0
  \]

- Petrelli and Tudorascu (2004)
  \[
  \frac{\partial u(x, t)}{\partial t} - \nabla \cdot (u\nabla \psi(x, t)) - \Delta f(t, u) = g(x, t, u)
  \]
Time-discretized gradient flows

1. Set up variational principle

Let \( h > 0 \) be the time step. Define the sequence \( \{u^h_k\}_{k \geq 0} \) recursively as follows: \( u^h_0 \) is the initial datum \( u^0 \); given \( u^h_{k-1} \), define \( u^h_k \) as the solution of the minimization problem

\[
\min_{u \in \mathcal{M}} \left\{ \frac{1}{2h} d(u^h_{k-1}, u)^2 + E(u) \right\} \tag{P}
\]

where \( d \), the Wasserstein metric, is defined as

\[
d(\mu^+, \mu^-)^2 := \inf_{\mu \in \mathcal{M}} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\mu(x, y) \right\}
\]

i.e. \( d \) is the least cost of Monge-Kantorovich mass reallocation of \( \mu^+ \) to \( \mu^- \) for \( c(x, y) = |x - y|^2 \).
2. Euler-Lagrange Equations

Use Variation of Domain method to recover E-L eqns.

\[
\int \int_{\mathbb{R}^d \times \mathbb{R}^d} (y - x) \cdot \xi(y) \, d\mu(x, y) - h \int_{\mathbb{R}^d} \phi(u^h_k) \nabla \cdot \xi \, dx = 0
\]

where \( \phi(s) =: e'(s)s - e(s) \)

or in Gradient Flow terms:

\[
\frac{u^h_k - u^h_{k-1}}{h} = -\text{grad} \, E(u^h_k)
\]

Then recover approximate E-L eqns., i.e.

\[
\left| \int_{\mathbb{R}^d} \left\{ \frac{1}{h} (u^h_k - u^h_{k-1}) \zeta - \phi(u^h_k) \Delta \zeta \right\} \, dx \right| \leq \frac{1}{2h} \| \nabla^2 \zeta \|_\infty d(u^h_k, u^h_{k-1})^2
\]
3. Linear time interpolation

Define \[ u^h(x, t) := u^h_k(x) \quad \text{if} \quad kh \leq t < (k + 1)h \]

After integration in each interval over time we obtain

\[
\left| \int_{[0,T] \times \mathbb{R}^d} \left\{ \frac{1}{h} (u^h(x, t + \tau) - u^h(x, t)) \zeta - \phi(u^h) \Delta \zeta \right\} dx dt \right| \leq C \sum_{k=1}^{n} d(u^h_k, u^h_{k-1})^2
\]

Necessary inequality:
\[
\sum_{k=1}^{n} d(u^h_k, u^h_{k-1})^2 \leq C h
\]
4. Convergence result as time step \( h \) goes to 0

- **Linear case**
  Through a Dunford-Pettis like criteria show existence of function \( u \) such that, up to a subsequence, \( u^h \rightharpoonup u \) in some \( L^p \) space.

- **Nonlinear case**
  Stronger convergence is needed, through precompactness result in \( L^1 \). Also needed discrete maximum principle:
  \[ u^0 \text{ bounded } \Rightarrow u^h \text{ bounded} \]

Then, passing to the limit in the general Euler-Lagrange equation shows that \( u \) is a “weak” solution of

\[
\frac{\partial u}{\partial t} = \nabla \cdot (u\nabla e'(u)) \left( \equiv \Delta \phi(u) \right)
\]
Nonlinear Diffusion Problems

\[
\begin{align*}
  u_t - \nabla \cdot (u \nabla \Psi(x,t)) - \Delta f(t,u) &= g(x,t,u) & \text{in } \Omega \times (0,T), \\
  (u \nabla \Psi + \nabla f(t,u)) \cdot \nu_x &= 0 & \text{on } \partial \Omega \times (0,T), \quad (NP) \\
  u(\cdot,0) &= u_0 \geq 0 & \text{in } \Omega.
\end{align*}
\]

**Theorem 4.** Assume (f1)-(f3), (g1)-(g4) and (Ψ), then the problem (NP) admits a nonnegative essentially bounded weak solution provided that Ω is bounded and convex and the initial data \(u^0\) is nonnegative and essentially bounded.
Hypothesis

1. \((u - v)(f(t, u) - f(t, v)) \geq c |u - v|^\omega \) for all \(u, v \geq 0\), \((f1)\)
2. \(f(\cdot, s)\) are Lipschitz continuous for \(s\) in bounded sets \((f2)\)
3. \(f(t, \cdot)\) differentiable, \(\frac{\partial f}{\partial s}\) positive and monotone in time \((f3)\)
4. \(g(x, \cdot, \cdot)\) nonnegative in \([0, \infty) \times [0, \infty)\) for all \(x \in \mathbb{R}^d\) \((g1)\)
5. \(g(x, t, u) \leq C(1 + u)\) locally uniformly w.r.t. \((x, t), t \geq 0\) \((g2)\)
6. \(g(x, t, \cdot)\) is continuous on \([0, \infty)\) \((g3)\)
7. \(\{g(x, \cdot, u)\}_{(x,u)}\) is equicontinuous on \([0, \infty)\) w.r.t. \((x, u)\) \((g4)\)
8. \(\Psi : \mathbb{R}^d \times [0, \infty) \rightarrow \mathbb{R}\) differentiable and locally Lipschitz in \(x \in \mathbb{R}^d\) \((\Psi)\)
Novelties

- Time-dependent potential $\Psi(\cdot, t)$ and diffusion coefficient $f(t, \cdot)$
- Non homogeneous forcing term $g(x, t, u)$
- Averaging in time for $\Psi, f$ and $g$, e.g.
  \[ \Psi^k := \frac{1}{h} \int_{kh}^{(k+1)h} \Psi(\cdot, t) \, dt \]
- New variational principle for $v_{k-1} := u_{k-1} + \int_{(k-1)h}^{kh} g(\cdot, t, u_{k-1}) \, dt$
  \[ \min_{u \in M} \left\{ \frac{1}{2h} d(v_{k-1}^h, u)^2 + E(u) \right\} \quad (P') \]
Lemma 5. If $0 \leq u^0 \leq M_0 < \infty$ a.e. in $\Omega$ for large enough $M_0$, then there exists $0 < M = M(M_0) < \infty$ such that $0 \leq u^h \leq M$ a.e. in $\Omega$, for all $h > 0$ if $f$ satisfies (f3), $\lim_{s \uparrow \infty} \phi_s(t, s) = \infty$ uniformly in $t > 0$ and for $s > 0$ large enough we have

$$\eta s \frac{\partial f}{\partial s}(t, \eta s + \eta - 1) - (\eta s + \eta - 1) \frac{\partial f}{\partial s}(t, s)$$

does not change sign for all $t > 0$, $\eta > 1$, being nonnegative if $\frac{\partial f}{\partial s}(\cdot, s)$ is increasing and nonpositive if decreasing.
New discrete maximum principle

\[ u^0 \text{ bounded } \Rightarrow u^h \text{ bounded} \]

Key inequality:

\[ v^h_{k-1} \leq U_k := (\phi')^{-1} \circ (M_k - \Psi^k) \Rightarrow u^h_k \leq U_k \]

where \( U_k \) is the solution of the k-th “homogeneous stationary” equation, i.e.

\[ -\nabla \cdot (u \nabla \Psi^k) - \Delta f^k(u) = 0 \]
Signed measures

\[
\begin{aligned}
\begin{cases}
  u_t - \nabla \cdot (u \nabla \Psi(x, t)) - \gamma \Delta u &= g(x, t) \quad \text{in } \Omega \times (0, T), \\
  (u \nabla \Psi + \gamma \nabla u) \cdot \nu_x &= 0 \quad \text{on } \partial \Omega \times (0, T), \\
  u(\cdot, 0) &= u_0
\end{cases}
\end{aligned}
\]

Let

\[
 u^k_{\pm} := \arg\min \left\{ \frac{1}{2} d(u, v^k_{\pm} - 1)^2 + h F_k(u) \right\} \quad \text{over all } u \in \mathcal{M}_{v^k_{\pm}}
\]

where

\[
 v^k_{\pm} := u^k_{\pm} + h g^k_{\pm} \quad \text{and} \quad g^k_{\pm}(x) := \frac{1}{h} \int_{h(k+1)}^{h(k+1)} g_{\pm}(x, t) \, dt
\]

Let \( u^{(k)} := u^k_{+} - u^k_{-} \) and define

\[
u^h(x, t) := u^{(k)}(x) \text{ for } kh \leq t < (k + 1)h
\[
\sum_{k=1}^{n-1} d(v^k_{+1}, u^k_{+})^2 + \sum_{k=1}^{n-1} d(v^k_{-1}, u^k_{-})^2 \leq C h
\]

**Theorem 5.** Given \( u^0 \in L^\infty(\Omega) \) and continuous functions \( g, \Psi : \mathbb{R}^d \times [0, \infty) \rightarrow \mathbb{R} \), such that \( \Psi \) satisfies \((\Psi)\) and \( g \) is Lipschitz in time uniformly in \( x \), then the problem (SMP) admits a solution \( u \in L^\infty(Q) \).
Why use gradient flows with Wasserstein metric?

- We can minimize directly in the weak topology.
Wasserstein metric convergence is equivalent to weak star convergence.

- There are no derivatives in the variational principle.
  This allows for use of discontinuous functions in approximation,
  for example step functions.

- We can construct new (convex) variational principles.
  For problems like the convection diffusion equation.

- We can recover new maximum principles.
  Fairly easily from the variational principles.