PDEs from Monge-Kantorovich Mass Transportation Theory

Luca Petrelli

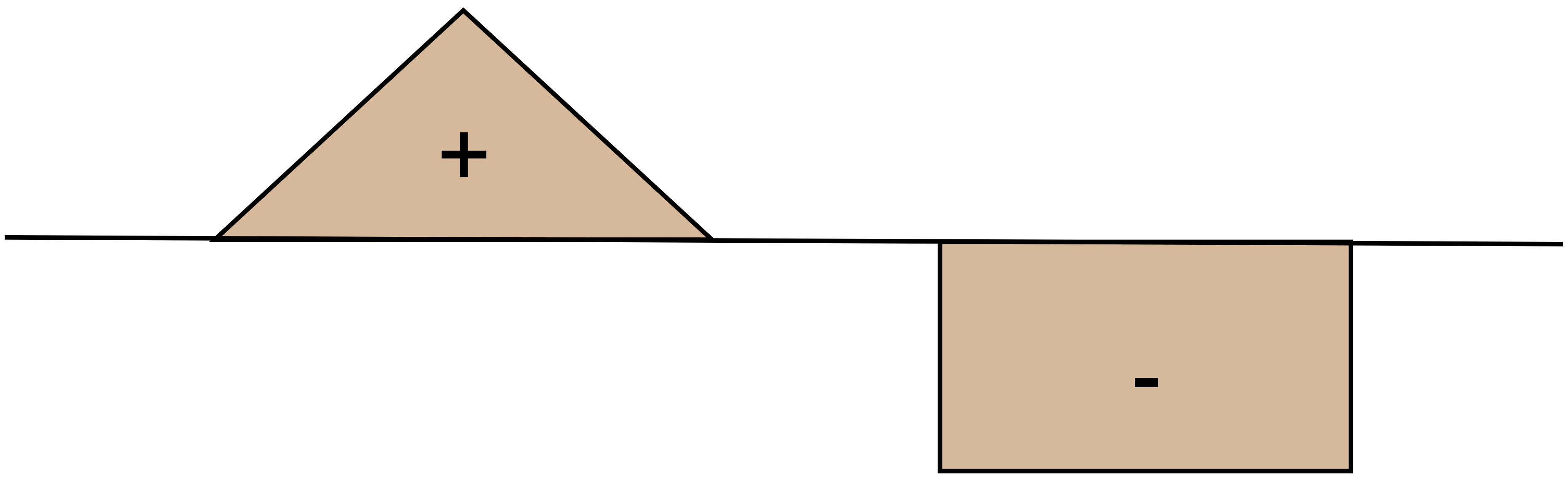
Math & Computer Science Dept.
Mount St Mary's University

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

$\mu^+, \quad \mu^-$ nonnegative Radon measures on $\Re^d$

$$\mu^+\left(\Re^d\right) = \mu^-\left(\Re^d\right) < \infty$$

$s: \ \Re^d \to \Re^d$ one-to-one mapping rearranging μ^+ into μ^-

$$s_\# \mu^+ = \mu^- \qquad\qquad (s_\#)$$

or

$$\int_X h(s(x))\, d\mu^+(x) = \int_Y h(y)\, d\mu^-(y) \quad \forall\, h \in C(\Re^d; \Re^d)$$

for $\quad X = spt(\mu^+),\ Y = spt(\mu^-)$

c(x,y) cost of moving a unit mass from $x \in \Re^d$ to $y \in \Re^d$

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

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

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

with $$\mathcal{A} = \{ s \,|\, s_{\#}(\mu^+) = \mu_- \}$$

Problem is too hard!

- Constraint is highly nonlinear!

$$\int_X h(s(x))\, d\mu^+(x) = \int_Y h(y)\, d\mu^-(y) \quad \forall\, h \in C(\Re^d; \Re^d)$$

- Hard to identify minimum!

$\{s_k\}_{k=1}^{\infty} \subset \mathcal{A}$ minimizing sequence such that $I[s_k] \to \inf_{s \in \mathcal{A}} I[s]$

Hard to find $\{s_{k_j}\}$ subsequence such that $s_{k_j} \to s^*$ 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:

$$\mathcal{M} := \left\{ \text{prob. meas. } \mu \text{ on } \Re^d \times \Re^d \;\middle|\; proj_x\, \mu = \mu^+,\, proj_y\, \mu = \mu^- \right\}$$

$$J[\mu] \;:=\; \int_{\Re^d \times \Re^d} c(x,y)\, d\mu(x,y)$$

Find $\mu^* \in \mathcal{M}$ such that $\qquad J[\mu^*] = \min_{\mu \in \mathcal{M}} J[\mu] \qquad\qquad (K)$

Motivation

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

$$\mu(E) := \mu^+ \left\{ x \in \Re^d \middle| (x, s(x)) \in E \right\} \qquad \left(E \subset \Re^d \times \Re^d,\ E \text{ Borel } \right)$$

Problem

μ^* need not be generated by any one-to-one mapping $s \in \mathcal{A}$

Solution

only look for "weak" or generalized solutions

Linear programming analogy

(Finite dimensional case)

$$\mu^+(x) \longrightarrow \mu_i^+ \qquad \mu^-(y) \longrightarrow \mu_j^-$$
$$\mu(x,y) \longrightarrow \mu_{i,j} \qquad c(x,y) \longrightarrow c_{i,j}$$
$$(i = 1, \cdots, n,\ j = 1, \cdots, m)$$

Mass Balance Condition $\displaystyle \sum_{i=1}^{n} \mu_i^+ = \sum_{j=1}^{m} \mu_j^- < \infty$

Constraints $\displaystyle \sum_{j=1}^{m} \mu_{i,j} = \mu_i^+, \quad \sum_{i=1}^{n} \mu_{i,j} = \mu_j^-, \quad \mu_{i,j} > 0$

Linear programming problem $\displaystyle \text{minimize} \ \sum_{i=1}^{n}\sum_{j=1}^{m} c_{i,j}\mu_{i,j}$

Then dual problem is

$$\text{maximize} \ \sum_{i=1}^{n} u_i \mu_i^+ + \sum_{j=1}^{m} v_j \mu_j^-$$

$$\text{subject to} \quad u_i + v_j \le c_{i,j}$$

Kantorovich's Dual Problem

Define:

$$\mathcal{L} := \left\{ (u,v) \,\middle|\, u,v : \Re^d \to \Re^+ \ \text{ continuous }, \ u(x) + v(y) \ \le \ c(x,y) \ \ \left(x, y \in \Re^d\right) \right\}$$

$$K(u,v) := \int_{\Re^d} u(x)\, d\mu^+(x) \ + \ \int_{\Re^d} v(y)\, d\mu^-(y)$$

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 $\dfrac{du}{dt} = -\operatorname{grad} E(u)$ is the gradient flow of E on $(\mathcal{M}, g)$.

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

Then $g_u(\dfrac{du}{dt}, s) + \operatorname{diff} E_{|u} \cdot s = 0$ for all vector fields s along u.

Main property of gradient flows:

- energy of system is decreasing along trajectories, i.e.

$$\frac{d}{dt} E(u) = \operatorname{diff} E_{|u} \cdot \frac{du}{dt} = -g_u(\frac{du}{dt}, \frac{du}{dt})$$

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 $\{ p \text{ measurable } \} / \sim$

via the elliptic equation $-\nabla \cdot (u \nabla p) = s$.

Define

$$g_u(s_1, s_2) = \int u \nabla p_1 \cdot \nabla p_2 \, dx \left(\equiv \int s_1 p_2 \, dx \right)$$

and $$E(u) = \int e(u) \, dx$$

Then

$$g_u(\frac{du}{dt}, s) + \operatorname{diff} E_{|u} \cdot s = \int \left(\frac{\partial u}{\partial t} p - \nabla \cdot (u \nabla p) \, e'(u) \right) dx =$$

$$= \int \left(\frac{\partial u}{\partial t} p + \nabla p \cdot (u \nabla e'(u)) \right) dx = \int p \left(\frac{\partial u}{\partial t} - \nabla \cdot (u \nabla e'(u)) \right) dx = 0$$

$$\Longrightarrow \qquad \frac{\partial u}{\partial t} = \nabla \cdot (u \nabla e'(u))$$

Examples of PDE that can be obtained as Gradient Flows

$e(u) = u \log u$	$\frac{\partial u}{\partial t} = \Delta u$	Heat Equation
$e(u) = u \log u + u V$	$\frac{\partial u}{\partial t} = \Delta u + \nabla \cdot (u \nabla V)$	Fokker-Planck Equation
$e(u) = \frac{1}{m-1} u^m$	$\frac{\partial u}{\partial t} = \Delta u^m$	Porous Medium Equation

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 *time-discretization* and then passing to the limit as the time step goes to 0.

- Jordan, Kinderlehrer and Otto (1998)

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

- Otto (1998)

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

- Kinderlehrer and Walkington (1999)

$$\frac{\partial u(x,t)}{\partial t} - \frac{\partial}{\partial x}\Big(u\nabla\psi(x) + K(u)_x\Big) = g(x,t)$$

- Agueh (2002)

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

- 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 $\{u_k^h\}_{k\geq 0}$ recursively as follows: u_0^h is the intial datum u^0; given u_{k-1}^h, define u_k^h as the solution of the minimization problem

$$\min_{u\in\mathcal{M}} \left\{ \frac{1}{2h} d\left(u_{k-1}^h, u\right)^2 + E(u) \right\} \qquad (P)$$

where d, the Wasserstein metric, is defined as

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

i.e. d is the least cost of Monge-Kantorovich mass reallocation of μ^+ to μ^-

for $c(x,y) = |x-y|^2$.

2. Euler-Lagrange Equations

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

$$\int\int_{\Re^d \times \Re^d} (y-x)\cdot\xi(y)\,d\mu(x,y) - h\int_{\Re^d}\phi(u_k^h)\,\nabla\cdot\xi\,dx \;=\; 0$$

where $\qquad \phi(s) \;=:\; e'(s)s - e(s)$

or in Gradient Flow terms:

$$\frac{u_k^h - u_{k-1}^h}{h} \;=\; -\mathrm{grad}\,E(u_k^h)$$

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

$$\left|\int_{\Re^d}\left\{\frac{1}{h}(u_k^h - u_{k-1}^h)\,\zeta - \phi(u_k^h)\Delta\zeta\right\}dx\right| \;\le\; \frac{1}{2h}\|\nabla^2\zeta\|_\infty d(u_k^h, u_{k-1}^h)^2$$

3. Linear time interpolation

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

After integration in each interval over time we obtain

$$\left| \int_{[0,T]\times\Re^d} \left\{ \frac{1}{h}\big(u^h(x,t+\tau) - u^h(x,t)\big)\,\zeta - \phi(u^h)\Delta\zeta \right\} dxdt \right| \leq C \sum_{k=1}^{n} d(u_k^h, u_{k-1}^h)^2$$

Necessary inequality: $\displaystyle \sum_{k=1}^{n} d\big(u_k^h, u_{k-1}^h\big)^2 \leq C\,h$

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^pspace.

- 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

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

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

- $(u-v)(f(t,u)-f(t,v)) \geq c|u-v|^{\omega}$ for all $u,\ v \geq 0,$ $(f1)$
- $f(\cdot,s)$ are Lipschitz continuous for s in bounded sets $(f2)$
- $f(t,\cdot)$ differentiable, $\dfrac{\partial f}{\partial s}$ positive and monotone in time $(f3)$
- $g(x,\cdot,\cdot)$ nonnegative in $[0,\infty)\times[0,\infty)$ for all $x \in \Re^d$ $(g1)$
- $g(x,t,u) \leq C(1+u)$ locally uniformly w.r.t. (x,t), $t \geq 0$ $(g2)$
- $g(x,t,\cdot)$ is continuous on $[0,\infty)$ $(g3)$
- $\{g(x,\cdot,u)\}_{(x,u)}$ is equicontinuous on $[0,\infty)$ w.r.t. (x,u) $(g4)$
- $\Psi : \Re^d \times [0,\infty) \to \Re$ diff.ble and locally Lipschitz in $x \in \Re^d$ (Ψ)

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 Ψ, f and g, e.g. $\Psi^k := \frac{1}{h}\int_{kh}^{(k+1)h} \Psi(\cdot, t)\, dt$

- New variational principle for $v_{k-1} := u_{k-1} + \int_{(k-1)h}^{kh} g(\cdot, t, u_{k-1})\, dt$

$$\min_{u \in \mathcal{M}} \left\{ \frac{1}{2h} d\left(v_{k-1}^h,\, u\right)^2 + E(u) \right\} \qquad (P')$$

- New discrete maximum principle

Lemma 5. If $0 \le u^0 \le M_0 < \infty$ a.e. in Ω for large enough M_0, then there exists $0 < M = M(M_0) < \infty$ such that $0 \le u^h \le M$ a.e. in Ω, 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*, η > 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: $v_{k-1}^h \le U_k := (\phi')^{(-1)} \circ (M_k - \Psi^k) \quad \Rightarrow \quad u_k^h \le U_k$

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

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

Let

$$u_\pm^k := \operatorname{argmin} \left\{ \frac{1}{2} d(u, v_\pm^{k-1})^2 + h \, F_k(u) \right\} \text{ over all } u \in \mathcal{M}_{v_\pm^{k-1}}$$

where $v_\pm^k := u_\pm^k + h \, g_\pm^k$ and $g_\pm^k(x) := \frac{1}{h} \int_{hk}^{h(k+1)} g_\pm(x,t) \, dt$

Let $u^{(k)} := u_+^k - u_-^k$ and define

$$u^h(x,t) := u^{(k)}(x) \ \text{ for } kh \le t < (k+1)h$$

$$\sum_{k=1}^{n-1} d(v_+^{k-1}, u_+^k)^2 + \sum_{k=1}^{n-1} d(v_-^{k-1}, u_-^k)^2 \ \leq \ C\, h$$

Theorem 5. Given $u^0 \in L^\infty(\Omega)$ and continuous functions $g, \Psi : \Re^d \times [0, \infty) \rightarrow \Re$, such that Ψ satisfies (Ψ) 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