Distributed Optimal Transport

Vishaal Krishnan and Sonia Martínez

Mechanical and Aerospace Enginering
University of California, San Diego

IEEE CDC 2018

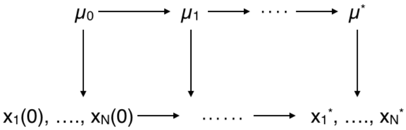
General Setting

Given:

- $\Omega \subset \mathbb{R}^d$ – Compact, $\mathcal{P}(\Omega)$ – Space of probability measures
- $\mu^* \in \mathcal{P}(\Omega)$ – Target distribution
- Directly sampling from μ^* is difficult

Objectives:

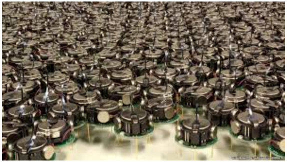
- **End goal:** Sample N points *optimally* from μ^*
- Sample from a known μ_0, transport samples to μ^*

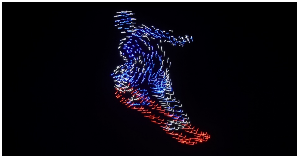


- Minimize net cost of transport – **Optimal Transport**
- Use a **distributed** algorithm to update samples $\{x_1(k), \ldots, x_N(k)\}$

Motivation

Engineering Applications: Deployment problems (sensor/robot networks)





Sampling algorithms (Markov Chain Monte Carlo)

- i.i.d sampling
- Realizations of a Markov chain
- **Decentralized**
- Not efficient w.r.t. transport cost

Optimal Transport

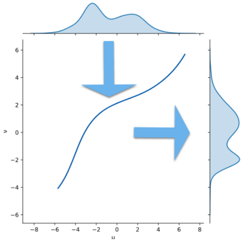
- Optimal sampling
- Mapped by an OT map
- **Centralized** computation
- Transport cost minimized

Outline

Outline

Optimal transport

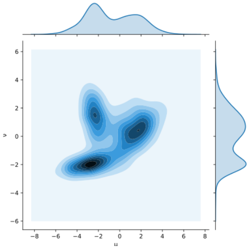
- $c(x, y)$ - Unit cost of transport from $x \in \Omega$ to $y \in \Omega$
- $\mu, \nu \in \mathcal{P}(\Omega)$



Monge (deterministic) formulation

Minimize cost over maps that transport μ to ν

$$C_M(\mu, \nu) = \inf_{\substack{T:\Omega \to \Omega \\ T_\# \mu = \nu}} \int_{\Omega \times \Omega} c(x, T(x)) d\mu(x)$$



Kantorovich (probabilistic) formulation

Minimize over probabilistic couplings of μ and ν

$$C_K(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) d\pi(x, y)$$

Question: When are they equivalent?

On Monge and Kantorovich formulations

Some assumptions:

- Cost c is continuous
- c is a metric on Ω
- c satisfies a **twist** condition:

  $\forall y_1, y_2,\ c(x, y_1) - c(x, y_2)$ has no critical point
- The measure μ is **atomless**:

  $\mu(\{x\}) = 0\ \forall x \in \Omega$

What we get:

- Solution to the Kantorovich problem **exists** and is **unique**
- Minimizer of Kantorovich solves the Monge problem:

  π^* concentrated over a T^* that solves Monge
- Allows us to work with the relaxation,

  i.e. the Kantorovich formulation

Kantorovich Duality

Primal problem

$$C_K(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} \int_{\Omega\times\Omega} c(x,y)d\pi(x,y)$$

Dual problem for metric costs c:

$$K(\mu,\nu) = \max_{\phi \in L^1_{\mu,\nu}(\Omega)} \int_\Omega \phi d\mu - \int_\Omega \phi d\nu$$

$$\text{s.t.} \quad |\phi(x) - \phi(y)| \leq c(x,y) \ \forall x,y \in \Omega$$

- **Strong duality:** $K(\mu,\nu) = C_K(\mu,\nu)$
- **Derivative:** $\frac{\delta K}{\delta \mu} = \bar{\phi}$ *The Kantorovich potential*

Relevant Literature

OT and Interpolation

- McCann, "*Existence and uniqueness of monotone measure- preserving maps*", 1995 (Parametrized family of transport maps)
- Benamou, Brenier, "*A Computational Fluid Mechanics solution to the M-K mass transfer problem*", 2000 (Transport PDE/Hamilton-Jacobi)
- Chen, Georgiou, Pavon, "*On the relation between optimal transport and Schrodinger bridges ...*", 2014 (Stochastic control)
- Cuturi, Doucet, *Fast Computation of Wasserstein Barycenters*, 2014 (Computing interpolants)

Applications in multi-robot systems

- Bandyopadhyay, Chung, Hadaegh, "*Probabilistic Swarm Guidance using Optimal Transport*", 2014 (Deployment problem)

Outline

Interpolation between measures

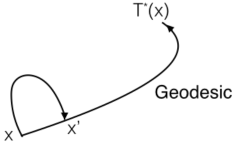
Objective: Construct a sequence $\mu_0 \to \mu_1 \to \ldots \to \mu^*$

Consider the scheme:

$$\mu_{k+1} \in \arg\min_{\nu} C(\mu_k, \nu) + C(\nu, \mu^*)$$
$$\text{s.t.} \quad C(\mu_k, \nu) \leq \epsilon$$

Corresponding sample update:

$$x^{k+1} \in \arg\min_{y} c(x^k, y) + c(y, T_k^*(x^k))$$
$$\text{s.t.} \quad c(x^k, y) \leq \epsilon$$



T_k^* is the OT map from μ_k to μ^*

But T_k^* is hard to compute $\Rightarrow$ **Use the dual formulation**

Sample update from Kantorovich Dual

Consider the scheme:

$$\mu_{k+1} \in \arg\min_{\nu} C(\mu_k, \nu) + K(\nu, \mu^*)$$
$$\text{s.t.} \quad C(\mu_k, \nu) \leq \epsilon$$

Transport vector field: $\mathbf{v} = -(Hc)^{-1} \nabla \phi_k$ (as $\epsilon \to 0$ above)

$$\left(\phi_k = \frac{\delta K(\cdot, \mu^*)}{\delta \mu}\bigg|_{\mu_k} \text{ and } Hc(x) = \nabla_2^2 c(x, x)\right)$$

- But c is **not differentiable** at (x, x) – due to assumptions on c

Sample update scheme:

$$x^{k+1} \in \arg\min_{z} c(x^k, z) + \phi_k(z)$$
$$\text{s.t.} \quad c(x^k, z) \leq \epsilon$$

Distributed algorithm to compute ϕ

- Samples $\mathbf{x} = (x_1, \ldots, x_N)$
- $\{\mathcal{V}_i\}_{i=1}^N$ – Voronoi partition of Ω
- Nearest-neighbor graph $\mathcal{G} = (V, E)$
- Approximate ϕ by a simple function $\phi = \sum_{i=1}^N \phi^i \ \mathbb{1}_{\mathcal{V}_i}$

Kantorovich Dual

$$\max_{(\phi^1,\ldots,\phi^N)} \sum_{i=1}^N \phi^i \left(\frac{1}{N} - \mu^*(\mathcal{V}_i)\right)$$

s.t. $|\phi^i - \phi^j| \leq c(x_i, x_j) \ \forall (i,j) \in E$

Primal-Dual Algorithm

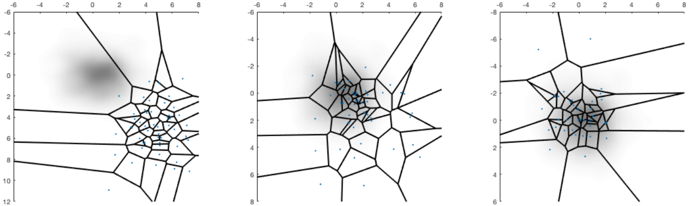
$$\phi^i(k+1) = \phi^i(k) - \tau \left(L_\lambda \phi\right)^i + \tau Q_i$$

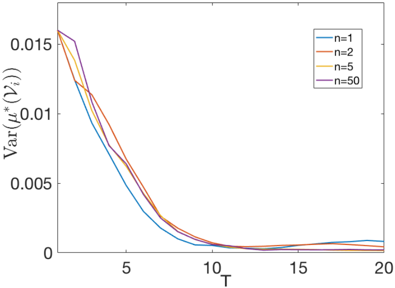
$$Q_i = \frac{1}{N} - \mu^*(\mathcal{V}_i)$$

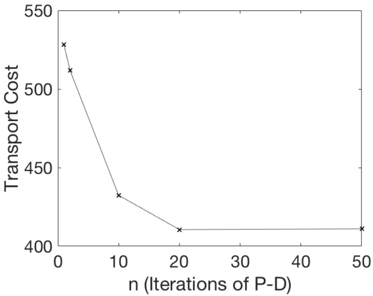
$$\lambda_{ij}(k+1) = [\lambda_{ij}(k) + \tau r_{ij}(k)]^+$$

$$r_{ij}(k) = |\phi^i(k) - \phi^j(k)|^2 - c^2(x_i, x_j)$$

Simulation Results







Outline

Continuum limit $N \to \infty$ and $\epsilon \to 0$

Transport equation:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0 \qquad (\rho \text{ is the density of measure } \mu)$$

Distributed Optimal Transport (DOT):

- Approximate c by a smooth $\tilde{c}$ around (x, x)
- **DOT velocity field:** $\mathbf{v} = (H\tilde{c})^{-1} \nabla \bar{\phi}$ (ideal)
- $\bar{\phi}$ – Steady state of primal-dual flow

Primal flow:

$$\partial_t \phi = \nabla \cdot (\lambda \nabla \phi) + (\rho - \rho^*),$$
$$\lambda \nabla \phi \cdot \mathbf{n} = 0 \text{ on } \partial\Omega$$

Dual flow (second-order):

$$\partial_t \lambda = [\theta]^+_\lambda,$$
$$\partial_t \theta = -\beta(t, x)\theta + \frac{1}{2}\left(|\nabla \phi|^2 - |\nabla \tilde{c}|^2\right),$$

DOT Flow

To investigate:

- Convergence of primal-dual flow
- Convergence of **on-the-fly implementation**: with $\mathbf{v} = -\left(H\tilde{c}\right)^{-1}\nabla\phi_t$

  (Use ϕ_t in place of $\bar{\phi}$ – Do not wait for p-d flow to converge)

Optimality conditions for Kantorovich dual problem:

$$-\nabla\cdot\left(\bar{\lambda}\nabla\bar{\phi}\right) = \rho - \rho^*,$$

$$\bar{\lambda}\nabla\bar{\phi}\cdot\mathbf{n} = 0, \qquad \text{on } \partial\Omega,$$

$$\bar{\lambda} \geq 0, \qquad |\nabla\bar{\phi}| \leq |\nabla\tilde{c}|, \qquad \bar{\lambda}(|\nabla\bar{\phi}| - |\nabla\tilde{c}|) = 0,$$

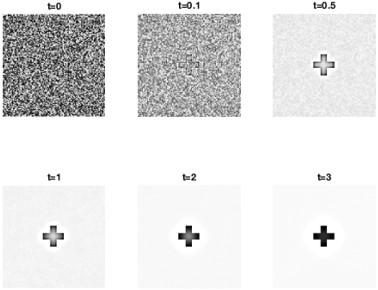
Lemma (Convergence of primal-dual flow)

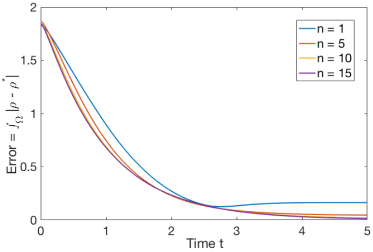
The solutions (ϕ_t, λ_t) to the primal-dual flow converge in the L^2-sense to the optimality conditions

DOT Flow

Theorem (Convergence of DOT flow)

Solutions ρ_t of the DOT flow with $\mathbf{v} = -\left(H\tilde{c}\right)^{-1}\nabla\phi_t$ converge in the L^2-norm to ρ^*





Thank You