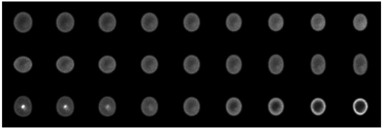
A Sparse Algorithm for Optimal Transport

Bernhard Schmitzer

DAUPHINE UNIVERSITÉ PARIS | CEREMADE UMR CNRS 7534

November 10, 2015

Applications of Numerical Optimal Transport



Versatile Tool

- computer vision & machine learning: histogram comparison [Rubner et al., 2000; Pele and Werman, 2009], ground metric learning [Wang and Guibas, 2012; Cuturi and Avis, 2014]
- imaging: interpolation [Maas et al., 2014], shape matching [Schmitzer and Schnörr, 2015], deformation analysis [Wang et al., 2012], denoising [Lellmann et al., 2014]
- optics [de Castro et al., 2014; Feng et al., 2015; Brix et al., 2015], physics [Frisch et al., 2002; Brenier, 2011], bakery logistics. . .

✗ computationally demanding

Solvers & Related Work

Discrete Solvers

- Hungarian method [Kuhn, 1955], Auction algorithm [Bertsekas, 1979], network simplex [Ahuja et al., 1993]

✓ numerically 'simple & robust', work on any cost-function

✗ scale poorly on large, dense problems

Solvers & Related Work

Discrete Solvers

- Hungarian method [Kuhn, 1955], Auction algorithm [Bertsekas, 1979], network simplex [Ahuja et al., 1993]

✓ numerically 'simple & robust', work on any cost-function

✗ scale poorly on large, dense problems

Continuous Solvers

- [Brenier, 1991; Haker et al., 2004; Carlier et al., 2010; Benamou et al., 2014], dynamic formulation [Benamou and Brenier, 2000]

✓ elegant theory, need only handle transport map

✗ restricted to particular ground costs, numerically more challenging

Solvers & Related Work

Discrete Solvers

- Hungarian method [Kuhn, 1955], Auction algorithm [Bertsekas, 1979], network simplex [Ahuja et al., 1993]

✓ numerically 'simple & robust', work on any cost-function

✗ scale poorly on large, dense problems

Continuous Solvers

- [Brenier, 1991; Haker et al., 2004; Carlier et al., 2010; Benamou et al., 2014], dynamic formulation [Benamou and Brenier, 2000]

✓ elegant theory, need only handle transport map

✗ restricted to particular ground costs, numerically more challenging

Approximations and Tricks

- approximations: cost thresholding [Pele and Werman, 2009], tangent space [Wang et al., 2012], entropy smoothing [Cuturi, 2013]
- multi-scale [Mérigot, 2011; Schmitzer and Schnörr, 2013], sparse iterations [Mérigot and Oudet, 2014; Schmitzer, 2015; Oberman and Ruan, 2015]

(Discrete) Optimal Transport in a Nutshell

Transport Plans / Couplings

$$\Pi(\mu,\nu) = \{\pi \in \mathrm{Prob}(X \times Y) \colon \mathrm{Proj}_{X\sharp}\pi = \mu,\ \mathrm{Proj}_{Y\sharp}\pi = \nu\}$$

(Discrete) Optimal Transport in a Nutshell

Transport Plans / Couplings

$$\Pi(\mu, \nu) = \{\pi \in \mathrm{Prob}(X \times Y) \colon \mathrm{Proj}_{X\sharp}\pi = \mu,\ \mathrm{Proj}_{Y\sharp}\pi = \nu\}$$

Primal Problem: $\inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y) \in X \times Y} c(x,y)\,\pi(x,y)$

(Discrete) Optimal Transport in a Nutshell

Transport Plans / Couplings

$$\Pi(\mu,\nu) = \{\pi \in \mathrm{Prob}(X \times Y) \colon \mathrm{Proj}_{X\sharp}\pi = \mu,\ \mathrm{Proj}_{Y\sharp}\pi = \nu\}$$

Primal Problem: $\displaystyle \inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y)\in X\times Y} c(x,y)\,\pi(x,y)$

Dual Problem: $\displaystyle \sup_{(\alpha,\beta)\in\mathbb{R}^{X\times Y}} \left[\sum_{x\in X} \alpha(x)\,\mu(x) + \sum_{y\in Y} \beta(y)\,\nu(y)\right]$

subject to $\alpha(x) + \beta(y) \leq c(x,y)$ for all $(x,y) \in X \times Y$.

(Discrete) Optimal Transport in a Nutshell

Transport Plans / Couplings

$$\Pi(\mu,\nu) = \{\pi \in \mathrm{Prob}(X \times Y) \colon \mathrm{Proj}_{X\sharp}\pi = \mu,\ \mathrm{Proj}_{Y\sharp}\pi = \nu\}$$

Primal Problem: $\inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y) \in X \times Y} c(x,y)\,\pi(x,y)$

Dual Problem: $\sup_{(\alpha,\beta) \in \mathbb{R}^{X \times Y}} \left[\sum_{x \in X} \alpha(x)\,\mu(x) + \sum_{y \in Y} \beta(y)\,\nu(y) \right]$

subject to $\alpha(x) + \beta(y) \leq c(x,y)$ for all $(x,y) \in X \times Y$.

(Discrete) Optimal Transport in a Nutshell

Transport Plans / Couplings

$$\Pi(\mu,\nu) = \{\pi \in \mathrm{Prob}(X \times Y) \colon \mathrm{Proj}_{X\sharp}\pi = \mu,\ \mathrm{Proj}_{Y\sharp}\pi = \nu\}$$

Primal Problem: $\inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y)\in X\times Y} c(x,y)\,\pi(x,y)$

Dual Problem: $\sup_{(\alpha,\beta)\in \mathbb{R}^{X\times Y}} \left[\sum_{x\in X} \alpha(x)\,\mu(x) + \sum_{y\in Y} \beta(y)\,\nu(y)\right]$

subject to $\alpha(x) + \beta(y) \leq c(x,y)$ for all $(x,y) \in X \times Y$.

PD Optimality Condition: $\pi(x,y) > 0 \Rightarrow \alpha(x) + \beta(y) = c(x,y)$

(Discrete) Optimal Transport in a Nutshell

Transport Plans / Couplings

$$\Pi(\mu,\nu) = \{\pi \in \mathrm{Prob}(X \times Y) \colon \mathrm{Proj}_{X\sharp}\pi = \mu,\ \mathrm{Proj}_{Y\sharp}\pi = \nu\}$$

Primal Problem: $\inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y)\in X\times Y} c(x,y)\,\pi(x,y)$

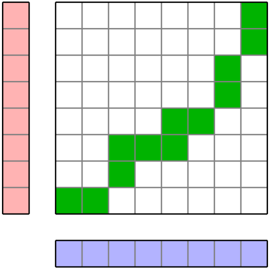
Dual Problem: $\sup_{(\alpha,\beta)\in\mathbb{R}^{X\times Y}} \left[\sum_{x\in X} \alpha(x)\,\mu(x) + \sum_{y\in Y} \beta(y)\,\nu(y)\right]$

subject to $\alpha(x) + \beta(y) \leq c(x,y)$ for all $(x,y) \in X \times Y$.

PD Optimality Condition: $\pi(x,y) > 0 \Rightarrow \alpha(x) + \beta(y) = c(x,y)$

Restricted Problem: $\mathcal{N} \subset X \times Y$

Multi-Scale Scheme

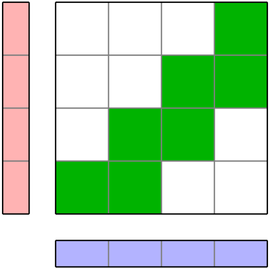


Intuition & Experience

- only small subset $\mathcal{N} \subset X \times Y$ relevant

How to select $\mathcal{N}$?

Multi-Scale Scheme



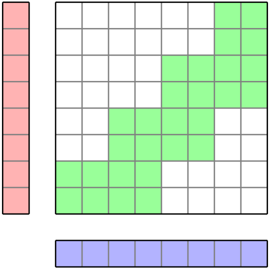
Intuition & Experience

- only small subset $\mathcal{N} \subset X \times Y$ relevant

How to select $\mathcal{N}$?

✓ multi-scale scheme

Multi-Scale Scheme



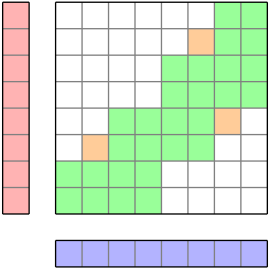
Intuition & Experience

- only small subset $\mathcal{N} \subset X \times Y$ relevant

How to select $\mathcal{N}$?

✓ multi-scale scheme

Multi-Scale Scheme



Intuition & Experience

- only small subset $\mathcal{N} \subset X \times Y$ relevant

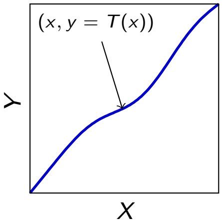
How to select $\mathcal{N}$?

✓ multi-scale scheme

How to guarantee global optimality?

- a priori estimates? ✗ very difficult
- a posteriori: quick verification, 'smart' updates of $\mathcal{N}$

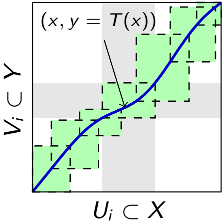
Polar Factorization & Local Optimality



Continuous Setting

- $X = Y = \mathbb{R}^n$, $c(x, y) = \|x - y\|^2$
- Optimal coupling induced by map $T : \mathbb{R}^n \to \mathbb{R}^n$: $\pi = (\mathrm{id}, T)_\sharp \mu$
- $T = \nabla\varphi$ for a convex potential $\varphi : \mathbb{R}^n \to \mathbb{R}$

Polar Factorization & Local Optimality

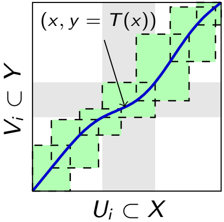


Continuous Setting

- $X = Y = \mathbb{R}^n$, $c(x, y) = \|x - y\|^2$
- Optimal coupling induced by map $T : \mathbb{R}^n \to \mathbb{R}^n$: $\pi = (\mathrm{id}, T)_\sharp \mu$
- $T = \nabla\varphi$ for a convex potential $\varphi : \mathbb{R}^n \to \mathbb{R}$

Local Optimality $\Rightarrow$ Global Optimality

Polar Factorization & Local Optimality



Continuous Setting

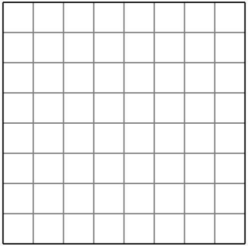
- $X = Y = \mathbb{R}^n$, $c(x, y) = \|x - y\|^2$
- Optimal coupling induced by map $T : \mathbb{R}^n \to \mathbb{R}^n$: $\pi = (\text{id}, T)_\sharp \mu$
- $T = \nabla\varphi$ for a convex potential $\varphi : \mathbb{R}^n \to \mathbb{R}$

Local Optimality $\Rightarrow$ Global Optimality

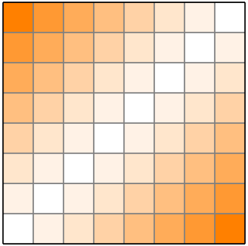
Discrete Equivalents:

- 1D: 'trivial': Monge property, 2D: not so much...

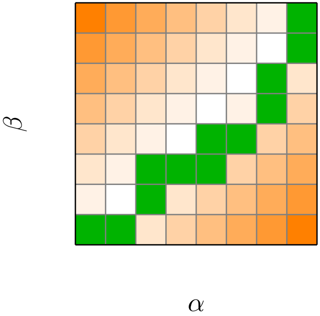
Short-Cuts



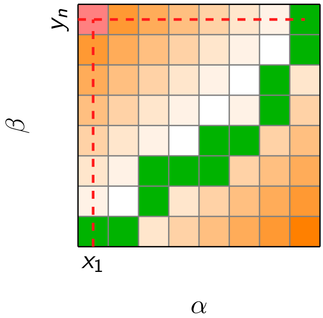
Short-Cuts



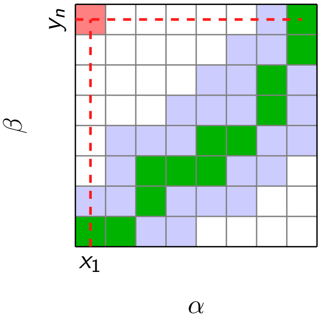
Short-Cuts



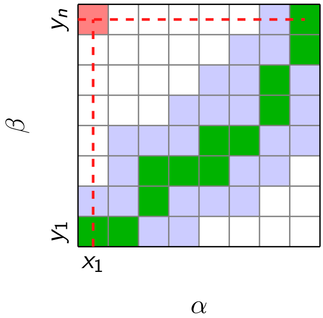
Short-Cuts



Short-Cuts

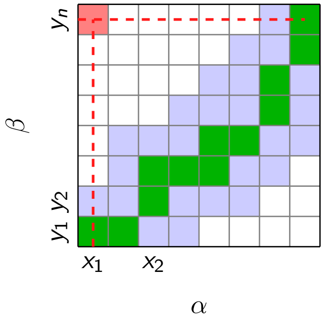


Short-Cuts



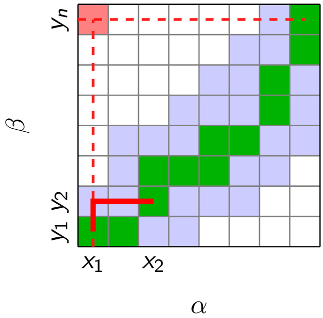
- $\alpha(x_1) + \beta(y_1) = c(x_1, y_1)$

Short-Cuts



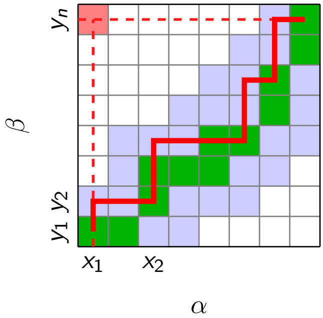
- $\alpha(x_1) + \beta(y_1) = c(x_1, y_1), \quad \alpha(x_1) + \beta(y_2) \leq c(x_1, y_2)$

Short-Cuts



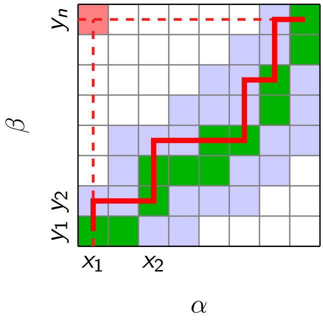
- $\alpha(x_1) + \beta(y_1) = c(x_1, y_1), \quad \alpha(x_1) + \beta(y_2) \leq c(x_1, y_2)$
- $\beta(y_2) \leq \beta(y_1) + [c(x_1, y_2) - c(x_1, y_1)]$

Short-Cuts



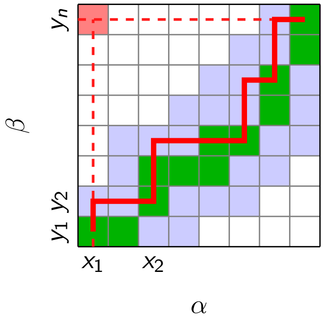
- $\alpha(x_1) + \beta(y_1) = c(x_1, y_1), \quad \alpha(x_1) + \beta(y_2) \leq c(x_1, y_2)$
- $\beta(y_2) \leq \beta(y_1) + [c(x_1, y_2) - c(x_1, y_1)]$
- $\alpha(x_1) + \beta(y_n) \leq c(x_1, y_2) + \sum_{i=2}^{n-1} [c(x_i, y_{i+1}) - c(x_i, y_i)]$

Short-Cuts



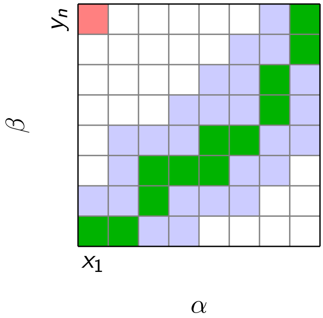
- $\alpha(x_1) + \beta(y_1) = c(x_1, y_1), \quad \alpha(x_1) + \beta(y_2) \leq c(x_1, y_2)$
- $\beta(y_2) \leq \beta(y_1) + [c(x_1, y_2) - c(x_1, y_1)]$
- $\alpha(x_1) + \beta(y_n) \leq c(x_1, y_2) + \sum_{i=2}^{n-1} [c(x_i, y_{i+1}) - c(x_i, y_i)] \leq c(x_1, y_n)?$

Short-Cuts

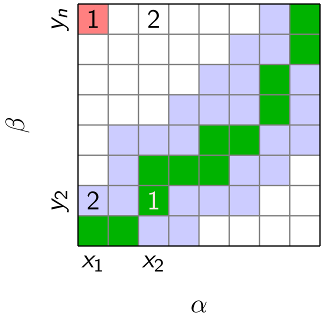


- $\alpha(x_1) + \beta(y_1) = c(x_1, y_1), \quad \alpha(x_1) + \beta(y_2) \leq c(x_1, y_2)$
- $\beta(y_2) \leq \beta(y_1) + [c(x_1, y_2) - c(x_1, y_1)]$
- $\alpha(x_1) + \beta(y_n) \leq c(x_1, y_2) + \sum_{i=2}^{n-1} [c(x_i, y_{i+1}) - c(x_i, y_i)] \leq c(x_1, y_n)?$
- continuum, $c(x, y) = \|x - y\|^2$: points along straight line are short-cuts

Shielding Condition

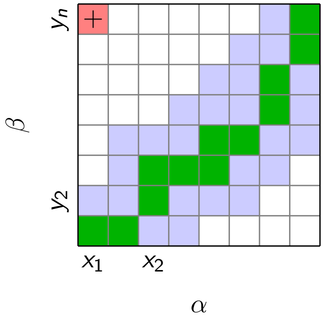


Shielding Condition



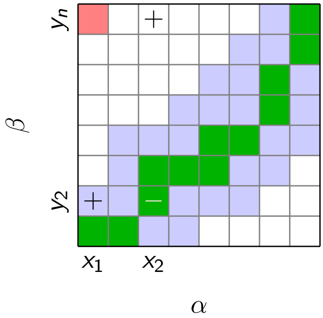
- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$

Shielding Condition



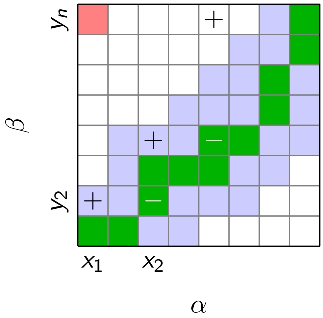
- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$

Shielding Condition



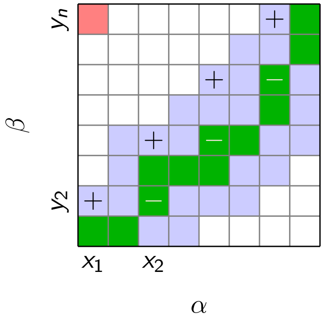
- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$

Shielding Condition



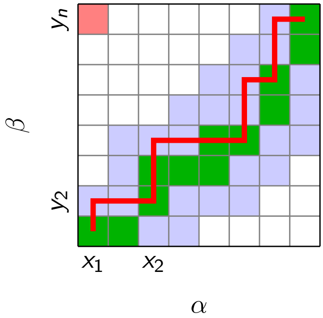
- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$

Shielding Condition



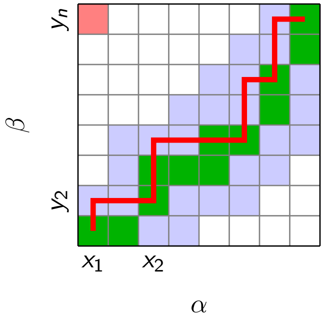
- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$

Shielding Condition



- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$

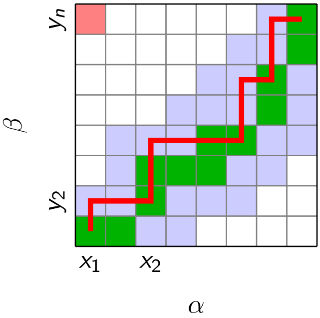
Shielding Condition



- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$
- shielding neighbourhood: always find a shielding cell

✓ existence of short-cuts follows

Shielding Condition



- $\underbrace{c(x_1, y_n) + c(x_2, y_2)}_{1} \geq \underbrace{c(x_1, y_2) + c(x_2, y_n)}_{2}$
- shielding neighbourhood: always find a shielding cell

✓ existence of short-cuts follows

✓ Local optimality + shielding neighbourhood ⇒ global optimality

A Sparse Algorithm

Ingredients

- sparse optimal transport solver $F : \mathcal{N} \mapsto \pi$
- construction of shielding neighbourhood $G : \pi \mapsto \mathcal{N}$

A Sparse Algorithm

Ingredients

- sparse optimal transport solver $F : \mathcal{N} \mapsto \pi$
- construction of shielding neighbourhood $G : \pi \mapsto \mathcal{N}$

Iteration

$$\pi_{k+1} = F(\mathcal{N}_k)$$
$$\mathcal{N}_{k+1} = G(\pi_{k+1})$$

until π_k is already locally optimal on $\mathcal{N}_k$.

A Sparse Algorithm

Ingredients

- sparse optimal transport solver $F : \mathcal{N} \mapsto \pi$
- construction of shielding neighbourhood $G : \pi \mapsto \mathcal{N}$

Iteration

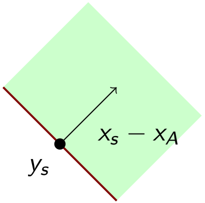
$$\pi_{k+1} = F(\mathcal{N}_k)$$
$$\mathcal{N}_{k+1} = G(\pi_{k+1})$$

until π_k is already locally optimal on $\mathcal{N}_k$.

Properties of Algorithm

- Calling F is fast when $\mathcal{N}_k$ is sparse. Any solver can be used as black box.
- When π_1 / $\mathcal{N}_1$ are good initial guesses: need only few iterations $\Rightarrow$ multi-scale scheme
- Design of G must exploit geometric structure of cost-function

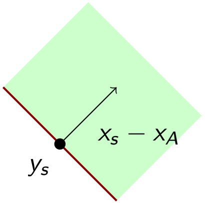
G for Squared Euclidean Distance

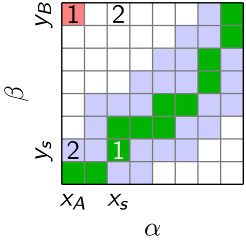


Shielding Condition for $c(x, y) = \|x - y\|^2$

$$\langle y_B - y_s, x_s - x_A \rangle > 0$$

G for Squared Euclidean Distance

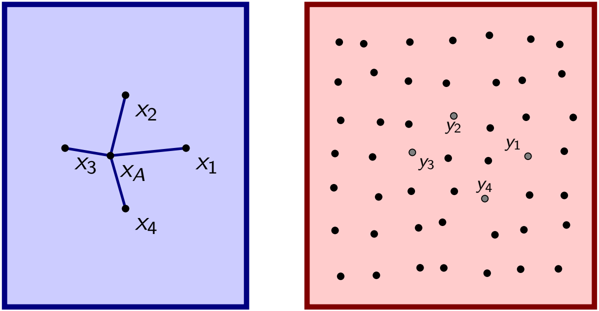




Shielding Condition for $c(x, y) = \|x - y\|^2$

$$\langle y_B - y_s, x_s - x_A \rangle > 0$$

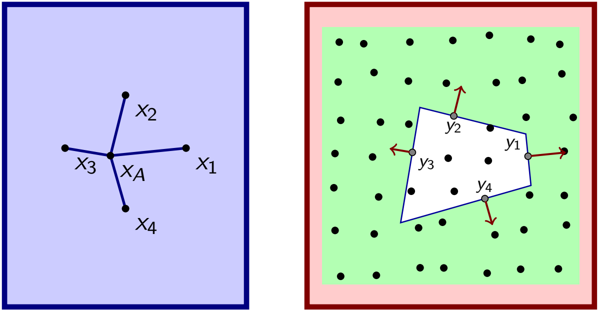
G for Squared Euclidean Distance



Shielding Condition for $c(x, y) = \|x - y\|^2$

$$\langle y_B - y_s, x_s - x_A \rangle > 0$$

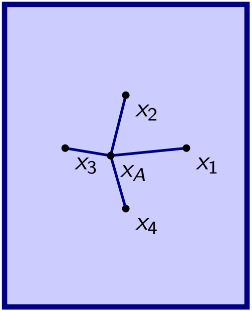
G for Squared Euclidean Distance

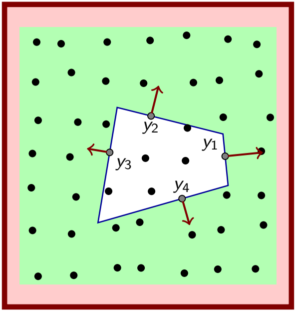


Shielding Condition for $c(x, y) = \|x - y\|^2$

$$\langle y_B - y_s, x_s - x_A \rangle > 0$$

G for Squared Euclidean Distance



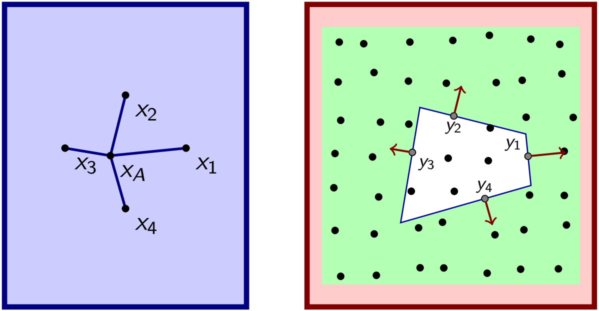


Shielding Condition for $c(x, y) = \|x - y\|^2$

$$\langle y_B - y_s, x_s - x_A \rangle > 0$$

✓ regular grids, ✓ point-clouds with tree structure

G for Squared Euclidean Distance



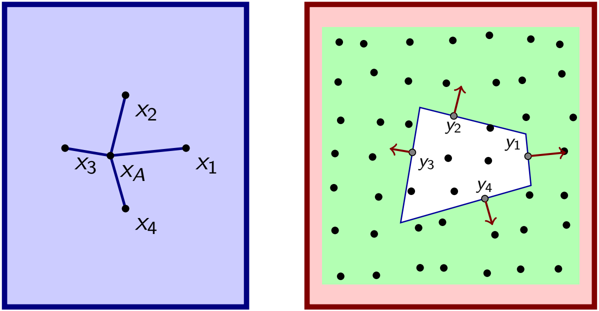
Shielding Condition for $c(x, y) = \|x - y\|^2$

$$\langle y_B - y_s, x_s - x_A \rangle > 0$$

✓ regular grids, ✓ point-clouds with tree structure

- mass assignment regular* $\Rightarrow |\mathcal{N}| = \mathcal{O}(|X|) \ll \mathcal{O}(|X \times Y|)$

G for Squared Euclidean Distance



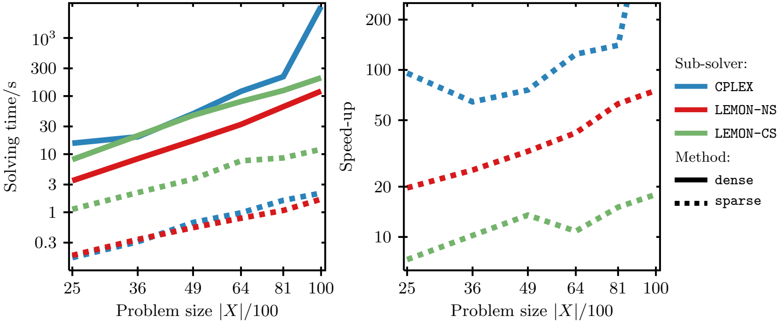
Shielding Condition for $c(x, y) = \|x - y\|^2 + \varepsilon(x, y)$

$$\langle y_B - y_s, x_s - x_A \rangle > 0$$

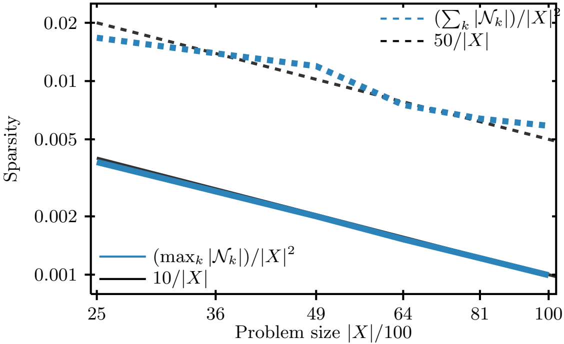
✓ regular grids, ✓ point-clouds with tree structure

- mass assignment regular* $\Rightarrow |\mathcal{N}| = \mathcal{O}(|X|) \ll \mathcal{O}(|X \times Y|)$
- can deal with noise
- more general costs …

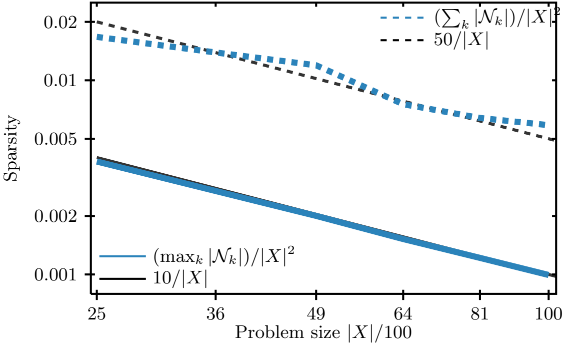
Numerical Results: Speed-up



Numerical Results: Sparsity

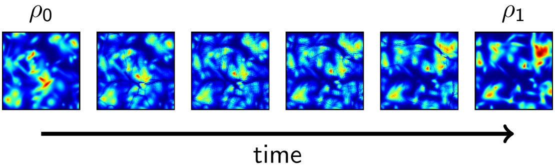


Numerical Results: Sparsity

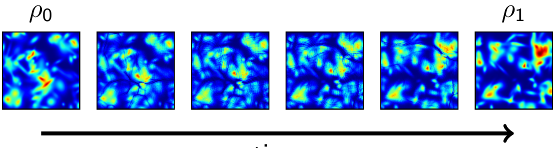


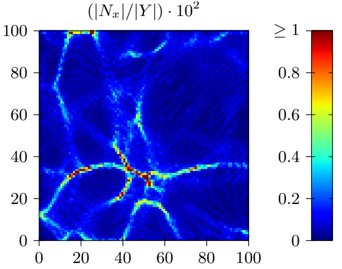
- 95% quantile of iteration numbers: 8

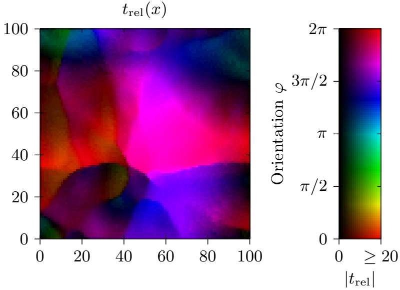
Numerical Results: Sparsity II



Numerical Results: Sparsity II

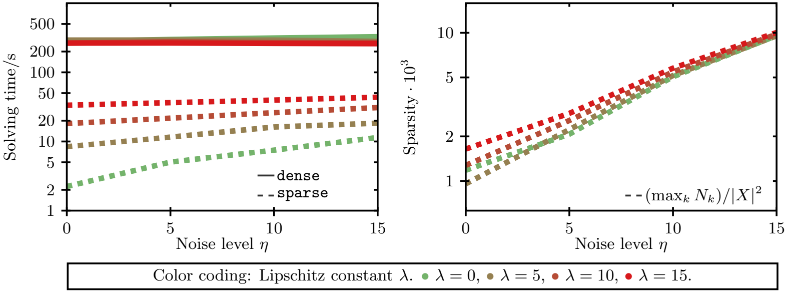






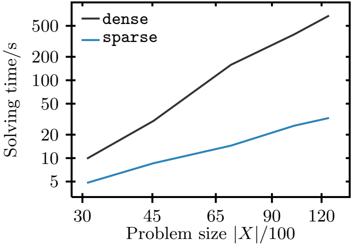
- $N_x = \{y \in Y : (x, y) \in \mathcal{N}\}$
- t_{rel}: Barycentric projection of relative transport map

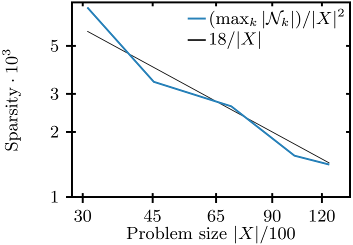
Numerical Results: Noisy Costs



- noise: random (η) + Lipschitz component (λ)
- slower with increasing noise (expected), ✓ no immediate breakdown

Shielding Neighbourhoods for More General Costs

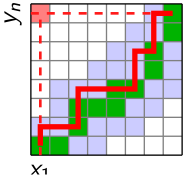




Preliminary Results

- ✗ more complicated,
  ✓ point-clouds with tree structure (⇒ multi-scale scheme)
- intuition: ✓ strictly convex costs, ✓ squared geodesic distance on sphere
- numerics: • speed, ✓ sparsity

Summary & Outlook



Summary

- ✓ verify global optimality locally ⇔ analogy to continuum
- ✓ basis for efficient sparsification of dense problems ⇒ combinatorial algorithms become applicable
- ✓ speed-up and saves memory

Open Questions

- closer look at other cost functions
- computational complexity
- code!

ArXiv: B. Schmitzer 'A Sparse Multi-Scale Algorithm for Dense Optimal Transport' 10/2015

References I

R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc., 1993.

J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3): 375–393, 2000.

J.-D. Benamou, B. D. Froese, and A. M. Oberman. Numerical solution of the optimal transportation problem using the Monge–Ampère equation. *Journal of Computational Physics*, 260(1):107–126, 2014.

D. P. Bertsekas. A distributed algorithm for the assignment problem. Technical report, Lab. for Information and Decision Systems Report, MIT, May 1979.

Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

Y. Brenier. A modified least action principle allowing mass concentrations for the early universe reconstruction problem. *Confluentes Mathematici*, 3(3):316–385, 2011.

K. Brix, Y. Hafizogullari, and A. Platen. Designing illumination lenses and mirrors by the numerical solution of Monge-Ampère equations. *Journal of the Optical Society of America A*, 32(11):2227–2236, 2015.

G. Carlier, A. Galichon, and F. Santambrogio. From Knothe's transport to Brenier's map and a continuation method for optimal transport. *SIAM J. Math. Anal.*, 41: 2554–2576, 2010.

References II

M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation distances. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 2292–2300, 2013. http://arxiv.org/abs/1306.0895.

M. Cuturi and D. Avis. Ground metric learning. *Journal of Machine Learning Research*, 15:533–564, 2014.

P. M. M. de Castro, Q. Mérigot, and B. Thibert. Far-field reflector problem and intersection of paraboloids. http://arxiv.org/abs/1403.0062, 2014.

Z. Feng, B. D. Froese, and R. Liang. A composite method for precise freeform optical beam shaping. *Applied Optics*, 54(31):9364–9369, 2015.

U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski. A reconstruction of the initial conditions of the universe by optimal mass transportation. *Nature*, 417: 260–262, 2002.

S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *Int. J. Comp. Vision*, 60:225–240, December 2004.

H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics*, 2:83–97, 1955.

J. Lellmann, D. A. Lorenz, C. Schönlieb, and T. Valkonen. Imaging with kantorovich-rubinstein discrepancy. *SIAM J. Imaging Sci.*, 7(4):2833–2859, 2014.

J. Maas, M. Rumpf, C. Schönlieb, and S. Simon. A generalized model for optimal transport of images including dissipation and density modulation. submitted, 2014.

References III

Q. Mérigot. A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5):1583–1592, 2011.

Q. Mérigot and E. Oudet. Discrete optimal transport: Complexity, geometry and applications. https://hal.archives-ouvertes.fr/hal-00604684/, 2014.

A. M. Oberman and Y. Ruan. An efficient linear programming method for optimal transportation. http://arxiv.org/abs/1509.03668, 2015.

O. Pele and W. Werman. Fast and robust Earth Mover's Distances. In *International Conference on Computer Vision (ICCV 2009)*, 2009.

Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int. J. Comp. Vision*, 40(2):99–121, 2000.

B. Schmitzer. A sparse algorithm for dense optimal transport. In *Scale Space and Variational Methods (SSVM 2015)*, pages 629–641, 2015.

B. Schmitzer and C. Schnörr. A hierarchical approach to optimal transport. In *Scale Space and Variational Methods (SSVM 2013)*, pages 452–464, 2013.

B. Schmitzer and C. Schnörr. Globally optimal joint image segmentation and shape matching based on Wasserstein modes. *Journal of Mathematical Imaging and Vision*, 52(3):436–458, 2015. doi: 10.1007/s10851-014-0546-8.

F. Wang and L. J. Guibas. Supervised earth mover's distance learning and its computer vision applications. In *European Conference on Computer Vision (ECCV 2012)*, pages 442–455, 2012.

References IV

W. Wang, D. Slepčev, S. Basu, J. A. Ozolek, and G. K. Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *Int. J. Comp. Vision*, 101:254–269, 2012.