Multi-scale algorithms for optimal transport

Bernhard Schmitzer

WWU MÜNSTER

June 2018

Overview

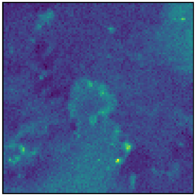
1. Introduction: optimal transport
2. Multi-scale methods
3. Shortcut algorithm
4. Sparse Sinkhorn algorithm

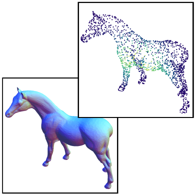
Overview

1. Introduction: optimal transport
2. Multi-scale methods
3. Shortcut algorithm
4. Sparse Sinkhorn algorithm

Metric measure spaces for data modelling

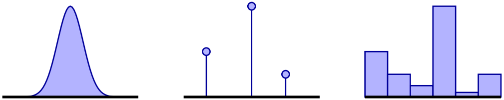
Comparing and understanding data





- ‘Are two samples similar?’

Language: positive Radon measures $\mathcal{M}_+(X)$ on metric space (X, d)

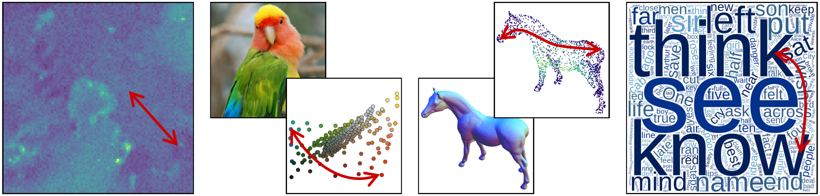


$\mathrm{d}\mu(x) = f(x)\,\mathrm{d}x$ $\qquad \mu = \sum_i \mu_i\,\delta_{x_i}$ $\qquad \mu \in \mathbb{R}^n_+$

- similarity of samples $\Leftrightarrow$ metric on $\mathcal{M}_+(X)$

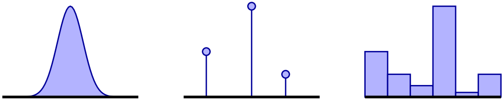
Metric measure spaces for data modelling

Comparing and understanding data



- 'Are two samples similar?'

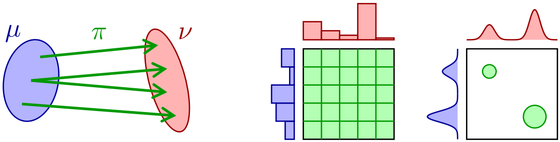
Language: positive Radon measures $\mathcal{M}_+(X)$ on metric space (X, d)



$\mathrm{d}\mu(x) = f(x)\,\mathrm{d}x$ $\quad\mu = \sum_i \mu_i\,\delta_{x_i}$ $\quad\mu \in \mathbb{R}^n_+$

- similarity of samples $\Leftrightarrow$ metric on $\mathcal{M}_+(X)$

Couplings and optimal transport



Couplings

- $\Pi(\mu,\nu) = \{\pi \in \mathcal{M}_+(X \times X) : \mathrm{P}_{1\sharp}\pi = \mu,\ \mathrm{P}_{2\sharp}\pi = \nu\}$
- **marginals**: $\mathrm{P}_{1\sharp}\pi(A) = \pi(A \times X)$, $\mathrm{P}_{2\sharp}\pi(B) = \pi(X \times B)$
- **rearrangement** of mass, generalization of **map**

Optimal transport [Kantorovich, 1942]

$$C(\mu,\nu) = \inf\left\{ \int_{X\times X} c(x,y)\,\mathrm{d}\pi(x,y) \,\middle|\, \pi \in \Pi(\mu,\nu) \right\}$$

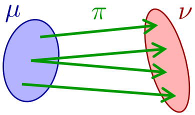
- **cost function** $c : X \times X \to \mathbb{R}$ for moving unit mass from x to y
- **convex problem**: linear program
- **minimizers exist** under mild assumptions

Wasserstein distance on probability measures $\mathcal{P}(X)$

$$W_p(\mu,\nu) = (C(\mu,\nu))^{1/p} \text{ for } c(x,y) = d(x,y)^p, \quad p \in [1,\infty)$$

Wasserstein distances: basic properties

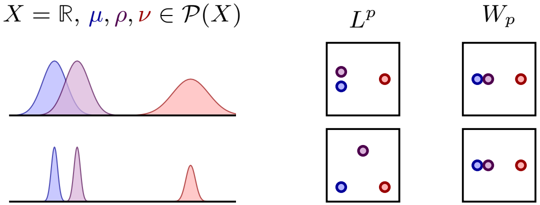
$$W_p(\mu,\nu) = \inf \left\{ \int_{X\times X} d(x,y)^p \, \mathrm{d}\pi(x,y) \,\middle|\, \pi \in \Pi(\mu,\nu) \right\}^{1/p}$$



Properties

- ✓ **intuitive:** minimal $\pi \Rightarrow$ optimal assignment
- ✓ 'respects' (X,d), **robust** to discretization errors, positional noise
- ✓ **flexible:** works for (almost) any metric space

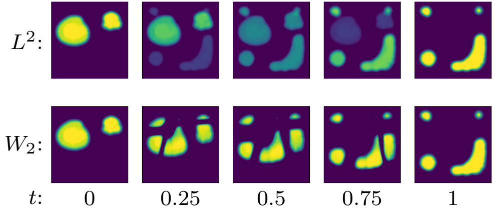
Comparison with L^p distances



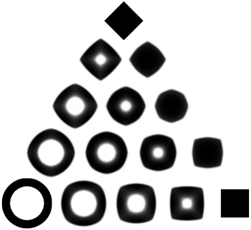
Wasserstein distances: advanced properties

Displacement interpolation

- (X, d) length space $\Rightarrow$ $(\mathcal{P}(X), W_p)$ is length space



Barycenter: weighted center of mass in $(\mathcal{P}(\mathbb{R}^d), W_2)$



- [Agueh and Carlier, 2011; Cuturi and Doucet, 2014; Benamou et al., 2015]

Wasserstein distances: summary

Attractive properties

- ✓ **intuitive, robust, flexible** metric for probability measures
- ✓ **rich geometric structure** (displacement interpolation, barycenters, gradient flows)
- ✓ accessible by **convex optimization**

⇒ Increasingly successful as numerical tool in data analysis

- [Rubner et al., 2000; Pele and Werman, 2009; Wang et al., 2012; Solomon et al., 2012; Cuturi and Avis, 2014; Peyré et al., 2016; Papadakis and Rabin, 2017; Mandad et al., 2017; Thorpe et al., 2017; Tameling et al., 2017]...

Limitations and open questions

- naive **numerical computation expensive**
- only for **probability measures**
- non-scalar data, spatial regularity, ...

Wasserstein distances: summary

Attractive properties

- ✓ **intuitive, robust, flexible** metric for probability measures
- ✓ **rich geometric structure** (displacement interpolation, barycenters, gradient flows)
- ✓ accessible by **convex optimization**

⇒ Increasingly successful as numerical tool in data analysis

- [Rubner et al., 2000; Pele and Werman, 2009; Wang et al., 2012; Solomon et al., 2012; Cuturi and Avis, 2014; Peyré et al., 2016; Papadakis and Rabin, 2017; Mandad et al., 2017; Thorpe et al., 2017; Tameling et al., 2017]. . .

Limitations and open questions

- naive **numerical computation expensive**
- only for **probability measures**
- non-scalar data, spatial regularity, . . .

Overview

1. Introduction: optimal transport
2. Multi-scale methods
3. Shortcut algorithm
4. Sparse Sinkhorn algorithm

Solvers and algorithms

"Classical" methods in a nutshell

- either ✓**flexible** but ✗**slow**
  - Hungarian method [Kuhn, 1955],
  - auction algorithm [Bertsekas, 1979],
  - network simplex [Ahuja et al., 1993]
- or ✓**efficient** but ✗**not very flexible**
  - [Aurenhammer et al., 1998; Haker et al., 2004; Benamou et al., 2014]

Entropy regularization and Sinkhorn algorithm

- [Wilson, 1969; Kosowsky and Yuille, 1994; Cuturi, 2013]
- **approximate method:** introduces some **blur** (✓/ ✗)

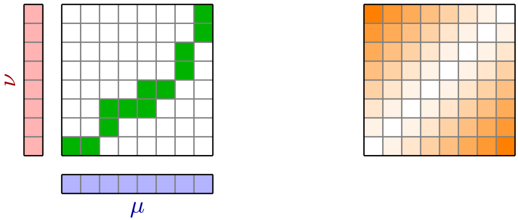
✓ **simple** algorithm, easy to **parallelize**

✓ **versatile:** generalizes to
  - **barycenters** [Benamou et al., 2015]
  - **gradient flows** [Peyré, 2015]
  - **unbalanced transport** [Chizat, Peyré, Schmitzer, and Vialard, 2016]

✗ **no free lunch**: slow for small regularization, memory intensive

Overview: G. Peyré and M. Cuturi. Computational Optimal Transport. arXiv:1803.00567, 2018.

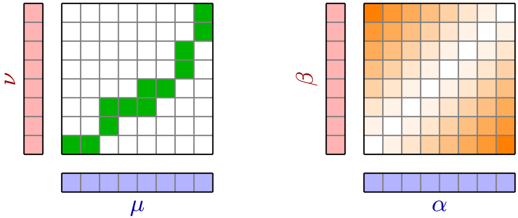
Kantorovich formulation in a nutshell



- **marginals** μ, $\nu \in \mathcal{P}(X)$, **ground cost** $c : X \times X \to \mathbb{R}$, **couplings** $\Pi(\mu, \nu)$

$$C(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y) \in X \times X} c(x, y)\, \pi(x, y)$$

Kantorovich formulation in a nutshell



- **marginals** $\mu, \nu \in \mathcal{P}(X)$, **ground cost** $c : X \times X \to \mathbb{R}$, **couplings** $\Pi(\mu, \nu)$

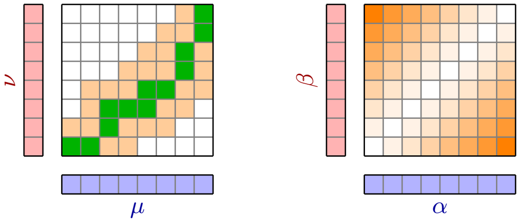
$$C(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y)\in X\times X} c(x,y)\,\pi(x,y)$$

- **dual problem**: prices $(\alpha, \beta) \in (\mathbb{R}^X, \mathbb{R}^X)$

$$C(\mu,\nu) = \sup_{(\alpha,\beta)} \sum_{x\in X} \alpha(x)\,\mu(x) + \sum_{y\in X} \beta(y)\,\nu(y) \quad \text{s.t. } \alpha(x) + \beta(y) \leq c(x,y)$$

- **PD optimality condition**: $[\pi(x,y) > 0] \Rightarrow [\alpha(x) + \beta(y) = c(x,y)]$

Kantorovich formulation in a nutshell



- **marginals** μ, $\nu \in \mathcal{P}(X)$, **ground cost** $c : X \times X \to \mathbb{R}$, **couplings** $\Pi(\mu, \nu)$

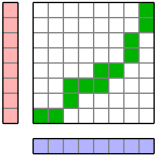
$$C(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \sum_{(x,y) \in X \times X} c(x, y)\, \pi(x, y)$$

- **dual problem**: prices $(\alpha, \beta) \in (\mathbb{R}^X, \mathbb{R}^X)$

$$C(\mu, \nu) = \sup_{(\alpha,\beta)} \sum_{x \in X} \alpha(x)\, \mu(x) + \sum_{y \in X} \beta(y)\, \nu(y) \quad \text{s.t. } \alpha(x) + \beta(y) \leq c(x, y)$$

- **PD optimality condition**: $[\pi(x, y) > 0] \Rightarrow [\alpha(x) + \beta(y) = c(x, y)]$
- **sparse sub-problem**: $\mathcal{N} \subset X \times X$

Multi-scale scheme

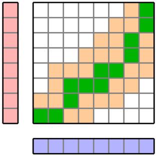


✓ Kantorovich: flexibility, simple discretization

✗ **high dimensionality** ($|X|^2$ variables)

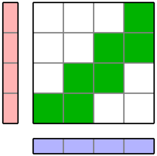
- often: optimal π has **sparse support**
  $\Rightarrow$ only sparse subset $\mathcal{N} \subset X \times X$ relevant ($|\mathcal{N}|$ variables)
- (related to polar factorization in continuum [Brenier, 1991])

Multi-scale scheme



- ✓ Kantorovich: flexibility, simple discretization
- ✗ **high dimensionality** ($|X|^2$ variables)
- often: optimal π has **sparse support**
  $\Rightarrow$ only sparse subset $\mathcal{N} \subset X \times X$ relevant ($|\mathcal{N}|$ variables)
- (related to polar factorization in continuum [Brenier, 1991])

Multi-scale scheme



✓ Kantorovich: flexibility, simple discretization

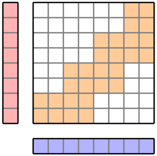
✗ **high dimensionality** ($|X|^2$ variables)

- often: optimal π has **sparse support**
  $\Rightarrow$ only sparse subset $\mathcal{N} \subset X \times X$ relevant ($|\mathcal{N}|$ variables)
- (related to polar factorization in continuum [Brenier, 1991])

Multi-scale scheme

- [Mérigot, 2011; Schmitzer and Schnörr, 2013; Glimm and Henscheid, 2013; Oberman and Ruan, 2015; Bartels and Schön, 2017; Bartels and Hertzog, 2017]
- **estimate** $\mathcal{N}$ on **coarser scale**

Multi-scale scheme



✓ Kantorovich: flexibility, simple discretization

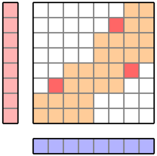
✗ **high dimensionality** ($|X|^2$ variables)

- often: optimal π has **sparse support**
  $\Rightarrow$ only sparse subset $\mathcal{N} \subset X \times X$ relevant ($|\mathcal{N}|$ variables)
- (related to polar factorization in continuum [Brenier, 1991])

Multi-scale scheme

- [Mérigot, 2011; Schmitzer and Schnörr, 2013; Glimm and Henscheid, 2013; Oberman and Ruan, 2015; Bartels and Schön, 2017; Bartels and Hertzog, 2017]
- **estimate** $\mathcal{N}$ on **coarser scale**

Multi-scale scheme



✓ Kantorovich: flexibility, simple discretization

✗ **high dimensionality** ($|X|^2$ variables)

- often: optimal π has **sparse support**
  $\Rightarrow$ only sparse subset $\mathcal{N} \subset X \times X$ relevant ($|\mathcal{N}|$ variables)
- (related to polar factorization in continuum [Brenier, 1991])

Multi-scale scheme

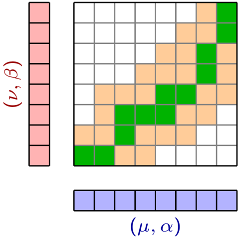
- [Mérigot, 2011; Schmitzer and Schnörr, 2013; Glimm and Henscheid, 2013; Oberman and Ruan, 2015; Bartels and Schön, 2017; Bartels and Hertzog, 2017]
- **estimate** $\mathcal{N}$ on **coarser scale**

Challenge: **rigorous guarantee of (near) global optimality**

- [Schmitzer and Schnörr, 2013; Schmitzer, 2016b]
- entropy regularization: [Schmitzer, 2016a]

Overview

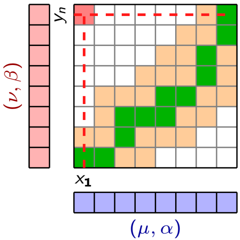
Shortcuts [Schmitzer, 2016b]



- given problem: μ, ν, cost c, neighbourhood $\mathcal{N}$
- **locally optimal primal & dual solution** on $\mathcal{N}$: (π, α, β)

$$\operatorname{spt}\pi \subset \mathcal{N}, \quad \alpha(x) + \beta(y) \begin{cases} \leq c(x,y) & \text{for } (x,y) \in \mathcal{N} \\ = c(x,y) & \text{for } (x,y) \in \operatorname{spt}\pi \end{cases} \qquad (*)$$

Shortcuts [Schmitzer, 2016b]

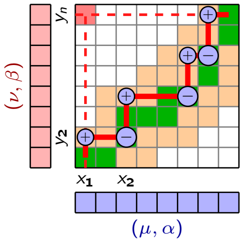


- given problem: μ, ν, cost c, neighbourhood $\mathcal{N}$
- **locally optimal primal & dual solution** on $\mathcal{N}$: (π, α, β)

$$\operatorname{spt}\pi \subset \mathcal{N}, \quad \alpha(x) + \beta(y) \begin{cases} \leq c(x,y) & \text{for } (x,y) \in \mathcal{N} \\ = c(x,y) & \text{for } (x,y) \in \operatorname{spt}\pi \end{cases} \tag{$*$}$$

- **global optimality** if all dual constraints satisfied

Shortcuts [Schmitzer, 2016b]



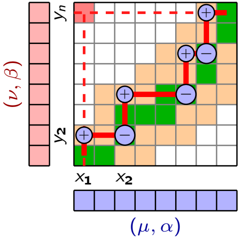
- given problem: μ, ν, cost c, neighbourhood $\mathcal{N}$
- **locally optimal primal & dual solution** on $\mathcal{N}$: (π, α, β)

$$\operatorname{spt} \pi \subset \mathcal{N}, \quad \alpha(x) + \beta(y) \begin{cases} \leq c(x,y) & \text{for } (x,y) \in \mathcal{N} \\ = c(x,y) & \text{for } (x,y) \in \operatorname{spt} \pi \end{cases} \qquad (*)$$

- **global optimality** if all dual constraints satisfied
- **Def: shortcut** for (x_1, y_n): tuple $((x_2, y_2), \ldots, (x_{n-1}, y_{n-1}))$ with $(x_i, y_i) \in \operatorname{spt} \pi$, $(x_i, y_{i+1}) \in \mathcal{N}$ and

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

Shortcuts [Schmitzer, 2016b]



- given problem: μ, ν, cost c, neighbourhood $\mathcal{N}$
- **locally optimal primal & dual solution** on $\mathcal{N}$: (π, α, β)

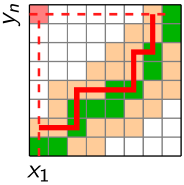
$$\operatorname{spt}\pi \subset \mathcal{N}, \quad \alpha(x) + \beta(y) \begin{cases} \leq c(x,y) & \text{for } (x,y) \in \mathcal{N} \\ = c(x,y) & \text{for } (x,y) \in \operatorname{spt}\pi \end{cases} \qquad (*)$$

- **global optimality** if all dual constraints satisfied
- **Def: shortcut** for (x_1, y_n): tuple $((x_2, y_2), \ldots, (x_{n-1}, y_{n-1}))$ with $(x_i, y_i) \in \operatorname{spt}\pi$, $(x_i, y_{i+1}) \in \mathcal{N}$ and

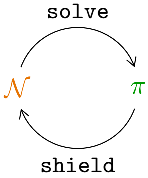
$$\alpha(x_1) + \beta(y_n) \overset{(*)}{\leq} c(x_1, y_2) + \sum_{i=2}^{n-1} [c(x_i, y_{i+1}) - c(x_i, y_i)] \leq c(x_1, y_n)$$

- **Lemma: local optimality + shortcuts $\Rightarrow$ global optimality**

Shortcut algorithm



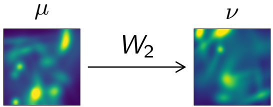
- search for shortcuts infeasible
- **shielding condition** for $(\mathcal{N}, \pi)$:
  local way to ensure existence of all shortcuts
- **algorithm:**
  until convergence:
  - $\pi \leftarrow$ solve local problem on $\mathcal{N}$
  - $\mathcal{N} \leftarrow$ generate shielding neighbourhood for π

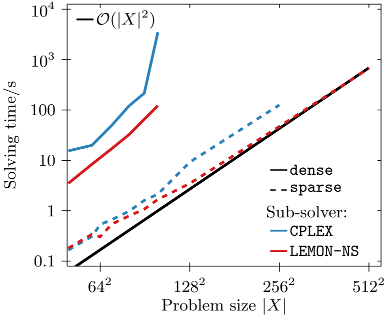


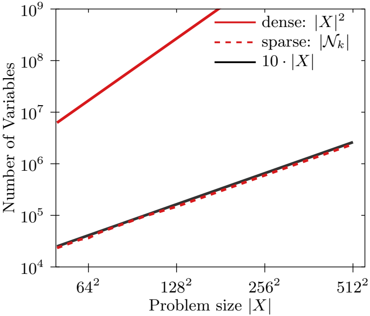
- **Thm:** returns **globally optimal solution**
- how to implement `shield`?
  - solved for "standard cases"
  - ongoing research

Numerical results

Example data:





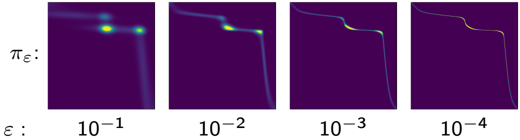


- ✓ significant speed-up
- ✗ run-time scaling approx. **quadratic**
- ✓ memory demand **linear** in marginal size

Overview

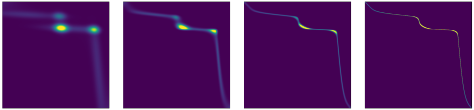
1. Introduction: optimal transport
2. Multi-scale methods
3. Shortcut algorithm
4. Sparse Sinkhorn algorithm

Sparse Sinkhorn algorithm



- **regularized problem:** $\pi_\varepsilon := \text{argmin}_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} c \, d\pi + \varepsilon \, \text{KL}(\pi|\rho)$
- ✗ π_ε is **dense**
- ✓ [Cominetti and San Martin, 1992]: $\pi_\varepsilon \to \pi_0$ exponentially for $\varepsilon \to 0$,
  π_0: sparse unregularized solution

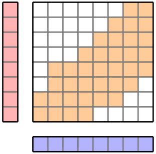
Sparse Sinkhorn algorithm



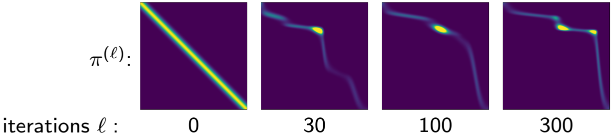
- **regularized problem:** $\pi_\varepsilon := \text{argmin}_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} c \, d\pi + \varepsilon \, \text{KL}(\pi|\rho)$
- ✗ π_ε is **dense**
- ✓ [Cominetti and San Martin, 1992]: $\pi_\varepsilon \to \pi_0$ exponentially for $\varepsilon \to 0$,
  π_0: sparse unregularized solution

Sparse approximation [Schmitzer, 2016a]

- for sparse subset $\mathcal{N} \subset X \times X$: **truncated coupling** $\pi \llcorner_{\mathcal{N}}$
- **Lemma:** bound for **truncation error** $\Delta(\pi \llcorner_{\mathcal{N}}, \alpha, \beta)$:
  small when little mass is truncated
- **sparse approximate algorithm:**
  - sparse Sinkhorn iterations on $\mathcal{N}$
  - update $\mathcal{N}$ when $\Delta(\pi \llcorner_{\mathcal{N}}, \alpha, \beta)$ too large
- more numerical tricks (e.g. ε-scaling, ...)



Sparse Sinkhorn algorithm



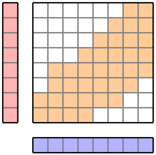
- **regularized problem:** $\pi_\varepsilon := \text{argmin}_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} c \, d\pi + \varepsilon \, \text{KL}(\pi|\rho)$

✗ π_ε is **dense**

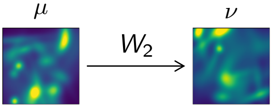
✓ [Cominetti and San Martin, 1992]: $\pi_\varepsilon \to \pi_0$ exponentially for $\varepsilon \to 0$, π_0: sparse unregularized solution

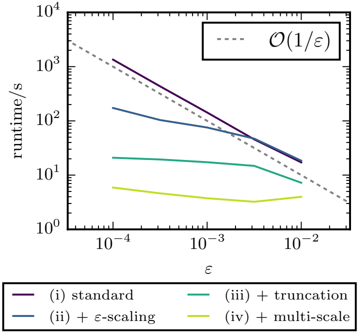
Sparse approximation [Schmitzer, 2016a]

- for sparse subset $\mathcal{N} \subset X \times X$: **truncated coupling** $\pi \llcorner \mathcal{N}$
- **Lemma:** bound for **truncation error** $\Delta(\pi \llcorner \mathcal{N}, \alpha, \beta)$: small when little mass is truncated
- **sparse approximate algorithm:**
  - sparse Sinkhorn iterations on $\mathcal{N}$
  - update $\mathcal{N}$ when $\Delta(\pi \llcorner \mathcal{N}, \alpha, \beta)$ too large
- more numerical tricks (e.g. ε-scaling, . . .)



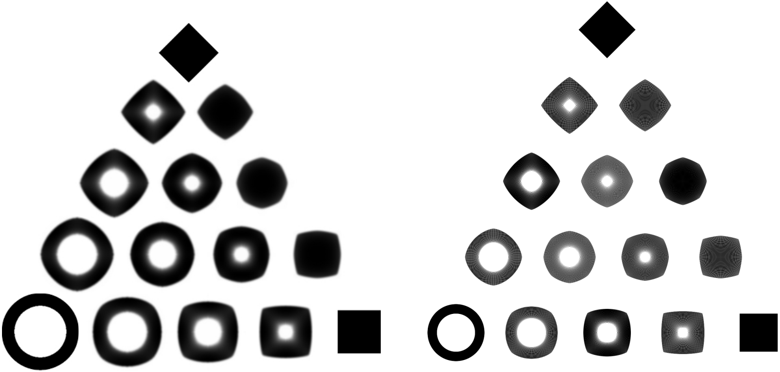
Numerical results: effect of computational adaptations

Example data: 



- standard Sinkhorn: number of iterations $\propto 1/\varepsilon$
- ✓ ε-scaling: reduces iterations
- ✓ acceleration with truncation (and multi-scale)

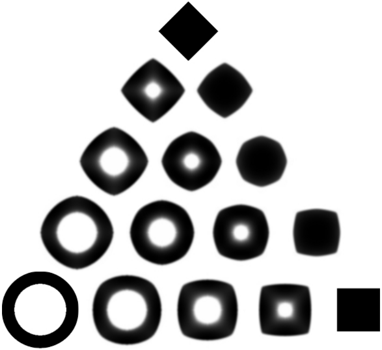
Numerical results: W_2 barycenter



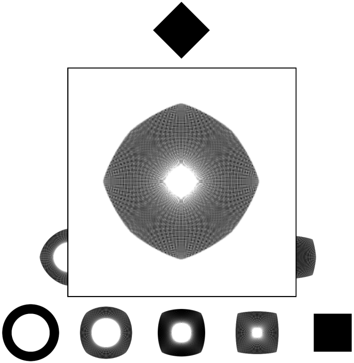
[Benamou et al., 2015] [Schmitzer, 2016a]

- ✓ adapted algorithm allows **smaller** $\varepsilon \Rightarrow$ **sharper results**
- ✓ **convergence of entropic regularization** for $\varepsilon \to 0$ [Carlier, Duval, Peyré, and Schmitzer, 2017]

Numerical results: W_2 barycenter



[Benamou et al., 2015]



[Schmitzer, 2016a]

- ✓ adapted algorithm allows **smaller** $\varepsilon \Rightarrow$ **sharper results**
- ✓ **convergence of entropic regularization** for $\varepsilon \to 0$ [Carlier, Duval, Peyré, and Schmitzer, 2017]

Overview

1. Introduction: optimal transport
2. Multi-scale methods
3. Shortcut algorithm
4. Sparse Sinkhorn algorithm

Conclusion

Optimal Transport

- ✓ flexible and robust tool for data analysis
- ✗ high naive computational complexity
  - ✓ steady development of more efficient methods
    - best algorithm depends on problem

Multi-scale methods

- Shortcut solver
- Sparse Sinkhorn algorithm
  - ✓ **versatile:** barycenters, gradient flows, unbalanced...

Further Reading

- B. Schmitzer. A sparse multi-scale algorithm for dense optimal transport. *J. Math. Imaging Vis.*, 56(2):238–259, 2016
- B. Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. arXiv:1610.06519, 2016.
- G. Peyré and M. Cuturi. Computational Optimal Transport. arXiv:1803.00567, 2018.

Code online: `https://github.com/bernhard-schmitzer`

References I

M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.

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

F. Aurenhammer, F. Hoffmann, and B. Aronov. Minkowski-type theorems and least-squares clustering. *Algorithmica*, 20(1):61–76, 1998. doi: 10.1007/PL00009187.

S. Bartels and S. Hertzog. Error bounds for discretized optimal transport and its reliable efficient numerical solution. arXiv:1710.04888, 2017.

S. Bartels and P. Schön. Adaptive approximation of the Monge–Kantorovich problem via primal-dual gap estimates. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(6):2237–2261, 2017.

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.

J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2): A1111–A1138, 2015. URL `https://hal.archives-ouvertes.fr/hal-01096124`.

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

References II

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

G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling algorithms for unbalanced transport problems. to appear in Mathematics of Computation, arXiv:1607.05816, 2016.

R. Cominetti and J. San Martin. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, 67:169–187, 1992.

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

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

M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning*, 2014.

T. Glimm and N. Henscheid. Iterative scheme for solving optimal transportation problems arising in reflector design. *ISRN Applied Mathematics*, 2013.

S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *Int. J. Comp. Vision*, 60(3):225–240, December 2004.

References III

L. V. Kantorovich. O peremeshchenii mass. *Doklady Akademii Nauk SSSR*, 37(7–8): 227–230, 1942.

J. Kosowsky and A. Yuille. The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks*, 7(3):477–490, 1994.

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

M. Mandad, D. Cohen-Steiner, L. Kobbelt, P. Alliez, and M. Desbrun. Variance-minimizing transport plans for inter-surface mapping. https://hal.inria.fr/hal-01519006/, 2017.

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

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

N. Papadakis and J. Rabin. Convex histogram-based joint image segmentation with regularized optimal transport cost. *J. Math. Imaging Vis.*, 2017.

O. Pele and W. Werman. Fast and robust earth mover's distances. In *International Conference on Computer Vision (ICCV 2009)*, 2009.

References IV

G. Peyré, M. Cuturi, and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning (ICML 2016)*, pages 2664–2672, 2016. URL `https://hal.archives-ouvertes.fr/hal-01322992`.

G. Peyré. Entropic approximation of Wasserstein gradient flows. *SIAM J. Imaging Sci.*, 8(4):2323–2351, 2015.

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. Stabilized sparse scaling algorithms for entropy regularized transport problems. arXiv:1610.06519, 2016a.

B. Schmitzer. A sparse multi-scale algorithm for dense optimal transport. *J. Math. Imaging Vis.*, 56(2):238–259, 2016b.

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

J. Solomon, A. Nguyen, A. Butscher, M. Ben-Chen, and L. Guibas. Soft maps between surfaces. *Computer Graphics Forum*, 31(5), 2012. ISSN 1467-8659. doi: 10.1111/j.1467-8659.2012.03167.x.

C. Tameling, M. Sommerfeld, and A. Munk. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. arXiv:1707.00973, 2017.

References V

M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev. A transportation Lp distance for signal analysis. *J. Math. Imaging Vis.*, 2017. doi: 10.1007/s10851-017-0726-4.

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.

A. G. Wilson. The use of entropy maximizing models, in the theory of trip distribution, mode split and route split. *Journal of Transport Economics and Policy*, pages 108–126, 1969.