Cost-Dependent Clustering: A General Multiscale Approach

Optimal Transport meets Probability, Statistics and Machine Learning, Oaxaca

Jörn Schrieber, Anita Schöbel, Dominic Schuhmacher

May 1, 2017

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Outline

Introduction

Cost-Dependent Clustering

Examples

Multiscale Approach for Discrete Optimal Transport

Setup:

- finite sets $X = \{x_1, \ldots, x_N\}$, $Y = \{y_1, \ldots, y_M\}$,
- measures μ on X, ν on Y,
- an arbitrary cost function $c \colon X \times Y \to \mathbb{R}_+$,
- as a matrix: $c_{i,j} = c(x_i, y_j)$

Multiscale Approach for Discrete Optimal Transport

Setup:

- finite sets $X = \{x_1, \dots, x_N\}$, $Y = \{y_1, \dots, y_M\}$,
- measures μ on X, ν on Y,
- an arbitrary cost function $c\colon X \times Y \to \mathbb{R}_+$,
- as a matrix: $c_{i,j} = c(x_i, y_j)$

Kantorovich formulation:

$$\mathrm{OT}(\mu,\nu,c) \quad \min \quad \sum_{i,j} c_{i,j}\pi_{i,j}$$
$$\text{s.t.} \quad \pi \in \Pi(\mu,\nu)$$

Multiscale Approach for Discrete Optimal Transport

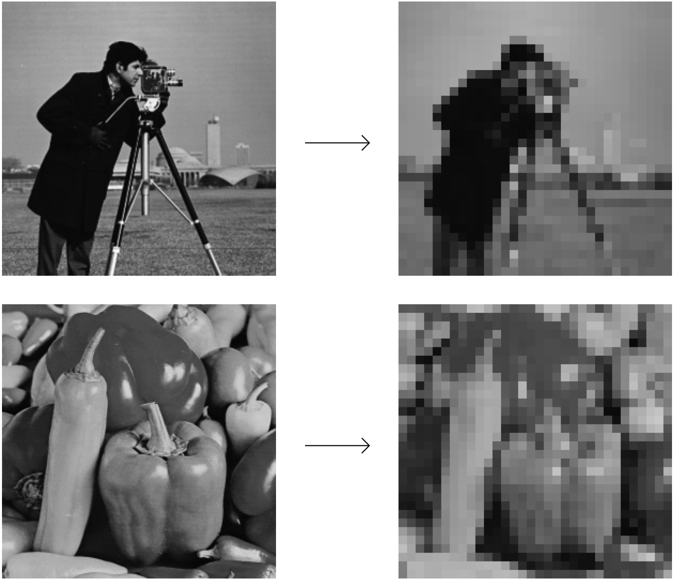
Methods for regular grids and point clouds

- [Oberman, Ruan, 2015]
- [Gerber, Maggioni, 2015]
- [Schmitzer, 2016]

- Idea: Find a hierarchical partition (clustering) for X and Y.
- Transfer a solution on a coarse scale to the next finer scale.

$$(X, Y) \underset{\text{coarsen}}{\overset{\text{transfer}}{\leftrightarrows}} (X^{(1)}, Y^{(1)}) \underset{\text{coarsen}}{\overset{\text{transfer}}{\leftrightarrows}} (X^{(2)}, Y^{(2)}) \quad \dots$$

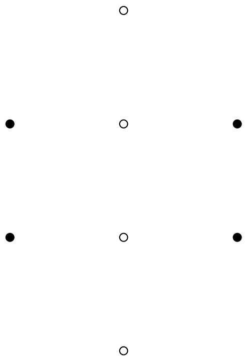
Standard Coarsening for Images



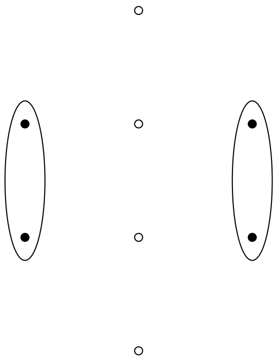
Three Main Goals

- extend multiscale schemes to arbitrary cost functions
- derive general lower and upper bounds on the optimal objective value
- evaluate the quality of a given clustering

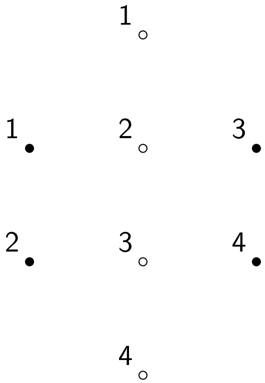
Example for Cost-Dependent Clustering



Example for Cost-Dependent Clustering



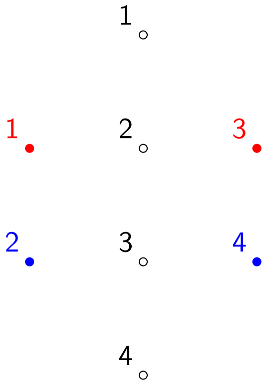
Example for Cost-Dependent Clustering



cost matrix
(squared Euclidean distance)

$$\begin{pmatrix} 2 & 1 & 2 & 5 \\ 5 & 2 & 1 & 2 \\ 2 & 1 & 2 & 5 \\ 5 & 2 & 1 & 2 \end{pmatrix}$$

Example for Cost-Dependent Clustering



Cost matrix
(squared Euclidean distance)

$$\begin{pmatrix} 2 & 1 & 2 & 5 \\ 5 & 2 & 1 & 2 \\ 2 & 1 & 2 & 5 \\ 5 & 2 & 1 & 2 \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 2 & 1 & 2 & 5 \\ 5 & 2 & 1 & 2 \end{pmatrix}$$

General Notation

- Let $X^{cl} = \{X_1^{cl}, \ldots, X_n^{cl}\}$ be a partition of X with $n < N$ subsets and $Y^{cl} = \{Y_1^{cl}, \ldots, Y_m^{cl}\}$ a partition of Y with $m < M$ subsets.
- Define measures μ^{cl} and ν^{cl} on X^{cl} and Y^{cl} respectively by $\mu^{cl}(X_k^{cl}) = \mu(X_k^{cl})$ and $\nu^{cl}(Y_l^{cl}) = \nu(Y_l^{cl})$.
- Define:
    - the minimum clustering cost matrix $c^{\min}$ and
    - the maximum clustering cost matrix $c^{\max}$ by

    $$c_{k,l}^{\min} = \min_{\substack{x_i \in X_k^{cl} \\ y_j \in Y_l^{cl}}} c_{i,j} \quad \text{and} \quad c_{k,l}^{\max} = \max_{\substack{x_i \in X_k^{cl} \\ y_j \in Y_l^{cl}}} c_{i,j}$$

    for $k = 1, \ldots, n$, $l = 1, \ldots, m$,
    - the clustering gap matrix $g = c^{\max} - c^{\min}$.

Lower and Upper Bounds

- Let c^* be the optimal objective value of $\mathrm{OT}(\mu, \nu, c)$.
- Let $\pi^{\min *}$ be an optimal coupling for $\mathrm{OT}(\mu^{cl}, \nu^{cl}, c^{\min})$.
- Then:

$$\sum_{k,l} c_{k,l}^{\min} \pi_{k,l}^{\min *} \leq c^* \leq \sum_{k,l} c_{k,l}^{\max} \pi_{k,l}^{\min *}$$

Don't want to compute $\pi^{\min *}$?

- Define

$$G_{row} = \sum_k \mu_k^{cl} \max_l g_{k,l} \quad \text{and} \quad G_{col} = \sum_l \nu_l^{cl} \max_k g_{k,l}$$

- Let $\tilde{c}$ be a cost function on $X^{cl} \times Y^{cl}$, such that $c^{\min} \leq \tilde{c} \leq c^{\max}$.
- Let $\tilde{c}^*$ be the optimal objective value for $\mathrm{OT}(\mu^{cl}, \nu^{cl}, \tilde{c})$.
- Then

$$|c^* - \tilde{c}^*| \leq \min\{G_{row}, G_{col}\}.$$

- The right hand side only depends on the clustering!

General Cost-Dependent Clustering Problem

- Given μ, ν and a general cost matrix c.
- For fixed $n, m \in \mathbb{N}$, find partitions of X and Y into at most n and m sets, respectively, such that $\min\{G_{row}, G_{col}\}$ is minimized.

General Cost-Dependent Clustering Problem

- Given μ, ν and a general cost matrix c.
- For fixed $n, m \in \mathbb{N}$, find partitions of X and Y into at most n and m sets, respectively, such that $\min\{G_{row}, G_{col}\}$ is minimized.

- This is a difficult combinatorial problem.
- Finding an optimal solution is often impractical.
- Fast heuristics are helpful.
- Any additional information about the sturcture of c might be useful to find a good heuristic.

A General Heuristic

- Sample n distinct elements $\hat{x}_1, \ldots, \hat{x}_n$ from μ.
- Sample m distinct elements $\hat{y}_1, \ldots, \hat{y}_m$ from ν.

A General Heuristic

- Sample n distinct elements $\hat{x}_1, \ldots, \hat{x}_n$ from μ.
- Sample m distinct elements $\hat{y}_1, \ldots, \hat{y}_m$ from ν.
- For $i = 1, \ldots, N$ assign x_i to one of $\hat{x}_1, \ldots, \hat{x}_n$ via

$$\min_{k=1,\ldots,n} \sum_{l=1}^{m} \nu(\hat{y}_l) |c(x_i, \hat{y}_l) - c(\hat{x}_k, \hat{y}_l)|.$$

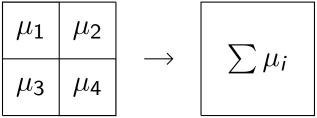
- For $j = 1, \ldots, M$ assign y_j to one of $\hat{y}_1, \ldots, \hat{y}_m$ via

$$\min_{l=1,\ldots,m} \sum_{k=1}^{n} \mu(\hat{x}_k) |c(\hat{x}_k, y_j) - c(\hat{x}_k, \hat{y}_l)|.$$

- Choose $\tilde{c}_{k,l} = c(\hat{x}_k, \hat{y}_l)$ or compute $c^{\min}$ and $c^{\max}$.

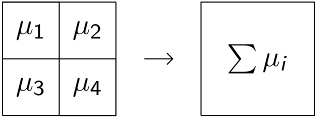
Example 1: Standard Image Coarsening Approach

- Consider the standard coarsening of images, where
  - 4 pixels are aggregated into one,
  - $c(x_i, y_j) = \|x_i - y_j\|$,
  - pixels have width h.



Example 1: Standard Image Coarsening Approach

- Consider the standard coarsening of images, where
  - 4 pixels are aggregated into one,
  - $c(x_i, y_j) = \|x_i - y_j\|$,
  - pixels have width h.



- Then we have
  - $\max_l g_{k,l} = 2\sqrt{2}h$ for all k,
  - $c^{\min} \leq \tilde{c} \leq c^{\max}$,
- therefore

$$|c^* - \tilde{c}^*| \leq G_{row} = 2\sqrt{2}h \cdot \mu(X).$$

Example 2: K-Means Clustering

- Consider a point cloud X in $\mathbb{R}^d$.
- Sample K random points of X distributed by μ.

Example 2: K-Means Clustering

- Consider a point cloud X in $\mathbb{R}^d$.
- Sample K random points of X distributed by μ.
- Use these points as an initial set of means for the K-means clustering algorithm, then alternate between
  - Assignment step: assign every point to the cluster it is closest to (in squared Euclidean distance)
  - Update step: compute new barycenters for each cluster
- until the same clustering occurs twice.

Example 2: K-Means Clustering

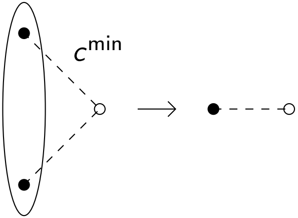
- Consider a point cloud X in $\mathbb{R}^d$.
- Sample K random points of X distributed by μ.
- Use these points as an initial set of means for the K-means clustering algorithm, then alternate between
  - Assignment step: assign every point to the cluster it is closest to (in squared Euclidean distance)
  - Update step: compute new barycenters for each cluster
- until the same clustering occurs twice.
- Let $X^{cl} = \{x_1^{cl}, \dots, x_K^{cl}\}$ be the set of barycenters.
- Repeat to obtain L clusters for Y.

Example 2: K-Means Clustering (cont.)

- If we choose $\tilde{c}_{k,l} = \|x_k^{cl} - y_l^{cl}\|^2$, then $\tilde{c} \leq c^{\max}$.
- But: $c^{\min} \leq \tilde{c}$ is not necessarily fulfilled!

Example 2: K-Means Clustering (cont.)

- If we choose $\tilde{c}_{k,l} = \|x_k^{cl} - y_l^{cl}\|^2$, then $\tilde{c} \le c^{\max}$.
- But: $c^{\min} \le \tilde{c}$ is not necessarily fulfilled!



- This is fixed by choosing

$$\tilde{c}_{k,l} = \max\{\|x_k^{cl} - y_l^{cl}\|^2, c_{k,l}^{\min}\}.$$

Open Questions

- Is the cost-dependent clustering problem NP-hard?
- How tight are the lower and upper bounds in practice?
- Is it worth it to compute $\pi^{\text{min}*}$ explicitly?
- How well does the general heuristic perform?

Open Questions

- Is the cost-dependent clustering problem NP-hard?
- How tight are the lower and upper bounds in practice?
- Is it worth it to compute $\pi^{\text{min}*}$ explicitly?
- How well does the general heuristic perform?

Thank you for your attention!