Transport aléatoire et optimal de mesures pour l'allocation de ressources et partition d'une ville en districts

Jérémie Bigot

Institut de Mathématiques de Bordeaux

Equipe Image, Optimisation et Probabilités (IOP)

Université de Bordeaux & Institut Universitaire de France

Joint work with **Bernard Bercu** (IMB, Bordeaux)

2ème Journée Aquitaine IA, RO et Data Science

Kedge Business School, Bordeaux, Février 2019

An example of a ressource allocation problem

Data at hand [1] :

- locations of Police stations in Chicago
- spatial locations of reported incidents of crime (with the exception of murders) in Chicago in 2014

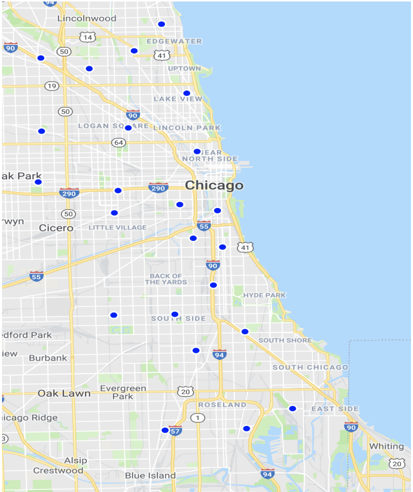
Questions (of interest ?) :

- given the location of a crime, which Police station should intervene ?
- how updating the answer in an “online fashion” along the year ?

1. Open Data from Chicago : `https://data.cityofchicago.org`

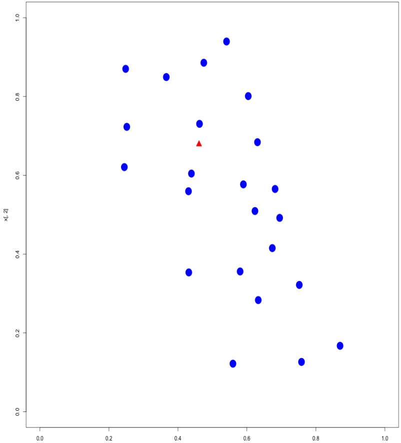
An example of a ressource allocation problem

Locations $y_1, \ldots, y_J$ of Police stations in Chicago



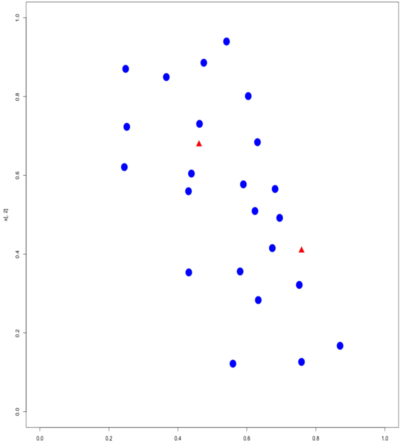
An example of a ressource allocation problem

Spatial location X_1 of the **first** reported incident of crime in Chicago in the year 2014



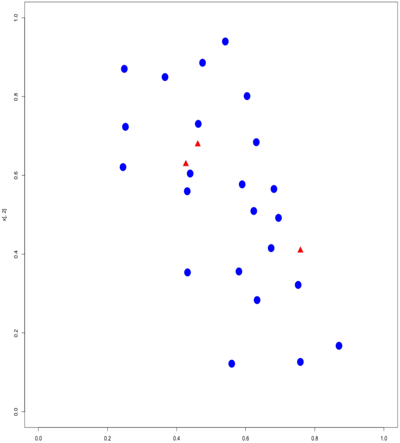
An example of a ressource allocation problem

Spatial locations X_1, X_2 of reported incidents of crime in Chicago in **chronological order**



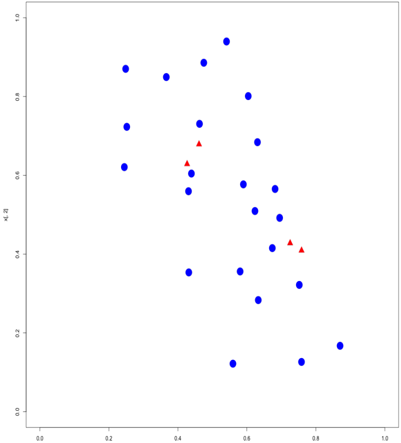
An example of a ressource allocation problem

Spatial locations X_1, X_2, X_3 of reported incidents of crime in Chicago in **chronological order**



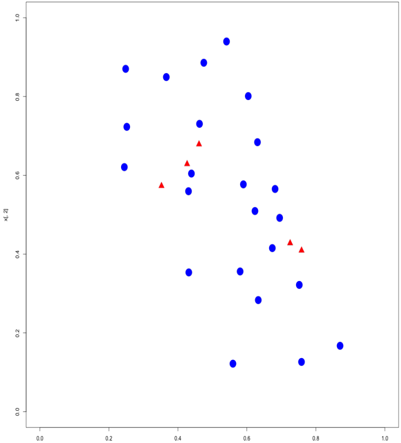
An example of a ressource allocation problem

Spatial locations $X_1, \ldots, X_4$ of reported incidents of crime in Chicago in **chronological order**



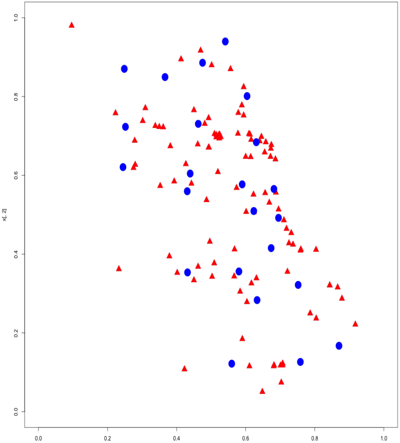
An example of a ressource allocation problem

Spatial locations $X_1, \ldots, X_5$ of reported incidents of crime in Chicago in **chronological order**



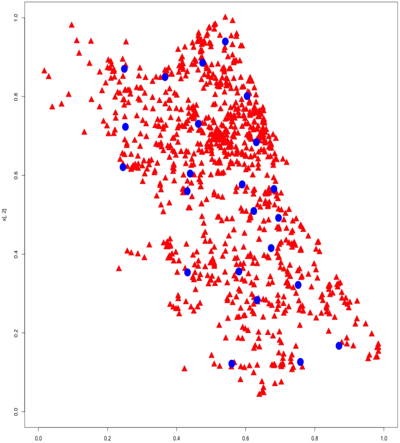
An example of a ressource allocation problem

Spatial locations of reported incidents of crime in Chicago in **chronological order** (first 100)



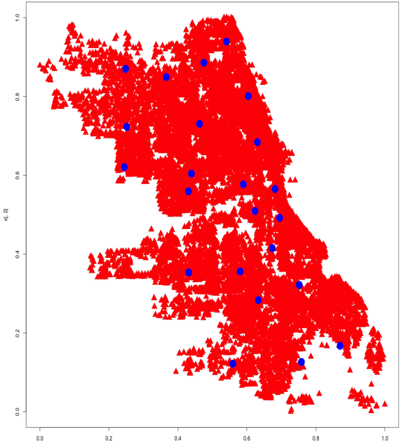
An example of a ressource allocation problem

Spatial locations of reported incidents of crime in Chicago in **chronological order** (first 1000)



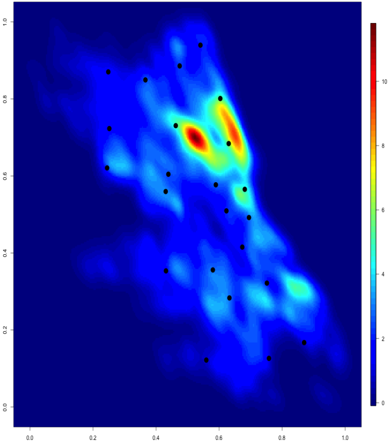
An example of a ressource allocation problem

Spatial locations $X_1, \ldots, X_N$ of reported incidents of crime in Chicago in **chronological order** (total $N = 16104$)



An example of a ressource allocation problem

Heat map (kernel density estimation) of spatial locations of reported incidents of crime in Chicago in 2014



1. Motivations from of a ressource allocation problem
2. Wassertein optimal transport
3. Regularized optimal transport and stochastic optimisation

Statistical approach to ressource allocation

Modeling assumptions :

- spatial locations of reported incidents of crime : a sequence of iid random variables

$$X_1, \ldots, X_n$$

  sampled from an **unknown** probability measure μ with support $\mathcal{X} \subset \mathbb{R}^2$

- locations of Police station : a **known and discrete** probability measure

$$\nu = \sum_{j=1}^{J} \nu_j \delta_{y_j}$$

  where

  - $y_j \in \mathbb{R}^2$ represent the spatial location of the j-th Police station
  - ν_j is a positive weight representing the "capacity" of each Police station (we took $\nu_j = 1/J$ that is uniform weights)

Statistical approach to ressource allocation

Point of view in this talk : ressource allocation can be solved by finding an optimal transportation map

$$T : \mathcal{X} \to \{y_1, \dots, y_J\}$$

which pushes forward μ onto $\nu = \sum_{j=1}^{J} \nu_j \delta_{y_j}$ (notation : $T\#\mu = \nu$), with respect to a given distance

$$c(x,y) = \|x - y\|_{\ell_p} = \left(\sum_{k=1}^{d} (x_k - y_k)^p \right)^{1/p}, \quad x, y \in \mathbb{R}^d \text{ (here } d = 2\text{)}$$

Question : how doing on-line estimation of such a map using the observations $X_1, \dots, X_n \sim_{iid} \mu$?

Optimal transport between probability measures

- Let $T : \mathcal{X} \to \{y_1, \ldots, y_J\}$ such that $T\#\mu = \nu$
- Let $\Pi(\mu, \nu)$ be the set of probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and ν

Definition

The optimal transport problem between μ and ν is

$$W_0(\mu, \nu) = \min_{T \,:\, T\#\mu=\nu} \int_{\mathcal{X}} c(x, T(x)) d\mu(x), \text{ (Monge's formulation)}$$

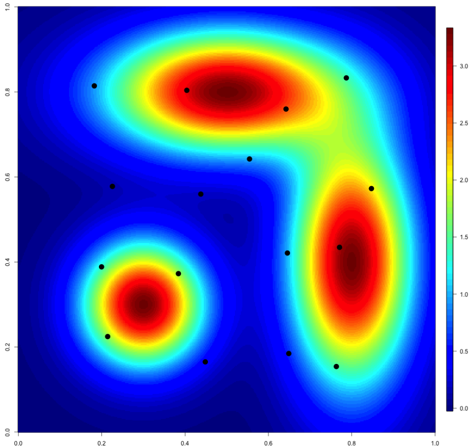
or

$$W_0(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x, y) d\pi(x, y), \text{ (Kantorovich's formulation)}$$

where $c(x, y)$ is the cost function of moving mass from x to y.

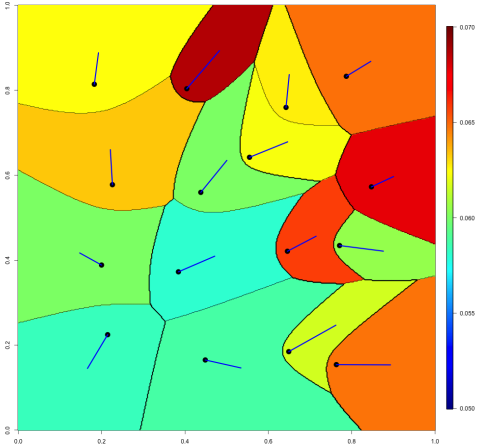
An example of semi-discrete optimal transport

Optimal transport of an absolutely continuous measure μ onto a discrete measure ν (black dots)



An example of semi-discrete optimal transport

Optimal transport of μ onto the discrete measure ν (black dots) - Optimal map T for the Euclidean cost $c(x,y) = \|x-y\|_{\ell_2}$



Optimal transport between probability measures

Problem : computational cost of optimal transport for data analysis [1]

Case of discrete measures : if

$$\mu = \sum_{i=1}^{K} \mu_i \delta_{x_i} \text{ and } \nu = \sum_{j=1}^{K} \nu_j \delta_{y_j}$$

then the cost to evaluate $W_0(\mu, \nu)$ (linear program) is generally

$$\mathcal{O}(K^3 \log K)$$

1. See the recent book by Cuturi & Peyré (2018)

Regularized optimal transport

Definition (Cuturi (2013))

Let μ and ν be any probability measures supported on $\mathcal{X}$. Then, the regularized optimal transport problem between μ and ν is

$$W_\varepsilon(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x,y)d\pi(x,y) + \varepsilon KL(\pi|\mu\otimes\nu),$$

where $\epsilon > 0$ (regularization parameter) and

$$KL(\pi|\xi) = \int_{\mathcal{X}\times\mathcal{X}} \left(\log\left(\frac{d\pi}{d\xi}(x,y)\right) - 1\right) d\pi(x,y), \text{ with } \xi = \mu\otimes\nu.$$

Case of discrete measures : for $\epsilon > 0$

- Sinkhorn algorithm (iterative scheme) to compute $W_\varepsilon(\mu,\nu)$
- computational cost of $\mathcal{O}(K^2)$ at each iteration

Stochastic optimal transport

Proposition (Genevay, Cuturi, Peyré and Bach (2016))

Let μ be any probability measure and $\nu = \sum_{j=1}^{J} \nu_j \delta_{y_j}$. For $\varepsilon \geq 0$, solve the **smooth concave maximization** *problem*

$$W_\varepsilon(\mu, \nu) = \max_{v \in \mathbb{R}^J} \underbrace{\mathbb{E}[h_\varepsilon(X, v)]}_{\textit{Stochastic optimization}}$$

where X is a random variable with distribution μ, and for $x \in \mathcal{X}$ and $v \in \mathbb{R}^J$,

$$h_\varepsilon(x, v) = \sum_{j=1}^{J} v_j \nu_j - \varepsilon \log\left(\sum_{j=1}^{J} \exp\left(\frac{v_j - c(x, y_j)}{\varepsilon} \right) \nu_j \right) - \varepsilon.$$

Stochastic algorithm [1]

For fixed $\epsilon > 0$, Robbins-Monro algorithm to compute a minimizer

$$v^* \in \arg\min_{v \in \mathbb{R}^J} \mathbb{E}[h_\varepsilon(X, v)]$$

Let $X_1, \ldots, X_n \sim_{iid} \mu$, choose $V_0 \in \mathbb{R}^J$ and a sequence γ_{n+1} of steps with $\sum_{n=1}^{\infty} \gamma_n = +\infty$ and $\sum_{n=1}^{\infty} \gamma_n^2 < +\infty$ and do

$$\widehat{V}_{n+1} = \widehat{V}_n + \gamma_{n+1} \nabla_v h_\varepsilon(X_{n+1}, \widehat{V}_n)$$

Easy computation of gradients for $\epsilon > 0$ (smooth optimization)

$$\nabla_v h_\varepsilon(x, v) = \nu - \pi(x, v)$$

where $\pi(x, v) \in \mathbb{R}^J$ with

$$\pi_j(x, v) = \left(\sum_{k=1}^{J} \nu_k \exp\left(\frac{v_k - c(x, y_k)}{\varepsilon} \right) \right)^{-1} \nu_j \exp\left(\frac{v_j - c(x, y_j)}{\varepsilon} \right)$$

1. Genevay, Cuturi, Peyré and Bach (2016)

Contribution in our work [1]

Main goal : estimation of the Wasserstein functional $W_\varepsilon(\mu,\nu)$ based on $X_1,\ldots,X_n \sim_{iid} \mu$ and assuming that ν is known

A simple recursive estimator :

$$\widehat{W}_n = \frac{1}{n}\sum_{k=1}^{n} h_\varepsilon(X_k, \widehat{V}_{k-1}).$$

Main results : asymptotic normality with a data-driven choice of the learning rate

$$\sqrt{n}\Big(\widehat{W}_n - W_\varepsilon(\mu,\nu)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma_\varepsilon^2(\mu,\nu)\big)$$

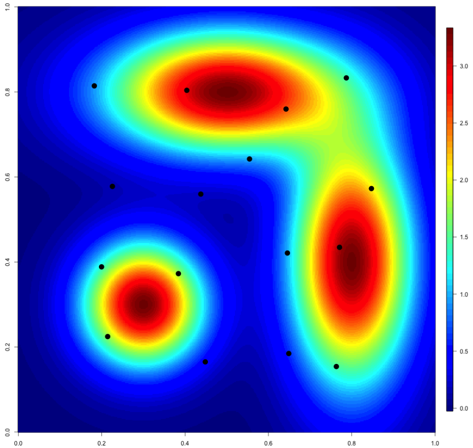
where the asymptotic variance $\sigma_\varepsilon^2(\mu,\nu)$ can also be estimated in a recursive manner

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n} h_\varepsilon^2(X_k, \widehat{V}_{k-1}) - \widehat{W}_n^2.$$

1. Bercu, B. & Bigot, J. (2018) ArXiv :1812.09150

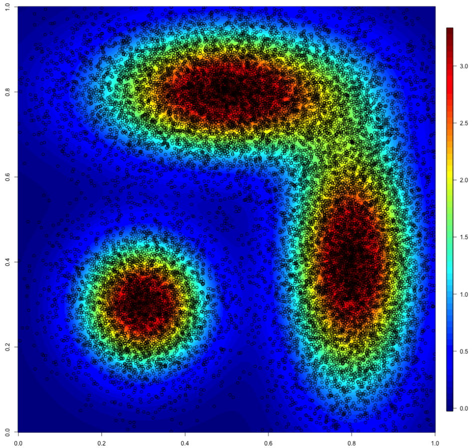
Numerical experiments - Simulated data

Optimal transport of an absolutely continuous measure μ onto a discrete measure ν (black dots)



Numerical experiments - Simulated data

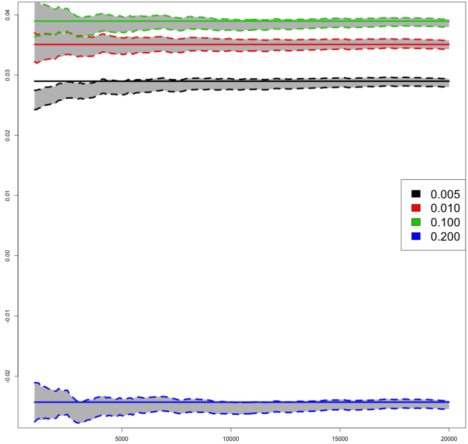
Samples $X_1, \ldots, X_N \sim_{iid} \mu$ (with $N = 20000$) and discrete measure ν (black dots)



Numerical experiments - Simulated data

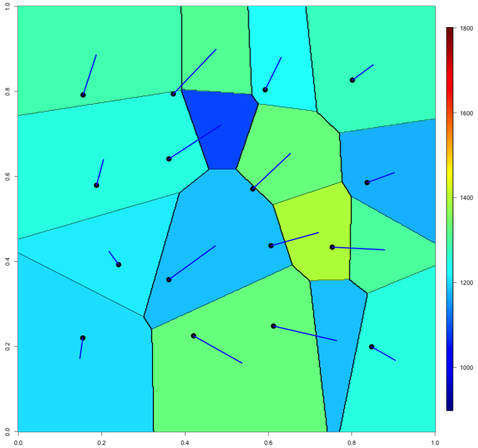
Convergence of the algorithm using the quadratic cost

$$c(x,y) = \|x-y\|_{\ell_2}^2$$



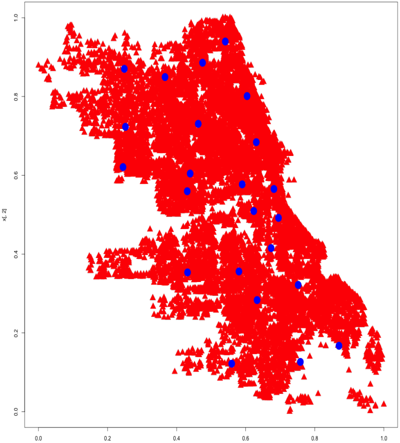
Numerical experiments - Simulated data

Optimal map T for the quadratic cost $c(x,y) = \|x-y\|_{\ell_2}^2$ after $N = 20000$ iterations and $\varepsilon = 0.005$



Numerical experiments - Real data

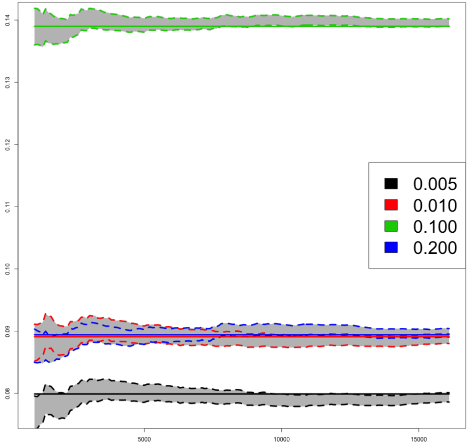
Spatial locations $X_1, \ldots, X_N$ of reported incidents of crime in Chicago in **chronological order** (total $N = 16104$)



Numerical experiments - Real data

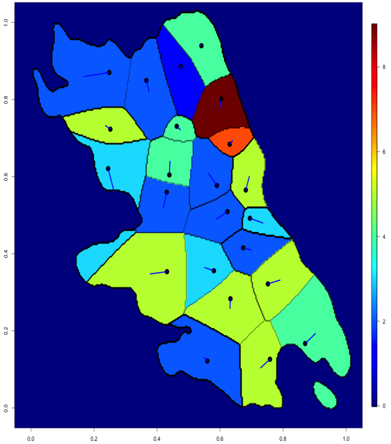
Convergence of the algorithm using the Euclidean cost

$$c(x,y) = \|x-y\|_{\ell_2}$$



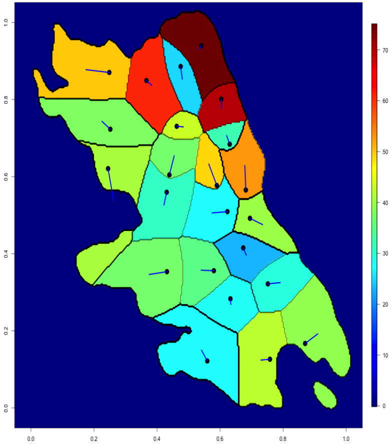
Numerical experiments - Real data

Optimal map T for the Euclidean cost $c(x, y) = \|x - y\|_{\ell_2}$ after $n = 100$ iterations and $\varepsilon = 0.005$



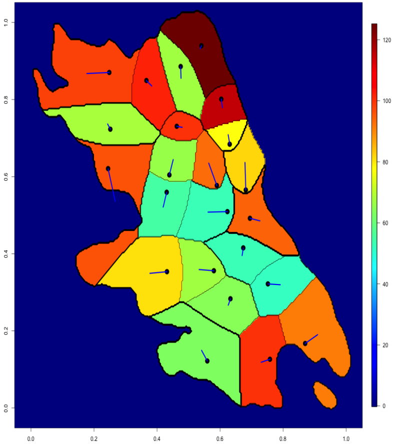
Numerical experiments - Real data

Optimal map T for the Euclidean cost $c(x,y) = \|x-y\|_{\ell_2}$ after $n = 1000$ iterations and $\varepsilon = 0.005$



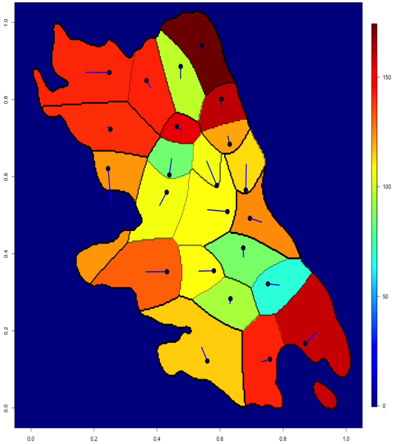
Numerical experiments - Real data

Optimal map T for the Euclidean cost $c(x,y) = \|x-y\|_{\ell_2}$ after $n = 2000$ iterations and $\varepsilon = 0.005$



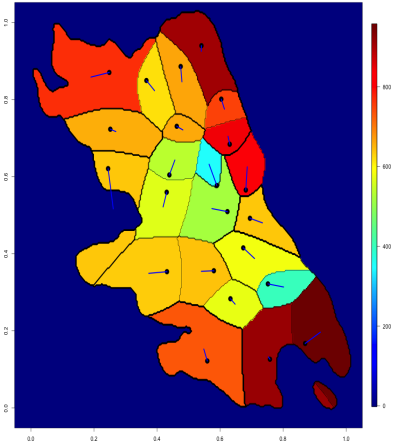
Numerical experiments - Real data

Optimal map T for the Euclidean cost $c(x,y) = \|x-y\|_{\ell_2}$ after $n = 3000$ iterations and $\varepsilon = 0.005$



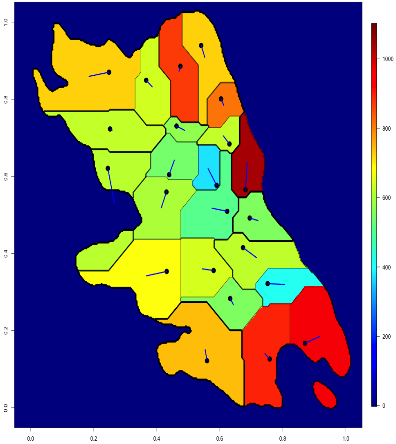
Numerical experiments - Real data

Optimal map T for the Euclidean cost $c(x,y) = \|x-y\|_{\ell_2}$ after $N = 16104$ iterations and $\varepsilon = 0.005$



Numerical experiments - Real data

Optimal map T for the ℓ_1 cost $c(x,y) = \|x-y\|_{\ell_1}$ after $N = 16104$ iterations and $\varepsilon = 0.005$



Numerical experiments - Real data

Optimal map T : Euclidean versus ℓ_1 cost