Optimal transport and redistricting: numerical experiments and a few questions

Nestor Guillen
University of Massachusetts

From Rebecca Solnit's **Hope in the dark**

(apropos of nothing in particular)

To hope is to give yourself to the future - and that commitment to the future is what makes the present inhabitable.

Traditional Districting Principles

1. **Equal Population**
2. Compactness
3. Contiguity
4. Respect for county/city boundaries
5. Respect for communities of interest
6. Voting Rights Act compliance

Traditional Districting Principles

1. **Equal Population**

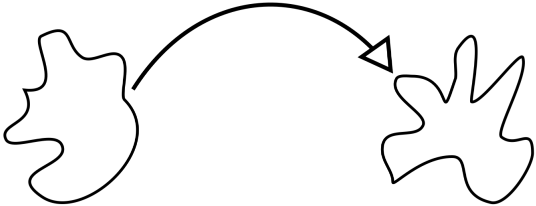
7. ???

…we shall retain just #1, and see how that interacts with **other** desirable criteria.

Optimal Transport

"Optimal transport" is a field of mathematics & an array of tools increasingly used throughout engineering and science.

It's concerned with maps that preserve volume and/or ways of "matching" two distributions together



and this "transport"/"matching" is optimal *in some sense*.

Optimal Transport

Goes back to the 19th century. Two names that will come up today:

1. Gaspard Monge (19th century, French physicist and mathematician)
2. Leonidas Kantorovich (20th century, only Soviet Nobel prize in economics!)

Optimal transport is partly a subfield of **linear programming and optimization**, partly a subfielf of Partial Differential Equations.

Optimal Transport

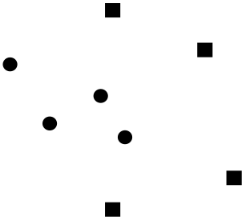
The field has blossomed since the early 90's and has become a fairly developed field whose methods are used in:

image processing, pattern recognition, fluid mechanics, weather modeling, geometric optics, probability, statistical inference, differential geometry, economics...

The Monge Problem

Setup

Two sets of points on the plane, X (squares) and Y (circles).



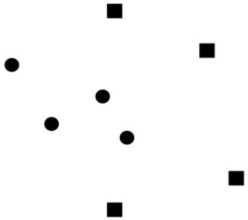
We are given the **transportation cost from x_i to y_j**, denoted

$$c(x_i, y_j)$$

The Monge Problem

Setup

Problem



Find a bijective map $T : X \to Y$ minimizing

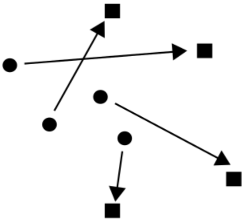
$$\sum_{i=1}^{N} c(x, T(x))$$

$(\textit{Total transportation cost of } T)$

The Monge Problem

Setup

Problem



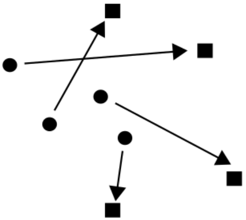
Find a bijective map $T : X \to Y$ minimizing

$$\sum_{i=1}^{N} c(x, T(x))$$

$(\textit{Total transportation cost of } T)$

The Monge Problem

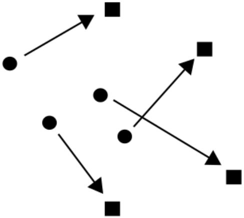
As X and Y are finite, there are finitely many choices, so at least one of them achieves the smallest value



However, this is a hard optimization problem when X and Y are sufficiently large.

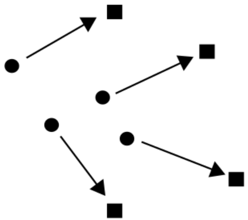
The Monge Problem

As X and Y are finite, there are finitely many choices, so at least one of them achieves the smallest value



The Monge Problem

As X and Y are finite, there are finitely many choices, so at least one of them achieves the smallest value



The Kantorovich Problem

Setup

A variation

We are now given "mass densities" f and g, representing

$$f_i = \text{ population at } x_i$$
$$g_j = \text{ capacity at } y_j$$

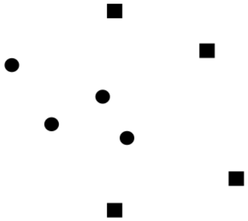
Their total masses are equal

$$\sum_{i=1}^{N} f_i = \sum_{j=1}^{M} g_j$$

The Kantorovich Problem

Setup

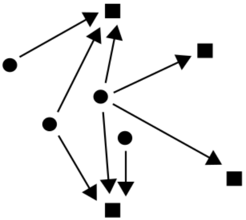
Instead of looking for $T : X \to Y$ we broaden our options by allowing ourselves **to split the mass** leaving each x_i



The Kantorovich Problem

Setup

Instead of looking for $T : X \to Y$ we broaden our options by allowing ourselves **to split the mass** leaving each x_i



The Kantorovich Problem

Setup

This splitting we encode by a matrix π, interpreted as

$$\pi_{ij} = \text{ population at } x_i \text{ assigned to } y_j$$

$$f_i = \sum_{j=1}^{M} \pi_{ij}, \quad g_j = \sum_{i=1}^{N} \pi_{ij}$$

Such a π is called a **transportation plan** between f and g.

The Kantorovich Problem

Setup

Then: among all plans π, find one minimizing the otal cost

$$\sum_{i=1}^{N} \sum_{j=1}^{M} c(x_i, y_j)\pi_{ij}$$

This is known as the Kantorovich problem.

This **is a linear optimization problem in** π: the set of admissible plans forms a convex set in $\mathbb{R}^{NM}$, and the objective functional is a linear function of π.

The Kantorovich Problem

In summary:
Find π which achieves the minimum value of

$$\sum_{i=1}^{N}\sum_{j=1}^{M} c_{ij}\pi_{ij}$$

while satisfying the constraints

$$\begin{aligned} f_i &= \sum_{j=1}^{M} \pi_{ij} \quad \text{for } i = 1, \ldots, N, \\ g_j &= \sum_{i=1}^{N} \pi_{ij} \quad \text{for } j = 1, \ldots, M, \\ \pi_{ij} &\geq 0, \qquad\quad \text{for all } i, j. \end{aligned}$$

The Monge-Kantorovich Problem

Monge versus Kantorovich

Monge:

- Highly nonlinear variational/optimization problem. How do we find a minimizer?
- Solutions are “nice”: you do not split a house in two.

Kantorovich:

- Amounts to a linear program! (easy to implement)
- But solutions may be “multi-valued” (mass-splitting!).
- Under extra assumptions, Kantorovich = Monge.

The Wasserstein distance

For probability distributions μ and ν, their Wasserstein dist. is

$$W_2(\mu,\nu) := \left(\min_{\pi} \int_{X\times Y} |x-y|^2 \pi(x,y)dxdy\right)^{\frac{1}{2}}$$

This is a widely studied metric space (it behaves a lot like an infinite dimensional manifold)

A special subset

The set $\mathcal{E}_M$ of empirical distributions made out of M points

$$\mathcal{E}_M = \{\nu \mid \nu = \frac{1}{M}\sum_{j=1}^{M} \delta_{y_j} \mid y_1,\ldots,y_M \in Y\}$$

OT: continuous source and discrete target

We look for a map in some region R

$$T : R \subset \mathbb{R}^2 \to \{y_1, \ldots, y_M\} = Y.$$

Minimizing

$$\int_{\mathbb{R}^2} c(x, T(x))\mu(x)\, dx$$

So the population distribution μ is a mass density function

OT: continuous source and discrete target

Theorem (cf. Brennier circa 1990)

There are numbers $\alpha_1, \ldots, \alpha_M$ *such that the function*

$$u(x) = \min_{1 \leq j \leq M} \{c(x, y_j) + \alpha_j\}$$

gives the optimal transport map, via

$$T(x) = y_k \Leftrightarrow u(x) = c(x, y_k) + \alpha_k$$

The sets $\{D_j\}$ *given by* $D_j = T^{-1}(y_j)$ *provide a partition of the population distribution.*

Geomtrically, what does this theorem say about the nature of the minimizers?

Discrete OT

So, we have seen that the pre-images $T^{-1}(y_k)$ are given by the **intersection** of sets of the form

$$\{x \mid -c(x, y_j) + c(x, y_k) \leq \alpha_j - \alpha_k\}$$

This is just as with convex polygons being the intersection of a number of half-spaces!

What happens when both source and target are discrete?

Discrete OT

Example: Cartesian Grid and Square Cost

"Everything is true and trivial in the Cartesian grid!"
– @iamcardib
(or @justinmsolomon, I don't remember rn)

Discrete OT

Example: Cartesian Grid and Square Cost

"Everything is true and trivial in the Cartesian grid!"
– @iamcardib
(or @justinmsolomon, I don't remember rn)

Let us take $c(x_i, y_j) = |x_i - y_j|^2$, where the population and the $y's$ are supported in $\mathbb{Z}^2$.
(Admittedly, here $\mathbb{Z}^2$ hardly plays any role)
We have the sets

$$\{x \in \mathbb{Z}^2 \mid -c(x, y_j) + c(x, y_k) \leq \alpha_j - \alpha_k\}$$

But expanding the squares, we get that

$$\{x \mid x \cdot (y_j - y_k) \leq \lambda\}$$

for some λ depending on α_j, α_k and $|y_j|, |y_k|$.

Discrete OT

Geometry of districts

More generally, observe that

$$\begin{aligned} u(x) &\leq c(x, y_k) + \alpha_k \; \forall \; k. \\ &= c(x, y_k) + \alpha_k \;\; \text{if } x \in T^{-1}(y_k). \end{aligned}$$

Therefore

$$T^{-1}(y_k) := \{u(x) \geq c(x, y_k) + \alpha_k\}.$$

Such a set is called a *c*-**section** (no, really!). They are the analogue of convex sets in the geometry induced by the cost c.

Optimal transport and partitioning

Problem:
Fix μ and K among all $\nu \in \mathcal{E}_K$, minimize the Wasserstein dist.

$$W_2(\mu, \nu)$$

This is a special instance of what is now known as "the Wasserstein barycenter problem".

In the math literature, this was first studied by Agueh and Carlier (2013). See Solomon's 2018 survey on numerical OT for an overall discussion of its uses in statistics, computer vision, ML, and more.

Optimal transport and partitioning

One can use OT (and linear programming in general) to provide an optimality criterium for partitions, and thus also for computer-assisted redistricting.

There is, for instance, this paper from 1965!

NONPARTISAN POLITICAL REDISTRICTING BY COMPUTER*

S. W. Hess and J. B. Weaver

Atlas Chemical Industries, Inc., Wilmington, Dela.

H. J. Siegfeldt, J. N. Whelan, and P. A. Zitlau

E. I. Du Pont de Nemours & Co., Wilmington, Dela.

(Received January 28, 1965)

Optimal transport and partitioning

Data:

- Cost function $c(x_i, y_j)$ between pairs of vertices.
- Population distribution f is given.
  (Total population is P)
- Number of districts K .

A **partition plan** π for f is one satisfying the following constraints:

First, the usual mass balance with the population

$$\sum_{j=1}^{M} \pi_{ij} = f_i \quad \text{for } i = 1, \ldots, N,$$

$$\pi_{ij} \geq 0, \qquad \text{for all } i, j.$$

Second, there are indices $j_1, \ldots, j_K$ such that

$$\sum_{i=1}^{N} \pi_{ij} = P/K \quad \text{if } j \text{ is not among } \{j_1, \ldots, j_K\},$$
$$\sum_{i=1}^{N} \pi_{ij} = 0 \qquad\quad \text{if } j \text{ is not among } \{j_1, \ldots, j_K\}.$$

Now the problem is to find $\{y_{j_1}, \ldots, y_{j_K}\}$ leading to the smallest optimal transport cost from μ

Optimal transport and partitioning

Partition Plans

Note:
One may think of the points $y_{j_1}, \ldots, y_{j_K}$ as virtual voting centers, or one may make them correspond to real ones, as may be convenient.

The point is that as they are not given a priori, the locations $y_{j_1}, \ldots, y_{j_K}$ are part of the **unknowns**.

(this is just like the Wasserstein barycente problem, except the cost is not quadratic)

Optimal transport and partitioning

Contrast this with: Voronoi diagrams

Consider a region Ω and points $\{x_1, \ldots, x_N\}$.

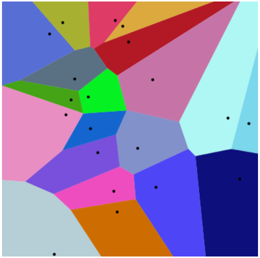


Image credit: Wikipedia

These points yield a "partition" of Ω into N regions, known as its Voronoi diagram, its cells being defined by

$$D_k = \{x \in \Omega \mid |x - x_k| \leq |x - x_j| \; j = 1, \ldots, N\}$$

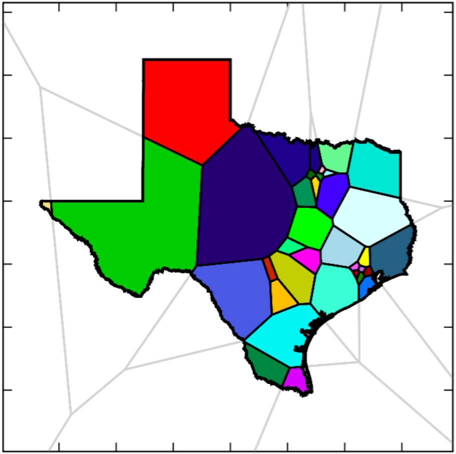
Optimal transport and partitioning

A related approach: Balanced Power Diagrams, as in Cohen-Addad, Klein, and Young (2018).

They turn census data into weighted point clouds in the plane, and compute a **balanced centroidal power diagram**.

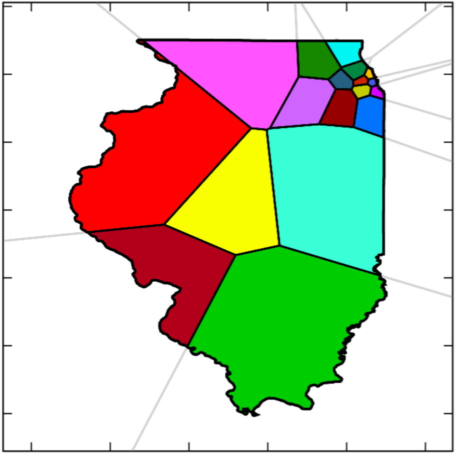
The setting of the problem in $\mathbb{R}^2$ leads to partitions given by convex polygons.

Optimal transport and partitioning



From Cohen-Addad, Klein, and Young (2018)

Optimal transport and partitioning



From Cohen-Addad, Klein, and Young (2018)

Optimal transport and partitioning

Strong geometry dependence on the cost function

Important: The resulting polygonal shapes are a consequence of our selection of a cost function c_{ij},

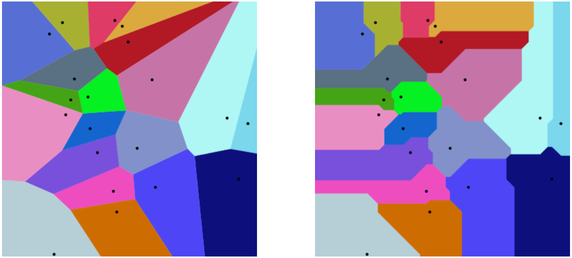
$$c(x, y) = |x - y|^2$$

Different cost functions, specially ones that incorporate specific policy principles, could yield law-complaint maps, while providing mathematical and rhetorical justification for them.

Optimal transport and partitioning

Strong geometry dependence on the cost function

The variation in map shape is best illustrated by Voronoi diagrams made using the Euclidean distance versus the "Manhattan" (cartesian grid) distance



Some numerical experiments

Let us do the following: fix a state (PA), which has 18 seats.

• Take say, the graph of **Census Tracts**. The pairwise **matrix of distances** between the centroids, and μ the state's **population distribution**.

• Sample graph locations (uniformly*): $\{y_1, \ldots, y_{18}\}$

• Compute the optimal transport map from μ to the uniform distribution given by the y_j.

Some numerical experiments

Ok, so what did we just see?

Although it did not happen with all plans, plenty of them had a couple of disconnected districts and **districts with holes**. Specially in the plans with 18 districts.

In the semi-discrete setting, there is large class of costs for which one can **prove** the above does **NOT** happen. The condition involves a **tensor** invariant proposed by Ma, Trudinger, and Wang in 2005.

The Ma-Trudinger-Wang condition

• A condition on the cost function for OT problems in the continuum setting.
• Boils down to a series of inequalities involving 4th order derivatives of $c(x, y)$.
• It is a a new type of curvature condition, with some relation to the Ricci curvature when $c(x, y) = d(x, y)^2$ (the squared geodesic distance).
• It is a necessary condition for differentiability of a related PDE –the Monge-Ampère equation, and in turn, for the continuity/differentiability of optimal transport maps.

The Ma-Trudinger-Wang condition

Suppose that c is C^4 in the sense that mixed derivatives of order 2 in both variables simultaneously are continuous, and fix local coordinate systems on M, $\bar{M}$ near $(x, \bar{x})$. Then for $V, W \in T_x M$ and $\eta, \zeta \in T_x^* M$, define

$$\mathrm{MTW}_{(x,\bar{x})}(V, W, \eta, \zeta) := -(c_{ij,\bar{r}\bar{s}} - c_{ij,\bar{t}} c^{\bar{t},s} c_{s,\bar{r}\bar{s}}) c^{\bar{r},k} c^{\bar{s},l}(x, \bar{x}) V^i W^j \eta_k \zeta_l.$$

Then, the MTW condition is: if $V \perp \eta$ then

$$\mathrm{MTW}_{(x,\bar{x})}(V, V, \eta, \eta) \geq 0.$$

So this expression is difficult to parse but it is related to a geometric property of “c-hyperplanes” (a family of surfaces arising from the cost)

The Ma-Trudinger-Wang condition

Question:

Is there a discrete analogue of the MTW condition for a cost function c_{ij} in a graph (V, E)? Can one show under such a condition that the resulting districting plans are

... path connected?

... "convex"?

In other words: when are c-sections for a given cost path-connected? how compact are these sets?

The Ma-Trudinger-Wang condition?

A good first step towards answering this question is the following.

Theorem (…maybe?, hopefully?)

For a graph (V, E), if $c(x_i, y_j)$ is just the graph distance, then optimizers should be path connected.

In progress

1. Update the “voting centers” to further reduce the cost (e.g. Lloyd’s algorithm).
2. Understand the lack of connectivity and lack of simple connectivity in some districts. Is it a side effect of the algorithm? Or does the underlying geometry cause this?.
3. Use this method to construct optimal plans according to the Transit Time Compactness metric.
4. Systematically produce “optimal maps” which are gerrymandered (consider a population mixture, impose relative size constrains district by district).
5. One potential use: computer-aided design of majority-minority districts.

More questions / things to do

1. How good are the compactness scores for the resulting districts? (i.e. compute lots plans for lots of states and tally various compactness scores).
2. Are there inequalities relating optimizers for the transport cost versus optimizers for the perimeter? (i.e. prove some theorems).
3. It would be worth combining this with a Markov chain for the "voting centers" which then determines the optimal transport plans. This generates a chain in the space of partitions. Does this chain have any interesting properties?.

🚚 🚚 🚚

Thank You!

Questions?

nguillen@math.umass.edu