

INTRINSIC SPARSITY OF KANTOROVICH SOLUTIONS

BAMDAD HOSSEINI AND STEFAN STEINERBERGER

ABSTRACT. Let X, Y be two finite sets of points having $\#X = m$ and $\#Y = n$ points with $\mu = (1/m) \sum_{i=1}^m \delta_{x_i}$ and $\nu = (1/n) \sum_{j=1}^n \delta_{y_j}$ being the associated uniform probability measures. A result of Birkhoff implies that if $m = n$, then the Kantorovich problem has a solution which also solves the Monge problem: optimal transport can be realized with a bijection $\pi : X \rightarrow Y$. This is impossible when $m \neq n$. We observe that when $m \neq n$, there exists a solution of the Kantorovich problem such that the mass of each point in X is moved to at most $n/\gcd(m, n)$ different points in Y and that, conversely, each point in Y receives mass from at most $m/\gcd(m, n)$ points in X .

Let μ and ν be two (probability) measures. A classical question, due to Monge, is to understand the optimal way of mapping μ to ν . If we denote the cost of transporting mass from x to y by $c(x, y)$, then the Monge problem asks for

$$\inf_T \left\{ \int_X c(x, T(x)) d\mu(x) : T_*(\mu) = \nu \right\} \quad (\text{Monge}),$$

where $T_*(\mu)$ denotes the push forward of μ by T . This problem may not be solvable because such transport maps T may simply not exist. Kantorovich proposed to relax the problem and instead try to minimize

$$\inf_{\gamma} \int_{X \times Y} c(x, y) d\gamma(x, y) \quad (\text{Kantorovich}),$$

where γ is a probability measure on $X \times Y$ having marginals μ and ν . There is a nice classical result linking these two problems in the discrete setting: if μ and ν are two uniform probability measures over two sets X and Y with n elements, then it is known that these two problems coincide.

Theorem (see e.g. [2, 4, 5]). *If $\mu = (1/n) \sum_{i=1}^n \delta_{x_i}$ and $\nu = (1/n) \sum_{i=1}^n \delta_{y_i}$, then there is a solution of the Kantorovich problem which also solves the Monge problem.*

The statement is independent of the transport costs $c(x_i, y_j)$. The argument is as follows: the Kantorovich problem can, in the discrete setting, be formulated as a linear program over bistochastic matrices. A theorem of Birkhoff [1] (also attributed to König [3] and von Neumann [6]) states that the bistochastic matrices are the convex hull of the permutation matrices. The minimum of a linear program in a non-empty polyhedron is attained in an extremal point.

2010 *Mathematics Subject Classification.* 49Q20, 90C46.

Key words and phrases. Kantorovich problem, Monge problem, discrete measures.

The authors gratefully acknowledge support from the Kantorovich Initiative. S.S. was partially supported by the NSF (DMS-2123224) and the Alfred P. Sloan Foundation.

No such statement can be true when $m \neq n$: the two sets have different cardinalities and no bijection is possible. The goal of this short note is to point out that there nonetheless exists a particularly simple solution of the Kantorovich problem.

Theorem. *Let $\mu = (1/m) \sum_{i=1}^m \delta_{x_i}$ and $\nu = (1/n) \sum_{i=1}^n \delta_{y_i}$. There is a solution of the Kantorovich problem such that mass from each point in X is moved to at most $n/\gcd(m, n)$ different points in Y and that each point in Y receives mass from at most $m/\gcd(m, n)$ points in X .*

Somewhat to our surprise, we were unable to find this simple but intriguing statement (illustrated in Fig. 1) in the literature. Besides its intrinsic appeal, it does seem like it could be potentially useful insofar as it guarantees the existence of ‘sparse’ solutions of the Kantorovich problem (with sparsity depending on m, n).

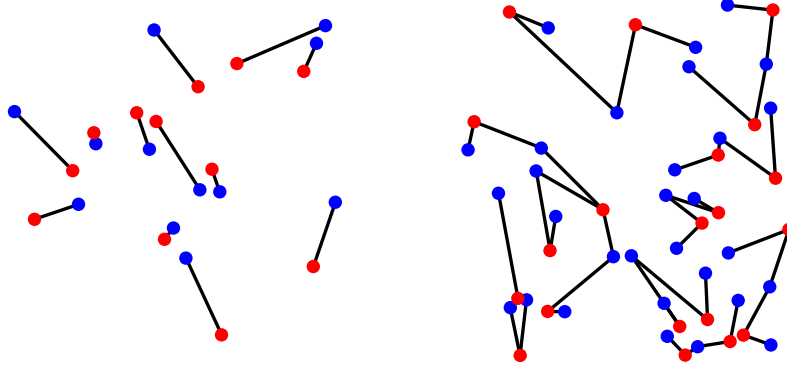


FIGURE 1. Left: $m = n$, the transport is a bijection. Right: $m = 20$ red points are sent to $n = 30$ blue points. Each red point is transported to at most $30/\gcd(20, 30) = 3$ blue points, each blue point receives mass from at most $20/\gcd(20, 30) = 2$ red points.

Proof. Suppose that

$$\mu = \frac{1}{m} \sum_{i=1}^m \delta_{x_i} \quad \text{and} \quad \nu = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

are two given measures. We replace each point x_i by $n/\gcd(m, n)$ identical points $x_{i,j}$ for $1 \leq j \leq n/\gcd(m, n)$ and, likewise, we replace each point y_i by $m/\gcd(m, n)$ identical points $y_{i,j}$ where $1 \leq j \leq m/\gcd(m, n)$. This allows us to write

$$\mu = \frac{\gcd(m, n)}{mn} \sum_{i=1}^m \sum_{j=1}^{n/\gcd(m, n)} \delta_{x_{i,j}} \quad \text{and} \quad \nu = \frac{\gcd(m, n)}{mn} \sum_{i=1}^n \sum_{j=1}^{m/\gcd(m, n)} \delta_{y_{i,j}}.$$

The problem can now be interpreted as finding a transport map from $mn/\gcd(m, n)$ points of the same weight to another set of $mn/\gcd(m, n)$ points of the same weight. Applying the classical result shows that there exists bijective map between the points that realizes the optimal Kantorovich cost. This corresponds into each point in X being split into at most $n/\gcd(m, n)$ equal parts and each point in Y being split into at most $m/\gcd(m, n)$ parts. \square

We conclude by observing that the same argument also applies to linear combinations of weighted Dirac measures as long as the weights are rational. Suppose

$$\mu = \sum_{i=1}^m \frac{a_i}{b_i} \delta_{x_i} \quad \text{where} \quad \sum_{i=1}^m \frac{a_i}{b_i} = 1$$

and where $a_i, b_i \in \mathbb{N}$ are positive rational weights. We note that it is possible to equivalently represent μ as a linear combination of a number of equally weighted Dirac measures. This number will depend on the least common multiple $\text{lcm}(b_1, \dots, b_m)$ of the denominators. This can be seen by writing

$$\mu = \frac{1}{\text{lcm}(b_1, \dots, b_m)} \sum_{i=1}^m \frac{a_i \text{lcm}(b_1, \dots, b_m)}{b_i} \delta_{x_i}$$

and noting that $a_i \text{lcm}(b_1, \dots, b_m)/b_i \in \mathbb{N}$. This implies the following corollary.

Corollary. *Let*

$$\mu = \sum_{i=1}^m \frac{a_i}{b_i} \delta_{x_i} \quad \text{and} \quad \nu = \sum_{i=1}^n \frac{c_i}{d_i} \delta_{y_i}$$

be two probability measures with positive rational weights and let

$$B = \text{lcm}(b_1, \dots, b_m) \quad \text{and} \quad D = \text{lcm}(d_1, \dots, d_n).$$

There exists a solution of the Kantorovich problem such that mass from each point in X is moved to at most $D/\text{gcd}(B, D)$ different points in Y and each point in Y receives mass from at most $B/\text{gcd}(B, D)$ different points in X .

We note that this reduces to the previous Theorem when $a_i = c_i = 1$, $b_i = m$ and $d_i = n$. We also observe that it is quite possible that $D/\text{gcd}(B, D) \gg n$ and that $B/\text{gcd}(B, D) \gg m$ (this will usually happen when the b_i, d_i have many different prime factors). In such a case, the statement would not say anything of interest.

REFERENCES

- [1] G. Birkhoff. Tres observaciones sobre el algebra lineal. Universidad Nacional de Tucumán Revista Series A, 5:147–151, 1946.
- [2] H. Brezis, Remarks on the Monge-Kantorovich problem in the discrete setting. C. R. Math. Acad. Sci. Paris 356 (2018), no. 2, 207–213.
- [3] D. König, Theorie der endlichen und unendlichen Graphen, Akademische. Verlags Gesellschaft, Leipzig, 1936.
- [4] Q. Merigot and B. Thibert. Optimal transport: discretization and algorithms. In Handbook of Numerical Analysis, vol. 22, pp. 133–212. Elsevier, 2021.
- [5] G. Peyré and M. Cuturi, Computational optimal transport: With applications to data science. Foundations and Trends in Machine Learning, 11(5-6), 355–607.
- [6] J. von Neumann, A certain zero-sum two-person game equivalent to an optimal assignment problem, Ann. Math. Studies 28:5–12, 1953.

DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
Email address: bamdadh@uw.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
Email address: steinerb@uw.edu