Mass transportation theory: Monge and Kantorovich formulations

Alberto González-Sanz*

In 1781, Gaspard Monge [6] proposed the following problem: Let $c(x, y)$ denote the cost of transporting a particle of mass from the location $x \in \mathbb{R}^d$ to $y \in \mathbb{R}^d$. If $n \in \mathbb{N}$ particles at locations $x_1, \ldots, x_n \in \mathbb{R}^d$ need to be transported to $y_1, \ldots, y_n \in \mathbb{R}^d$ in a bijective way, what is the optimal assignment of locations? In mathematical terms, Monge problem can be formulated as

$$\operatorname*{argmin}_T \sum_{i=1}^n c(x_i, T(x_i)) \quad \text{s.t. } T : \{x_1, \ldots, x_n\} \to \{y_1, \ldots, y_n\} \text{ is bijective.} \tag{M}$$

Solving (M) is not straightforward, as it involves a nonconvex optimization. However, if the locations are univariate measurements (i.e., elements of $\mathbb{R}$) and $c(x, y) = (x - y)^2$, the optimal solution can be computed by matching the order statistics—an essential tool for nonparametric statistical inference—of the locations. That is, we relabel $x_1, \ldots, x_n$ as $x_{(1)}, \ldots, x_{(n)}$ so that $x_{(1)} < \cdots < x_{(n)}$. We do the same for the y-locations. Then the optimal assignment is $x_{(i)} \mapsto y_{(i)}$ (see Fig. 1). But how is this done for multiple variables?

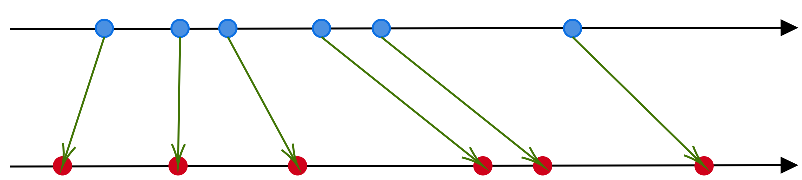

Figure 1: Monge problem for univariate data.

It took around 150 years until L.V. Kantorovich, later awarded the Nobel Prize in Economics, shed some light on Monge problem. In 1942, he formulated in [5] a generalization of (M) allowing for particle splitting—that is, a fraction of each particle can be mapped to different locations. In mathematical terms, we want to find

$$\operatorname*{argmin}_\gamma \sum_{i,j=1}^n c(x_i, y_j)\gamma_{i,j} \quad \text{s.t. } \sum_{k=1}^n \gamma_{i,k} = \sum_{k=1}^n \gamma_{k,j} = 1 \text{ for all } i, j \in \{1, \ldots, n\}. \tag{K}$$

This indicates that the $\gamma_{i,j}$-portion of the particle at location x_i is mapped to location y_j. The Kantorovich optimization problem has a larger feasibility set, since permutation matrices—those having all entries equal to 0 except for a single 1 in each row and column—define

*Department of Statistics, Columbia University, ag4855@columbia.edu

the feasibility set of the Monge problem. The main advantage of Kantorovich's relaxation is that (K) is a convex optimization problem. Indeed, it is a linear program over the polytope Ω_n of doubly stochastic matrices. Hence, the solutions of (K) form a face Γ of Ω_n (recall that the vertices of Γ are, in particular, vertices of Ω_n). A celebrated result by Birkhoff (see [1]) states that the vertices of Ω_n are precisely the permutation matrices. Therefore, the solution set of (K) can be characterized as the convex hull of the solutions of (M).

To conclude, Monge and Kantorovich problems are generally formulated in the continuous version: Given two random variables $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}^d$ with distributions P and Q, Monge problem minimizes $\mathbb{E}[c(X, T(X))]$ among Borel maps $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T(X)$ follows the distribution Q, while Kantorovich problem minimizes $\mathbb{E}[c(U, V)]$ among the random variables (U, V) with marginal distributions P and Q. The relationship between the Kantorovich and Monge problems in the continuous setting has been explored in the works of Yann Brenier [2], Juan Antonio Cuesta-Albertos and Carlos Matrán Bea [3], Ludger Rüschendorf and Svetlozar Rachev [8], as well as Wilfrid Gangbo and Robert J. McCann [4]. Of course, that's just the tip of the iceberg—there's a whole history behind this, and the standard references (see e.g. [7, 9, 10]) tell the rest of the story.

References

[1] G. Birkhoff. Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucumán. Revista A*, 5:147–151, 1946. Contains the statement of what is now called the Birkhoff–von Neumann theorem.

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

[3] J. A. Cuesta and C. Matrán. Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.*, 17(3):1264–1276, 1989.

[4] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[5] L. Kantorovitch. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[6] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. Histoire de l'Académie Royale des Sciences de Paris, Paris, 1781. Presented to the Académie Royale des Sciences, Paris, 1781.

[7] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[8] L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32(1):48–54, 1990.

[9] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[10] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.