C. R. Acad. Sci. Paris, Ser. I 356 (2018) 207–213

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Functional analysis/Mathematical analysis

Remarks on the Monge–Kantorovich problem in the discrete setting

Remarques sur le problème de Monge–Kantorovich dans le cas discret

Haïm Brezis [a,b,c]

[a] *Department of Mathematics, Hill Center, Busch Campus, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854, USA*
[b] *Departments of Mathematics and Computer Science, Technion, Israel Institute of Technology, 32000 Haifa, Israel*
[c] *Laboratoire Jacques-Louis-Lions, Université Pierre-et-Marie-Curie, 4, place Jussieu, 75252 Paris cedex 05, France*

A R T I C L E I N F O

Article history:
Received 19 December 2017
Accepted 19 December 2017
Available online 5 January 2018

Presented by Haïm Brezis

A B S T R A C T

In Optimal Transport theory, three quantities play a central role: the minimal cost of transport, originally introduced by Monge, its relaxed version introduced by Kantorovich, and a dual formulation also due to Kantorovich. The goal of this Note is to publicize a very elementary, self-contained argument extracted from [9], which shows that all three quantities coincide in the discrete case.

R É S U M É

En théorie du transport optimal, trois quantités jouent un rôle central : le coût minimal de transport, introduit par Monge, sa version relaxée, introduite par Kantorovich, et la formulation duale, due aussi à Kantorovich. L'objet de cette note est de mettre en avant une démonstration totalement élémentaire, extraite de [9], du fait que ces trois quantités coïncident dans le cas discret ; cette preuve ne requiert aucune connaissance préalable.

1. Introduction

Consider two sets X, Y consisting of m points (P_i) and (N_i), $1 \leq i \leq m$, i.e.

$$X = \{P_i, P_2, \dots, P_m\} \text{ and } Y = \{N_1, N_2, \dots, N_m\}.$$

Let $c : X \times Y \to \mathbb{R}$ be any function (c stands for "cost"). We introduce three quantities. The first one denoted M (for Monge) is defined by

E-mail address: brezis@math.rutgers.edu.

https://doi.org/10.1016/j.crma.2017.12.008
1631-073X/

$$M := \min_{\sigma \in \mathcal{S}_m} \sum_{i=1}^{m} c(P_i, N_{\sigma(i)}), \tag{1}$$

where the minimum is taken over the set $\mathcal{S}_m$ of all permutations of the integers $\{1, 2, \dots, m\}$. The second one, denoted K (for Kantorovich), is defined by

$$K := \min_{A} \left\{ \sum_{i,j=1}^{m} a_{ij} c(P_i, N_j);\ A = (a_{ij}) \text{ is doubly stochastic} \right\}. \tag{2}$$

Recall that a matrix $A = (a_{ij})$ is doubly stochastic if

$$a_{ij} \geq 0\, \forall i, j,\ \sum_{i=1}^{m} a_{ij} = 1\ \forall j,\ \text{ and } \sum_{j=1}^{m} a_{ij} = 1\ \forall i. \tag{3}$$

Finally define D (for duality) by

$$D := \sup_{\substack{\psi: Y \to \mathbb{R} \\ \varphi: X \to \mathbb{R}}} \left\{ \sum_{i=1}^{m} (\varphi(P_i) - \psi(N_i));\ \varphi(x) - \psi(y) \leq c(x, y),\ \forall x \in X, \forall y \in Y \right\}. \tag{4}$$

Theorem 1.1. *We have*

$$M = K = D. \tag{5}$$

Moreover the "sup" in (4) is achieved.

Equality $K = D$ in Theorem 1.1 is at the heart of Kantorovich's pioneering discovery concerning the Monge problem (see [19] and [20]). Equality $M = K$ makes totally transparent the connection between Kantorovich's formulation and Monge's original goal (see item (3) in Section 4 below). The purpose of this note is to advertise the MK (= Monge–Kantorovich) theory in its most elementary (but in itself striking and useful!) setting, as it appears, e.g., in Brezis–Coron–Lieb [11] (see Section 3 and item (1) in Section 4 below). This "primitive" case illuminates the foundations of the MK saga which has "exploded" in recent years; see, e.g., the remarkable works of [2], [3], [7], [15], [16], [17], [23], [24], [29], [34], [35], etc. I reproduce in Section 2 an elementary self-contained proof of Theorem 1.1 (accessible to first-year students), extracted from a presentation of [11] that I gave in 1985 (see [9]).

2. Proof of Theorem 1.1

Choosing for A in (2) a permutation matrix yields

$$K \leq M. \tag{6}$$

On the other hand, assume that φ and ψ are as in (4). Let $A = (a_{ij})$ be a doubly stochastic matrix. Multiplying the inequalities $\varphi(P_i) - \psi(N_j) \leq c(P_i, N_j)$ by a_{ij} and summing over i, j yields

$$\sum_{i=1}^{m} (\varphi(P_i) - \psi(N_i)) \leq \sum_{i,j=1}^{m} a_{ij} c(P_i, N_j). \tag{7}$$

Minimizing over A and maximizing over φ, ψ gives

$$D \leq K. \tag{8}$$

In view of (6) and (8), it suffices to establish that

$$M \leq D. \tag{9}$$

Proof of (9). Without loss of generality we may relabel the points (N_j) so that

$$M = \sum_{i=1}^{m} c(P_i, N_i) \leq \sum_{j=1}^{m} c(P_j, N_{\sigma(j)}) \quad \forall \sigma \in \mathcal{S}_m. \tag{10}$$

By (4) it remains to show that there exist functions $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ such that

$$\sum_{i=1}^{m}(\varphi(P_i) - \psi(N_i)) \geq M \tag{11}$$

and

$$\varphi(P_i) - \psi(N_j) \leq c(P_i, N_j) \quad \forall i, j. \tag{12}$$

Set

$$d_i = c(P_i, N_i) \quad \forall i \tag{13}$$

and

$$b_{ij} = c(P_i, N_j) - d_i = c(P_i, N_j) - c(P_i, N_i) \quad \forall i, j. \tag{14}$$

Consider the numbers $\lambda_i = \psi(N_i)$, $1 \leq i \leq m$, as being the *unknowns*. Once the λ_i's have been determined set

$$\varphi(P_i) = \psi(N_i) + d_i = \lambda_i + d_i \quad \forall i. \tag{15}$$

(This choice is dictated by (10), (11), and (12) applied with $i = j$.) From (15), (13) and (10) we see that (11) holds. We now rewrite (12) as

$$\lambda_i - \lambda_j \leq b_{ij} \quad \forall i, j. \tag{16}$$

Note that by (13) and (14)

$$b_{ii} = 0 \quad 1 \leq i \leq m, \tag{17}$$

and that the hypothesis (10) reads

$$\sum_{j=1}^{m} b_{j\sigma(j)} \geq 0 \quad \forall \sigma \in \mathcal{S}_m. \tag{18}$$

We complete the proof of (12) (and thus the existence of functions φ and ψ satisfying (11)–(12)) via the next lemma essentially due to Afriat [1].

Lemma 2.1. *Assume that (b_{ij}) is a general matrix satisfying (17)–(18). Then the system of inequalities (16) admits a solution.*

Proof. (Copied from [9], inspired by [1]). We first propose an ansatz for the λ_i's and then prove that this ansatz has all the required properties. A *chain* K connecting i to j is a finite sequence $K = (i_1, \dots, i_k)$ such that $k \geq 2$, $i_l \in \{1, \dots, m\}$ $\forall l$, $i_1 = i$, and $i_k = j$. (We do not assume that $i_1, \dots, i_k$ are distinct.)

Given a chain K connecting $i = i_1$ to $j = i_k$, set

$$S_K := b_{i_1 i_2} + b_{i_2 i_3} + \dots + b_{i_{k-1} i_k}. \tag{19}$$

Suppose now that a solution (λ_i) to (16) exists and consider a chain K connecting i to j. We have

$$\begin{aligned} \lambda_{i_1} - \lambda_{i_2} &\leq b_{i_1 i_2}, \\ \lambda_{i_2} - \lambda_{i_3} &\leq b_{i_2 i_3}, \\ &\dots \\ \lambda_{i_{k-1}} - \lambda_{i_k} &\leq b_{i_{k-1} i_k}. \end{aligned}$$

Adding these inequalities yields

$$\lambda_i - \lambda_j \leq S_K, \tag{20}$$

and in particular

$$\lambda_i - \lambda_1 \leq \inf_K \left\{ S_K;\, K \text{ is a chain connecting } i \text{ to } 1 \right\}. \tag{21}$$

We now turn to the *existence* of a solution (λ_i) to (16). Since the λ_i's are defined modulo an additive constant it is tempting, in view of (21), to set, for every $1 \leq i \leq m$,

$$\lambda_i := \inf_K \left\{ S_K ; K \text{ is a chain connecting } i \text{ to } 1 \right\}. \tag{22}$$

(A priori, it may happen that $\lambda_i = -\infty$, but this will be excluded below.) Fix $1 \le i, j \le m$ and let $K = (i_1, \dots, i_k)$ be any chain connecting j to 1, then $\tilde{K} := (i, K)$ connects i to 1 and therefore (by (22))

$$\lambda_i \le S_{\tilde{K}} = b_{ij} + S_K. \tag{23}$$

Taking the inf over K in (23) we obtain

$$\lambda_i \le b_{ij} + \lambda_j \quad \forall 1 \le i, j \le m. \tag{24}$$

This corresponds to the desired inequality (16) provided we establish that $\lambda_j \ne -\infty \; \forall j$; assumptions (17) and (18) enter here. We will prove that

$$\lambda_1 = 0. \tag{25}$$

Then, combining (24) and (25) we deduce that

$$0 = \lambda_1 \le b_{1j} + \lambda_j \quad \forall j$$

and thus $\lambda_j \ne -\infty \; \forall j$. We now turn to the proof of (25). First, we choose the chain $K = (1, 1)$ in (22) and obtain

$$\lambda_1 \le b_{11} = 0. \tag{26}$$

Next we establish that $\lambda_1 \ge 0$. We start with some terminology. A chain K connecting i to $j = i$ is called a *cycle* (or a loop). A cycle is *simple* if $i_1, \dots, i_{k-1}$ are distinct. We claim that, for every cycle K,

$$S_K \ge 0. \tag{27}$$

Indeed when K is a simple cycle (27) follows from (18) (and (17)) applied to the permutation $i_1 \to i_2 \cdots \to i_k$ (the other integers are invariant). By decomposing a general cycle into simple cycles we find that (27) holds for all cycles. Applying (27) to any chain connecting 1 to 1, we deduce from (22) that $\lambda_1 \ge 0$. $\square$

Remark 2.1. The above proof provides in fact a *necessary and sufficient* condition for the existence of a solution to (16). It reads as follows.

$$\sum_{j \in B} b_{j\sigma(j)} \ge 0, \quad \forall \sigma \in \mathcal{S}_k, \tag{28}$$

for every integer $1 \le k \le m$ and for every subset B of $\{1, \dots, m\}$ containing k distinct elements, where the permutations σ act only on B. This result appears already in [1] as a consequence of Theorems 3.1 and 7.2 in [1]. Unfortunately, the proofs in [1] are obscured by a flurry of definitions!

3. When the cost *c* is a distance

We now present a simple consequence of Theorem 1.1 when the cost c is a distance, which corresponds to the setting of [20]. Let $d(x, y)$ be a pseudometric (i.e. the distance between two distinct points can be zero) on a set Z. Let $(P_i), (N_i), 1 \le i \le m$ be points in Z such that $P_i \ne N_j \quad \forall i, j$ (but it may happen that $P_i = P_j$ or $N_i = N_j$ for some $i \ne j$).

Corollary 3.1. *We have*

$$\begin{aligned} D_{\text{Lip}} &:= \sup_\zeta \left\{ \sum_{i=1}^m (\zeta(P_i) - \zeta(N_i)); \zeta : Z \to \mathbb{R}, |\zeta(x) - \zeta(y)| \le d(x, y) \, \forall x, y \in Z \right\} \\ &= \min_{\sigma \in S_m} \sum_{i=1}^m d(P_i, N_{\sigma(i)}) = M. \end{aligned}$$

Proof. Clearly $D_{\text{Lip}} \le M$. After relabeling the points (N_j) we may assume, as in (10), that

$$M = \sum_{i=1}^m d(P_i, N_i) \le \sum_{j=1}^m d(P_j, N_{\sigma(j)}) \quad \forall \sigma \in \mathcal{S}_m. \tag{29}$$

Applying Theorem 1.1 with $X = \{P_1, P_2, \dots, P_m\}$, $Y = \{N_1, N_2, \dots, N_m\}$ and $c(x, y) = d(x, y)$ we know that $M = D$; thus we obtain functions $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ such that

$$\varphi(P_i) - \psi(N_i) = d(P_i, N_i) \quad \forall i, \tag{30}$$

and

$$\varphi(P_i) - \psi(N_j) \le d(P_i, N_j) \quad \forall i, j. \tag{31}$$

We claim that

$$|\psi(N_i) - \psi(N_j)| \le d(N_i, N_j) \quad \forall i, j. \tag{32}$$

Indeed, by (30), (31), and the triangle inequality we have

$$\psi(N_i) = \varphi(P_i) - d(P_i, N_i) \le \psi(N_j) + d(P_i, N_j) - d(P_i, N_i) \le \psi(N_j) + d(N_i, N_j),$$

which implies (32). Set, for $z \in Z$,

$$\zeta_0(z) = \inf_j \left\{ \psi(N_j) + d(z, N_j) \right\}, \tag{33}$$

so that

$$|\zeta_0(x) - \zeta_0(y)| \le d(x, y) \quad \forall x, y \in Z.$$

From (32) we see that

$$\zeta_0(N_i) = \psi(N_i) \quad \forall i. \tag{34}$$

On the other hand, we have, by (33) and (31), $\zeta_0(P_i) \ge \varphi(P_i) \quad \forall i$, while taking $j = i$ in (33), and using (30), yields

$$\zeta_0(P_i) \le d(P_i, N_i) + \psi(N_i) = \varphi(P_i) \quad \forall i.$$

Therefore,

$$\zeta_0(P_i) = \varphi(P_i) \quad \forall i. \tag{35}$$

Choosing $\zeta = \zeta_0$ in the definition of D_{Lip} and applying (30), (34), and (35) yields $D_{\text{Lip}} \ge M$. □

4. Final comments

(1) Corollary 3.1 is taken from [11]. Equality $M = D_{\text{Lip}}$ plays an important role in proving that the "least energy required to produce prescribed singularities" (in liquid crystals) coincides with the "length of a minimal connection connecting these singularities" (for subsequent developments see, e.g., [5], [8], [12], [13] and [28]). The proof of Corollary 3.1 in [11] takes a few lines, but it relies heavily on three nontrivial tools. The equality $K = D_{\text{Lip}}$ is derived from Kantorovich's duality (see item (2) below). While the equality $M = K$ relies on Birkhoff's theorem [6] on doubly stochastic matrices (also called Birkhoff–von Neumann's theorem because von Neumann [36] rediscovered it independently a few years later). It asserts that the extreme points of the convex set A of doubly stochastic matrices are precisely the permutation matrices. Applying the Krein–Milman theorem, one deduces that any matrix in A is a convex combination of permutation matrices, and consequently $K \ge M$. (For recent developments related to Birkhoff's theorem, I refer the reader to [22] and [14].) By contrast, the above proof of Corollary 3.1 is elementary and self-contained. No prerequisite is needed and moreover it yields the two equalities $M = K$ and $K = D$ in a single shot!

(2) Equality $K = D$ in Theorem 1.1 is at the heart of Kantorovich's discovery (dating back to the late 1930s – see the references in [33]) and goes *far beyond* the discrete setting considered here. Note that K and D involve the minimization (resp. maximization) of linear functionals on convex sets. The most common way to show that $K = D$ is via duality, either in the sense of linear programming or in the sense of conjugate convex functions (applying for example the theorem of Fenchel–Rockafellar; see, e.g., Theorem I.12 in [10]). I refer the reader to [2], [3], [15], [17], [23], [24], [29], [34], [35], etc.

(3) According to A. M. Vershik (personal communication), equality $M = K$ was known to L. V. Kantorovich, based on Birkhoff's theorem (see item (1) above), which was published around the same time as [19]-[20]. Apparently Birkhoff's ideas were in the air since a precursor of Birkhoff's theorem appeared already in 1931 (see the historical note on p. 25 of [14]). Surprisingly, Birkhoff's theorem is hardly ever mentioned in the vast MK literature. The reason for it being that the MK community has been mostly preoccupied with the equality $M = K$ in the *non-atomic* case; in this setting, the Monge formulation was not even precisely stated until the 1970s when it was posed explicitly in modern terms by A. M. Vershik [32] (see also [7]). In their rush to the continuum case, the MK aficionados paid little attention to the discrete case — which is in itself striking and useful!!

(4) As already mentioned, our elementary proof of Theorem 1.1 does *not* require any of the tools described in items (1) and (2) above. Instead, it relies on the construction (22) (copied from [9]) involving "chains" and "cycles". This device is reminiscent of Rockafellar's celebrated theorem [26] on cyclically monotone operators. The same construction appears subsequently, at the suggestion of Rockafellar, in [25] in the context of Mathematical Economics, and then in [31] in the MK context. In [31], Smith and Knott introduced the terminology "c-cyclical monotonicity", which has become very fashionable in the MK community, see [2], [3], [4], [15], [17], [21], [23], [24], [27], [30], [29], [34], [35], etc. In the literature, one can find two distinct definitions. The original definition says that if X, Y are arbitrary sets and $c : X \times Y \to \mathbb{R}$ is any function, then a set $\Gamma \subset X \times Y$ is c-cyclically monotone if for every integer n, and for any finite sequence (x_i, y_i), $1 \le i \le n$, of points in Γ (not necessarily distinct), one has

$$\sum_{i=1}^{n} \{c(x_i, y_{i+1}) - c(x_i, y_i)\} \ge 0, \tag{36}$$

where $y_{n+1} := y_1$. In another definition, (36) is replaced by

$$\sum_{i=1}^{n} \left\{c(x_i, y_{\sigma(i)}) - c(x_i, y_i)\right\} \ge 0 \quad \forall \sigma \in \mathcal{S}_n. \tag{37}$$

In fact, the two definitions are equivalent. Clearly, (37) implies (36) (just choose $\sigma(i) = i + 1$ when $1 \le i \le n - 1$ and $\sigma(n) = 1$). For the reverse implication, we return to the *proof* of Lemma 2.1 with $b_{ij} = c(x_i, y_j) - c(x_i, y_i)$. We claim that for every cycle $K = (i_1, i_2, \ldots, i_{k-1}, i_1)$, one has $S_K \ge 0$ (so that the conclusion of Lemma 2.1 holds, and clearly implies (37)). Applying (36) to $(x_{i_1}, y_{i_1}), \ldots, (x_{i_{k-1}}, y_{i_{k-1}})$ (instead of (x_i, y_i)) yields $S_K \ge 0$. If we take $\Gamma = (P_i, N_i)$, $1 \le i \le m$, as in the setting of Theorem 1.1, assumption (37) seems (at least formally) stronger than assumption (18) in Lemma 2.1 because (37) is assumed for *all* finite sequences (x_i, y_i) in Γ, and moreover these points are not necessarily distinct — but the conclusions are the same and thus the two assumptions are a posteriori equivalent! Finally, observe that if $X = Y = H$ is a Hilbert space and $c(x, y) = |x - y|^2$, then a set $\Gamma \subset H \times H$ is c-cyclically monotone if and only if it is cyclically monotone in the usual sense (coined by Rockafellar).

(5) E. Ghys [18] and A. Vershik [33] tell the fascinating stories of the Monge and Kantorovich discoveries. I highly recommend these papers.

Acknowledgements

I thank P. Mironescu for catching a small oversight in [9] and L. Ambrosio for providing useful references. I am very grateful to A.M. Vershik for sharing with me informations concerning the history and later developments of the MK theory, and to L.C. Evans for his warm encouragements.

References

[1] S.N. Afriat, The system of inequalities $a_{rs} \ge X_r - X_s$, Math. Proc. Camb. Philos. Soc. 59 (1963) 125–133.
[2] L. Ambrosio, Lecture notes on optimal transport problems, in: P. Colli, J.F. Rodrigues (Eds.), Mathematical Aspects of Evolving Interfaces, in: Lect. Notes Math., vol. 1812, Springer, 2003, pp. 1–52.
[3] L. Ambrosio, N. Gigli, A user's guide to optimal transport, in: Modelling and Optimisation of Flows on Networks, in: Lect. Notes Math., vol. 2062, Springer, 2013, pp. 1–155.
[4] M. Beiglböck, Cyclical monotonicity and the ergodic theorem, Ergod. Theory Dyn. Syst. 35 (2015) 710–713.
[5] F. Bethuel, H. Brezis, J.-M. Coron, Relaxed energies for harmonic maps, in: H. Berestycki, J.-M. Coron, I. Ekeland (Eds.), Variational Problems, Proceedings of a Conference Held in Paris in 1988, Birkhäuser, 1990, pp. 37–52.
[6] G. Birkhoff, Tres observaciones sobre el algebra lineal, Rev. Univ. Nac. Tucumán Ser. A, Mat. Fis. Teor. 5 (1946) 147–150.
[7] V.I. Bogachev, A.V. Kolesnikov, The Monge–Kantorovich problem: achievements, connections and perspectives, Usp. Mat. Nauk 67 (2012) 3–110, English translation: Russ. Math. Surv. 67 (2012) 785–890.
[8] J. Bourgain, H. Brezis, P. Mironescu, $H^{1/2}$ maps with values into the circle: minimal connections, lifting, and the Ginzburg–Landau equation, Publ. Math. Inst. Hautes Études Sci. 99 (2004) 1–115.
[9] H. Brezis, Liquid crystals and energy estimates for $\mathbb{S}^2$-valued maps, in: J. Ericksen, D. Kinderlehrer (Eds.), Theory and Applications of Liquid Crystals, Proceedings of a Conference Held in Minneapolis, MN, USA, in 1985, Springer, 1987, pp. 31–52. See also article 112L1 in http://sites.math.rutgers.edu/~brezis/publications.html.
[10] H. Brezis, Functional Analysis, Sobolev Spaces and PDEs, Springer, 2011.
[11] H. Brezis, J.-M. Coron, E. Lieb, Harmonic maps with defects, Commun. Math. Phys. 107 (1986) 649–705.
[12] H. Brezis, P. Mironescu, Sobolev Maps with Values into the Circle, Birkhäuser, in preparation.
[13] H. Brezis, P. Mironescu, I. Shafrir, Distances between classes in $W^{1,1}(\Omega; \mathbb{S}^1)$, Calc. Var. Partial Differ. Equ. (2017), https://doi.org/10.1007/s00526-017-1280-z, to appear.
[14] R. Burkard, M. Dell Amico, S. Martello, Assignment Problems, revised edition, SIAM, 2012.
[15] L.C. Evans, Partial differential equations and Monge–Kantorovich mass transfer, in: S.T. Yau (Ed.), Current Developments in Mathematics, Lectures delivered at Harvard in 1997, International Press, Cambridge, MA, USA, 1999, pp. 65–126. See also "Survey of applications of PDE methods to Monge–Kantorovich mass transfer problems" https://math.berkeley.edu/evans/.
[16] L.C. Evans, W. Gangbo, Differential equations methods for the Monge–Kantorovich mass transfer problem, Mem. Amer. Math. Soc. 137 (1999).
[17] W. Gangbo, R. McCann, The geometry of optimal transportation, Acta Math. 177 (1996) 113–161.

[18] É. Ghys, Gaspard Monge. Le mémoire sur les déblais et les remblais, CNRS, Images des mathématiques, 2012. See http://images.math.cnrs.fr/Gaspard-Monge,1094.html?lang=fr.
[19] L.V. Kantorovitch, On the translocation of masses, Dokl. Akad. Nauk SSSR 37 (1942) 227–229, English translation: J. Math. Sci. 133 (2006) 1381–1382.
[20] L.V. Kantorovitch, On a problem of Monge, Usp. Mat. Nauk 3 (1948) 225–226, English translation: J. Math. Sci. 133 (2006) 1383.
[21] M. Knott, C. Smith, On a generalization of cyclic monotonicity and distances among random vectors, Linear Algebra Appl. 199 (1994) 363–371.
[22] N. Linial, Z. Luria, On the vertices of the d-dimensional Birkhoff polytope, Discrete Comput. Geom. 51 (2014) 161–170.
[23] R.J. McCann, N. Guillen, Five lectures on optimal transportation: geometry, regularity and applications, in: G. Dafni, et al. (Eds.), Analysis and Geometry of Metric Measure Spaces, Lecture Notes of the "Séminaire de mathématiques supérieures" (SMS), 2011, American Mathematical Society, 2013, pp. 145–180. See also [46] in http://www.math.toronto.edu/mccann/publications.
[24] S. Rachev, L. Rüschendorf, Mass Transportation Problems, Springer, 1998.
[25] J.-C. Rochet, A necessary and sufficient condition for rationalizability in a quasi-linear context, J. Math. Econ. 16 (1987) 191–200.
[26] R.T. Rockafellar, Characterization of the subdifferentials of convex functions, Pac. J. Math. 17 (1966) 497–510.
[27] L. Rüschendorf, On c-optimal random variables, Stat. Probab. Lett. 27 (1996) 267–270.
[28] E. Sandier, Ginzburg–Landau minimizers from $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$ and minimal connections, Indiana Univ. Math. J. 50 (2001) 1807–1844.
[29] F. Santambrogio, Optimal Transport for Applied Mathematicians, Birkhäuser, 2015.
[30] C.S. Smith, M. Knott, Note on the optimal transportation of distributions, J. Optim. Theory Appl. 52 (1987) 323–329.
[31] C.S. Smith, M. Knott, On Hoeffding–Fréchet bounds and cyclic monotone relations, J. Multivar. Anal. 40 (1992) 328–334.
[32] A.M. Vershik, Some remarks on the infinite-dimensional problems of linear programming, Usp. Mat. Nauk 25 (1970) 117–124, English translation: Russ. Math. Surv. 25 (1970) 117–124.
[33] A.M. Vershik, Long history of the Monge–Kantorovich transportation problem, Math. Intell. 35 (2013) 1–9.
[34] C. Villani, Topics in Optimal Transportation, American Mathematical Society, 2003.
[35] C. Villani, Optimal Transport. Old and New, Springer, 2009.
[36] J. von Neumann, A certain zero-sum two-person game equivalent to the optimal assignment problem, in: Contrib. Theory Games, vol. 2, Princeton University Press, Princeton, NJ, USA, 1953, pp. 5–12.