Ergod. Th. & Dynam. Sys. (2015), **35**, 710–713 © *Cambridge University Press, 2013*
doi:10.1017/etds.2013.75

Cyclical monotonicity and the ergodic theorem

MATHIAS BEIGLBÖCK

Fakultät für Mathematik, Universität Wien, Nordbergstraße 15, 1090 Wien, Austria
(e-mail: mathias.beiglboeck@univie.ac.at)

(*Received* 9 *December* 2012 *and accepted in revised form* 18 *August* 2013)

Abstract. It is well known that *optimal* transport plans are *cyclically monotone*. The reverse implication that cyclically monotone transport plans are optimal needs some assumptions and the proof is non-trivial even if the costs are given by the squared euclidean distance on $\mathbb{R}^n$. We establish this result as a corollary to the ergodic theorem.

1. *Introduction*

We consider the *Monge–Kantorovich transport problem* for probabilities μ, ν on Polish spaces X, Y (cf. [**Vil03, Vil09**]). The set $\Pi(\mu, \nu)$ of *transport plans* consists of all measures on $X \times Y$ with X-marginal μ and Y-marginal ν. Associated to a *cost function* $c : X \times Y \to [0, \infty]$ and $\pi \in \Pi(\mu, \nu)$ are the *transport costs*

$$\langle c, \pi \rangle = \int_{X \times Y} c(x, y)\, d\pi(x, y).$$

The Monge–Kantorovich problem is then to determine the value

$$P_c := \inf\{\langle c, \pi \rangle : \pi \in \Pi(\mu, \nu)\} \tag{1}$$

and to identify an *optimal* transport plan $\hat{\pi} \in \Pi(\mu, \nu)$, i.e. a minimizer of (1).

A basic and important goal is to characterize minimizers through a tractable property of their support sets: a Borel set $\Gamma \subseteq X \times Y$ is *cyclically monotone* if and only if

$$c(x_1, y_2) - c(x_1, y_1) + \cdots + c(x_{n-1}, y_n) - c(x_{n-1}, y_{n-1}) + c(x_n, y_1) - c(x_n, y_n) \geq 0$$

whenever $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \Gamma$.

A transport plan π is *cyclically monotone* if it assigns full measure to some cyclically monotone set Γ.

Concerning the origins of cyclical monotonicity in convex analysis and the study of the relation to optimality we mention [**GM96, KS92, Roc66, Rüs96**]. Intuitively speaking, cyclically monotone transport plans resist improvement by means of cyclical rerouting and optimal transport plans are expected to have this property. Indeed, Ambrosio and Pratelli [**AP03**] show for general lower semi-continuous $c : X \times Y \to [0, \infty]$ that every optimal transport plan is cyclically monotone.

https://doi.org/10.1017/etds.2013.75 Published online by Cambridge University Press

The reverse implication is more intricate. Ambrosio and Pratelli find that there exist cyclically monotone transport plans which are *not* optimal; notably the construction given in [**AP03**, Example 3.5] makes heavy use of the fact that c may attain the value ∞. Villani [**Vil03**, Problem 2.25] asked whether cyclically monotone transport plans are always optimal in the case where c is the squared euclidean distance on $\mathbb{R}^n$. This problem was resolved by Pratelli [**Pra08**] and Schachermayer and Teichmann [**ST09**]: every cyclically monotone transport plan is optimal provided that c is continuous or merely lower semi-continuous and finitely valued†. We give a new proof to the following theorem.

THEOREM 1. *Let $c : X \times Y \to [0, \infty)$ be a measurable cost function and $\pi \in \Pi(\mu, \nu)$ a finite cost transport plan which is cyclically monotone. Then π is optimal.*

2. *Theorem 1 as consequence of the ergodic theorem*

The novelty of our approach is to connect cyclical monotonicity to the pointwise *ergodic theorem* which we restate here: let (Z, κ) be a probability space and $\sigma : Z \to Z$ measure-preserving, i.e. $\sigma(\kappa) = \kappa$. Then for every $f \in L^1(\kappa)$ the limit

$$f^* = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ \sigma^k \tag{2}$$

exists almost surely and in $L^1(\kappa)$ (see, for example, [**Kal02**, Theorem 9.6]).

Proof of Theorem 1. Let $\pi, \tilde{\pi} \in \Pi(\mu, \nu)$ be finite cost transport plans and $\pi(\Gamma) = 1$ for some cyclically monotone set Γ. We have to show that π leads to lower costs than $\tilde{\pi}$. To this end, we specify $Z = (X \times Y)^{\mathbb{N}}$ and consider the *shift mapping*

$$\sigma : Z \to Z, \quad (x_i, y_i)_{i=1}^{\infty} \mapsto (x_{i+1}, y_{i+1})_{i=1}^{\infty}.$$

Moreover, we define the projections $P, Q : Z \to X \times Y$ by

$$P((x_i, y_i)_{i=1}^{\infty}) := (x_1, y_1) \quad \text{and} \quad Q((x_i, y_i)_{i=1}^{\infty}) := (x_1, y_2).$$

The transport plans π and $\tilde{\pi}$ give rise to a natural σ-invariant measure on Z.

LEMMA 2.1. *There is a measure κ on Z such that $\sigma(\kappa) = \kappa$, $P(\kappa) = \pi$, $Q(\kappa) = \tilde{\pi}$.*

Proof. Let $(\pi^y)_{y \in Y}$ be the disintegration of π with respect to (Y, ν) and let $(\tilde{\pi}_x)_{x \in X}$ be the disintegration of $\tilde{\pi}$ with respect to (X, μ). Then $R(y, A) := \pi^y(A \times \{y\})$ and $S(x, B) := \tilde{\pi}_x(\{x\} \times B)$ constitute transition kernels from Y to X and from X to Y, respectively. We identify Z with the product

$$Y^{(1)} \times X^{(1)} \times Y^{(2)} \times X^{(2)} \times \cdots$$

where $X^{(n)}$, $Y^{(n)}$, $n \geq 1$, are copies of X and Y. Then we consider the discrete time Markov process with initial distribution ν, and the probabilities of moving from $y \in Y^{(n)}$ to $A \subseteq X^{(n)}$ and from $x \in X^{(n)}$ to $B \subseteq Y^{(n+1)}$ given by R and S, respectively.

The resulting distribution κ on the space of paths Z has the desired properties. (We refer to [**Kal02**, Ch. 7] for more details on the definition of κ.) □

† We refer to [**BC10, BGMS09**] for more general results, in particular lower semi-continuity of the cost function is not important in either implication.

https://doi.org/10.1017/etds.2013.75 Published online by Cambridge University Press

The map $f := c \circ Q - c \circ P \in L^1(\kappa)$ satisfies

$$\int_{X\times Y} c\, d\tilde{\pi} - \int_{X\times Y} c\, d\pi = \int_Z f\, d\kappa.$$

The crucial step of our proof is that, applying the ergodic theorem to the function f and integrating over (2), we have

$$\int_Z f\, d\kappa = \int_Z f^*\, d\kappa = \int_Z \lim_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ \sigma^k\, d\kappa.$$

Rewriting this expression in terms of c, we obtain

$$\int_{X\times Y} c\, d\tilde{\pi} - \int_{X\times Y} c\, d\pi = \int_Z \left[\lim_n \frac{1}{n} \sum_{k=1}^{n} c(x_k, y_{k+1}) - c(x_k, y_k) \right] d\kappa(x_i, y_i)_i. \quad (3)$$

To conclude the proof it is sufficient to show that the integrand on the right-hand side of (3) is κ-almost surely non-negative. Note that $\kappa(\Gamma \times (X \times Y)^{\mathbb{N}}) = \pi(\Gamma) = 1$. Since κ is σ-invariant and $\Gamma^{\mathbb{N}} = \bigcap_{n\geq 1}(\sigma^{-n}(\Gamma \times (X \times Y)^{\mathbb{N}}))$ we have $\kappa(\Gamma^{\mathbb{N}}) = 1$. Thus it suffices to argue on sequences $(x_n, y_n)_{n=1}^{\infty}$ with $(x_n, y_n) \in \Gamma$, $n \geq 1$.

We note that the proof of Theorem 1 is immediate under the additional assumption that c is bounded: in this case cyclical monotonicity of Γ trivially implies

$$\liminf_n \frac{1}{n} \sum_{k=1}^{n} c(x_k, y_{k+1}) - c(x_k, y_k) \geq 0 \quad \text{whenever } (x_n, y_n)_{n=1}^{\infty} \in \Gamma^{\mathbb{N}},$$

thus π is optimal.

To prove Theorem 1 in the general case fix $(\bar{x}, \bar{y}) \in \Gamma$. Given $(x_i, y_i)_{i=1}^{\infty} \in \Gamma^{\mathbb{N}}$, cyclical monotonicity of Γ implies that

$$c(\bar{x}, y_1) - c(\bar{x}, \bar{y}) + \left(\sum_{k=1}^{n} c(x_k, y_{k+1}) - c(x_k, y_k) \right) + c(x_{n+1}, \bar{y}) - c(x_{n+1}, y_{n+1})$$

is non-negative for each n. As c takes values in $[0, \infty)$ this further yields

$$\liminf_n \left[\frac{1}{n} \left(\sum_{k=1}^{n} (c(x_k, y_{k+1}) - c(x_k, y_k) \right) + \frac{c(x_{n+1}, \bar{y})}{n} \right] \geq 0. \quad (4)$$

Setting $g((x_i, y_i)_{i=1}^{\infty}) := c(x_1, \bar{y})$ we have $c(x_{n+1}, \bar{y}) = g \circ \sigma^n((x_i, y_i)_{i=1}^{\infty})$. As g is a finite function, g/n tends to 0 in measure (with respect to κ) and because σ is measure preserving, also $\lim_{n\to\infty} g \circ \sigma^n / n = 0$ in measure. Passing to a subsequence if necessary, we may assume that convergence holds κ-almost surely. Together with (4) this proves that (3) is non-negative and thus concludes the proof of Theorem 1. □

3. *Comments*

We have used the pointwise ergodic theorem, which may be somewhat tricky for readers not accustomed to ergodic theory. We note that the simpler mean ergodic theorem is also sufficient. It asserts that $\lim_n (1/n) \sum_{k=0}^{n-1} f \circ \sigma^k$ exists in L^p whenever $f \in L^p$, $p \geq 1$; see, for example, [**Kal02**, Exercise 9.14]. In particular, this implies that the limit in (2)

https://doi.org/10.1017/etds.2013.75 Published online by Cambridge University Press

exists almost surely along a subsequence $(n_m)_{m=1}^{\infty}$. The above proof of Theorem 1 then still applies if we argue on n_m rather than n.

Using similar arguments as in [**BC10, BGMS09**], our approach can be used to prove even more general versions of Theorem 1. As this does not require new ideas nor strengthen the results of [**BC10, BGMS09**] we do not elaborate.

References

[AP03] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. *Optimal Transportation and Applications (Martina Franca, 2001) (Lecture Notes in Mathematics, 1813)*. Springer, Berlin, 2003, pp. 123–160.

[BC10] S. Bianchini and L. Caravenna. On optimality of c-cyclically monotone transference plans. *C. R. Math. Acad. Sci. Paris* **348**(11–12) (2010), 613–618.

[BGMS09] M. Beiglböck, M. Goldstern, G. Maresch and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.* **256**(6) (2009), 1907–1927.

[GM96] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.* **177**(2) (1996), 113–161.

[Kal02] O. Kallenberg. *Foundations of Modern Probability (Probability and its Applications)*, 2nd edn. Springer, New York, 2002.

[KS92] M. Knott and C. Smith. On Hoeffding–Fréchet bounds and cyclic monotone relations. *J. Multivariate Anal.* **40**(2) (1992), 328–334.

[Pra08] A. Pratelli. On the sufficiency of c-cyclical monotonicity for optimality of transport plans. *Math. Z.* **258**(3) (2008), 677–690.

[Roc66] R. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.* **17** (1966), 497–510.

[Rüs96] L. Rüschendorf. On c-optimal random variables. *Statist. Probab. Lett.* **27**(3) (1996), 267–270.

[ST09] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge–Kantorovich problem. *Proc. Amer. Math. Soc.* **137**(2) (2009), 519–529.

[Vil03] C. Villani. *Topics in Optimal Transportation (Graduate Studies in Mathematics, 58)*. American Mathematical Society, Providence, RI, 2003.

[Vil09] C. Villani. *Optimal Transport: Old and New (Grundlehren der mathematischen Wissenschaften, 338)*. Springer, Berlin, 2009.

https://doi.org/10.1017/etds.2013.75 Published online by Cambridge University Press