# Proof of convergence of Brenier's maps to Knothe's transport

G. Carlier, A. Galichon, F. Santambrogio

Medha Agarwal and Garrett Mulcahy

October 11, 2022

UNIVERSITY of WASHINGTON

# Contents

# Preliminaries

# Monge Problem

- Let $\mu : \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}^+$ be the source measure defined on the Borel $\sigma$-algebra of $\mathbb{R}^d$ and let $\nu : \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}^+$ be the target measure we wish to transport $\mu$ to. Then the transport map $T$ pushes $\mu$ to $\nu$ if $T_\# \mu = \nu$, i.e. for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$T_\# \mu(A) := \mu(T^{-1}(A)) = \nu(A). \tag{1}$$

- The transport map is characterized by the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ which leads to Monge problem

$$\inf_{T : T_\# \mu = \nu} \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x)\,. \tag{2}$$

The solution to the constrained problem in (2) is the **optimal transport** from $\mu$ to $\nu$.

# Brenier's Theorem

- Brenier maps arise from a relaxation of (2) called Monge-Kantorovich problem and defined as

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) d\pi(x,y), \tag{3}$$

where $\Pi(\mu,\nu)$ is the subspace of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ (the set of Borel probability measures on $\mathbb{R}^d \times \mathbb{R}^d$) with marginals equal to $\mu$ and $\nu$, respectively.

## Brenier's Theorem (Villani (2003), Theorem 2.12)

Under the following assumptions

- the measures $\mu, \nu$ have finite second moments,
- $\mu$ does not give mass to small sets, and
- $c$ represents the quadratic cost, i.e. $c(x,y) = |x-y|^2$,

there exists a unique optimal solution $\pi^*$ to (3) of the form $\pi^* = (\mathrm{Id} \times \nabla\varphi)_{\#}\mu$ where $\nabla\varphi$ is the unique (up to $\mu$-a.e. equivalence) gradient of a convex function that pushes forward $\mu$ to $\nu$, i.e. $\nu = \nabla\varphi_{\#}\mu$.

# 1-D Optimal Transport

As a consequence of the uniqueness clause of Brenier's Theorem, we have the following result

### 1-D Optimal Transport

Let $\mu$ and $\nu$ be probability measures on $\mathbb{R}$ with finite second moments and distribution functions $F$ and $G$, respectively. Additionally, assume $\mu$ is absolutely continuous. Then $G^{-1} \circ F$ is the Brenier map transporting $\mu$ to $\nu$, where $G^{-1}$ is the generalized inverse of $G$ on $[0, 1]$, i.e. $G^{-1}(a) = \inf\{x \in \mathbb{R} : G(x) > a\}$.

The key ingredients in this proof are that:

(1) If $X \sim \mu$ then $F(X) \sim \text{Unif}[0, 1]$. This gives that $G^{-1} \circ F$ transports $\mu$ to $\nu$

(2) A increasing function $\mathbb{R} \to \mathbb{R}$ is the derivative of a convex function. This allows us to invoke Brenier's Theorem.

# Knothe-Rosenblatt

Knothe (1957) and Rosenblatt (1952) proposed Knothe-Rosenblatt (KR) coupling separately that transports $\mu$ to $\nu$ in a triangular manner when $\mu$ is absolutely continuous. The map is of the form $T(x) = (T^1(x), T^2(x_{2:d}), \ldots, T^d(x_d))$ and is defined as

- Let $\mu^d$ and $\nu^d$ be the $d$th marginals of $\mu$ and $\nu$. Then $T^d$ is the monotone non-decreasing mapping that transports $\mu^d$ to $\nu^d$.
- Let $\mu^k_{(x_{k+1:d})}$ (respectively $\nu^k_{(x_{k+1:d})}$) be the conditional measure of $\mu$ (respectively $\nu$) wrt to the $k$th coordinate given $x_{k+1:d}$. Then $T^k$ is the monotone non-decreasing mapping that transports $\mu^k_{(x_{k+1:d})}$ to $\nu^k_{(x_{k+1:d})}$.

If $\mu$ and $\nu$ are absolutely continuous measures, then KR map is the unique triangular map satisfying (1).

# Assumptions & Lemmas

## Assumptions on $\mu$ and $\nu$

We have two major assumptions on the marginals of the conditional measures of $\mu$ and $\nu$. The reason for the conditions on $\mu$ is to be able to apply Brenier's Theorem. We additonally require conditions on $\nu$ to obtain invertibility of the transport maps from $\mu$ to $\nu$.

### Assumption (H-source)

The marginal one-dimensional measure $\mu^d$ ($d$th marginal measure of $\mu$) has no atoms. Similarly for $k \geq 1$, the one-dimensional conditional measures $\mu^k_{x_{d:k+1}}$ for $\mu^d -$ a.e. $x_d$, $\mu^{d-1}_{x_d} -$ a.e. $x_{d-1}, \ldots \mu^{k+1}_{x_{d:k+1}}$ have no atoms.

### Assumption (H-target)

The marginal one-dimensional measure $\nu^d$ has no atoms. Similarly for $k \geq 2$, the one-dimensional conditional measures $\nu^k_{y_{d:k+1}}$ for
$\nu^d -$ a.e. $y_d, \nu^{d-1}_{y_d} -$ a.e. $y_{d-1}, \ldots \nu^{k+1}_{x_{d:k+1}}$ have no atoms.

We note that (H-target) is slightly less restrictive than (H-source). Moreover, both conditions are satisfied if, say, $\mu, \nu << m$.

# Lemmas

We state the following quick lemmas.

### Lemma 1

Let $\mu, \nu$ be Borel probability measures on $\mathbb{R}^d$. If $\int \phi(x)d\mu(x) = \int \phi(x)d\nu(x)$ for all $\phi \in C_c(\mathbb{R}^d)$, then $\mu = \nu$ as measures.

This is a consequence of the Riesz Representation Theorem.

### Lemma 2

If two measurable functions $f, g$ on $\mathbb{R}^d$ are equal $\pi$-almost everywhere, then $\int f(x) - g(x)d\pi(x) = 0$.

# Main Result

# Statement

This presentation is concerned with the following result

**Theorem (Carlier et al. (2010), Theorem 2.1)**

Let $\mu$ and $\nu$ be two probability measures on $\mathbb{R}^d$ satisfying (H-source) and (H-target), respectively, with finite second moments, and $\gamma_\epsilon$ be an optimal transport plan for the cost $c_\epsilon(x, y) = \sum_{i=1}^d \lambda_i(\epsilon)(x_i - y_i)^2$, for some weights $\lambda_k(\epsilon) > 0$. Suppose that for all $k \in \{1, \dots, d-1\}$, $\lambda_k(\epsilon)/\lambda_{k+1}(\epsilon) \to 0$ as $\epsilon \to 0$. Let $T_K$ be the Knothe-Rosenblatt transport from $\mu$ to $\nu$ and $\gamma_K \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ be the associated transport plan (i.e. $\gamma_K := (id \times T_K)_{\#}\mu$). Then $\gamma_\epsilon \rightharpoonup \gamma_K$ as $\epsilon \to 0$.
Moreover, should the plans $\gamma_\epsilon$ be induced by transport maps $T_\epsilon$, then these maps would converge to $T_K$ in $L^2(\mu)$ as $\epsilon \to 0$.

This results gives a limiting procedure in which the optimal $\gamma_\epsilon$ converge to the KR map. It is interesting to note that the sense in which are $\gamma_\epsilon$ are optimal differs from the sense in which $\gamma_K$ is optimal, yet this theorem gives a way of connecting them.

# Proof of main result

# Proof Idea

- The proof is constructed in a recursive manner after a weak limit $\gamma_\epsilon$ is given (denoted by $\gamma$).
- The main strategy of the proof is to use the optimality and uniqueness of KR maps (denoted $\gamma_K$) to show that the $d$th marginal of $\gamma$ and $\gamma_K$ are equal, i.e. $\gamma_K^d = \gamma^d$.
- The equality of marginals is used to show that $\gamma_K^{d-1} = \gamma^{d-1}$ (the marginal on $(x_{d-1}, x_d), (y_{d-1}, y_d)$).
- Following the same idea, we recursively show that $\gamma^h = \gamma_K^h$ if $\gamma^{h+1} = \gamma_K^{h+1}$ for all $1 \leq h < d$.

# Getting started

- Dividing by $\lambda_d(\epsilon)$ if necessary, we may assume that $\lambda_d(\epsilon) = 1$ for all $\epsilon > 0$. This implies $\lambda_i(\epsilon) \to 0$ as $\epsilon \to 0$ for $i = 1, \ldots, d-1$.
- It follows from Theorem 4.1 in Villani (2009) that optimal $\gamma_\epsilon \in \Pi(\mu, \nu)$ exists for each $\epsilon > 0$.
- By the sequential compactness of $\Pi(\mu, \nu)$ in the weak topology, there exists $\gamma \in \Pi(\mu, \nu)$ such that, up to a subsequence (Bolzano-Weierstrass), $\gamma_\epsilon \rightharpoonup \gamma$ as $\epsilon \to 0$. We will show that $\gamma = \gamma_K$.
- The optimality of $\gamma_\epsilon$ w.r.t $c_\epsilon$ gives for all $\epsilon > 0$ that

$$\int c_\epsilon d\gamma_\epsilon \leq \int c_\epsilon d\gamma_K.$$

As a consequence of $\mu$ and $\nu$ having finite second moments, $\lambda_i(\epsilon) \int |x_i - y_i| d\gamma_\epsilon \to 0$ as $\epsilon \to 0$ for all $i \in \{1, \ldots, d-1\}$. Letting $\epsilon \to 0$ gives that:

$$\int |x_d - y_d|^2 \, d\gamma \leq \int |x_d - y_d|^2 \, d\gamma_K.$$

# Equality of $d$th marginal

- Since $|x_d - y_d|^2$ is a function of $(x_d, y_d)$, letting $\pi_d(x, y) = (x_d, y_d)$ we can rewrite the previous inequality as

$$\int |x_d - y_d|^2 \, d(\pi_d)_{\#}\gamma \leq \int |x_d - y_d|^2 \, d(\pi_d)_{\#}\gamma_K.$$

- However, by construction of the KR map $(\pi_d)_{\#}\gamma_K$ is the optimal map from $\mu_d$ to $\nu_d$ w.r.t. quadratic cost. By the uniqueness result of Brenier's Theorem, we must have that $(\pi_d)_{\#}\gamma = (\pi_d)_{\#}\gamma_K$.
- We denote this common measure $\gamma^d$.

# Equality of $(d-1, d)$th marginals

The goal know is the build off of equality of the $d$th marginals to demonstrate equality of $\gamma^{d-1}$ and $\gamma_K^{d-1}$, i.e. the marginals on $((x_{d-1}, x_d), (y_{d-1}, y_d))$. We will do this by showing for $\gamma^d$-a.e. $(x_d, y_d)$ that $\gamma^{d-1}_{(x_d,y_d)} = \gamma^{d-1}_{(x_d,y_d),K}$.
Now, for the quadratic cost $|x_d - y_d|^2$ we have that $(\pi_d)_\#\gamma_K$ is the optimal map. For $\epsilon > 0$ we then have

$$\begin{aligned}
\int |x_d - y_d|^2 \, d(\pi_d)_\#\gamma_K(x_d, y_d) + \sum_{i=1}^{d-1} \lambda_i(\epsilon) \int (x_i - y_i)^2 d\gamma_\epsilon \\
&\leq \int |x_d - y_d|^2 \, d(\pi_d)_\#\gamma_\epsilon(x_d, y_d) + \sum_{i=1}^{d-1} \lambda_i(\epsilon) \int (x_i - y_i)^2 d\gamma_\epsilon \\
&= \int c_\epsilon d\gamma_\epsilon \\
&\leq \int c_\epsilon d\gamma_K \\
&= \int |x_d - y_d|^2 \, d(\pi_d)_\#\gamma_K(x_d, y_d) + \sum_{i=1}^{d-1} \lambda_i(\epsilon) \int (x_i - y_i)^2 d\gamma_K .
\end{aligned}$$

# Equality of $(d-1,d)$ marginal

Eliminating common terms, dividing by $\lambda_{d-1}(\epsilon)$, and letting $\epsilon \to 0$ we get

$$\int |x_{d-1} - y_{d-1}|^2 \, d\gamma \leq \int |x_{d-1} - y_{d-1}|^2 \, d\gamma_K.$$

We let $\rho_{d-1}(x,y) = ((x_{d-1}, x_d), (y_{d-1}, y_d))$ and set $\gamma^{d-1} = (\rho_{d-1})_\# \gamma$, $\gamma_K^{d-1} = (\rho_{d-1})_\# \gamma_K$. Recalling that $\gamma^d = \gamma_K^d$, by disintegrating we can rewrite the above inequality as

$$\int \left( \int |x_{d-1} - y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d)}(x_{d-1}, y_{d-1}) \right) d\gamma^d(x_d, y_d) \tag{4}$$

$$\leq \int \left( \int |x_{d-1} - y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d),K}(x_{d-1}, y_{d-1}) \right) d\gamma^d(x_d, y_d) \tag{5}$$

The goal here is to invoke the optimality and uniqueness of the 1-d transport map $\gamma^{d-1}_{(x_d,y_d),K}$. However, to do so we must first verify that $\gamma^{d-1}_{(x_d,y_d),K}$ and $\gamma^{d-1}_{(x_d,y_d)}$ have the same marginals, which will show that they are couplings for the same source and target measures.

# Equality of $(d-1, d)$th marginal

In light of Lemma 1, showing that $\gamma^{d-1}_{(x_d,y_d),K}$ and $\gamma^{d-1}_{(x_d,y_d)}$ have same marginal on $x_{d-1}$ requires showing for all $\phi \in C_c(\mathbb{R})$ that

$$\int \phi(x_{d-1}) d\gamma^{d-1}_{(x_d,y_d)} = \int \phi(x_{d-1}) d\gamma^{d-1}_{(x_d,y_d),K}$$

as well as an analogous expression $\phi(y_{d-1})$ for the other marginal. It suffices to have this equality for $\gamma^d$-a.e. $(x_d, y_d)$, so by applying Lemma 1 again we will be satisfied if for all $\psi \in C_c(\mathbb{R} \times \mathbb{R})$ we have

$$\int \psi(x_d, y_d) \left( \int \phi(x_{d-1}) d\gamma^{d-1}_{(x_d,y_d)} \right) d\gamma^d = \int \psi(x_d, y_d) \left( \int \phi(x_{d-1}) d\gamma^{d-1}_{(x_d,y_d),K} \right) d\gamma^d.$$

Equivalently, this means

$$\int \psi(x_d, y_d)\phi(x_{d-1}) d\gamma^{d-1} = \int \psi(x_d, y_d)\phi(x_{d-1}) d\gamma^{d-1}_K.$$

But since $y_d = T_d(x_d)$ and $\gamma^{d-1}$ and $\gamma^{d-1}_K$ have same marginal on $(x_{d-1}, x_d)$, namely that $(d-1, d)$ marginal of $\mu$, we have the desired equality. For $y_{d-1}$, (H-target) gives that $T_d$ is invertible, so we make a similar argument.

# Equality of $(d-1,d)$th marginal

To summarize, we have for $\gamma^d$-a.e. $(x_d,y_d)$ that $\gamma^{d-1}_{(x_d,y_d)}$ and $\gamma^{d-1}_{(x_d,y_d),K}$ have the same marginals. By the optimality in the KR map construction, this gives for $\gamma^d$-a.e. $(x_d,y_d)$ that

$$\int |x_{d-1}-y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d),K}(x_{d-1},y_{d-1}) - \int |x_{d-1}-y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d)}(x_{d-1},y_{d-1}) \leq 0.$$

Integrating with respect to $\gamma^d$ gives

$$\int \left( \int |x_{d-1}-y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d),K} - \int |x_{d-1}-y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d)} \right) d\gamma^d \leq 0.$$

But (4) demonstrated that the LHS must be nonnegative. Hence, the integral is equal to 0. Since the integrand is nonnegative, we have for $\gamma^d$-a.e. $(x_d,y_d)$ that

$$\int |x_{d-1}-y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d),K}(x_{d-1},y_{d-1}) = \int |x_{d-1}-y_{d-1}|^2 \, d\gamma^{d-1}_{(x_d,y_d)}(x_{d-1},y_{d-1})$$

Thus, by the uniqueness of Brenier's map we have $\gamma^{d-1}_{(x_d,y_d),K} = \gamma^{d-1}_{(x_d,y_d)}$ for $\gamma^d$-a.e. $(x_d,y_d)$. This gives that $\gamma^{d-1} = \gamma^{d-1}_K$.

# General step

Assuming the optimality of $(\pi_k)_{\#}\gamma_K$ for $k \geq h$, we get the following inequalities

$$\begin{aligned}
&\sum_{k\geq h}\int \lambda_k(\epsilon)|x_k - y_k|^2\, d(\pi_k)_{\#}\gamma_K(x_k, y_k) + \sum_{k<h}\lambda_k(\epsilon)\int (x_k - y_k)^2 d\gamma_\epsilon(x_k, y_k)\\
&\quad\leq \sum_{k\geq h}\int |x_k - y_k|^2\, d(\pi_k)_{\#}\gamma_\epsilon(x_k, y_k) + \sum_{k<h}\lambda_k(\epsilon)\int (x_k - y_k)^2 d\gamma_\epsilon(x_k, y_k)\\
&\quad= \int c_\epsilon d\gamma_\epsilon \leq \int c_\epsilon d\gamma_K\\
&\quad= \sum_{k\geq h}\int |x_k - y_k|^2\, d(\pi_k)_{\#}\gamma_K(x_k, y_k) + \sum_{k<h}\lambda_k(\epsilon)\int (x_k - y_k)^2 d\gamma_K.
\end{aligned}$$

As before, by getting rid of the common terms, dividing by $\lambda_{h-1}(\epsilon)$, and passing to the limit gives

$$\int c^{(h-1)} d\gamma \leq \int c^{(h-1)} d\gamma_K\,.$$

# General step

Again, we disintegrate the marginals $\gamma^{h-1}$ and $\gamma_K^{h-1}$ in a sequence of conditionals until we finally have the integral with respect to $(x_{h-1}, y_{h-1})$ through the measure $\gamma^{h-1}_{(x_{d:h}, y_{d:h})}$ and $\gamma^{h-1}_{(x_{d:h}, y_{d:h}),K}$.
To prove that the two conditional measures have same marginal distribution along $x_{h-1}$ and $y_{h-1}$, we integrate them against test functions of the form

$$\psi(x_h, \ldots, x_d, y_h, \ldots, y_d)\phi(x_{h-1})$$

and

$$\psi(x_h, \ldots, x_d, y_h, \ldots, y_d)\phi(y_{h-1})\,.$$

To get the above expression in terms of $x_{h-1}, x_h, \ldots, x_d$ and $y_{h-1}, y_h, \ldots, y_d$ (resp), we replace the $(y_h, \ldots, y_d)$ variables by $(T_K(x_h), \ldots, T_K(x_d))$. In the other case, we invert the KR map, which is straightforward as each $y_k$ for $k \in \{h, \ldots, d\}$, is a 1-d monotone transformation of $x_k$ conditioned on $x_{k+1:d}$. As before, in the end we obtain that $\gamma^{h-1} = \gamma_K^{h-1}$.
Using this recursive routine, we ultimately prove that $\gamma = \gamma_K$.

# References

Carlier, G., Galichon, A., and Santambrogio, F. (2010). From Knothe's transport to Brenier's map and a continuation method for optimal transport. *SIAM Journal on Mathematical Analysis*, 41(6):2554–2576.

Knothe, H. (1957). Contributions to the theory of convex bodies. *Michigan Mathematical Journal*, 4(1):39–52.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

Villani, C. (2003). *Topics in optimal transportation*, volume 58. American Mathematical Soc.

Villani, C. (2009). *Optimal transport: old and new*, volume 338. Springer.