Eötvös Loránd University
Faculty of Science

Optimal transport and Arens-Eells spaces

Csaba Fodor

BSc Thesis in applied mathematics
Supervisor: Titkos Tamás (Rényi Institute)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Budapest, 2023

Table of Contents

Acknowledgements

I would like to extend my sincere gratitude to my supervisor, Tamás Titkos, for his guidance and feedback throughout the process of writing this thesis. I grateful to my teachers, who made it possible for me to get here. I am also thankful to my friends and family for their unwavering support.

CHAPTER 1

Introduction

The goal of this thesis is to investigate Arens–Eells spaces and to show that there is a connection to the optimal transport problem, by demonstrating that the Arens–Eells norm has a close connection to the 1-Wasserstein distance, a widely used norm in the theory of optimal transport.

In the first part we will discuss Arens–Eells spaces. The space of Lipschitz functions mapping from M to $\mathbb{R}$ is a vector space, but it is not a proper metric space since every constant function has a Lipschitz norm of 0. This problem can be rectified by choosing a point, and only considering functions that are 0 at this point, since this way only the constant 0 function will have a Lipschitz norm of 0. This way we can define a metric space of these Lipschitz functions, called $\mathrm{Lip}_0(M)$. A question that naturally arises is whether this space has a predual; we will see that it in fact does, this being the Arens–Eells space.

To illustrate one of the most important concepts concerning the theory of Arens-Eells spaces, the so-called molecule, let's consider $C[0,1]$ with the sup norm. We can define $\delta(t) : C[0,1] \to \mathbb{R}$ functions, called evaluation functionals, by

$$\delta(t)(f) = f(t) = \langle f, \delta(t) \rangle$$

for all $f \in C[0,1]$. This delta function is in the dual of $C[0,1]$, and its norm is 1. Moreover, the supremum norm can be written as

$$\|f\|_\infty = \sup\{|f(x)| : x \in [0,1]\} = \sup\{|\langle f, \delta(x)\rangle| : x \in [0,1]\}. \tag{1.1}$$

Now consider $\mathrm{Lip}_0(M)$, and let us define the so-called elementary molecules:

$$m_{xy} = \delta(x) - \delta(y) \qquad (x, y \in M).$$

Similarly to $C([0,1])$ case, m_{xy} is in the dual of $\mathrm{Lip}_0(M)$, and the norm of $\mathrm{Lip}_0(M)$ can be written as

$$
(1.2) \qquad \begin{aligned}
\|f\|_{\mathrm{Lip}} &= \sup \left\{ \frac{f(x) - f(y)}{\varrho(x,y)} : x, y \in [0,1], x \neq y \right\} \\
&= \sup \left\{ \frac{\langle f, m_{xy} \rangle}{\varrho(x,y)} : x, y \in [0,1], x \neq y \right\}.
\end{aligned}
$$

We can see the obvious analogy between (1.1) and (1.2). With the help of molecules, we will define the Arens–Eells spaces, and we will then present some examples of it, and we will prove some of its key properties.

In the second part we will give an overview of the problem of optimal transport and introduce the Wasserstein–space. The optimal transport problem originates from not long before the French Revolution, when Gaspard Monge published some results on moving soil in an optimal way. Since then it grew into an extensive field, investigating how one probability measure can be transformed into an other most efficiently, with respect to a predetermined cost function. It can be used anywhere from meteorology [**1**], [**2**] and economics [**3**], [**4**], [**5**] to image processing [**6**], [**7**], [**8**].

In the third part, we will investigate how the Arens–Eells norm is related to the Wasserstein distance.

CHAPTER 2

Arens–Eells spaces

2.1. Construction

In this chapter, we will construct the Arens–Eells space with the help of the previously introduced elementary molecules. What follows is mainly based on the 3rd chapter of [**9**].

Definition 2.1. A pointed metric space (M, ρ, e) is a metric space with a distinguished element $e \in M$, called the base point. We will see in section 2.3, that the choice of this base point e, doesn't change the structure of the Arens–Eells space.

Definition 2.2. Let (M, ρ) and $(\tilde{M}, \tilde{\rho})$ be two metric spaces, and $f : M \to \tilde{M}$ be a mapping between the two. The Lipschitz constant of f is defined by

$$\|f\|_{\mathrm{Lip}} = \sup \left\{ \frac{\tilde{\rho}(f(x), f(y))}{\rho(x, y)} : x \neq y \in M \right\}.$$

If this Lipschitz constant is finite, we call f a Lipschitz function.
The collection of real-valued $f : M \to \mathbb{R}$ Lipschitz functions is denoted by $\mathrm{Lip}(M)$.
Let (M, ρ, e) a pointed metric space. We denote the collection of real-valued Lipschitz functions that vanish at e by

$$\mathrm{Lip}_0(M) = \{ f \in \mathrm{Lip}(M) : f(e) = 0 \}.$$

One of the key notions of this thesis is the evaluation functional on $Lip_0(M)$.

Definition 2.3. For a point $x \in M$ and a function $f \in \mathrm{Lip}_0(M)$ we define the evaluation functional $\delta(x) : M \to \mathrm{Lip}_0(M)$ by $\delta(x) := f(x).$

We immediately remark that the evaluation at e vanishes so $\delta(e)$ is the zero functional.

Definition 2.4. We define the support of a real-valued $f : M \to \mathbb{R}$ function, denoted as $\operatorname{supp}(f)$ as the set of points, where it is non-zero. That is

$$\operatorname{supp}(f) = \{x \in M : f(x) \neq 0\}.$$

Definition 2.5. A molecule of a set M is defined to be a function $m : M \to \mathbb{R}$ such that its support is finite and $\sum_{x \in M} m(x) = 0$.
We call molecules in the form of $(x) - (y)$ elementary molecules and we denote them as m_{xy}.

If (M, ρ, e) is a pointed metric space, with base point e, then it is customary to write m_{ey} as m_y. We can always write m as a linear combination of these elementary molecules: $\sum_{x \in M} a_x m_x$, where $a_x = m(x)$.

We define the duality between a real-valued function f and a molecule m by

$$\langle f, m \rangle = \sum_{x \in M} f(x)m(x). \tag{2.1}$$

Definition 2.6. Let (M, ρ, e) a pointed metric space. We define the Arens-Eells norm of a molecule as

$$\|m\| = \inf \left\{ \sum_{i=1}^{n} |a_i| \rho(x_i, y_i) : m = \sum_{i=1}^{n} a_i m_{x_i y_i} \right\}.$$

In other words we write m as a combination of molecules in all possible ways and we take the one with the smallest possible "cost". We will expand on what exactly we mean by cost in chapter 4.
We call the completion of the space of molecules of M with this norm the Arens–Eells space over M, and we denote it by Æ(M). This space is also known as the Lipschitz-free space over M.

To prove that this is indeed a norm we need a lemma:

Lemma 2.7. *Let M be a metric space, m be a molecule, and let $f : M \to \mathbb{R}$ be a Lipschitz function. If m can be written in the form of $\sum_{i=1}^{n} a_i m_{x_i y_i}$, then $|\langle f, m \rangle| \leq \|f\|_{\text{Lip}} \sum_{i=1}^{n} |a_i| \rho(x_i, y_i)$.*

Proof. By definition, and the triangle inequality we have

$$\langle f, m\rangle = \langle f, \sum_{i=1}^{n} a_i m_{x_i y_i}\rangle = \sum_{i=1}^{n} a_i(f(x_i) - f(y_i)) \le \sum_{i=1}^{n} |a_i||f(x_i) - f(y_i)|.$$

Since the Lipschitz norm is defined by a supremum of the numbers

$$\frac{|f(x_i) - f(y_i)|}{\rho(x_i, y_i)},$$

we know that

$$\frac{|f(x_i) - f(y_i)|}{\rho(x_i, y_i)} \le \|f\|_{\mathrm{Lip}},$$

and thus multiplying this inequality by $\rho(x_i, y_i)$ and substituting $|f(x_i) - f(y_i)|$ by $\|f\|_{\mathrm{Lip}} \cdot \rho(x_i, y_i)$ we get

$$\langle f, m\rangle \le \sum_{i=1}^{n} |a_i||f(x_i) - f(y_i)| \le \|f\|_{\mathrm{Lip}} \sum |a_i|\rho(x_i, y_i).$$

Lemma 2.8. *Let $m = \sum_{x \in M} a_x m_x$ be a molecule. The expression defined by*

$$\|m\| = \inf\left\{ \sum_{i=1}^{n} |a_i|\rho(x_i, y_i) : m = \sum_{i=1}^{n} a_i m_{x_i y_i} \right\}$$

is a norm.

Proof. The infimum above is finite, since every molecule has at least one representation as a linear combination of elementary molecules. It is obvious that $\|m\| \ge 0$ for all m. Next we show that $\|m\| = 0$ if and only if $m = 0$. Assume that $m \neq 0$. According to Lemma 2.7 $\langle f, m\rangle \le \|f\|_{\mathrm{Lip}} \sum |a_i|\rho(x_i, y_i)$, for all $\|f\|_{\mathrm{Lip}} \ge 0$, so we only need to show that if $m \neq 0$ then $\exists f : \langle f, m\rangle \neq 0$. Since $m \neq 0$, there is at least one $\bar{x} \in M$ such that $m(\bar{x}) \neq 0$. Furthermore, the support m is finite, therefore there exists an $a > 0$ such that for all $x \in \mathrm{supp}(m), x \neq \bar{x} : \rho(x, \bar{x}) > a$. Now define f as $f(x) = \max\{a - \rho(x, \bar{x}), 0\}$. By the construction, for every x' from the support, $\rho(x', \bar{x}) > a$, therefore $f(x') = 0$. In other words, f peaks at $\bar{x}$, and vanishes on all other points of the support. For this f, we have $|\langle f, m\rangle| = f(\bar{x})m(\bar{x}) = am(\bar{x}) > 0$. Next we show that $\|\lambda m\| = |\lambda|\|m\|$ for all $\lambda \in \mathbb{R}$ and $m \in \text{Æ}(M)$. This holds since $\lambda m = \sum \lambda a_i m_{x_i y_i}$, if $m = \sum a_i m_{x_i y_i}$.

Thus

$$\begin{aligned}\|\lambda m\| &= \inf\left\{\sum_{i=1}^{n}|\lambda a_i|\varrho(x_i,y_i) : \lambda m = \sum_{i=1}^{n} a_i m_{x_i y_i}\right\}\\ &= \inf\left\{|\lambda|\sum_{i=1}^{n}|a_i|\varrho(x_i,y_i) : m = \sum_{i=1}^{n} a_i m_{x_i y_i}\right\} = |\lambda|\,\|m\|.\end{aligned}$$

To show the triangle inequality, let m and m' be two molecules, and consider all possible representations of m, m', and $m+m'$

$$A_m := \left\{\sum_{i=1}^{n} a_i m_{x_i,y_i} \,\middle|\, m = \sum_{i=1}^{n} a_i m_{x_i,y_i}\right\};\ A_{m'} := \left\{\sum_{i=1}^{n} a'_i m_{x'_i,y'_i} \,\middle|\, m = \sum_{i=1}^{n} a'_i m_{x'_i,y'_i}\right\}$$

$$A_{m+m'} := \left\{\sum_{i=1}^{n} a''_i m_{x''_i,y''_i} \,\middle|\, m+m' = \sum_{i=1}^{n} a''_i m_{x''_i,y''_i}\right\}$$

For any two representations from A_m and A'_m, their sum will be a representation of $m+m'$, but of course, there can be representations of $m+m'$ that cannot be written in this form. This means that

$$A_m + A_{m'} \subseteq A_{m+m'},$$

and therefore the infimum over $A_{m+m'}$ will be smaller than the sum of the infimums over A_m and $A_{m'}$, which implies $\|m+m'\| \leq \|m\| + \|m'\|$.

Another way of constructing the Arens–Eells space, as seen in [**10**], is to use the evaluation functionals $\delta(x)$. Then the Arens–Eells space, in this construction often denoted by $F(M)$, comes from the closed linear span of these evaluation functionals, a subspace of $\mathrm{Lip}_0(M)^*$:

$$F(M) = \overline{\mathrm{span}}\{\delta(x) : x \in M\} \subseteq \mathrm{Lip}_0(M)^*.$$

The norm then becomes the restriction of the norm on $\mathrm{Lip}_0(M)^*$ to $F(M)$.

2.2. Examples

In this section, we will present a few examples of Arens–Eells spaces, based on [**9**] and [**10**].

Example 2.9. Let $M = \{0, 1\}$, $\varrho(0, 1) = 1$, and the base point be 0. Then $\mathrm{Lip}_0(M) = \{ f : \|f\|_{\mathrm{Lip}} < \infty; f(0) = 0 \} = \mathbb{R}$, because

$$\frac{f(1) - f(0)}{\varrho(1, 0)} = f(1) < \infty, \text{ for all } f : M \to \mathbb{R};$$

and $Æ(M) = \overline{\mathrm{span}}\{\delta(0), \delta(1)\} = \overline{\mathrm{span}}\{\delta(1)\} = \mathbb{R}$.

Example 2.10. Let $M = [0; 1]$, $\varrho(x, y) = |x - y|$, and the base point be 0. We will show that $\mathrm{Lip}_0(M) = L_\infty[0; 1]$ and that the maps $T_1 : L_\infty([0; 1]) \to \mathrm{Lip}_0([0; 1])$ and $T_2 : \mathrm{Lip}_0([0; 1]) \to L_\infty([0; 1])$ are non-expansive, such that they are each other's inverses. This will imply that both are isometries, as the inverse of a non-expansive function is non-contractive, and if a function is both non-expansive and non-contractive, then it is an isometry.

Let $T_1 : L_\infty([0; 1]) \to \mathrm{Lip}_0([0; 1])$, where for all $f \in L_\infty([0; 1])$

$$(T_1 f)(x) = \int_0^x f(t)\, dt.$$

Then for all $x < y \in [0; 1]$

$$(T_1 f)(y) - (T_1 f)(x) = \int_x^y f(t)\, dt \le \|f\|_\infty (y - x).$$

Dividing by $y - x$, we have $\|T_1 f\|_{\mathrm{Lip}} \le \|f\|_\infty$, thus $(T_1 f) \in \mathrm{Lip}_0([0; 1])$.
Let $T_2 : \mathrm{Lip}_0([0; 1]) \to L_\infty([0; 1])$; $(T_2 f) = f'$. Since f is a Lipschitz function, we can use the fundamental theorem of calculus, so we have

$$f(y) - f(x) = \int_x^y f'(t)\, dt;$$

where f' exists almost everywhere and $x < y \in [0; 1]$. Based on the definition of the derivative and the Lipschitz norm, if $x < y \in [0; 1]$ and f' exists, we have

$$|f'(x)| = \lim_{y \to x} \frac{f(y) - f(x)}{y - x} = \lim_{y \to x} \left| \frac{f(y) - f(x)}{y - x} \right| \le \|f\|_{\mathrm{Lip}}.$$

So $\|T_2 f\|_\infty \le \|f\|_{\mathrm{Lip}}$ and $T_2 f \in L_\infty([0; 1])$.
We also have

$$\forall x : (T_2 T_1 f)(x) = \left(\int_0^x f(t)\, dt \right)' = f(x) \text{ and}$$

$$\forall x : (T_1T_2 f)(x) = \int_0^x f'(t)\,dt = f(x);$$

thus T_1 and T_2 are each others inverses. So we conclude that $\mathrm{Lip}_0(M)$ and $L_\infty[0,1]$ are isometrically isomorphic. Now we use that $Æ(M)^* = \mathrm{Lip}_0(M)$ (we will prove this in the next section), and that $L_1([0,1])$ is the unique predual of $L_\infty[0,1]$ therefore $Æ([0,1]) = L_1([0,1])$.

We will mention a few more examples without much explanation. First, we mention that $Æ(\mathbb{N}) = \ell^1(\mathbb{N}) = Æ(\mathbb{Z})$. For the second example, let $M = ([0,1], |\cdot|)$, and define M^e as the metric space, where we add the point e to $[0,1]$, and define its distance from every other point as 1. Then $Æ(M^e) = \ell^1(M)$. Last, if M is a closed subset of $\mathbb{R}$, with a measure of 0, then $Æ(M) = \ell^1(M)$.

2.3. Duality

In this section we will present a proof from [9] which shows that the dual of the Arens–Eells space is linearly isometrically isomorphic to the $\mathrm{Lip}_0(M)$ space, meaning there exists a bijective linear map from one to the other, that preserves norms.

Theorem 2.11. *Let (M, ϱ, e) be a pointed metric space. Then $Æ(M)^* = \mathrm{Lip}_0(M)$.*

Proof. To prove that $Æ(M)^*$ and $\mathrm{Lip}_0(M)$ are isometrically isomorphic, we define two maps $T_1 : Æ(M)^* \to \mathrm{Lip}_0(M)$ and $T_2 : \mathrm{Lip}_0(M) \to Æ(M)^*$ which are both non-expansive, and they are eachother's inverses. This will imply that T_1 and T_2 are isometries.

Let φ be an element of $Æ(M)^*$ and define the map $T_1 : Æ(M)^* \to \mathrm{Lip}_0(M)$ by

$$(T_1\varphi)(x) = \varphi(m_x):$$

What we get, is an element of $\mathrm{Lip}_0(M)$. Indeed

$$|(T_1\varphi)(x) - (T_1\varphi)(y)| = |\varphi(m_x) - \varphi(m_y)| = |\varphi(m_{xy})| \le \|\varphi\|\|m_{xy}\| \le \|\varphi\|\varrho(x,y);$$

therefore

$$\frac{|(T_1\varphi)(x) - (T_1\varphi)(y)|}{\varrho(x,y)} \le \|\varphi\|$$

so we get that $T_1\varphi$ is a Lipschitz function, and that

$$\|T_1\varphi\|_{Lip} \le \|\varphi\|.$$

Moreover $T_1\varphi$ vanishes on e, and thus $T_1\varphi \in Lip_0(M)$:

$$(T_1\varphi)(e) = \varphi(m_e) = \varphi(\delta(e)) = 0.$$

Now define $T_2 : Lip_0(M) \to Æ(M)^*$. For any $f \in Lip_0(M)$ we need to show that $T_2 f$ is a bounded linear functional on $Æ(M)$. We define $T_2 f$ on a dense subspace by

$$(T_2 f)(m) = \langle f; m\rangle.$$

Since the set of molecules is dense, it is enough to show that $T_2 f$ is bounded on the set of molecules. Based on Lemma 2.7:

$$|(T_2 f)(m)| = |\langle f; m\rangle| \le \|f\|_{Lip}\,\|m\|_{Æ}; \tag{2.2}$$

so $\|T_2 f\| = \sup \frac{|(T_2 f)(m)|}{\|m\|} \le \|f\|_{Lip}$, i.e. T_2 is a non-expansive map.

Next we show that $(T_2)^{-1} = T_1$, i.e., $T_2 T_1\varphi = \varphi$ for all $\varphi \in Æ(M)^*$. For all $p \in M$, we have

$$(T_2 T_1\varphi)(m_p) = \langle T_1\varphi; m_p\rangle = (T_1\varphi)(p) = \varphi(m_p).$$

Let $f \in Lip_0(M)$ and $x \in M$, then

$$(T_1 T_2 f)(x) = (T_2 f)(x) = \langle f; m_x\rangle = \sum_{y \in M} f(y)m_x(y) = f(x).$$

We have seen that $T_2T_1\varphi = \varphi$ and $T_1T_2 f = f$, therefore T_1 and T_2 are each other's inverses.

An important consequence of the theorem is that the definition of the Arens–Eells space doesn't depend on the choice of base point.

2.4. Universal extension property

The property says that any Lipschitz map between a pointed metric space M and a Banach–space V can be replaced by a combination of a map from M to

$Æ(M)$ and a unique linear map from $Æ(M)$ to V. To prove this we once again rely on [9]. First we need a lemma:

Lemma 2.12. For any metric space M, the equality $\|m_{xy}\|_{Æ} = \rho(x, y)$ holds. Moreover for all $x, y \in M$ $\|\cdot\|_{Æ}$ is the largest of all seminorms $\|\cdot\|_0$ on the space of molecules, which satisfy $\|m_{xy}\|_0 \leq \rho(x, y)$ for all $x, y \in M$.

Proof. We know from the definition that $\|m_{xy}\|_{Æ} \leq \rho(x, y)$. To see the converse inequality, for a fixed $y \in M$ let us define the function $\rho_y(x) = \rho(x, y)$. It is a Lipschitz function and its Lipschitz norm is 1, and therefore, according to (2.2) we have

$$\|m_{x,y}\|_{Æ} \geq 1 \cdot |\langle \rho_y(x), m_{x,y}\rangle| = \rho(x, y).$$

Let $\|\cdot\|_0$ be a seminorm on the space molecules, such that for each pair of points $x, y \in M$ we have $\|m_{xy}\|_0 \leq \rho(x, y)$. Using the triangle-inequality first, then the assumption, we have

$$\|m\|_0 = \left\| \sum a_i m_{x_i y_i} \right\|_0 \leq \sum |a_i| \|m_{x_i y_i}\|_0 \leq \sum |a_i| \rho(x_i, y_i);$$

for all $m = \sum a_i m_{x_i y_i}$. Since the Arens–Eells norm is defined by the infimum of the right-hand side, taking the infimum of this expression we get

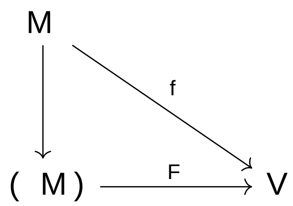
$$\|m\|_0 \leq \|m\|_{Æ}.$$

With this lemma we can now prove the universal extension property:

Theorem 2.13. Let M be a pointed metric space, V be a Banach-space, and $f : M \to V$ be a Lipschitz-map sending the base point to the zero element of V. Then the map $\iota : x \mapsto m_x$ isometrically embeds M into $Æ(M)$, moreover there exists a unique bounded linear map $F : Æ(M) \to V$, such that

$$F \circ \iota = f; \quad \text{and} \quad \|F\| = \|f\|_{Lip}.$$

The theorem can be visualized with the following commutative diagram:



Proof. First we have

$$\|\delta(x) - \delta(y)\| = \|m_x - m_y\| = \|m_{xy}\| = \rho(x;y)$$

for all $x;y \in M$ as demonstrated by Lemma 2.12, proving this is indeed an isometric embedding.

To prove the universality property, define $F : Æ(M) \to V$ by

$$F(m) = \sum a_x f(x);$$

where $m = \sum a_x m_x$. First, we need to prove that F is bounded, that is $\exists c > 0$ for $\forall m \in Æ(M)$ such that $\|Fm\| \le c\|m\|_{Æ(M)}$. Let $\|\cdot\|_0 = \frac{1}{\|f\|_{Lip}}\|F\cdot\|$. For molecules in form of $\delta(x) - \delta(y)$ we have

$$\begin{aligned}\|m_{xy}\|_0 &= \frac{1}{\|f\|_{Lip}}\|F(m_{xy})\| \\ &= \frac{1}{\|f\|_{Lip}}\|F(\delta(x) - \delta(y))\| = \frac{1}{\|f\|_{Lip}}\|f(x) - f(y)\| \le \rho(x;y)\end{aligned}$$

Thus by Lemma 2.12 we have

$$\|m\|_0 = \frac{1}{\|f\|_{Lip}}\|Fm\| \le \|m\| :$$

Multiplying by $\|f\|_{Lip}$ we have shown F is bounded and $\|F\| \le \|f\|_{Lip}$.

If we show the reverse inequality too, we get $\|F\| = \|f\|_{Lip}$.

$$\|F\| = \sup_{m \neq \tilde{m}} \left\{ \frac{\|F(m) - F(\tilde{m})\|_V}{\|m - \tilde{m}\|} \right\}$$

$$\geq \sup_{\substack{\delta(x) = m \\ \delta(y) = \tilde{m}}} \left\{ \frac{\|F(\delta(x)) - F(\delta(y))\|_V}{\|\delta(x) - \delta(y)\|} \right\}$$

$$= \sup_{x \neq y} \left\{ \frac{\|F(\delta(x)) - F(\delta(y))\|_V}{\varrho(x; y)} \right\}$$

$$= \sup_{x \neq y} \left\{ \frac{\|f(x) - f(y)\|_V}{\varrho(x; y)} \right\} = \|f\|_{Lip}:$$

CHAPTER 3

p-Wasserstein–spaces

In this chapter we introduce a special metric space, the so-called Wasserstein space, which is closely related to the theory of optimal transportation. First, let us introduce some important notions we will use in the following chapter.

Definition 3.1. Let (M, ϱ) be a metric space. Denote the sigma-algebra generated by the open sets in M by $\mathcal{B}$. A set-function $\;: \mathcal{B} \to \mathbb{R}$ is called a measure if it is σ-additive. We say that is finite if (M) is finite. If $(M) = 1$ and ≥ 0, we call it a probability measure. We denote the set of probability measures over M by $\mathcal{P}(M)$.

Definition 3.2. We say that a measure is finitely supported on M, if there exists a finite set $\{x_1, x_2, \ldots, x_n\} \subseteq M$, such that $= \sum_{i=1}^{n} {}_i \, {}_{x_i}$, where $\sum_{i=1}^{n} {}_i = 1$ and ${}_i > 0$ for all $i \in \{0, 1, \ldots, n\}$.

The set $\{x_1, x_2, \ldots, x_n\}$ is called the support of and is denoted by supp().

The following notion will play a key role in the theory of optimal transport.

Definition 3.3. Let $, \in \mathcal{P}$. We call $\in \mathcal{P}(M \times M)$ is a coupling for and if, for all $A, B \in \mathcal{B}$, if their first marginal is , its second marginal is , that is,

(1) $(A \times M) = (A)$ and
(2) $(M \times B) = (B)$:

for all $A, B \in \mathcal{B}$. The set of couplings will be denoted by $\mathcal{C}(,)$.

Definition 3.4. Let be a measure and T be a measurable map. The measure called the push-forward of by T, denoted by $T_{\#}$, is defined by

$$(T_{\#} \;)(A) = \;(T^{-1}(A))$$

for every $A \in \mathcal{B}$.

Dirac measures will play a key role in this section, so we recall that if $x \in M$ then a Dirac measure concentrated to x is denoted by δ_x and is defined by

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \qquad \text{for all } A \in \mathcal{B}.$$

3.1. A short introduction to Optimal transport

The Optimal Transport problem was first introduced by Gaspard Monge, when he proposed some results on moving soil from one point to another in an optimal way. Later, Leonid Kantorovich also studied the problem, motivated by the issue of distributing resources optimally.

Gaspard Monge was a French mathematician, he worked on many different things, ranging from differential geometry [11] and descriptive geometry [12] to cannon making [13].

Figure 1. Gaspard Monge, photo source: Wikipedia

In 1781, he proposed some results on moving soil from one location to another in an optimal way, in both $\mathbb{R}^2$ and $\mathbb{R}^3$ [14]. Rather than Monge's original formulation of the problem, we will use a more modern one as introduced in the first chapter of [15].

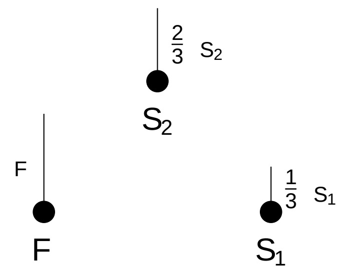
We are given two probability measures: $\mu \in P(M)$ and $\nu \in P(N)$ and a cost function $c : M \times N \to \mathbb{R}^+$ and we need to find a map T such that the transport cost $\int c(x; T(x))\, dx$ is minimal, i.e., we want to find a map T which realizes the minimum of the set

$$\left\{ \int c(x; T(x))\, dx : T_{\#} = ; \right\}$$

where $T_{\#}$ is the push-forward of . In other words, we are looking for the cheapest way to transform into .

As we will see, the problem is ill-posed, as such T a might not exist, and even if it exists, non-uniqueness can occur. First we show an example where the transport problem cannot be solved.

Example 3.5. Consider a factory producing cars, and two stores selling them, one requiring $\frac{1}{3}$ of the factory's production, the other $\frac{2}{3}$. This can be modelled by setting $= {}_F$ and $= \frac{1}{3}\,{}_{S_1} + \frac{2}{3}\,{}_{S_2}$.



We then need to transform into . This isn't possible since for any possible solution (i.e., transport map) $T : M \; ! \; M$, we have:

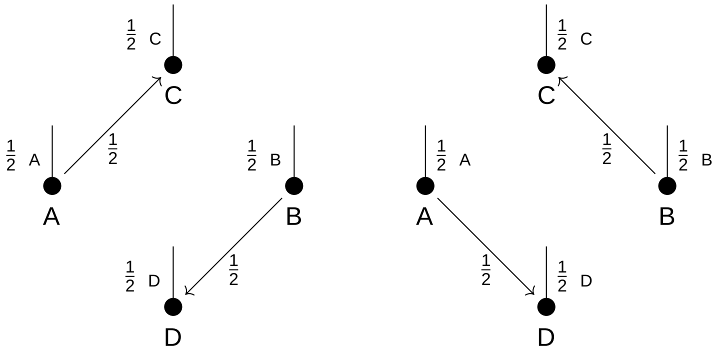
$$T_{\#}(\;_x)(A) = \;_x(T^{\;1}(A)) = \begin{cases} 1, & \text{if } x \,2\, T^{\;1}(A) \\ 0, & \text{if } x \,2\!\!\!/\, T^{\;1}(A) \end{cases}$$

$$= \begin{cases} 1, & \text{if } T(x) \,2\, (A) \\ 0, & \text{if } T(x) \,2\!\!\!/\, (A) \end{cases} = \;_{T(x)}(A):$$

This means that $T_{\#}(\;_x) = \;_{T(x)}$, so the push-forward of a Dirac measure is a Dirac measure, but is not a Dirac measure.

Next we present an elementary example where the optimal solution is not unique.

Example 3.6. Condider 4 points, $A; B; C; D$ arranged in a square. Let $= \frac{1}{2}\,{}_A + \frac{1}{2}\,{}_B$ and $= \frac{1}{2}\,{}_C + \frac{1}{2}\,{}_D$. To transform into , for the rst map we have $T_1(A) = C$ and $T_1(B) = D$ and for the second we have $T_2(A) = D$ and $T_2(B) = C$.



Here we can see that it doesn't matter which plan we choose, the cost will be the same for both maps, as the distance along the edges of the square are equal.

The solutions to the problem of optimal transport can greatly differ for different choices of the cost function. We will illustrate this through the next example.

Example 3.7. We have n books displayed one after the other on a long bookshelf, starting from the left. We buy a new book and want to display it on the very left. We will have to move some books in order to liberate that place, so we formulate a few T_i transport maps. In the first one, we move the book on the left to the very end, after the last book, so $T_1(1) = n+1$ and $T_1(k) = k$ for the rest. In the second, we move every book to the left by one place, so $T_2(k) = k+1$. In the third, we move the fist book to the place of the third one, and move the third one to the end of the row, so $T_3(1) = 3; T_3(3) = n+1$ and $T_3(k) = k$ for the rest. We can model this by using weighted Dirac-measures. In the beginning we have n books in the first n places, we can write this as a sum of n weighted Dirac-measures $\mu = \sum_{i=1}^{n} \frac{1}{n}\delta_i$, after moving the book we will have n books in the first $n+1$ place, except the first, so we will have $\nu = \sum_{j=2}^{n+1} \frac{1}{n}\delta_j$.

Depending on the cost function, it will matter which map we choose.

First let the cost function $c_1(x;y) = \sqrt{|y-x|}$.

In this case the cost of the first transport map is

$$C_{c_1}(T_1) = \int c_1(x;T_1(x))\,d\mu(x) = \frac{1}{n}\,c_1(x;T_1(x)) = \frac{1}{n}\sqrt{(n+1)-1} = \frac{\sqrt{n}}{n}:$$

For the second transport map, we have:

$$C_{c_1}(T_2) = \int c_1(x;T_2(x))\,d\mu(x) = \int \sqrt{(x+1)-x}\,d\mu(x) = \int 1\,d\mu(x) = \sum_{i=1}^{n+1} \frac{1}{k} = 1:$$

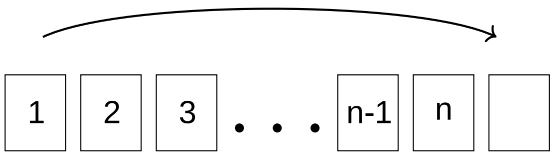


Figure 2. First map

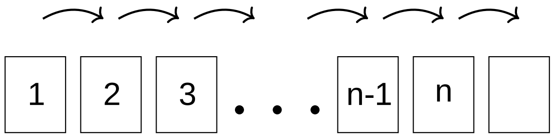


Figure 3. Second map

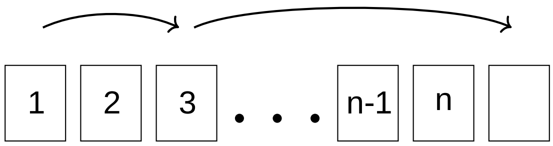


Figure 4. Third map

And for the third transport map:

$$C_{c_1}(T_3) = \int c_1(x;T_3(x))\,d\mu(x) = \frac{1}{n}\int c_1(x;T_3(x))\,d\mu_1(x) + \frac{1}{n}\int c_1(x;T_3(x))\,d\mu_3(x)$$
$$= \frac{1}{n}\sqrt{3-1} + \frac{1}{n}\sqrt{(n+1)-3} = \frac{\sqrt{2}}{n} + \frac{\sqrt{n-2}}{n} > \frac{\sqrt{n}}{n} \text{ for } n \geq 3.$$

So in this case the first map is the best out of the three. It can be shown that if the cost function is concave, then a transport with fewer but longer movements is cheaper than a transport with many small movements.

Next let the cost function be $c(x;y) = |y - x|$. With similar calculations we get:

$$C_{c_2}(T_1) = \int c_2(x;T_1(x))\,d\mu(x) = \frac{1}{n}|(n+1) - 1| = 1;$$
$$C_{c_2}(T_2) = \int c_2(x;T_2(x))\,d\mu(x) = \int |(x+1) - x|\,d\mu(x) = \sum_{i=1}^{n}\frac{1}{n} = 1;$$
$$C_{c_2}(T_3) = \int c_2(x;T_3(x))\,d\mu(x) = \frac{1}{n}|3 - 1| + \frac{1}{n}|(n+1) - 3| = 1.$$

So in this case it doesn't matter which map we choose. This will be analogous to the 1-Wasserstein distance we will introduce in the next section.

Last let the cost function be $c(x; y) = |y - x|^2$. Then we get:

$$C_{c_3}(T_1) = \int c_3(x; T_1(x))\, d\mu(x) = \frac{1}{n} |(n+1) - 1|^2 = n;$$

$$C_{c_3}(T_2) = \int c_3(x; T_2(x))\, d\mu(x) = \sum_{i=1}^{n} \frac{1}{n} |(k+1) - k|^2 = 1;$$

$$C_{c_3}(T_3) = \int c_3(x; T_3(x))\, d\mu(x) = \frac{1}{n} |3 - 1|^2 + \frac{1}{n} |(n+1) - 3|^2$$

$$= \frac{2^2 + (n-2)^2}{n}.$$

In this case, the second map is the best. Again, it can be shown that in the case of convex cost functions, lots of small movements tend to be better than fewer but longer movements.

A long time passed without significant development in the field. Then came Leonid Kantorovich during the Second World War and revolutionized the theory of optimal transport.

Kantorovich was a Russian mathematician and economist. Shortly before the Second World War, he laid the foundations of linear programming in the publication [16], to find a solution to distributing resources to maximize output. The war would interrupt his studies in this field, as he focused on tackling challenges for the Red Army, such as during the siege of Leningrad, calculating how far each vehicle had to be, to be able to safely transport supplies to the city over the frozen lake. This effort was called the Road of Life [17]. He was still able to work on economics-related problems, and after the war, he would become an influential economist [18]. In 1975 Leonid Kantorovich and Tjalling C. Koopmans received a Nobel-prize in economics "for their contributions on the theory of optimum allocation of resources" [19].

Figure 5. Leonid Kantorovich, photo source: Nobel Foundation archive

He generalized the problem by getting rid of the mass splitting obstacle, introducing the concept of coupling to the problem [20]. In this context, they are called transport plans. With these, the Kantorovich formulation of the Optimal Transport problem is

$$\inf \left\{ \int_{M_1 \quad M_2} c\,d \ : \quad 2\,C(\ ; \ ) \right\} :$$

We can look at $(A \quad B)$ as the mass moving from A to B, thus eliminating the constraint of the Monge formulation about the discrete particles. It can be shown that if the Kantorovich formulation has convex constraints, then we can calculate this in mum by solving for the dual problem (in the operations research sense).

Example 3.8. Going back to the previous example, we can solve the allocation of car production in the Kantorovich formulation. We can write down the coupling between $\ = 1 \quad _F$ and $\ = \frac{1}{3} \ _{S_1} + \frac{2}{3} \ _{S_2}$ with a table, here is de ned as $(F; S_1) = \frac{1}{3}; \ (F; S_2) = \frac{2}{3}$ and $(\ ; \ ) = 0$ elsewhere:

	F	S_1	S_2
F	0	$\frac{1}{3}$	$\frac{2}{3}$
S_1	0	0	0
S_2	0	0	0

This time the in mum only has one element, so it is easy to calculate:

$$\int c(x; y)\, d\,(x; y) = c(F; S_1) \ (F; S_1) + c(F; S_2) \ (F; S_2) = \frac{1}{3} c(F; S_1) + \frac{2}{3} c(F; S_2):$$

3.2. Definition of p-Wasserstein–spaces

With Kantorovich's formulation, i.e., with couplings, we can define the so-called p-Wasserstein–space, where the role of the cost function is played by the p-th power of the distance.

Definition 3.9. Let $(M;\varrho)$ be a complete, separable metric space, and $p \geq 1$ a real number. Then we denote the set of probability measures with a finite p-th moment by $P_p(M)$. That is,

$$P_p(M) = \left\{ \mu \in P(M) : \exists x_0 \in M \text{ such that } \int_M \varrho(x;x_0)^p \, d\mu(x) < \infty \right\}.$$

On the set $P_p(M)$ we can introduce the so-called p-Wasserstein distance as follows.

Definition 3.10. For $p \geq 1$, the p-Wasserstein distance on $P_p(M)$ is defined by

$$d_{W_p}(\mu;\nu) = \left(\inf_{\pi \in \Pi(\mu;\nu)} \int_{M \times M} \varrho(x;y)^p \, d\pi(x;y) \right)^{\frac{1}{p}}.$$

The set call $P_p(M)$ endowed with the distance d_{W_p} is called the p-Wasserstein–space.

It is known that the Wasserstein distance is a metric on $P_p(M)$ (for the details see chapter 6 of [21]), that is

(1) $d_{W_p}(\mu;\nu) = 0 \Leftrightarrow \mu = \nu$
(2) $d_{W_p}(\mu;\nu) = d_{W_p}(\nu;\mu)$
(3) $\forall \mu;\nu;\eta : d_{W_p}(\mu;\nu) \leq d_{W_p}(\mu;\eta) + d_{W_p}(\eta;\nu)$.

Now we collect a few important observations about the basic properties of the Wasserstein–space:

- If we take the optimal transport problem, with cost function $c(x;y) = \varrho(x;y)^p$, the Wasserstein distance becomes the cost of the optimal transport plan in the Kantorovich formulation.
- If one of μ or ν is the Dirac measure, then the set of transport plans has only one element, the product measure. In this case it is easy to calculate

the Wasserstein distance:

$$d_{W_p}(\mu;\delta_{x_0}) = \left(\int_{M\times M} \varrho(x;y)^p \, d(\mu\otimes\delta_{x_0})(x;y)\right)^{\frac{1}{p}}$$
$$= \left(\int \varrho(x;x_0)^p \, d\mu(x)\right)^{\frac{1}{p}}.$$

If both μ and ν are the Dirac measures, then their Wasserstein distance is just the distance of their supporting points in the underlying space:

$$d_{W_p}(\delta_x;\delta_{x_0}) = \varrho(x;x_0).$$

It can be shown, that every measure in the p-Wasserstein–space can be approximated by a convex combination of Dirac measures. That is, for all $\mu \in W_p(M)$ and for all $\varepsilon > 0$ there exists a probability measure $\nu = \sum_{i=1}^{n} \lambda_i \delta_{x_i}$ such that $d_{W_p}(\mu;\nu) < \varepsilon$.

If $\mu = \sum_{i=1}^{k} \lambda_i \delta_{x_i}$, $\nu = \sum_{j=1}^{l} \tilde{\lambda}_j \delta_{y_j}$ have a finite support, then their optimal coupling π has also a finite support. The Wasserstein distance then becomes

$$d_{W_p}(\mu;\nu) = \left(\int \varrho(x;y)^p \, d\pi(x;y)\right)^{\frac{1}{p}} = \left(\sum\sum \varrho(x;y)^p \, \pi(x;y)\right)^{\frac{1}{p}}.$$

CHAPTER 4

The 1-Wasserstein distance and the Arens-Eells norm

In this chapter, we will talk about how optimal transport comes into the picture with the Arens–Eells spaces, and we will see how the 1-Wasserstein distance and the Arens–Eells norm are related. We will rely on chapter 3 of [9].

Let $\mu, \nu \in W_1(M)$ be two finitely supported measures, and consider $\mu - \nu$. Then $(\mu - \nu)(M) = \mu(M) - \nu(M) = 1 - 1 = 0$, and it has a finite support. So we can analyse it through the Arens–Eells space, by treating it as a molecule. We will see that $\|\mu - \nu\|_{Æ} = d_{W_1}(\mu, \nu)$.

To find the analogue to the concept of coupling, let's think of molecules in the following way: let m be a molecule, and let us think of $x \in M$ as either a factory or a store depending on whether $m(x)$ is greater or smaller than 0. We have some factories and stores with a specified amount of production and demand, now we need to get the goods from the factories to the stores. Define $h : M^2 \to \mathbb{R}^+$ such that

$$m = \sum_{x,y \in M} h(x,y) m_{xy}.$$

This is equivalent to saying that for all $x \in M$ we have

$$m(x) = \sum_{y \in M} (h(x,y) - h(y,x)).$$

We will call this h function the transport plan, and the number $h(x,y)$ represents the amount of goods that need to be transported from x to y. We define the cost of the transport plan as $\sum h(x,y)\rho(x,y)$. Thus the minimal cost of the transport plan becomes the Arens-Eells norm.

We will see that in this case there always exists an optimal transport plan. We saw in example 3.7, that the optimum may not be unique, but we can choose an optimal one where there is no point we both ship to and ship from.

Theorem 4.1. *For every molecule m on a metric space M, there exists a nitely supported function $\hat{h}$, such that $\hat{h} : M^2 \to \mathbb{R}^+$ and*

(1) $m = \sum_{xy} \hat{h}(x,y) m_{xy}$
(2) $\|\hat{h} \varrho\|_1 = \|m\|$
(3) $\hat{h}(x,y) > 0 \Rightarrow m(x) > 0$ *and* $m(y) < 0$
(4) $\|\hat{h}\|_1 = \frac{1}{2}\|m\|_1$

Proof. First we prove (1) and (2). Let A be the support of m. Then identifying a function $h : A^2 \to \mathbb{R}^+$ with a point in $\mathbb{R}^{|A|^2}$, we can see that the set

$$C = \{h : m = \sum_{x,y} h(x,y) m_{xy}\}.$$

is nonempty and closed. It is nonempty because every molecule has at least one representation, and it is closed because if we take a sequence $h_i \in C$, meaning for each h_i : $m = \sum_{x,y} h_i(x,y) m_{xy}$, then their limit h must also satisfy $m = \sum_{x,y} h(x,y) m_{xy}$.
Since $\sum_{x,y} |h(x,y)| \varrho(x,y)$ is continuous in h and tends to infinity as $h \to \infty$, and C is nonempty and closed, it follows that it attains its minimum on C, i. e. there exists a h such that

$$\sum_{x,y} |h(x,y)| \varrho(x,y) = \|h \varrho\|_1 = \|m\| .$$

It can be shown, that the norm on Æ(A) is equal to the norm on Æ(M) (as seen in Theorem 3.7 in [**9**]), thus replacing the minimising function h, with $\tilde{h} = h^+(x,y) + h(x,y)$, we get a $A^2 \to \mathbb{R}^+$ function satisfying (1) and (2).

Now we'll prove (3) and (4). Consider the set of those positive functions that satisfy (1) and (2). Since this set is closed, there is an element h_0 with minimal ^{1} norm. Then for all $x \in M$ we have $h_0(x,x) = 0$, because if $h_0(x,x) > 0$ for some $x \in M$, then changing it to $h_0(x,x) = 0$ would still satisfy (1) and (2), but the l^1 norm would be lowered. So we have $h_0(x,x)$ for all x. Now assume that we have three points $p,q,r \in M$ such that $h(p,q) > 0$ and $h(q,r) > 0$. Then we can set h_1 as

$$
\begin{aligned}
a &:= \min\{h_0(p,q), h_0(q,r)\} \\
h_1(p,q) &= h_0(p,q) - a, \\
h_1(q,r) &= h_0(q,r) - a, \\
h_1(p,r) &= h_0(p,r) + a
\end{aligned}
$$

and $h_0 = h_1$ elsewhere. Then one can show that $\|h_1 - \varrho\|_1 \le \|h_0 - \varrho\|_1$. Indeed,

$$
\begin{aligned}
\|h_1 - \varrho\|_1 &= \sum_{x,y} |h_1(x,y) - \varrho(x.y)| \\
&= \sum_{\substack{(x,y)\notin(p,q)\\(x,y)\notin(q,r)\\(x,y)\notin(r,p)}} |h_1(x,y) - \varrho(x,y)| + |h_1(p,q) - \varrho(p,q)| \\
&\quad + |h_1(q,r) - \varrho(q,r)| + |h_1(p,r) - \varrho(p,r)| \\
&= \sum_{\substack{(x,y)\notin(p,q)\\(x,y)\notin(q,r)\\(x,y)\notin(r,p)}} |h_0(x,y) - \varrho(x,y)| + |(h_0(p,q) - a) - \varrho(p,q)| \\
&\quad + |(h_0(q,r) - a) - \varrho(q,r)| + |(h_0(p,r) - a) - \varrho(p,r)| \\
&= \sum_{\substack{(x,y)\notin(p,q)\\(x,y)\notin(q,r)\\(x,y)\notin(r,p)}} h_0(x,y) - \varrho(x,y) + (h_0(p,q) - a) - \varrho(p,q) \\
&\quad + (h_0(q,r) - a) - \varrho(q,r) + (h_0(p,r) + a) - \varrho(p,r) \\
&= \sum_{\substack{(x,y)\notin(p,q)\\(x,y)\notin(q,r)\\(x,y)\notin(r,p)}} h_0(x,y) - \varrho(x,y) + h_0(p,q) - \varrho(x,y) \\
&\quad + h_0(q,r) - \varrho(q,r) + h_0(p,r) - \varrho(p,r) \\
&\quad - a \cdot (\varrho(p,q) + \varrho(q,r) - \varrho(p,r))
\end{aligned}
$$

As per the triangle inequality, we have:

$$
\begin{aligned}
&\sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} h_0(x,y)\cdot\rho(x,y) + h_0(p,q)\cdot\rho(x,y)\\
&\quad + h_0(q,r)\cdot\rho(q,r) + h_0(p,r)\cdot\rho(p,r)\\
&\quad - a\cdot(\rho(p,q)+\rho(q,r)-\rho(p,r))\\
\le&\sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} h_0(x,y)\cdot\rho(x,y) + h_0(p,q)\cdot\rho(x,y)\\
&\quad + h_0(q,r)\cdot\rho(q,r) + h_0(p,r)\cdot\rho(p,r)\\
=&\sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} |h_0(x,y)\cdot\rho(x,y)| + |h_0(p,q)\cdot\rho(x,y)|\\
&\quad + |h_0(q,r)\cdot\rho(q,r)| + |h_0(p,r)\cdot\rho(p,r)| = \|h_0\cdot\rho\|_1.
\end{aligned}
$$

But $\|h_1\|_1 < \|h_0\|_1$, here we can omit the absolute values again:

$$
\begin{aligned}
\|h_1\|_1 &= \sum h_1(x,y) = \sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} h_1(x,y) + h_1(p,q) + h_1(q,r) + h_1(r,p)\\
&= \sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} h_1(x,y) + h_1(p,q) + h_1(q,r) + h_1(r,p)\\
&= \sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} h_0(x,y) + (h_0(p,q) - a) + (h_0(q,r) - a) + (h_0(r,p) + a)\\
&= \sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} h_0(x,y) + h_0(p,q) + h_0(q,r) + h_0(r,p) - a\\
&< \sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} h_0(x,y) + h_0(p,q) + h_0(q,r) + h_0(r,p)\\
&= \sum_{\substack{(x,y)\neq(p,q)\\(x,y)\neq(q,r)\\(x,y)\neq(r,p)}} |h_0(x,y)| + |h_0(p,q)| + |h_0(q,r)| + |h_0(r,p)| = \|h_0\|_1.
\end{aligned}
$$

This contradicts the minimality property of h, thus there aren't $p, q, r \in M$, such that $h(p, q) > 0$ and $h(q, r) > 0$, that is to say there isn't a point where we both ship to and ship from, proving *(3)*. Since h_0 is non-negative, we have $m^+(x) = \sum_y h_0(x, y)$ for all $x \in M$, thus $\|h_0\|_1 = \|m^+\|_1 = \frac{1}{2}\|m\|_1$, proving *(4)*.

Consider two finitely supported measures and . Denote the union of their supports by X. Then the coupling Π between the two is also finitely supported. We can describe the transformation between the two measures by a coupling, which is in this case a table representing $X \quad X$, just as in example 3.8. Let $_{ij}$ be the table's entries, meaning $_{ij}$ is the amount we want to transport from x_i to x_j. Then the coupling can be written as

$$\Pi = \sum_{(x_i, x_j) \in X \quad X} \quad _{ij} \quad _{x_i x_j}.$$

Now, as in the beginning of the chapter, consider . We discussed, that it can be treated as a molecule, let us denote the difference by m. What the first proposition of theorem 4.1 tells us, is that we can always find an optimal function $\hat{h}$, such that

$$m = \sum_{(x_i, x_j) \in X \quad X} \hat{h}(x_i, x_j) m_{x_i x_j}.$$

Next, if we set $_{ij} := \hat{h}(x_i, x_j)$, and we calculate the cost of the transport from to with respect to Π, and use the second proposition of Theorem 4.1, we get:

$$\begin{aligned}
\mathrm{d}_{W_1}(\; , \;) &= \int_{X \quad X} \rho(u, v)\, \mathrm{d}\Pi(u, v) \\
&= \int_{X \quad X} \rho(u, v)\, \mathrm{d} \sum_{(i,j) \in X \quad X} \quad _{ij} \quad _{x_i y_j} \quad (u, v) \\
&= \sum_{(i,j) \in X \quad X} \quad _{ij} \int_{X \quad X} \rho(u, v)\, \mathrm{d} \quad _{x_i y_j}(u, v) = \sum_{(i,j) \in X \quad X} \quad _{ij}\rho(x_i, x_j) \\
&= \|\hat{h} \quad \rho\|_1 = \|m\| \quad = \| \qquad \| .
\end{aligned}$$

According to Theorem 4.1, we have that this cost is optimal, thus

$$\| \qquad \| = \int_{X \times X} \varrho(u;v)\,\mathrm{d}\Pi(u;v) = \mathrm{d}_{W_1}(\ ;\ );$$

meaning that we found an optimal coupling through the Arens–Eells space.

Bibliography

[1] C. J. Budd, M. J. P. Cullen, & E. J. Walsh, *Monge{Ampere based moving mesh methods for numerical weather prediction, with applications to the Eady problem* Journal of Computational Physics, 236, pages 247-270., 2013., DOI URL

[2] M. Cullen, H. Maroo , *The Fully Compressible Semi-Geostrophic System from Meteorology*, Arch. Rational Mech. Anal. 167, pages 309{336. 2003., DOI URL

[3] M. Beiglbock, P. Henry-Labordere, and F. Penkner, *Model-independent Bounds for Option Prices: A Mass Transport Approach*, arXiv:1106.5929 [q- n.PR], DOI URL

[4] P.-A. Chiappori, R. J. McCann, and L. P. Nesheim, *Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness*, Econom. Theory vol. 42 (2010), no. 2, pages 317{354., Feb. 2010.

[5] Dolinsky, Yan and Soner, Halil Mete, *Martingale Optimal Transport and Robust Hedging in Continuous Time*, Swiss Finance Institute Research Paper No. 13-13, January 22, 2013., DOI URL

[6] J. Rabin, J. Delon, and Y. Gousseau, *Regularization of transportation maps for color and contrast transfer*, 2010 17th IEEE International Conference on Image Processing (ICIP), IEEE, 2010, pages 1933{1936.

[7] Y. Rubner, C. Tomasi, & L.J. Guibas, *The Earth Mover's Distance as a Metric for Image Retrieval* International Journal of Computer Vision 40, pages 99{121., 2000., DOI URL

[8] Nicolas Bonneel, Julie Digne, *A survey of Optimal Transport for Computer Graphics and Computer Vision*, Computer Graphics Forum, Vol. 42, Issue 2, pages 439-460., 2023., DOI URL

[9] Nik Weaver, *Lipschitz Algebras*, Second edition, World Scienti c, 2018.

[10] Ramon Jose Aliaga Varea, *Geometry and structure of Lipschitz-free spaces and their biduals*, PhD thesis, Universitat Politencia De Valencia, 2020., URL

[11] Gaspard Monge, *Application de l'analyse a la geometrie*, Bachelier, Paris, 1850.

[12] Gaspard Monge, *Geometrie descriptive*, Baudouin, Paris, 1798-1799.

[13] Gaspard Monge, *Description de l'art de fabriquer des canons*, Imprimerie du comite de salut publique, Paris, 1973.

[14] Gaspard Monge, *Memoire sur la theorie des deblais et des remblais*, Histoire de l'Academie Royale des Sciences de Paris, avec les Memoires de Mathematique et de Physique pour la même annee, pages 666{704, 1781.

[15] Filippo Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of variations, pdes, and modeling*, Birkhauser, 2015.

[16] L. Kantorovich, *The Mathematical Method of Production Planning and Organization*, Leningrad University Press, Leningrad, 1939.

[17] B. Gustafsson, *Scienti c Computing: A Historical Perspective, Texts in Computational Science and Engineering*, Springer, 2018.

[18] Leonid Vitaliyevich Kantorovich *Biographical* NobelPrize.org, Nobel Prize Outreach AB 2023. Fri. 22 Dec 2023. URL

[19] Press release. NobelPrize.org. Nobel Prize Outreach AB 2023. Thu. 21 Dec 2023. URL

[20] Leonid Kantorovich, *On the translocation of masses*, C.R. (Doklady) Acad. Sci. URSS (N.S.), 37:199{201, 1942.

[21] C. Villani, *Optimal transport, old and new*, Springer, Berlin, 2008.