# Optimal Transport for Data Analysis

Bernhard Schmitzer

2017-05-02

## 1 Reminders on Measure Theory

### 1.1 Foundations

Reference: Ambrosio, Fusco, Pallara: Functions of Bounded Variation and Free Discontinuity Problems, Chapters 1 & 2, [Ambrosio et al., 2000].

**Definition 1.1** ($\sigma$-algebra)**.** A collection $\mathcal{E}$ of subsets of a set $X$ is called $\sigma$-algebra if

(i) $\emptyset \in \mathcal{E}$; $[A \in \mathcal{E}] \Rightarrow [X \setminus A \in \mathcal{E}]$;

(ii) for a sequence $A_n \in \mathcal{E} \Rightarrow \bigcup_{n=0}^{\infty} A_n \in \mathcal{E}$.

---

Comment: Closed under finite unions, intersections and countable intersections. $A \cap B = X \setminus ((X \setminus A) \cup (X \setminus B))$.

---

Comment: Elements of $\mathcal{E}$: 'measurable sets'. Pair $(X, \mathcal{E})$: 'measure space'.

---

**Example 1.2.** Borel algebra: smallest $\sigma$ algebra containing all open sets of a topological space.

Comment: Intersection of two $\sigma$-algebras is again $\sigma$-algebra. 'smallest' is well-defined.

---

**Definition 1.3** (Positive measure and vector measure)**.** For measure space $(X, \mathcal{E})$ a function $\mu : \mathcal{E} \mapsto [0, +\infty]$ is called 'positive measure' if

(i) $\mu(\emptyset) = 0$;

(ii) for pairwise disjoint sequence $A_n \in \mathcal{E} \Rightarrow \mu\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mu(A_n)$

For measure space $(X, \mathcal{E})$ and $\mathbb{R}^m$, $m \geq 1$, a function $\mu : \mathcal{E} \mapsto \mathbb{R}^m$ is called 'measure' if $\mu$ satisfies (i) and (ii) with absolute convergence.

---

Comment: Measures are vector space, measures are finite, positive measures may be infinite.

---

**Example 1.4.** Examples for measures:

1. counting measure: $\#(A) = |A|$ if $A$ finite, $+\infty$ else.
2. Dirac measure: $\delta_x(A) = 1$ if $x \in A$, 0 else.
3. Lebesgue measure $\mathcal{L}([a, b]) = b - a$ for $b \geq a$.
4. Scaled measures: positive measure $\mu$, function $f \in L^1(\mu)$, new measure $\nu = f \cdot \mu$. $\nu(A) \stackrel{\text{def.}}{=} \int_A f(x) \, \mathrm{d}\mu(x)$.

5. Weak gradient of discontinuous function $f$, $\mu = D\,f$.

$$\int_\Omega \varphi(x) \cdot \mathrm{d}\mu(x) = - \int_\Omega \operatorname{div} \varphi(x)\, f(x)\, \mathrm{d}x$$

for $\varphi \in C^1(\Omega)$.

**Definition 1.5** (Total variation)**.** For measure $\mu$ on $(X, \mathcal{E})$ the total variation $|\mu|$ of $A \in \mathcal{E}$ is

$$|\mu|(A) = \sup \left\{ \sum_{n=0}^{\infty} |\mu(A_n)| \middle| A_n \in \mathcal{E}, \text{ pairwise disjoint}, \bigcup_{n=0}^{\infty} A_n = A \right\} .$$

$|\mu|$ is finite, positive measure on $(X, \mathcal{E})$.

---

Comment: Careful with nomenclature in image analysis.

---

**Definition 1.6.** A set $N \subset X$ is $\mu$-negligible if $\exists\ A \in \mathcal{E}$ with $N \subset A$ and $\mu(A) = 0$. Two functions $f$, $g : X \to Y$ are identical '$\mu$-almost everywhere' when $\{x \in X | f(x) \neq g(x)\}$ is $\mu$-negligible.

**Example 1.7.** Null sets are Lebesgue-negligible sets.

## 1.2 Measures and Maps

**Definition 1.8** (Measurable functions, push-forward)**.** Let $(X, \mathcal{E})$, $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is 'measurable' if $f^{-1}(A) \in \mathcal{E}$ for $A \in \mathcal{F}$.
For measure $\mu$ on $(X, \mathcal{E})$ the 'push-forward' of $\mu$ under $f$ to $(Y, \mathcal{F})$, we write $f_\sharp\mu$, is defined by $f_\sharp\mu(A) = \mu(f^{-1}(A))$ for $A \in \mathcal{F}$.
Change of variables formula:

$$\int_X g(f(x))\, \mathrm{d}\mu(x) = \int_Y g(y)\, \mathrm{d}f_\sharp\mu(y)$$

---

**Sketch:** Varying densities.

---

**Example 1.9** (Marginal)**.** Let $\mathrm{proj}_i : X \times X \to X$, $\mathrm{proj}_i(x_0, x_1) = x_i$. Marginals of measure $\gamma$ on $X \times X$:

$$\mathrm{proj}_{0\,\sharp}\gamma(A) = \gamma(A \times X)\,, \qquad \mathrm{proj}_{1\,\sharp}\gamma(A) = \gamma(X \times A)\,.$$

---

**Sketch:** Discuss pre-images of $\mathrm{proj}_i$.

---

## 1.3 Comparison, Decomposition

**Definition 1.10** (Absolute continuity, singularity)**.** Let $\mu$ be positive measure, $\nu$ measure on measurable space $(X, \mathcal{E})$. $\nu$ is 'absolutely continuous' w.r.t. $\mu$, we write $\nu \ll \mu$, if $[\mu(A) = 0] \Rightarrow [\nu(A) = 0]$.

---

**Sketch:** Density $\ll$ Lebesgue, density $\not\ll$ density when support different, Dirac measures $\not\ll$ Lebesgue, mixed measures $\not\ll$ density, mixed measures $\ll$ mixed measures when Diracs coincide.

---

Positive measures $\mu$, $\nu$ are 'mutually singular', we write $\mu \perp \nu$, if $\exists\ A \in \mathcal{E}$ such that $\mu(A) = 0$, $\mu(X \setminus A) = 0$. For general measures replace $\mu$, $\nu$ by $|\mu|$, $|\nu|$.

**Definition 1.11** ($\sigma$-finite)**.** A positive measure $\mu$ is called $\sigma$-finite if $X = \bigcup_{n=0}^{\infty} A_n$ for sequence $A_n \in \mathcal{E}$ with $\mu(A_n) < +\infty$.

**Example 1.12.** Lebesgue measure is not finite but $\sigma$-finite.

**Theorem 1.13** (Radon–Nikodym, Lebesgue decomposition [Ambrosio et al., 2000, Theorem 1.28])**.** Let $\mu$ be $\sigma$-finite positive measure. $\nu$ general measure.
Radon–Nikodym: For $\nu \ll \mu$ there is a function $f \in L^1(\mu)$ such that $\nu = f \cdot \mu$.
Lebesgue decomposition: there exist unique measures $\nu_a$, $\nu_s$ such that

$$\nu = \nu_a + \nu_s, \qquad \nu_a \ll \mu, \qquad \nu_s \perp \mu\,.$$

Note: $\nu_a = f \cdot \mu$ for some $f \in L^1(\mu)$.

**Corollary 1.14.** A real-valued measure $\nu$ can be decomposed into $\nu = \nu_+ - \nu_-$ with $\nu_+$, $\nu_-$ mutually singular positive measures.

*Proof.* Since $\nu \ll |\nu|$ there exists $f \in L^1(|\nu|)$ with $\nu = f \cdot |\nu|$. Set $A_+ = f^{-1}((0, +\infty))$, $A_- = f^{-1}((-\infty, 0))$ and set $\nu_\pm(B) = |\nu(B \cap A_\pm)|$. □

---

Comment: $f$ is only unique $|\nu|$-almost everywhere.

---

## 1.4 Duality

References: Kurdila, Zabarankin: Convex functional analysis [Kurdila and Zabarankin, 2005]. For Hilbert spaces: Bauschke, Combettes: Convex Analysis and Monotone Operator Theory in Hilbert Spaces [Bauschke and Combettes, 2011]

**Definition 1.15** (Dual space)**.** For normed vector space $(X, \|\cdot\|_X)$ its topological dual space is given by

$$X^* = \{y : X \to \mathbb{R} \,|\, y \,\text{linear, continous, i.e.}\ \exists\, C < \infty,\ |y(x)| \leq C\,\|x\|_X\, \forall\, x \in X\}\,.$$

Norm on $X^*$:

$$\|y\|_{X^*} = \sup\,\{|y(x)||x \in X,\ \|x\|_X \leq 1\}$$

$(X^*, \|\cdot\|_{X^*})$ is Banach space. For $y(x)$ one often writes $\langle y, x\rangle$ or $\langle y, x\rangle_{X^*,X}$.

---

Comment: Linear not necessarily continuous in infinite dimensions. Dual norm is operator norm.

---

**Definition 1.16** (Weak convergence)**.** A sequence $x_n$ in $X$ converges weakly to $x \in X$ if $y(x_n) \to y(x)$ for all $y \in X^*$. We write $x_n \rightharpoonup x$.

**Definition 1.17** (Weak* convergence)**.** A sequence $y_n$ in $X^*$ converges weakly to $y \in X^*$ if $y_n(x) \to y(x)$ for all $x \in X$. We write $y_n \overset{*}{\rightharpoonup} y$.

Application to measures:

**Definition 1.18** (Radon measures)**.** Let $(X, d)$ be compact metric space, let $\mathcal{E}$ be Borel-$\sigma$-algebra. A finite measure (positive or vector valued) is called a ‘Radon measure’. Write:

- $\mathcal{M}_+(X)$: positive Radon measures,

- $\mathcal{P}(X) \subset \mathcal{M}_+(X)$: Radon probability measures (total mass = 1),
- $\mathcal{M}(X)^m$: (vector valued) Radon measures.

**Theorem 1.19** (Regularity [Ambrosio et al., 2000, Proposition 1.43])**.** For positive Radon measures on $(X, \mathcal{E})$ one has for $A \in \mathcal{E}$

$$\mu(A) = \sup \{\mu(B) \,|\, B \in \mathcal{E},\, B \subset A,\, B \text{ compact}\} = \inf \{\mu(B) \,|\, B \in \mathcal{E},\, A \subset B,\, B \text{ open}\} .$$

**Theorem 1.20** (Duality [Ambrosio et al., 2000, Theorem 1.54])**.** Let $(X, d)$ be compact metric space. Let $C(X)^m$ be space of continuous functions from $X$ to $\mathbb{R}^m$, equipped with sup-norm. The topological dual of $C(X)^m$ can be identified with the space $\mathcal{M}(X)^m$ equipped with the total variation norm $\|\mu\|_{\mathcal{M}} \stackrel{\text{def.}}{=} |\mu|(X)$. Duality pairing for $\mu \in \mathcal{M}(X)^m$, $f \in C(X)^m$:

$$\mu(f) = \langle \mu, f \rangle_{\mathcal{M},C} = \int_X f(x) \, \mathrm{d}\mu(x)$$

**Corollary 1.21.** Two measures $\mu$, $\nu \in \mathcal{M}(X)^m$ with $\mu(f) = \nu(f)$ for all $f \in C(X)^m$ coincide.

**Theorem 1.22** (Banach–Alaoglu [Kurdila and Zabarankin, 2005, Theorem 2.4.4])**.** Let $X$ be a separable normed space. Any bounded sequence in $X^*$ has a weak$*$ convergent subsequence.

---

Comment: Since $C(X)$ is separable, any bounded sequence in $\mathcal{M}(X)$ has a weak$*$ convergent subsequence.

---

# 2 Monge formulation of optimal transport

---

Comment: Gaspard Monge: French mathematician and engineer, 18$^{\text{th}}$ century. Studied problem of optimal allocation of resources to minimize transport cost.

---

**Sketch:** Bakeries and cafes

---

**Example 2.1** (According to Villani)**.** Every morning in Paris bread must be transported from bakeries to cafes for consumption. Every bakery produces prescribed amount of bread, every cafe orders prescribed amount. Assume: total amounts identical. Look for most economical way to distribute bread.

Mathematical model:

- $\Omega \subset \mathbb{R}^2$: area of Paris
- $\mu \in \mathcal{P}(\Omega)$: distribution of bakeries and produced amount of bread,
- $\nu \in \mathcal{P}(\Omega)$: distribution of cafes and consumed amount of bread
- Cost function $c : \Omega \times \Omega \to \mathbb{R}_+$. $c(x, y)$ gives cost of transporting 1 unit of bread from bakery at $x$ to cafe at $y$.
- Describe transport by map $T : \Omega \to \Omega$. Bakery at $x$ will deliver bread to cafe at $T(x)$. Consistency condition: $T_\sharp \mu = \nu$.

---

  Comment: Each cafe receives precisely ordered amount of bread.

---

- Total cost of transport map

$$C_M(T) = \int_\Omega c(x, T(x))\, \mathrm{d}\mu(x)$$

---

Comment: For bakery at location $x$ pay $c(x,T(x)) \cdot \mu(x)$. Sum (i.e. integrate) over all bakeries.

---

**Definition 2.2.** Monge optimal transport problem: find $T$ that minimizes $C_M$.

Problems:

- Do maps $T$ with $T_\sharp\mu = \nu$ exist? Can not split mass.

---

**Sketch:** Splitting of mass.

---

- Does minimal $T$ exist? Non-linear, non-convex constraint and objective.

---

Comment: $\Rightarrow$ problem remained unsolved for long time.

---

## 3 Kantorovich formulation of optimal transport

---

Comment: Leonid Kantorovich: Russian mathematician, $20^{\text{th}}$ century. Founding father of linear programming, proposed modern formulation of optimal transport. (Nobel prize in economics 1975.)

---

Do not describe transport by map $T$, but by positive measure $\pi \in \mathcal{M}_+(\Omega \times \Omega)$.

**Definition 3.1** (Coupling / Transport Plan)**.** Let $\mu$, $\nu \in \mathcal{P}(\Omega)$. Set of 'couplings' or 'transport plans' $\Pi(\mu,\nu)$ is given by

$$\Pi(\mu,\nu) = \left\{\pi \in \mathcal{P}(\Omega \times \Omega) \,\middle|\, {\mathrm{proj}_0}_\sharp\pi = \mu,\ {\mathrm{proj}_1}_\sharp\pi = \nu\right\}\ .$$

**Example 3.2.** $\Pi(\mu,\nu) \neq \emptyset$, contains at least product measure $\mu \otimes \nu \in \Pi(\mu,\nu)$. $(\mu\otimes\nu)(A \times B) = \mu(A)\cdot\nu(B)$ for measurable $A$, $B \subset \Omega$.

**Definition 3.3.** For compact metric space $(\Omega, d)$, $\mu$, $\nu \in \mathcal{P}(\Omega)$, $c \in C(\Omega \times \Omega)$ the Kantorovich optimal transport problem is given by

$$C(\mu,\nu) = \inf\left\{\int_{\Omega\times\Omega} c(x,y)\,\mathrm{d}\pi(x,y) \,\middle|\, \pi \in \Pi(\mu,\nu)\right\} \tag{1}$$

---

Comment: Linear (continuous) objective, affine constraint set.

---

Comment: Language of measures covers finite dimensional and infinite dimensional case.

---

**Theorem 3.4.** Minimizers of (1) exist.

*Proof.*

- Let $\pi_n$ be minimizing sequence. Since $\pi_n \in \mathcal{P}(\Omega\times\Omega)$ have $\|\pi_n\|_\mathcal{M} = 1$. By Banach-Alaoglu (Theorem 1.22) $\exists$ converging subsequence. After extraction of subsequence have convergent minimizing sequence $\pi_n \overset{*}{\rightharpoonup} \pi$.

- Positivity: $\pi$ is a positive measure. Otherwise find function $\phi \in C(\Omega\times\Omega)$ with $\int_{\Omega\times\Omega}\phi\,\mathrm{d}\pi < 0$ (use Corollary 1.14 and Theorem 1.19 for construction) which contradicts weak$*$ convergence.

- Unit mass: $\pi(\Omega\times\Omega)=\int_{\Omega\times\Omega}1\,\mathrm{d}\pi=\lim_{n\to\infty}\int_{\Omega\times\Omega}1\,\mathrm{d}\pi_n=\pi_n(\Omega\times\Omega)=1$
- Marginal constraint: For every $\phi\in C(\Omega)$

$$\begin{aligned}\int_\Omega \phi\,\mathrm{d}{\mathrm{proj}_0}_\sharp\pi &= \int_{\Omega\times\Omega}\phi\circ\mathrm{proj}_0\,\mathrm{d}\pi\\ &= \lim_{n\to\infty}\int_{\Omega\times\Omega}\phi\circ\mathrm{proj}_0\,\mathrm{d}\pi_n=\lim_{n\to\infty}\int_\Omega\phi\,\mathrm{d}{\mathrm{proj}_0}_\sharp\pi_n=\int_\Omega\phi\,\mathrm{d}\mu\end{aligned}$$

  So ${\mathrm{proj}_0}_\sharp\pi=\mu$. Analogous: ${\mathrm{proj}_1}_\sharp\pi=\nu$.
- So: $\pi\in\Pi(\mu,\nu)$.
- Since $c\in C(\Omega\times\Omega)$ and $\pi_n\stackrel{*}{\rightharpoonup}\pi$ have

$$\int_{\Omega\times\Omega}c\,\mathrm{d}\pi=\lim_{n\to\infty}\int_{\Omega\times\Omega}c\,\mathrm{d}\pi_n\,.$$

  Therefore, $\pi$ is minimizer.

□

---

Comment: For proof under more general conditions see for instance [Villani, 2009, Chapter 4]

---

Proof two additional useful results to get some practice and intuition.

**Proposition 3.5** (Restriction [Villani, 2009, Theorem 4.6])**.** Let $\mu$, $\nu\in\mathcal{P}(\Omega)$, $c\in C(\Omega\times\Omega)$, let $\pi$ be optimizer for $C(\mu,\nu)$. Let $\tilde\pi\in\mathcal{M}_+(\Omega\times\Omega)$, $\tilde\pi(\Omega\times\Omega)>0$, $\tilde\pi(A)\le\pi(A)$ for all measurable $A\subset\Omega\times\Omega$. Set $\pi'=\frac{\tilde\pi}{\tilde\pi(\Omega\times\Omega)}$, $\pi'\in\mathcal{P}(\Omega\times\Omega)$. Let $\mu'={\mathrm{proj}_0}_\sharp\pi'$, $\nu'={\mathrm{proj}_1}_\sharp\pi'$. Then $\pi'$ is minimal for $C(\mu',\nu')$.

**Example 3.6.** $\tilde\pi(A)\stackrel{\text{def.}}{=}\pi(A\cap(\Omega_0\times\Omega_1))$ for $\Omega_0$, $\Omega_1\subset\Omega$.

---

**Sketch:** Restriction to subset. More general restriction.

---

*Proof.*
- Assume $\pi'$ is not optimal. Then there is a measure $\pi''\in\Pi(\mu',\nu')$ with strictly better cost.
- Consider the measure $\hat\pi=\pi-\tilde\pi+\tilde\pi(\Omega\times\Omega)\cdot\pi''$. $\hat\pi$ is a positive measure since $\tilde\pi\le\pi$. $\hat\pi\in\mathcal{P}(\Omega\times\Omega)$ since $\pi''\in\mathcal{P}(\Omega\times\Omega)$.

$$\begin{aligned}{\mathrm{proj}_0}_\sharp\hat\pi &= {\mathrm{proj}_0}_\sharp\pi-{\mathrm{proj}_0}_\sharp\tilde\pi+\tilde\pi(\Omega\times\Omega)\cdot{\mathrm{proj}_0}_\sharp\pi''\\ &=\mu-\tilde\pi(\Omega\times\Omega)\cdot\big(\mu'-\mu'\big)=\mu\end{aligned}$$

  So $\hat\pi\in\Pi(\mu,\nu)$.
- $\hat\pi$ has lower transport cost than $\pi$:

$$\int_{\Omega\times\Omega}c\,\mathrm{d}\hat\pi=\int_{\Omega\times\Omega}c\,\mathrm{d}\pi-\tilde\pi(\Omega\times\Omega)\int_{\Omega\times\Omega}c\,\mathrm{d}\pi'+\tilde\pi(\Omega\times\Omega)\int_{\Omega\times\Omega}c\,\mathrm{d}\pi''<\int_{\Omega\times\Omega}c\,\mathrm{d}\pi$$

- So $\pi$ is not optimal which is a contradiction. Therefore $\pi'$ must be optimal.

□

**Proposition 3.7** (Convexity [Villani, 2009, Theorem 4.8])**.** The function $\mathcal{P}(\Omega)^2\to\mathbb{R}$, $(\mu,\nu)\mapsto C(\mu,\nu)$ is convex.

*Proof.*

- Let $\mu_0$, $\mu_1$, $\nu_0$, $\nu_1 \in \mathcal{P}(\Omega)$. Let $\pi_i$ be corresponding minimizers in $C(\mu_i, \nu_i)$, $i \in \{0, 1\}$.
- For $\lambda \in (0, 1)$ set

$$\hat{\mu} = (1-\lambda)\,\mu_0 + \lambda\,\mu_1\,, \qquad \hat{\nu} = (1-\lambda)\,\nu_0 + \lambda\,\nu_1\,, \qquad \hat{\pi} = (1-\lambda)\,\pi_0 + \lambda\,\pi_1\,.$$

- $\hat{\pi} \in \Pi(\hat{\mu}, \hat{\nu})$ since

$$\mathrm{proj}_{0\,\sharp}\hat{\pi} = (1-\lambda)\,\mathrm{proj}_{0\,\sharp}\pi_0 + \lambda\,\mathrm{proj}_{0\,\sharp}\pi_1 = (1-\lambda)\,\mu_0 + \lambda\,\mu_1 = \hat{\mu}\,.$$

- Convexity:

$$C(\hat{\mu}, \hat{\nu}) \leq \int_{\Omega\times\Omega} c\,d\hat{\pi} = (1-\lambda)\int_{\Omega\times\Omega} c\,d\pi_0 + \lambda\int_{\Omega\times\Omega} c\,d\pi_1 = (1-\lambda)\,C(\mu_0, \nu_0) + \lambda\,C(\mu_1, \nu_1)$$

□

# 4 Kantorovich duality

## 4.1 More duality

**Definition 4.1** (Topologically paired spaces)**.** Two vector spaces $X$, $X^*$ with locally convex Hausdorff topology are called *topologically paired spaces* if all continuous linear functionals on one space can be identified with all elements of the other.

**Example 4.2.** Let $(\Omega, d)$ be a compact metric space. $C(X)$ and $\mathcal{M}(X)$ with the sup-norm topology and the weak-$*$ topology are topologically paired spaces.
Any continuous linear functional on $C(X)$ can be identified with an element in $\mathcal{M}(X)$ by construction. If $\Phi$ is a weak-$*$ continuous linear functional on $\mathcal{M}(X)$ it can be identified with the continuous function $\varphi : x \mapsto \Phi(\delta_x)$.

**Definition 4.3** (Fenchel–Legendre conjugates)**.** Let $X$, $X^*$ be topologically paired spaces. Let $f : X \to \mathbb{R} \cup \{\infty\}$. Its Fenchel–Legendre conjugate $f^* : X^* \to \mathbb{R} \cup \{\infty\}$ is given by

$$f^*(y) = \sup\{\langle y, x\rangle - f(x) | x \in X\}\,.$$

$f^*$ is convex, lsc on $X^*$. Likewise, for $g : X^* \to \mathbb{R} \cup \{\infty\}$ define conjugate $g^*$. If $f$, $g$ convex, lsc then $f = f^{**}$, $g = g^{**}$.

---

Comment: Lsc: lower semicontinuous, $[x_n \to x] \Rightarrow [f(x) \leq \liminf_{n\to\infty} f(x_n)]$

---

**Theorem 4.4** (Fenchel–Rockafellar [Rockafellar, 1967])**.** Let $(X, X^*)$, $(Y, Y^*)$ be two pairs of topologically paired spaces. Let $f : X \to \mathbb{R} \cup \{\infty\}$, $g : Y \to \mathbb{R} \cup \{\infty\}$, $A : X \to Y$ linear, continuous. Assume $\exists\, x \in X$ such that $f$ finite at $x$, $g$ finite and continuous at $A\,x$. Then

$$\inf\{f(x) + g(A\,x) | x \in X\} = \max\{-f^*(-A^*z) - g^*(z) | z \in Y^*\}\,.$$

Maximizer of rhs exists. $A^* : Y^* \to X^*$ is adjoint of $A$ defined by $\langle z, A\,x\rangle_{Y^*,Y} = \langle A^*z, x\rangle_{X^*,X}$.

---

Comment: Can sometimes be used 'in both directions' to establish existence of both primal and dual problem.

---

## 4.2 Dual Kantorovich Problem

**Theorem 4.5.** Given the setting of Definition 3.3 one finds

$$C(\mu,\nu) = \sup\left\{\int_\Omega \alpha\,\mathrm{d}\mu + \int_\Omega \beta\,\mathrm{d}\nu \middle| \alpha,\,\beta \in C(\Omega), \alpha(x)+\beta(y) \leq c(x,y) \text{ for all } (x,y)\in\Omega^2\right\} \quad (2)$$

*Proof.*

- Problem (2) can be written as

$$C(\mu,\nu) = -\inf\left\{f(\alpha,\beta) + g(A(\alpha,\beta))\middle|(\alpha,\beta)\in C(\Omega)^2\right\}$$

  with

$$\begin{aligned}
&f : C(\Omega)^2 \to \mathbb{R}, && (\alpha,\beta) \mapsto -\int_\Omega \alpha\,\mathrm{d}\mu - \int_\Omega \beta\,\mathrm{d}\nu \\
&g : C(\Omega^2) \to \mathbb{R}\cup\{\infty\}, && \psi \mapsto \begin{cases} 0 & \text{if } \psi(x,y) \leq c(x,y) \text{ for all } (x,y)\in\Omega^2 \\ +\infty & \text{else.}\end{cases} \\
&A : C(\Omega)^2 \to C(\Omega^2), && [A(\alpha,\beta)](x,y) = \alpha(x)+\beta(y)\,.
\end{aligned}$$

- $f$, $g$ are convex, lsc. $A$ is bounded, linear.
- Let $(\alpha,\beta)$ be two constant, finite functions with $\alpha(x)+\beta(y) < \min\{c(x',y')|(x',y')\in\Omega^2\}$. Then $f(\alpha,\beta) < \infty$, $g(A(\alpha,\beta)) < \infty$ and $g$ is continuous at $A(\alpha,\beta)$. Thus, with Theorem 4.4 (and Example 4.2)

$$C(\mu,\nu) = \min\left\{f^*(-A^*\pi) + g^*(\pi)\middle|\pi\in\mathcal{M}(\Omega^2)\right\}.$$

- One obtains:

$$\begin{aligned}
f^*(-\rho,-\sigma) &= \sup\left\{-\int_\Omega \alpha\,\mathrm{d}\rho - \int_\Omega \beta\,\mathrm{d}\sigma + \int_\Omega \alpha\,\mathrm{d}\mu + \int_\Omega \beta\,\mathrm{d}\nu \middle|(\alpha,\beta)\in C(\Omega)^2\right\} \\
&= \begin{cases} 0 & \text{if } \rho=\mu,\ \sigma=\nu\,, \\ +\infty & \text{else.}\end{cases}
\end{aligned}$$

  (Reasoning similar than for positivity of limit $\pi$ in proof of Theorem 3.4.)

$$\begin{aligned}
g^*(\pi) &= \sup\left\{\int_{\Omega^2} \psi\,\mathrm{d}\pi \middle| \psi\in C(\Omega^2), \psi(x,y)\leq c(x,y) \text{ for all } (x,y)\in\Omega^2\right\} \\
&= \begin{cases} \int_{\Omega^2} c\,\mathrm{d}\pi & \text{if } \pi\in\mathcal{M}_+(\Omega^2), \\ +\infty & \text{else.}\end{cases}
\end{aligned}$$

□

So far we have not yet proven existence of dual maximizers. For this we need some additional arguments. We follow the presentation in [Santambrogio, 2015, Section 1.2].

**Definition 4.6** (c-transform)**.** For $\psi \in C(\Omega)$ define its $c$-transform $\psi^c \in C(\Omega)$ by

$$\psi^c(y) = \inf\{c(x,y) - \psi(x)|x\in\Omega\}$$

and its $\bar{c}$-transform $\psi^{\bar{c}} \in C(\Omega)$ by

$$\psi^{\bar{c}}(x) = \inf \{c(x,y) - \psi(y) | y \in \Omega\}$$

A function $\psi$ is called $\bar{c}$-concave if it can be written as $\psi = \phi^c$ for some $\phi \in C(\Omega)$. Analogously, $\psi$ is $c$-concave if it can be written as $\psi = \phi^{\bar{c}}$.

---

Comment: Setting $\beta = \alpha^c$ (or $\alpha = \beta^{\bar{c}}$) in (2) corresponds to optimization over $\beta$ for fixed $\alpha$ (and vice versa). In general alternating optimization of (2) in $\alpha$ and $\beta$ does not yield an optimal solution.

---

**Lemma 4.7.** The set of $c$-concave and $\bar{c}$-concave functions are equicontinuous.

*Proof.*

- Since $c \in C(\Omega \times \Omega)$ and $(\Omega, d)$ compact there is a continuous function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ such that $|c(x,y) - c(x,y')| \leq \omega(d(y,y'))$.
- Let $\psi = \phi^c$. Set $\phi_x : y \mapsto c(x,y) - \phi(x)$. For every $x \in \Omega$ have $|\phi_x(y) - \phi_x(y')| \leq \omega(d(y,y'))$. One finds

$$\psi(y) \leq \phi_x(y) \leq \phi_x(y') + \omega(d(y,y'))$$

  for all $x, y, y' \in \Omega$. Taking the infimum over $x$ one gets $\psi(y) \leq \psi(y') + \omega(d(y,y'))$ and by symmetry $|\psi(y) - \psi(y')| \leq \omega(d(y,y'))$. This implies equicontinuity of $\bar{c}$-concave functions.
- Argument for $\phi^{\bar{c}}$ analogous.

$\square$

**Theorem 4.8** (Arzelà-Ascoli [Rudin, 1986, Thm. 11.28])**.** If $(\Omega, d)$ is a compact metric space and $(f_n)_n$ is a sequence of uniformly bounded, equicontinuous functions in $C(\Omega)$ then $(f_n)_n$ has a uniformly converging subsequence.

**Theorem 4.9** ([Santambrogio, 2015, Prop. 1.11])**.** Maximizers of (2) exist.

*Proof.*

- For feasible $(\alpha, \beta)$ with finite score in (2) one can always replace $\beta$ by $\alpha^c$ and subsequently $\alpha$ by $(\alpha^c)^{\bar{c}}$ which are still feasible and do not decrease the functional value. Hence, we may impose the additional constraint that $(\alpha, \beta)$ in (2) are $(c, \bar{c})$-concave.
- Replacing feasible $(\alpha, \beta)$ in (2) by

$$(x \mapsto \alpha(x) - C, y \mapsto \beta(y) + C) \qquad \text{with } C = \min_{x' \in \Omega} \alpha(x')$$

  does not change the functional value or affect the constraints.
- Arguing as in Lemma 4.7 one finds for $c$-concave $\alpha$ with $\min_x \alpha(x) = 0$ that $\alpha(x) \in [0, \omega(\operatorname{diam} \Omega)]$ and for the corresponding $\beta = \alpha^c$ that $\beta(y) \in [\min c - \omega(\operatorname{diam} \Omega), \max c]$.
- Hence, we may consider maximizing sequences of (2) that are uniformly bounded and equicontinuous. By the Arzelà-Ascoli Theorem there exists a uniformly converging subsequence. Since the objective (and the constraints) of (2) are upper semicontinuous (see proof of Theorem 4.5), the limit must be a maximizer.

$\square$

**Corollary 4.10** (Primal-Dual Optimality Condition)**.** If $\pi$ solves (1) and $(\alpha, \beta)$ solve (2) then $\alpha(x) + \beta(y) = c(x, y)$ $\pi$-almost everywhere.

*Proof.*

$$\int_{\Omega\times\Omega} c(x,y)\,\mathrm{d}\pi(x,y) = \int_\Omega \alpha\,\mathrm{d}\mu + \int_\Omega \beta\,\mathrm{d}\nu = \int_{\Omega\times\Omega} [\alpha(x)+\beta(y)]\mathrm{d}\pi(x,y) \leq \int_{\Omega\times\Omega} c(x,y)\,\mathrm{d}\pi(x,y)$$

$\square$

**Remark 4.11** (Economic Interpretation of Kantorovich Duality)**.** Bakeries and cafes hire a third-party company to do the transportation and agree to split the transport cost. When picking up bread at bakery $x$ in the morning, the company charges an advance payment $\alpha(x)$ per unit of bread for the transport. Upon delivery at a cafe at $y$ it charges a final payment $\beta(y)$ per unit of bread from the cafe.
The total payment to the company will be $\int_\Omega \alpha\,\mathrm{d}\mu + \int_\Omega \beta\,\mathrm{d}\nu$. It is left to the company to decide which bread to deliver where. And they will want to minimize the total transport cost, i.e. to find the *global minimum* of $\int_{\Omega\times\Omega} c\,\mathrm{d}\pi$.
But it can never charge more than $c(x, y) - \alpha(x)$ when dropping of bread from $x$ at $y$, otherwise the cafe $y$ may complain and try to hire another company to get bread from bakery $x$ at a lower price. When every cafe receives bread from its 'subjectively cheapest' bakery (and similarly each bakery delivers to its 'subjectively cheapest' cafe), the transport plan is said to be at equilibrium: no party will attempt to change its partner in a local attempt to reduce its costs.
Kantorovich duality states that for the optimal transport model the global minimum and equilibrium coincide.

## References


L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.* Oxford mathematical monographs. Oxford University Press, 2000.

H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics. Springer, 1st edition, 2011.

A. J. Kurdila and M. Zabarankin. *Convex functional analysis*, volume 1 of *Systems and Control: Foundations and Applications.* Birkhäuser, 2005.

R. T. Rockafellar. Duality and stability in extremum problems involving convex functions. *Pacific J. Math*, 21(1):167–187, 1967.

W. Rudin. *Real and complex analysis.* McGraw-Hill Book Company, 3rd edition, 1986.

F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications.* Birkhäuser Boston, 2015.

C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften.* Springer, 2009.