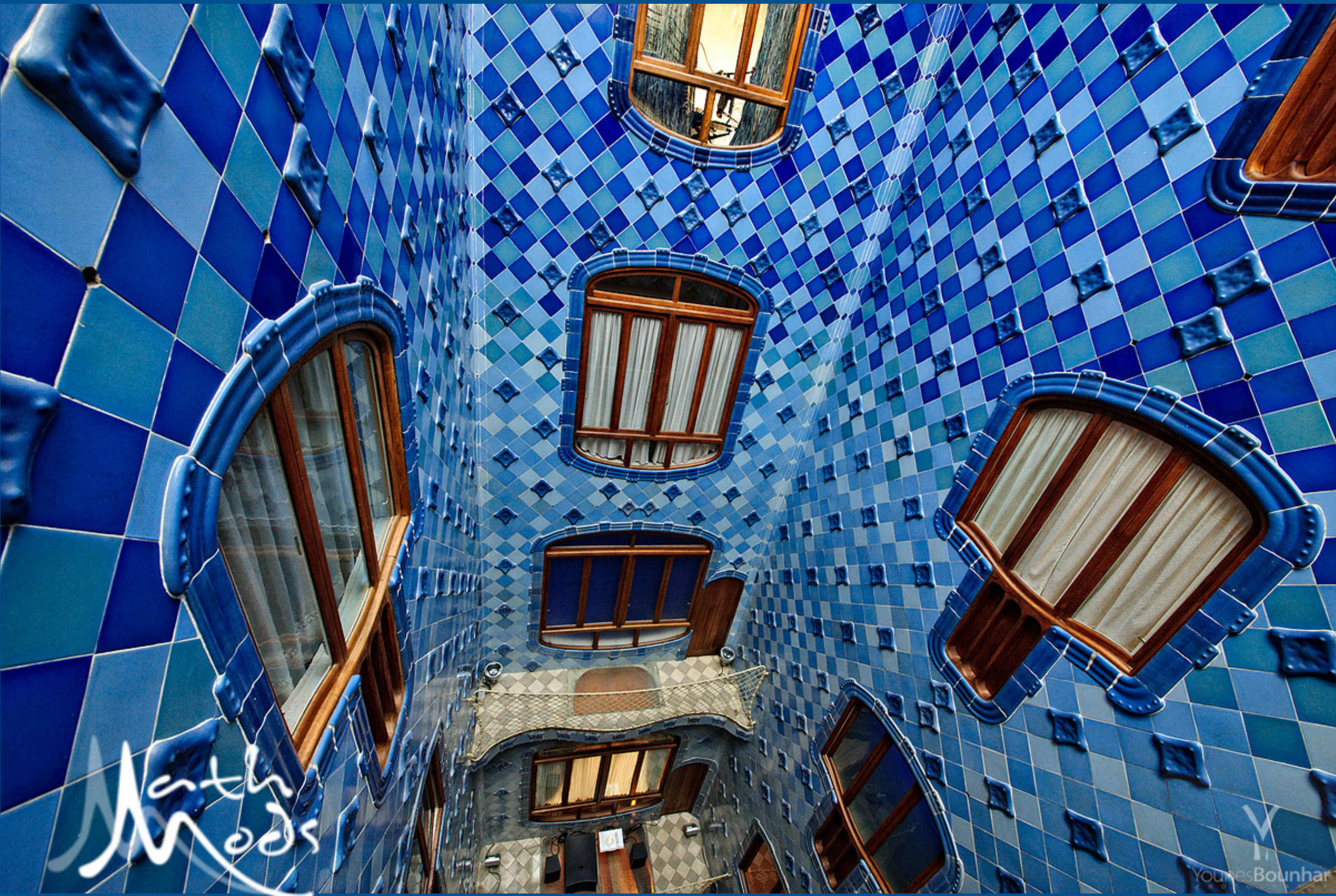
The Optimal Transport Problem

Master Thesis

Oscar Ramirez



The Optimal Transport Problem

Master Thesis

by

Oscar Ramirez

to obtain the degree of Master of Science
in Mathematical Modelling and Engineering,
to be defended publicly on September, 2018.

Project duration:	September, 2016 – September, 2018
Thesis committee:	Prof. Juan Enrique Martinez Legaz, UAB, supervisor

Preface

Preface…

Oscar Ramirez
Barcelona, September 2018

Contents

1

Preliminaries.

Notation

$\mathbb{R}$	Real numbers field.
$\bar{\mathbb{R}}$	$\mathbb{R} \cup \{+\infty\}$. That is $[-\infty, \infty]$
$\mathbb{R}_+$	The set of nonnegative real numbers, that is the interval $[0, \infty)$.
$\bar{\mathbb{R}}_+$	The set of nonnegative extended real numbers, that is the interval $[0, \infty]$
δ_x	The Dirac mass at point x.
$\mathbb{R}^d$	The d-dimensional Euclidean space.
$\mathcal{P}(X)$	Space of probabilities on X.
$\mu \ll \nu$	The measure is absolutely continuous with respect to ν.
$\mathbb{1}_\Omega$	Indicator function of a set Ω. If $x \in \Omega$ then $\mathbb{1}_\Omega(x) = 1$. If $x \in \Omega^c$, we have $\mathbb{1}_\Omega(x) = 0$.
$\mu \llcorner A$	A measure μ restricted to a set A.
ω_d	The Measure of the unite ball in $\mathbb{R}^d$.
$\wedge$	The min operator, that is $a \wedge b := \min\{a, b\}$.
$\vee$	The max operator, that is $a \vee b := \max\{a, b\}$.
$T_\# \mu$	The image measure of μ through the map T.
$f_{\mid\Omega}$	The restriction of a function f to a set Ω.
$\Pi(\mu, \nu)$	The set of transport plans from μ to ν.
$\frac{\delta F}{\delta \rho}$	First variation of $F : \mathcal{P}(X) \to \mathbb{R}$, that is $\frac{\mathrm{d}}{\mathrm{d}\epsilon} F(\rho + \epsilon\chi)\Big\vert_{\epsilon=0} = \int \frac{\delta F}{\delta \rho} \mathrm{d}\chi$
W_p	Wasserstein distance of order p.
$\mathbb{W}_p$	Wasserstein space of order p.
γ_T	The transport plan in $\Pi(\mu, \nu)$.
$M(T)$	Monge cost of a map T.
$K(\gamma)$	Kantorovich cost of a plan γ.
$\mu \otimes \nu$	The product measure of μ and ν such that $\mu \otimes \nu(A \times B) = \mu(A)\nu(B)$.
$M^{k \times h}$	The set of real matrices with k rows and h columns.
$M^\top$	Transpose of a matrix M.
i.i.d.	Independent and identical probability distributions.
l.s.c.	Lower semicontinuous.

Definitions.

Definition 1 (Lower Semicontinuity.)**.** *On a complete metric space X, a function $f : X \to \bar{\mathbb{R}}$ is said to be lower semi-continuous (l.s.c.) if for every sequence $(x_n)_{n \in \mathbb{N}}$ converging to $x \in X$, we have*

$$f(x) \leq \liminf_{n \in \mathbb{N}} f(x_n)$$

Definition 2 (Sequentially compact.)**.** *A subset K of a metric space X is said to be compact if from any sequence x_n, we can extract a converging subsequence $x_{n_k} \to x \in K$.*

We can see from the above definition that any continuous function is lower-semicontinuous.

Definition 3 (Compactness.). *A subset K of a metric space X is compact if every open cover of K has a finite subcover.*

Theorem 1. *A subset of a metric space is compact if and only if it is sequentially compact.*

Theorem 2. *Maxima and Minima Let X be a compact metric space and $f : X \to \mathbb{R}$ is continuous, real-valued function. Then f is bounded on X and attains its maximum and minimum. That is, there are x, y belonging to X such that,*

$$f(x) = \inf_{z \in X} f(z) \qquad \text{and} \qquad f(y) = \sup_{z \in X} f(z)$$

Continuity is a strong requirement. Luckily we can assure the existence of a minimizer on lower-semicontinuous functions (or maximizer on upper-semicontinuous). The usual procedure to prove existence of a minimizer is making use of Weierstrass' criterion, we take a minimizer sequence and we prove the space in which we are trying to find a minimizer element is compact.

Theorem 3. *Weierstrass' criterion for existence of minimizers. If $f : X \to \bar{\mathbb{R}}$ is lower semi-continuous and X is compact, then there exists $\hat{x} \in X$.*

Proof. Define $l := \inf\{f(x) : x \in X\} \in \bar{\mathbb{R}}$, notice that $l = +\infty$ only if f is identically $+\infty$, then this case is trivial since any point minimizes f. By compactness there exists a minimizing sequence x_n, that is $f(x_n) \to l$. By compactness we can extract a subsequence converging to some $\hat{x}$ such that $\hat{x} \in X$. By lower-semicontinuity of f, we have that $f(\hat{x}) \liminf_n f(x_n) = l$. Since l is the infimum $l \leq f(\hat{x})$. This proves that $l = f(\hat{x}) \in \mathbb{R}$. □

We can apply the above analysis using a notion of upper-semicontinuity and compactness to find the maximum.

Definition 4. *Topological dual*

Definition 5. *Weak compactness in dual spaces A sequence x_n in a Banach space X is said to be weakly convergin to x, and we write $x_n \rightharpoonup x$, if for every $\xi \in X^*$. We have $\langle \xi, x_n \rangle \to \langle \xi, x \rangle$. A sequence*

Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be two spaces with finite nonnegative measures. On the space $X_1 \times X_2$ we consider sets of the form $\mathcal{A}1 \times \mathcal{A}_2$, where $A_i \in \mathcal{A}_i$, called measurable rectangles. Let $\mu_1 \times \mu_2 (A_1 \times A_2)$:= μ 1 (A 1)μ 2 (A 2). Extending the function μ 1 ×μ 2 by additivity to finite unions of pairwise disjoint measurable rectangles we obtain a finitely additive function on the algebra R generated by such rectangles. We observe that such an extension of μ 1 ×μ 2 to R is well-defined (is independent of partitions of the set into pairwise disjoint measurable rectangles), which is obvious by the additivity of μ 1 and μ 2 . Fi- nally, let A 1 □A 2 denote the σ-algebra generated by all measurable rectangles; this σ-algebra is called the product of the σ-algebras A 1 and A 2 .

Theorem 4. *The set function $\mu_1 \times \mu_2$is countably additive on the algebra generated by all measurable rectangles and uniquely extends to a countably additive measure, denoted by $\mu_1 \otimes \mu_2$, on the Lebesgue completion of this algebra denoted by $\mathcal{A}1 \otimes \mathcal{A}2$*

2

Basics in Convex Analysis.

Proposition 1. *Let $f : X \to \bar{\mathbb{R}}$ be a convex and lower-semicontinuous function. Assume that there exists $x_0 \in X$ such that $f(x_0) = -\infty$. Then f is nowhere finite on X.*

3

Linear Programming

Interior Methods

4

Optimal Transport Theory

Problem 1. *Given two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and a cost function $c : X \times Y \to \{0, +\infty\}$, solve*

$$\inf\left\{M(T) := \int c(x, T(x))\mathrm{d}\mu(x) : \quad T_{\#}\mu = \nu\right\} \tag{MP}$$

Definition 6. *coupling*

Problem 2. *Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to [0, +\infty]$, we consider the problem*

$$\inf\left\{K(\gamma) := \int_{X\times Y} c\mathrm{d}\gamma : \gamma \in \Pi(\mu, \nu)\right\} \tag{KP}$$

where $\Pi(\mu, \nu)$ is the set of transport plans.

Kantorovich formulation as relaxation

5

Computational Optimal Transport

Linear Programming Formulation.

Sinkhorn-Knopp Algorithm.

Simplex Method Algorithm.

Simulated Annealing.

Continuous Formulation.

6

Applications

Nash Equilibrium.
Track of a Dynamic.
Domain Adaptation.
Isoperimetric Inequality.
Barycenter of a Fourier Power Spectrum.

Bibliography