Progress in Nonlinear Differential Equations
and Their Applications
87

Filippo Santambrogio

# Optimal Transport for Applied Mathematicians

## Calculus of Variations, PDEs, and Modeling

Birkhäuser

Filippo Santambrogio

# Optimal Transport for Applied Mathematicians

## Calculus of Variations, PDEs, and Modeling



Birkhäuser

Filippo Santambrogio
Laboratoire de Mathématiques d'Orsay
Université Paris-Sud
Orsay, France

ISSN 1421-1750 ISSN 2374-0280 (electronic)
Progress in Nonlinear Differential Equations and Their Applications
ISBN 978-3-319-20827-5 ISBN 978-3-319-20828-2 (eBook)
DOI 10.1007/978-3-319-20828-2

Library of Congress Control Number: 2015946595

Mathematics Subject Classification (2010): 49J45, 90C05, 35J96, 35K10, 39B62, 90B06, 91A13, 90B20, 26B25, 49N15, 52A41, 49M29, 76M25

Springer Cham Heidelberg New York Dordrecht London


Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Birkhäuser

# Progress in Nonlinear Differential Equations and Their Applications

Volume 87



More information about this series at http://www.springer.com/series/4889

*To my sweet wife Anna,*
*who is furious for my book was conceived*
*after hers but is ready long before…*
*but she seems to love me anyway*

# Preface

## Why this book?

Why a new book on optimal transport? Were the two books by Fields Medalist Cédric Villani not enough? And what about the Bible of Gradient Flows, the book by Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, which also contains many advanced and general details about optimal transport?

The present text, very vaguely inspired by the classes that I gave in Orsay in 2011 and 2012 and by two short introductory courses that I gave earlier in a summer school in Grenoble [274, 275], would like to propose a different point of view and is partially addressed to a different audience. There is nowadays a large and expanding community working with optimal transport as a tool to do applied mathematics. We can think in particular of applications to image processing, economics, and evolution PDEs, in particular when modeling population dynamics in biology or social sciences, or fluid mechanics. More generally, in applied mathematics, optimal transport is both a technical tool to perform proofs, do estimates, and suggest numerical methods and a modeling tool to describe phenomena where distances, paths, and costs are involved.

For those who arrive at optimal transport from this side, some of the most important issues are how to handle numerically the computations of optimal maps and costs, which are the different formulations of the problem, so as to adapt them to new models and purposes; how to obtain in the most concrete ways the main results connecting transport and PDEs; and how the theory is used in the approach of existing models. We cannot say that the answers to these questions are not contained in the existing books, but it is true that probably none of them have been written with this purpose.

The first book by C. Villani [292] is the closest to our intentions. It is a wide introduction to the topic and its applications, suitable for every audience. Yet, some of the subjects that I decided to deal with here are unfortunately not described in [292] (e.g., the minimal flow problems discussed in Chapter 4 of this book). Also, since 2003, the theory has enormously advanced. Also the books by ST Rachev and L Ruschednorf (two volumes, [257, 258]) should not be forgotten, as they cover many applications in probability and statistics. But their scopes diverge quite soon from ours, and we will not develop most of the applications nor the language developed in [257]. Indeed, we will mainly stick to a deterministic framework and to a variational taste.

If we look at what happened after [292], we should mention at least two beautiful books which appeared since then. The new book by C. Villani [293] has expanded the previous presentation into an almost-thousand-page volume, where most of the extra content actually deals with geometrical issues, in particular the notion of curvature. Optimal transport for the quadratic cost on a manifold becomes a central tool, and it is the starting point to study the case of metric measure spaces. The other reference book is the one by L. Ambrosio, N. Gigli, and G. Savaré [15], devoted to the study of gradient flow evolution in metric space and in particular in the Wasserstein space of probability measures. This topic has many interesting applications (e.g., the heat equation, Fokker-Planck equation, porous media, etc.), and the tools that are developed are very useful. Yet, the main point of view of the authors is the study of the hidden differential structures in these abstract spaces; modeling and applied mathematics were probably not their first concerns.

Some of the above references are very technical and develop powerful abstract machineries that can be applied in very general situations but could be difficult to use for many readers. As a consequence, some shorter surveys and more tractable lecture notes have appeared (e.g., [11] for a simplified version of some parts of [15] or [9] as a short introduction to the topic by L. Ambrosio and N. Gigli). Yet, a deeper analysis of the content of [9] shows that the first half deals with the general theory of optimal transport, with some variations, while the rest is devoted to gradient flows in metric spaces in their generality and to metric measure spaces with curvature bounds. We also mention two very recent survey, [63] and [228]: the former accounts for the main achievements on the theory on the occasion of the centennial of the birth of L. V. Kantorovich, and the second gives a general presentation of the topic with connections with geometry and economics.

In the meantime, many initiatives took place underlining the increasing interest in the applied side of optimal transport: publications,[1] schools, workshops, research projects, etc. The community is very lively since the 2010s in particular in France but also in Canada, Italy, Austria, the UK, etc. All these considerations suggested that a dedicated book could have been a good idea, and the one that you have in your hands now is the result of the work for the last three years.

[1]A special issue of ESAIM M2AN on "Optimal Transport in Applied Mathematics" is in preparation and some of the bibliographical references of the present book are taken from such an issue.

## What about this book?

This book contains a rigorous description of the theory of optimal transport and of some neglected variants and explains the most important connections that it has with many topics in evolution PDEs, image processing, and economics.

I avoided as much as possible the most general frameworks and concentrated on the Euclidean case, except where statements in Polish spaces did not cost any more and happened to help in making the picture clearer. I skipped many difficulties by choosing to add compactness assumptions every time that this simplified the exposition without reducing too much the interest of the statement (to my personal taste). Also, in many cases, I first started from the easiest cases (e.g., with compactness or continuity) before generalizing.

When a choice was possible, I tried to prefer more "concrete" proofs, which I think are easier to figure out for the readers. As an example, the existence of velocity fields for Lipschitz curves in Wasserstein spaces has been proven by approximation and not via abstract functional analysis tools, as in [15], where the main point is a clever use of the Hahn-Banach theorem.

I did not search for an exhaustive survey of all possible topics, but I structured the book into eight chapters, more or less corresponding to one (long) lecture each. Obviously, I added a lot of material to what one could usually deal with during one single lecture, but the choice of the topics and their order really follows an eight-lecture course given in Orsay in 2011 (only exceptions: Chapter 1 and Chapter 5 took one lecture and a half, and the subsequent ones were shortened to half a lecture in 2011). The topics which are too far from those of the eight "lectures" have been omitted from the main body of the chapters. On the other hand, every chapter ends with a discussion section, where extensions, connections, side topics, and different points of view are presented. In some cases (congested and branched transport), these discussion sections correspond to a mini-survey on a related transport model. They are more informal; and sometimes statements are proven while at other times they are only sketched or evoked.

In order to enhance the readership and allow as many people as possible to access the content of this book, I decided to explain some notations in detail that I could have probably considered as known (but it is always better to recall them). Throughout the chapters, some notions are recalled via special boxes called *Memo* or *Important Notion*.[2] A third type of box, called *Good to Know!*, provides extra notions that are not usually part of the background of nonspecialized graduate students in mathematics. The density of the boxes, and of explanatory figures as well, decreases as the book goes on.

For the sake of simplicity, I also decided not to insist too much on at least one important issue: measurability. You can trust me that all the sets, functions, and maps

[2]The difference between the two is just that in the second case, *I* would rather consider it as a Memo, but *students* usually do not agree.

that are introduced throughout the text are indeed measurable as desired, but I did not underline this explicitly. Yet, actually, this is the only concession to sloppiness: proofs are rigorous (at least, I hope) throughout the book and could be used for sure by pure or applied mathematicians looking for a reference on the corresponding subjects. The only chapter where the presentation is a little bit informal is Chapter 6, on numerical methods, in the sense that we do not give proofs of convergence or precise implementation details.

Last but not least, there is a chapter on numerical methods! In particular those that are most linked to PDEs (continuous methods), while the most combinatorial and discrete ones are briefly described in the discussion section.

## For whom is this book?

This book has been written with the point of view of an applied mathematician, and applied mathematicians are supposed to be the natural readership for it. Yet, the ambition is to speak to a much wider public. Pure mathematicians (whether this distinction between pure and applied makes sense is a matter of personal taste) are obviously welcome. They will find rigorous proofs, sometimes inspired by a different point of view. They could be interested in discovering where optimal transport can be used and how and to bring their own contributions.

More generally, the distinction can be moved to the level of people working *with* optimal transport rather than *on* optimal transport (instead of pure vs applied). The former are the natural readership, but the latter can find out they are interested in the content too. In the opposite direction, can we say that the text is also addressed to nonmathematicians (physicists, engineers, theoretical economists, etc.)? This raises the question of the mathematical background that the readers need in order to read it. Obviously, if they have enough mathematical background and if they work on fields close enough to the applications that are presented, it could be interesting for them to see what is behind those applications.

The question of how much mathematics is needed also concerns students. This is a graduate text in mathematics. Even if I tried to give tools to review the required background, it is true that some previous knowledge is required to fully profit from the content of the book. The main prerequisites are measure theory and functional analysis. I deliberately decided not to follow the advice of an "anonymous" referee, who suggested to include an appendix on measure theory. The idea is that mathematicians who want to approach optimal transport should already know something on these subjects (what is a measurable function, what is a measure, which are the conditions for the main convergence theorems, what about $L^p$ and $W^{1,p}$ functions, what is weak convergence, etc.). The goal of the Memo Boxes is to help readers to not get lost. For nonmathematicians reading the book, I hope that the choice of a more concrete approach could help them in finding out what kind of properties is important and reasonable. On the other hand, these readers are also expected to know some mathematical language, and for sure, they will need to put in extra effort to fully profit from it.

Concerning readership, the numerical part (Chapter 6) deserves being discussed a little bit more. Comparing in detail the different methods, their drawbacks, and their strengths, the smartest tricks for their implementation, and discussing the most recent algorithms are beyond the scopes of this book. Hence, this chapter is probably useless for people already working in this field. On the other hand, it can be of interest for people working with optimal transport without knowing numerical methods or for numericists who are not into optimal transport.

Also, I am certain that it will be possible to use some of the material that I present here for a graduate course on these topics because of the many boxes recalling the main background notions and the exercises at the end of the book.

## What is in this book?

After this preface and a short introduction to optimal transport (where I mainly present the problem, its history, and its main connections with other part of mathematics), this book contains eight chapters. The two most important chapters, those which constitute the general theory of optimal transport, are Chapters 1 and 5. In the structure of the book, the first half of the text is devoted to the problem of the optimal way of transporting mass from a given measure to another (in the Monge-Kantorovich framework and then with a minimal flow approach), and Chapter 1 is the most important. Then, in the second half, I move to the case where measures vary, which is indeed the case in Chapter 5 and later in Chapters 7 and 8. Chapter 6 comes after Chapter 5 because of the connections of the Benamou-Brenier method with geodesics in the Wasserstein space.

Chapter 1 presents the relaxation that Kantorovich did of the original Monge problem and its duality issues (Kantorovich potentials, $c$-cyclical monotonicity, etc.). It uses these tools to provide the first theorem of existence of an optimal map (Brenier theorem). The discussion section as well mainly stems from the Kantorovich interpretation and duality.

Chapter 2 focuses on the unidimensional case, which is easier and already has many consequences. Then, the Knothe map is presented; it is a transport map built with 1D bricks, and its degenerate optimality is discussed. The main notion here is that of monotone transport. In the discussion section, 1D and monotone maps are used for applications in mathematics (isoperimetric inequalities) and outside mathematics (histogram equalization in image processing).

Chapter 3 deals with some limit cases, not covered in Chapter 1. Indeed, from the results of the first chapter, we know how to handle transport costs of the form $|x-y|^p$ for $p \in (1, +\infty)$, but not $p = 1$, which was the original question by Monge. This requires to use some extra ideas, in particular selecting a special minimizer via a secondary variational problem. Similar techniques are also needed for the other limit case, i.e., $p = \infty$, which is also detailed in the chapter. In the discussion section we present the main challenges and methods to tackle the general problem of convex costs of the form $h(y - x)$ (without strict convexity and with possible infinite values), which has been a lively research subject in the last few years, and later we consider the case $0 < p < 1$, i.e., costs which are concave in the distance.

Chapter 4 presents alternative formulations, more Eulerian in spirit: how to describe a transportation phenomenon via a flow, i.e., a vector field $\mathbf{w}$ with prescribed divergence, and minimize the total cost via functionals involving $\mathbf{w}$. When we minimize the $L^1$ norm of $\mathbf{w}$, this turns out to be equivalent to the original problem by Monge. The main body of the chapter provides the dictionary to pass from Lagrangian to Eulerian frameworks and back and studies this minimization problem and its solutions. In the discussion section, two variants are proposed: traffic congestion (with strictly convex costs in $\mathbf{w}$) and branched transport (with concave costs in $\mathbf{w}$).

Chapter 5 introduces another essential tool in the theory of optimal transport: the distances (called Wasserstein distances) induced by transport costs on the space of measures. After studying their properties, we study the curves in these spaces, and in particular geodesics, and we underline the connection with the continuity equation. The discussion section makes a comparison between Wasserstein distances and other distances on probabilities and finally describes an application in terms of barycenters of measures.

Chapter 6 starts from the ideas presented in the previous chapter and uses them to propose numerical methods. Indeed, in the description of the Wasserstein space, one can see that finding the optimal transport is equivalent to minimizing a kinetic energy functional among solutions of the continuity equation. This provided the first numerical method for optimal transport called the *Benamou-Brenier* method. In the rest of the chapter, two other "continuous" numerical methods are described, and the discussion section deals with discrete and semidiscrete methods.

Chapter 7 contains a "bestiary" of useful functionals defined over measures and studies their properties, not necessarily in connection with optimal transport (convexity, semicontinuity, computing the first variation, etc.). The idea is that these functionals often appear in modeling issues accompanied by transport distances. Also, the notion of displacement convexity (i.e., convexity along geodesics of the Wasserstein space) is described in detail. The discussion section is quite heterogeneous, with applications to geometrical inequalities but also equilibria in urban regions.

Chapter 8 gives an original presentation of one of the most striking applications of optimal transport: gradient flows in Wasserstein spaces, which allow us to deal with many evolution equations, in particular of the parabolic type. The general framework is presented, and the Fokker-Planck equation is studied in detail. The discussion section presents other equations which have this gradient flow structure and also other evolution equations where optimal transport plays a role, without being gradient flows.

Before the detailed bibliography and the index which conclude the book, there is a list of 69 exercises from the various chapters and of different levels of difficulties. From students to senior researchers, the readers are invited to play with these exercises and enjoy the taste of optimal transport.

Orsay, France
May 2015

Filippo Santambrogio

## Introduction to optimal transport

The history of optimal transport began a long time ago in France, a few years before the revolution, when Gaspard Monge proposed the following problem in a report that he submitted to the *Académie des Sciences* [239].[3] Given two densities of mass $f, g \geq 0$ on $\mathbb{R}^d$, with $\int f(x)\,\mathrm{d}x = \int g(y)\,\mathrm{d}y = 1$, find a map $\mathrm{T} : \mathbb{R}^d \to \mathbb{R}^d$, pushing the first one onto the other, i.e. such that

$$\int_A g(y)\,\mathrm{d}y = \int_{\mathrm{T}^{-1}(A)} f(x)\,\mathrm{d}x \quad \text{for any Borel subset } A \subset \mathbb{R}^d, \tag{1}$$

and minimizing the quantity

$$M(\mathrm{T}) := \int_{\mathbb{R}^d} |\mathrm{T}(x) - x| f(x)\,\mathrm{d}x$$

among all the maps satisfying this condition. This means that we have a collection of particles, distributed according to the density $f$ on $\mathbb{R}^d$, that have to be moved so that they form a new distribution whose density is prescribed and is $g$. The movement has to be chosen so as to minimize the average displacement. In the description of Monge, the starting density $f$ represented a distribution of sand that had to be moved to a target configuration $g$. These two configurations correspond to what was called in French *déblais* and *remblais*. Obviously, the dimension of the space was only supposed to be $d = 2$ or 3. The map T describes the movement (that we must choose in an optimal way), and $\mathrm{T}(x)$ represents the destination of the particle originally located at $x$.

In the following, we will often use the image measure of a measure $\mu$ on $X$ (measures will indeed replace the densities $f$ and $g$ in the most general formulation of the problem) through a measurable map $\mathrm{T} : X \to Y$: it is the measure on $Y$ denoted by $\mathrm{T}_{\#}\mu$ and characterized, as in (1), by

$$(\mathrm{T}_{\#}\mu)(A) = \mu(\mathrm{T}^{-1}(A)) \quad \text{for every measurable set } A$$

$$\text{or} \quad \int_Y \phi \,\mathrm{d}\,(\mathrm{T}_{\#}\mu) = \int_X (\phi \circ \mathrm{T})\,\mathrm{d}\mu \quad \text{for every measurable function } \phi.$$

More generally, we can consider the problem

$$(\mathrm{MP}) \qquad \min\{M(\mathrm{T}) := \int c(x, \mathrm{T}(x))\,\mathrm{d}\mu(x) \;:\; \mathrm{T}_{\#}\mu = \nu\},$$

for a more general transport cost $c : X \times Y \to \mathbb{R}$.

[3]This happened in 1781, but we translate his problem into modern mathematical language.

When we stay in the Euclidean setting, with two measures $\mu, \nu$ induced by densities $f, g$, it is easy – just by a change-of-variables formula – to transform the equality $\nu = T_{\#}\mu$ into the PDE

$$g(T(x)) \det(DT(x)) = f(x), \tag{2}$$

if we suppose $f, g$ and T to be regular enough and T to be injective.

Yet, this equation is highly nonlinear in T, and this is one of the difficulties preventing an easy analysis of the Monge problem. For instance: how do we prove the existence of a minimizer? Usually, what one does is the following: take a minimizing sequence $T_n$, find a bound on it giving compactness in some topology (here, if the support of $\nu$ is compact, the maps $T_n$ take value in a common bounded set, $\mathrm{spt}(\nu)$, and so one can get compactness of $T_n$ in the weak-* $L^\infty$ convergence), take a limit $T_n \rightharpoonup T$, and prove that T is a minimizer. This requires semicontinuity of the functional $M$ with respect to this convergence (which is true in many cases, for instance, if $c$ is convex in its second variable): we need $T_n \rightharpoonup T \Rightarrow \liminf_n M(T_n) \geq M(T)$), but we also need that the limit T still satisfies the constraint. Yet, the nonlinearity of the PDE prevents us from proving this stability when we only have weak convergence (the reader can find an example of a weakly converging sequence such that the corresponding image measures do not converge as an exercise; it is actually **Ex**(1) in the list of exercises).

In [239], Monge analyzed fine questions on the geometric properties of the solution to this problem, and he underlined several important ideas that we will see in Chapter 3: the fact that transport rays do not meet, that they are orthogonal to a particular family of surfaces, and that a natural choice along transport rays is to order the points in a monotone way. Yet, he did not really solve the problem. The question of the existence of a minimizer was not even addressed. In the next 150 years, the optimal transport problem mainly remained intimately French, and the *Académie des Sciences* offered a prize on this question. The first prize was won by P. Appell [21] with a long mémoire which improved some points but was far from being satisfactory (and did not address the existence issue[4]).

The problem of Monge has stayed with no solution (does a minimizer exist? how to characterize it?) until progress was made in the 1940s. Indeed, only with the work by Kantorovich (1942, see [200]), it was inserted into a suitable framework which gave the possibility to attack it and, later, to provide solutions and study them. The problem has then been widely generalized, with very general cost functions $c(x, y)$ instead of the Euclidean distance $|x - y|$ and more general measures and spaces. The main idea by Kantorovich is that of looking at Monge's problem as connected to linear programming. Kantorovich indeed decided to change the point of view, and to describe the movement of the particles via a measure $\gamma$ on $X \times Y$, satisfying $(\pi_x)_{\#}\gamma = \mu$ and $(\pi_y)_{\#}\gamma = \nu$. These probability measures over $X \times Y$

[4]The reader can see [181] – in French, sorry – for more details on these historical questions about the work by Monge and the content of the papers presented for this prize.

are an alternative way to describing the displacement of the particles of $\mu$: instead of giving, for each $x$, the destination $T(x)$ of the particle originally located at $x$, we give for each pair $(x, y)$ the number of particles going from $x$ to $y$. It is clear that this description allows for more general movements, since from a single point $x$, particles can a priori move to different destinations $y$. If multiple destinations really occur, then this movement cannot be described through a map T. The cost to be minimized becomes simply $\int_{X\times Y} c \, d\gamma$. We have now a linear problem, under linear constraints. It is possible to prove existence of a solution and to characterize it by using techniques from convex optimization, such as *duality*, in order to characterize the optimal solution (see Chapter 1).

In some cases, and in particular if $c(x, y) = |x - y|^2$ (another very natural cost, with many applications in physical modeling because of its connection with kinetic energy), it is even possible to prove that the optimal $\gamma$ does not allow this splitting of masses. Particles at $x$ are only sent to a unique destination $T(x)$, thus providing a solution to the original problem by Monge. This is what is done by Brenier in [82], where he also proves a very special form for the optimal map: the optimal T is of the form $T(x) = \nabla u(x)$, for a convex function $u$. This makes, by the way, a strong connection with the Monge-Ampère equation. Indeed, from (2), we get

$$\det(D^2 u(x)) = \frac{f(x)}{g(\nabla u(x))},$$

which is an (degenerate and nonlinear) elliptic equation exactly in the class of convex functions. Brenier also uses this result to provide an original *polar factorization* theorem for vector maps (see Section 1.7.2): vector fields can be written as the composition of the gradient of a convex function and of a measure-preserving map. This generalizes the fact that matrices can be written as the product of a symmetric positive-definite matrix and an orthogonal one.

The results by Brenier can be easily adapted to other costs, strictly convex functions of the difference $x-y$. They have also been adapted to the squared distance on Riemannian manifolds (see [231]). But the original cost proposed by Monge, the distance itself, was much more difficult.

After the French school, it was time for the Russian mathematicians. From the precise approach introduced by Kantorovich, Sudakov [290] proposed a solution for the original Monge problem (MP). The optimal transport plan $\gamma$ in the Kantorovich problem with cost $|x - y|$ has the following property: the space $\mathbb{R}^d$ can be decomposed in an essentially disjoint union of segments that are preserved by $\gamma$ (i.e., $\gamma$ is concentrated on pairs $(x, y)$ belonging to the same segment). These segments are built from a Lipschitz function, whose level sets are the surfaces "foreseen" by Monge. Then, it is enough to reduce the problem to a family of 1D problems. If $\mu \ll \mathscr{L}^d$, the measures that $\mu$ induces on each segment should also be absolutely continuous and have no atoms. And in dimension one, as soon as the source measure has no atoms, one can define a monotone increasing transport, which is optimal for any convex cost of the difference $x - y$.