

Grundlehren der
mathematischen Wissenschaften 338

A Series of Comprehensive Studies in Mathematics

Series editors

M. Berger B. Eckmann P. de la Harpe
F. Hirzebruch N. Hitchin L. Hörmander
A. Kupiainen G. Lebeau M. Ratner
D. Serre Ya. G. Sinai N.J.A. Sloane
A.M. Vershik M. Waldschmidt

Editor-in-Chief

A. Chenciner J. Coates S.R.S. Varadhan

Cédric Villani

Optimal Transport

Old and New

 Springer

Cédric Villani

Unité de Mathématiques Pures et Appliquées (UMPA)

École Normale Supérieure de Lyon

46, allée d'Italie

69364 Lyon CX 07

France

cvillani@umpa.ens-lyon.fr

ISSN 0072-7830

ISBN 978-3-540-71049-3

e-ISBN 978-3-540-71050-9

Library of Congress Control Number: 2008932183

Mathematics Subject Classification Numbers (2000): 49-xx, 53-xx, 60-xx

© 2009 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: WMX Design GmbH, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Do mo chuisle mo chroí, Aëlle

Preface

When I was first approached for the 2005 edition of the Saint-Flour Probability Summer School, I was intrigued, flattered and scared.¹ Apart from the challenge posed by the teaching of a rather analytical subject to a probabilistic audience, there was the danger of producing a remake of my recent book *Topics in Optimal Transportation*.

However, I gradually realized that I was being offered a unique opportunity to rewrite the whole theory from a different perspective, with alternative proofs and a different focus, and a more probabilistic presentation; plus the incorporation of recent progress. Among the most striking of these recent advances, there was the rising awareness that John Mather's minimal measures had a lot to do with optimal transport, and that both theories could actually be embedded within a single framework. There was also the discovery that optimal transport could provide a robust synthetic approach to Ricci curvature bounds. These links with dynamical systems on one hand, differential geometry on the other hand, were only briefly alluded to in my first book; here on the contrary they will be at the basis of the presentation. To summarize: more probability, more geometry, and more dynamical systems. Of course there cannot be more of everything, so in some sense there is less analysis and less physics, and also there are fewer digressions.

So the present course is by no means a reduction or an expansion of my previous book, but should be regarded as a complementary reading. Both sources can be read independently, or together, and hopefully the complementarity of points of view will have pedagogical value.

¹ Fans of Tom Waits may have identified this quotation.

Throughout the book I have tried to optimize the results and the presentation, to provide complete and self-contained proofs of the most important results, and comprehensive bibliographical notes — a dauntingly difficult task in view of the rapid expansion of the literature. Many statements and theorems have been written specifically for this course, and many results appear in rather sharp form for the first time. I also added several appendices, either to present some domains of mathematics to non-experts, or to provide proofs of important auxiliary results. All this has resulted in a rapid growth of the document, which in the end is about six times (!) the size that I had planned initially. So **the non-expert reader is advised to skip long proofs at first reading**, and concentrate on explanations, statements, examples and sketches of proofs when they are available.

About terminology: For some reason I decided to switch from “transportation” to “transport”, but this really is a matter of taste.

For people who are already familiar with the theory of optimal transport, here are some more serious changes.

Part I is devoted to a qualitative description of optimal transport. The dynamical point of view is given a prominent role from the beginning, with Robert McCann’s concept of displacement interpolation. This notion is discussed before any theorem about the solvability of the Monge problem, in an abstract setting of “Lagrangian action” which generalizes the notion of length space. This provides a unified picture of recent developments dealing with various classes of cost functions, in a smooth or nonsmooth context.

I also wrote down in detail some important estimates by John Mather, well-known in certain circles, and made extensive use of them, in particular to prove the Lipschitz regularity of “intermediate” transport maps (starting from some intermediate time, rather than from initial time). Then the absolute continuity of displacement interpolants comes for free, and this gives a more unified picture of the Mather and Monge–Kantorovich theories. I rewrote in this way the classical theorems of solvability of the Monge problem for quadratic cost in Euclidean space. Finally, this approach allows one to treat change of variables formulas associated with optimal transport by means of changes of variables that are Lipschitz, and not just with bounded variation.

Part II discusses optimal transport in Riemannian geometry, a line of research which started around 2000; I have rewritten all these applications in terms of Ricci curvature, or more precisely curvature-

dimension bounds. This part opens with an introduction to Ricci curvature, hopefully readable without any prior knowledge of this notion.

Part III presents a synthetic treatment of Ricci curvature bounds in metric-measure spaces. It starts with a presentation of the theory of Gromov–Hausdorff convergence; all the rest is based on recent research papers mainly due to John Lott, Karl-Theodor Sturm and myself.

In all three parts, noncompact situations will be systematically treated, either by limiting processes, or by restriction arguments (the restriction of an optimal transport is still optimal; this is a simple but powerful principle). The notion of approximate differentiability, introduced in the field by Luigi Ambrosio, appears to be particularly handy in the study of optimal transport in noncompact Riemannian manifolds.

Several parts of the subject are not developed as much as they would deserve. Numerical simulation is not addressed at all, except for a few comments in the concluding part. The regularity theory of optimal transport is described in Chapter 12 (including the remarkable recent works of Xu-Jia Wang, Neil Trudinger and Grégoire Loeper), but without the core proofs and latest developments; this is not only because of the technicality of the subject, but also because smoothness is not needed in the rest of the book. Still another poorly developed subject is the Monge–Mather–Mañé problem arising in dynamical systems, and including as a variant the optimal transport problem when the cost function is a distance. This topic is discussed in several treatises, such as Albert Fathi’s monograph, *Weak KAM theorem in Lagrangian dynamics*; but now it would be desirable to rewrite everything in a framework that also encompasses the optimal transport problem. An important step in this direction was recently performed by Patrick Bernard and Boris Buffoni. In Chapter 8 I shall provide an introduction to Mather’s theory, but there would be much more to say.

The treatment of Chapter 22 (concentration of measure) is strongly influenced by Michel Ledoux’s book, *The Concentration of Measure Phenomenon*; while the results of Chapters 23 to 25 owe a lot to the monograph by Luigi Ambrosio, Nicola Gigli and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*. Both references are warmly recommended complementary reading. One can also consult the two-volume treatise by Svetlozar Rachev and Ludger Rüschendorf, *Mass Transportation Problems*, for many applications of optimal transport to various fields of probability theory.

While writing this text I asked for help from a number of friends and collaborators. Among them, Luigi Ambrosio and John Lott are the ones whom I requested most to contribute; this book owes a lot to their detailed comments and suggestions. Most of Part III, but also significant portions of Parts I and II, are made up with ideas taken from my collaborations with John, which started in 2004 as I was enjoying the hospitality of the Miller Institute in Berkeley. Frequent discussions with Patrick Bernard and Albert Fathi allowed me to get the links between optimal transport and John Mather's theory, which were a key to the presentation in Part I; John himself gave precious hints about the history of the subject. Neil Trudinger and Xu-Jia Wang spent vast amounts of time teaching me the regularity theory of Monge–Ampère equations. Alessio Figalli took up the dreadful challenge to check the entire set of notes from first to last page. Apart from these people, I got valuable help from Stefano Bianchini, François Bolley, Yann Brenier, Xavier Cabré, Vincent Calvez, José Antonio Carrillo, Dario Cordero-Erausquin, Denis Feyel, Sylvain Gallot, Wilfrid Gangbo, Diogo Aguiar Gomes, Nathaël Gozlan, Arnaud Guillin, Nicolas Juillet, Kazuhiro Kuwae, Michel Ledoux, Grégoire Loeper, Francesco Maggi, Robert McCann, Shin-ichi Ohta, Vladimir Oliker, Yann Ollivier, Felix Otto, Ludger Rüschendorf, Giuseppe Savaré, Walter Schachermayer, Benedikt Schulte, Theo Sturm, Josef Teichmann, Anthon Thalmaier, Hermann Thorisson, Süleyman Üstünel, Anatoly Vershik, and others.

Short versions of this course were tried on mixed audiences in the Universities of Bonn, Dortmund, Grenoble and Orléans, as well as the Borel seminar in Leysin and the IHES in Bures-sur-Yvette. Part of the writing was done during stays at the marvelous MFO Institute in Oberwolfach, the CIRM in Luminy, and the Australian National University in Canberra. All these institutions are warmly thanked.

It is a pleasure to thank Jean Picard for all his organization work on the 2005 Saint-Flour summer school; and the participants for their questions, comments and bug-tracking, in particular Sylvain Arlot (great bug-tracker!), Fabrice Baudoin, Jérôme Demange, Steve Evans (whom I also thank for his beautiful lectures), Christophe Leuridan, Jan Oblój, Erwan Saint Loubert Bié, and others. I extend these thanks to the joyful group of young PhD students and maîtres de conférences with whom I spent such a good time on excursions, restaurants, quantum ping-pong and other activities, making my stay in Saint-Flour truly wonderful (with special thanks to my personal driver, Stéphane Loisel, and my table tennis sparring-partner and adversary, François Simenhaus). I will cherish my visit there in memory as long as I live!

Typing these notes was mostly performed on my (now defunct) faithful laptop Torsten, a gift of the Miller Institute. Support by the Agence Nationale de la Recherche and Institut Universitaire de France is acknowledged. My eternal gratitude goes to those who made fine typesetting accessible to every mathematician, most notably Donald Knuth for $\text{T}_{\text{E}}\text{X}$, and the developers of $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$, $\text{BIB}_{\text{T}}\text{E}_{\text{X}}$ and XFig . Final thanks to Catriona Byrne and her team for a great editing process.

As usual, I encourage all readers to report mistakes and misprints. **I will maintain a list of errata, accessible from my Web page.**

Lyon, June 2008

Cédric Villani

Contents

Preface	VII
Conventions	XVII
<hr/>	
Introduction	1
<hr/>	
1 Couplings and changes of variables	5
2 Three examples of coupling techniques	21
3 The founding fathers of optimal transport	29
<hr/>	
Part I Qualitative description of optimal transport	39
<hr/>	
4 Basic properties	43
5 Cyclical monotonicity and Kantorovich duality	51
6 The Wasserstein distances	93
7 Displacement interpolation	113
8 The Monge–Mather shortening principle	163
9 Solution of the Monge problem I: Global approach ...	205
10 Solution of the Monge problem II: Local approach ...	215

11	The Jacobian equation	273
12	Smoothness	281
13	Qualitative picture	333

Part II Optimal transport and Riemannian geometry 353

14	Ricci curvature	357
15	Otto calculus	421
16	Displacement convexity I	435
17	Displacement convexity II	449
18	Volume control	493
19	Density control and local regularity	505
20	Infinitesimal displacement convexity	525
21	Isoperimetric-type inequalities	545
22	Concentration inequalities	567
23	Gradient flows I	629
24	Gradient flows II: Qualitative properties	693
25	Gradient flows III: Functional inequalities	719

Part III Synthetic treatment of Ricci curvature 731

26	Analytic and synthetic points of view	735
27	Convergence of metric-measure spaces	743
28	Stability of optimal transport	773

29 Weak Ricci curvature bounds I: Definition and Stability	795
30 Weak Ricci curvature bounds II: Geometric and analytic properties	847
<hr/>	
Conclusions and open problems	903
<hr/>	
References	915
List of short statements	957
List of figures	965
Index	967
Some notable cost functions	971

Conventions

Axioms

I use the classical axioms of set theory; not the full version of the axiom of choice (only the classical axiom of “countable dependent choice”).

Sets and structures

Id is the identity mapping, whatever the space. If A is a set then the function 1_A is the indicator function of A : $1_A(x) = 1$ if $x \in A$, and 0 otherwise. If F is a formula, then 1_F is the indicator function of the set defined by the formula F .

If f and g are two functions, then (f, g) is the function $x \mapsto (f(x), g(x))$. The composition $f \circ g$ will often be denoted by $f(g)$.

\mathbb{N} is the set of *positive* integers: $\mathbb{N} = \{1, 2, 3, \dots\}$. A sequence is written $(x_k)_{k \in \mathbb{N}}$, or simply, when no confusion seems possible, (x_k) .

\mathbb{R} is the set of real numbers. When I write \mathbb{R}^n it is implicitly assumed that n is a positive integer. The Euclidean scalar product between two vectors a and b in \mathbb{R}^n is denoted interchangeably by $a \cdot b$ or $\langle a, b \rangle$. The Euclidean norm will be denoted simply by $|\cdot|$, independently of the dimension n .

$M_n(\mathbb{R})$ is the space of real $n \times n$ matrices, and I_n the $n \times n$ identity matrix. The trace of a matrix M will be denoted by $\text{tr } M$, its determinant by $\det M$, its adjoint by M^* , and its Hilbert–Schmidt norm $\sqrt{\text{tr}(M^*M)}$ by $\|M\|_{\text{HS}}$ (or just $\|M\|$).

Unless otherwise stated, Riemannian manifolds appearing in the text are finite-dimensional, smooth and complete. If a Riemannian manifold M is given, I shall usually denote by n its dimension, by d the geodesic distance on M , and by vol the volume (= n -dimensional

Hausdorff) measure on M . The tangent space at x will be denoted by $T_x M$, and the tangent bundle by TM . The norm on $T_x M$ will most of the time be denoted by $|\cdot|$, as in \mathbb{R}^n , without explicit mention of the point x . (The symbol $\|\cdot\|$ will be reserved for special norms or functional norms.) If S is a set without smooth structure, the notation $T_x S$ will instead denote the tangent cone to S at x (Definition 10.46).

If Q is a quadratic form defined on \mathbb{R}^n , or on the tangent bundle of a manifold, its value on a (tangent) vector v will be denoted by $\langle Q \cdot v, v \rangle$, or simply $Q(v)$.

The open ball of radius r and center x in a metric space \mathcal{X} is denoted interchangeably by $B(x, r)$ or $B_r(x)$. If \mathcal{X} is a Riemannian manifold, the distance is of course the geodesic distance. The closed ball will be denoted interchangeably by $B[x, r]$ or $B_{r]}(x)$. The diameter of a metric space \mathcal{X} will be denoted by $\text{diam}(\mathcal{X})$.

The closure of a set A in a metric space will be denoted by \bar{A} (this is also the set of all limits of sequences with values in A).

A metric space \mathcal{X} is said to be *locally compact* if every point $x \in \mathcal{X}$ admits a compact neighborhood; and *boundedly compact* if every closed and bounded subset of \mathcal{X} is compact.

A map f between metric spaces (\mathcal{X}, d) and (\mathcal{X}', d') is said to be C -Lipschitz if $d'(f(x), f(y)) \leq C d(x, y)$ for all x, y in \mathcal{X} . The best admissible constant C is then denoted by $\|f\|_{\text{Lip}}$.

A map is said to be locally Lipschitz if it is Lipschitz on bounded sets, *not necessarily compact* (so it makes sense to speak of a locally Lipschitz map defined almost everywhere).

A curve in a space \mathcal{X} is a continuous map defined on an interval of \mathbb{R} , valued in \mathcal{X} . For me the words “curve” and “path” are synonymous. The time- t evaluation map e_t is defined by $e_t(\gamma) = \gamma_t = \gamma(t)$.

If γ is a curve defined from an interval of \mathbb{R} into a metric space, its length will be denoted by $\mathcal{L}(\gamma)$, and its speed by $|\dot{\gamma}|$; definitions are recalled on p. 119.

Usually geodesics will be *minimizing, constant-speed* geodesic curves. If \mathcal{X} is a metric space, $\Gamma(\mathcal{X})$ stands for the space of all geodesics $\gamma : [0, 1] \rightarrow \mathcal{X}$.

Being given x_0 and x_1 in a metric space, I denote by $[x_0, x_1]_t$ the set of all t -barycenters of x_0 and x_1 , as defined on p. 393. If A_0 and A_1 are two sets, then $[A_0, A_1]_t$ stands for the set of all $[x_0, x_1]_t$ with $(x_0, x_1) \in A_0 \times A_1$.

Function spaces

$C(\mathcal{X})$ is the space of continuous functions $\mathcal{X} \rightarrow \mathbb{R}$, $C_b(\mathcal{X})$ the space of bounded continuous functions $\mathcal{X} \rightarrow \mathbb{R}$; and $C_0(\mathcal{X})$ the space of continuous functions $\mathcal{X} \rightarrow \mathbb{R}$ converging to 0 at infinity; all of them are equipped with the norm of uniform convergence $\|\varphi\|_\infty = \sup |\varphi|$. Then $C_b^k(\mathcal{X})$ is the space of k -times continuously differentiable functions $u : \mathcal{X} \rightarrow \mathbb{R}$, such that all the partial derivatives of u up to order k are bounded; it is equipped with the norm given by the supremum of all norms $\|\partial u\|_{C_b}$, where ∂u is a partial derivative of order at most k ; $C_c^k(\mathcal{X})$ is the space of k -times continuously differentiable functions with compact support; etc. When the target space is not \mathbb{R} but some other space \mathcal{Y} , the notation is transformed in an obvious way: $C(\mathcal{X}; \mathcal{Y})$, etc.

L^p is the Lebesgue space of exponent p ; the space and the measure will often be implicit, but clear from the context.

Calculus

The derivative of a function $u = u(t)$, defined on an interval of \mathbb{R} and valued in \mathbb{R}^n or in a smooth manifold, will be denoted by u' , or more often by \dot{u} . The notation d^+u/dt stands for the upper right-derivative of a real-valued function u : $d^+u/dt = \limsup_{s \downarrow 0} [u(t+s) - u(t)]/s$.

If u is a function of several variables, the partial derivative with respect to the variable t will be denoted by $\partial_t u$, or $\partial u/\partial t$. *The notation u_t does not stand for $\partial_t u$, but for $u(t)$.*

The gradient operator will be denoted by grad or simply ∇ ; the divergence operator by div or $\nabla \cdot$; the Laplace operator by Δ ; the Hessian operator by Hess or ∇^2 (so ∇^2 *does not* stand for the Laplace operator). The notation is the same in \mathbb{R}^n or in a Riemannian manifold. Δ is the divergence of the gradient, so it is typically a nonpositive operator. The value of the gradient of \underline{f} at point x will be denoted either by $\nabla_x f$ or $\nabla f(x)$. The notation $\tilde{\nabla}$ stands for the approximate gradient, introduced in Definition 10.2.

If T is a map $\mathbb{R}^n \rightarrow \mathbb{R}^n$, ∇T stands for the Jacobian matrix of T , that is the matrix of all partial derivatives $(\partial T_i/\partial x_j)$ ($1 \leq i, j \leq n$).

All these differential operators will be applied to (smooth) functions but also to measures, by duality. For instance, the Laplacian of a measure μ is defined via the identity $\int \zeta d(\Delta\mu) = \int (\Delta\zeta) d\mu$ ($\zeta \in C_c^2$). The notation is consistent in the sense that $\Delta(f\text{vol}) = (\Delta f)\text{vol}$. Similarly, I shall take the divergence of a vector-valued measure, etc.

$f = o(g)$ means $f/g \rightarrow 0$ (in an asymptotic regime that should be clear from the context), while $f = O(g)$ means that f/g is bounded.

\log stands for the natural logarithm with base e .

The positive and negative parts of $x \in \mathbb{R}$ are defined respectively by $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$; both are nonnegative, and $|x| = x_+ + x_-$. The notation $a \wedge b$ will sometimes be used for $\min(a, b)$. All these notions are extended in the usual way to functions and also to signed measures.

Probability measures

δ_x is the Dirac mass at point x .

All measures considered in the text are Borel measures on **Polish spaces**, which are complete, separable metric spaces, equipped with their Borel σ -algebra. I shall usually not use the completed σ -algebra, except on some rare occasions (emphasized in the text) in Chapter 5.

A measure is said to be finite if it has finite mass, and locally finite if it attributes finite mass to compact sets.

The space of Borel probability measures on \mathcal{X} is denoted by $P(\mathcal{X})$, the space of finite Borel measures by $M_+(\mathcal{X})$, the space of signed finite Borel measures by $M(\mathcal{X})$. The total variation of μ is denoted by $\|\mu\|_{\text{TV}}$.

The integral of a function f with respect to a probability measure μ will be denoted interchangeably by $\int f(x) d\mu(x)$ or $\int f(x) \mu(dx)$ or $\int f d\mu$.

If μ is a Borel measure on a topological space \mathcal{X} , a set N is said to be μ -negligible if N is included in a Borel set of zero μ -measure. Then μ is said to be concentrated on a set C if $\mathcal{X} \setminus C$ is negligible. (If C itself is Borel measurable, this is of course equivalent to $\mu[\mathcal{X} \setminus C] = 0$.) By abuse of language, I may say that \mathcal{X} has full μ -measure if μ is concentrated on \mathcal{X} .

If μ is a Borel measure, its support $\text{Spt } \mu$ is the smallest *closed* set on which it is concentrated. The same notation Spt will be used for the support of a continuous function.

If μ is a Borel measure on \mathcal{X} , and T is a Borel map $\mathcal{X} \rightarrow \mathcal{Y}$, then $T_{\#}\mu$ stands for the image measure² (or push-forward) of μ by T : It is a Borel measure on \mathcal{Y} , defined by $(T_{\#}\mu)[A] = \mu[T^{-1}(A)]$.

The law of a random variable X defined on a probability space (Ω, \mathbb{P}) is denoted by law (X) ; this is the same as $X_{\#}\mathbb{P}$.

The weak topology on $P(\mathcal{X})$ (or topology of weak convergence, or narrow topology) is induced by convergence against $C_b(\mathcal{X})$, i.e. *bounded*

² Depending on the authors, the measure $T_{\#}\mu$ is often denoted by $T\#\mu$, $T_*\mu$, $T(\mu)$, $T\mu$, $\int \delta_{T(a)} \mu(da)$, $\mu \circ T^{-1}$, μT^{-1} , or $\mu[T \in \cdot]$.

continuous test functions. If \mathcal{X} is Polish, then the space $P(\mathcal{X})$ itself is Polish. Unless explicitly stated, I do not use the weak-* topology of measures (induced by $C_0(\mathcal{X})$ or $C_c(\mathcal{X})$).

When a probability measure is clearly specified by the context, it will sometimes be denoted just by \mathbb{P} , and the associated integral, or expectation, will be denoted by \mathbb{E} .

If $\pi(dx dy)$ is a probability measure in two variables $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, its marginal (or projection) on \mathcal{X} (resp. \mathcal{Y}) is the measure $X_{\#}\pi$ (resp. $Y_{\#}\pi$), where $X(x, y) = x$, $Y(x, y) = y$. If (x, y) is random with law $(x, y) = \pi$, then the conditional law of x given y is denoted by $\pi(dx|y)$; this is a measurable function $\mathcal{Y} \rightarrow P(\mathcal{X})$, obtained by disintegrating π along its y -marginal. The conditional law of y given x will be denoted by $\pi(dy|x)$.

A measure μ is said to be absolutely continuous with respect to a measure ν if there exists a measurable function f such that $\mu = f \nu$.

Notation specific to optimal transport and related fields

If $\mu \in P(\mathcal{X})$ and $\nu \in P(\mathcal{Y})$ are given, then $\Pi(\mu, \nu)$ is the set of all joint probability measures on $\mathcal{X} \times \mathcal{Y}$ whose marginals are μ and ν .

$C(\mu, \nu)$ is the optimal (total) cost between μ and ν , see p. 80. It implicitly depends on the choice of a cost function $c(x, y)$.

For any $p \in [1, +\infty)$, W_p is the Wasserstein distance of order p , see Definition 6.1; and $P_p(\mathcal{X})$ is the Wasserstein space of order p , i.e. the set of probability measures with finite moments of order p , equipped with the distance W_p , see Definition 6.4.

$P_c(\mathcal{X})$ is the set of probability measures on \mathcal{X} with compact support.

If a reference measure ν on \mathcal{X} is specified, then $P^{\text{ac}}(\mathcal{X})$ (resp. $P_p^{\text{ac}}(\mathcal{X})$, $P_c^{\text{ac}}(\mathcal{X})$) stands for those elements of $P(\mathcal{X})$ (resp. $P_p(\mathcal{X})$, $P_c(\mathcal{X})$) which are absolutely continuous with respect to ν .

\mathcal{DC}_N is the displacement convexity class of order N (N plays the role of a dimension); this is a family of convex functions, defined on p. 443 and in Definition 17.1.

U_ν is a functional defined on $P(\mathcal{X})$; it depends on a convex function U and a reference measure ν on \mathcal{X} . This functional will be defined at various levels of generality, first in equation (15.2), then in Definition 29.1 and Theorem 30.4.

$U_{\pi, \nu}^\beta$ is another functional on $P(\mathcal{X})$, which involves not only a convex function U and a reference measure ν , but also a coupling π and a distortion coefficient β , which is a nonnegative function on $\mathcal{X} \times \mathcal{X}$: See again Definition 29.1 and Theorem 30.4.

The Γ and Γ_2 operators are quadratic differential operators associated with a diffusion operator; they are defined in (14.47) and (14.48).

$\beta_t^{(K,N)}$ is the notation for the distortion coefficients that will play a prominent role in these notes; they are defined in (14.61).

$\text{CD}(K, N)$ means “curvature-dimension condition (K, N) ”, which morally means that the Ricci curvature is bounded below by Kg (K a real number, g the Riemannian metric) and the dimension is bounded above by N (a real number which is not less than 1).

If $c(x, y)$ is a cost function then $\check{c}(y, x) = c(x, y)$. Similarly, if $\pi(dx dy)$ is a coupling, then $\tilde{\pi}$ is the coupling obtained by swapping variables, that is $\tilde{\pi}(dy dx) = \pi(dx dy)$, or more rigorously, $\tilde{\pi} = S_{\#}\pi$, where $S(x, y) = (y, x)$.

Assumptions **(Super)**, **(Twist)**, **(Lip)**, **(SC)**, **(locLip)**, **(locSC)**, **(H ∞)** are defined on p. 234, **(STwist)** on p. 299, **(Cut $^{n-1}$)** on p. 303.

Introduction

To start, I shall recall in Chapter 1 some basic facts about couplings and changes of variables, including definitions, a short list of famous couplings (Knothe–Rosenblatt coupling, Moser coupling, optimal coupling, etc.); and some important basic formulas about change of variables, conservation of mass, and linear diffusion equations.

In Chapter 2 I shall present, without detailed proofs, three applications of optimal coupling techniques, providing a flavor of the kind of applications that will be considered later.

Finally, Chapter 3 is a short historical perspective about the foundations and development of optimal coupling theory.