V. I. Bogachev, A. V. Kolesnikov, The Monge–Kantorovich problem: achievements, connections, and perspectives, *Russian Mathematical Surveys*, 2012, Volume 67, Issue 5, 785–890

DOI: 10.1070/RM2012v067n05ABEH004808

Russian Math. Surveys **67**:5 785–890

Uspekhi Mat. Nauk **67**:5 3–110 DOI 10.1070/RM2012v067n05ABEH004808

Dedicated to the centenary of the birth of Leonid Vital'evich Kantorovich

The Monge–Kantorovich problem: achievements, connections, and perspectives

V. I. Bogachev and A. V. Kolesnikov

Abstract. This article gives a survey of recent research related to the Monge–Kantorovich problem. Principle results are presented on the existence of solutions and their properties both in the Monge optimal transportation problem and the Kantorovich optimal plan problem, along with results on the connections between both problems and the cases when they are equivalent. Diverse applications of these problems in non-linear analysis, probability theory, and differential geometry are discussed.

Bibliography: 196 titles.

Keywords: Monge problem, Kantorovich problem, optimal transportation, transport inequality, Kantorovich–Rubinshtein metric.

Contents

This work was supported by the Russian Foundation for Basic Research (grant nos. 10-01-00518, 11-01-90421-Укр-ф-а, 11-01-12104-офи-м, 12-01-33009) and the project SFB 701 at Bielefeld University. Some of the results used in the paper were obtained in the work on project no. 11-01-0175 carried out in 2012–2013, in the framework of the Higher School of Economics Academic Fund Program.

AMS 2010 *Mathematics Subject Classification.* Primary 28C20, 35J96, 49Q20, 60B05.

Introduction

In this year of the centenary of the birth of Leonid Vital'evich Kantorovich, it is 70 years since the publication of one of his shortest but undoubtedly outstanding papers [1] (with an even shorter continuation [2]), the ideas of which were fated to a long life; some of them were developed in the two joint papers [3] and [4] by Kantorovich and Rubinshtein. As noted by Vershik in [5], "the beauty and naturalness of the formulation of the problem, the fundamental character of the main theorem (a criterion for optimality), and finally, the wealth of applications (only partly realized, but new ones are being discovered in areas coming to light only now), all this lets us to rank these papers among the classics of 20th-century mathematics".

The aim of our survey is to give an account of the state of the art of research connected conceptually with the paper [1] as well as with the considerably older problem of Monge [6], about which Kantorovich learned after the appearance of [1] (see [2]); interesting materials of a biographical nature are collected in [7] and [8], and one can expect new publications in this centenary year.

Although there are a number of thorough surveys and monographs devoted to this subject (see [9]–[17]), it seems reasonable to briefly retell the known history of its origin and development. In modern terms the Monge problem can be described as follows (in general form this formulation of the problem is due to Vershik [18]). We are given two probability spaces $(X,\mathscr{A},\mu)$ and $(Y,\mathscr{B},\nu)$ and a non-negative measurable function h on $X\times Y$, called a cost function (often denoted by c). It is required to find a map $T\colon X\to Y$ that is measurable with respect to the pair $(\mathscr{A},\mathscr{B})$ (that is, $T^{-1}(B)\in\mathscr{A}$ for all $B\in\mathscr{B}$), takes the measure μ into ν, and minimizes the expression

$$M(\mu,\nu,T)=M_h(\mu,\nu,T):=\int_X h(x,T(x))\,\mu(dx)$$

among all such maps. The condition that μ is taken to ν means that ν coincides with the image of μ under the map T, as given by the formula $\mu\circ T^{-1}(B):=\mu(T^{-1}(B))$, $B\in\mathscr{B}$. If some $T\in T(\mu,\nu)$ gives a minimum, then this T is called an *optimal map* of the measure μ to the measure ν, or an *optimal transportation*.

In the interpretation of Monge the problem was concerned with the most economical transportation of soil for construction works, that is, both measures were the standard volumes, the cost function was the usual distance, and the work performed had to be minimized (under the condition that the work of transporting a mass element ΔV over a distance Δr is $\Delta V\Delta r$). By appearance the problem looks applied, but it is fairly clear that it need not be such, although at the time the French academicians were often called on to solve genuinely applied problems of absolutely concrete practical nature (see [19]). Of course, this is not the only possible 'economic' formulation of the transportation problem. One can speak, for instance, of transporting goods from stores to delivery addresses or in general of distributing certain resources (a model example in some papers was the delivery of breads and croissants from bakeries to Parisian cafés), and discrete variants of the Monge and Kantorovich problems arise in so many modern areas that we do not even need to invent artificial applications. Gaspard Monge was a very universal scholar who made significant contributions to descriptive geometry, differential geometry, engineering, organization of science, and higher education. He was an active participant in the French Revolution (as the Naval Minister he signed the decision of the Court to execute Louis XVI), a comrade-in-arms of Napoleon, a participant in his Egyptian expedition, and the person who ensured Napoleon's election to the Academy of Sciences (in those years, as now, membership in the Academy did not necessarily require scientific achievements). Among Monge's students there were such renowned scientists as Cauchy, Poisson, Meusnier, Carnot (the Carnot cycle was an invention of his son S. Carnot), Poncelet, and Coriolis. Later, after the fall of Napoleon and the French 'perestroika', he was expelled from the Academy (by adroit colleagues who were quick in changing course), but much later the Academy of Sciences announced prizes for a solution of the optimal mass transfer problem posed by him for the concrete function $h(x,y) = |x-y|$ (though the very existence of a solution was tacitly assumed, and the problem was concerned with the study of some of its geometric properties). The problem was solved by Appel [20], who received such a prize. However, also his solution did not give a proof of the existence of an optimal map, but only established certain properties of such a map. Not only was the existence problem itself not solved, but it was not even precisely formulated until the 1970s, when it was posed explicitly in modern terms by Vershik in the paper [18] cited above and when the very important (and now classical in measure theory) paper [21] of Sudakov appeared and was believed to settle the matter. However, 20 years later a gap was found in it, which fortunately was eventually filled (see [9], [22]–[24], and more precise comments below). One can say that it was only by the beginning of this millennium that the special Monge problem was solved and also that many related subtleties previously invisible came to be recognized. Over the last decade research in this area has notably intensified. This increased activity has been connected both with the consideration of diverse problems arising with one or another choice of the cost function h and of the class of measures μ and ν being transformed, and with the determination of relations between these kinds of problems and the most diverse directions in non-linear analysis, geometry, partial differential equations, and stochastic analysis. It is these newly discovered relations that have attracted so many specialists from so many different areas to the questions considered here.

However, all this could hardly have been envisaged when in the severe wartime autumn of 1942 Major L. V. Kantorovich, then head of the department of mathematics at the Higher Technical Engineering College of the Soviet Navy and involved in applied military projects, submitted his note [1]. By that time he was already a fairly well-known mathematician who had published several papers in the theory of functions of real and complex variables, serious investigations in descriptive set theory, partially ordered vector spaces, and linear problems of functional analysis closely related to measure theory and integral representations of linear functionals. The creative work of Kantorovich, especially in the indicated areas, was certainly influenced by the fact that he was a student of G. M. Fichtenholz, a remarkable mathematician and teacher, and an eminent expert in the theory of the integral (see [25] on his works). Another teacher of Kantorovich was V. I. Smirnov. In [1] Kantorovich formulated a problem very close to Monge's problem (even with similar examples of an 'applied' nature about the transfer of soil), but with a fundamentally important nuance: in the Kantorovich problem instead of searching for a map T (an 'optimal transportation') it is proposed to find just an 'optimal plan of transportation', that is, a probability measure $\widehat{\sigma}$ on $(X\times Y, \mathscr{A}\otimes\mathscr{B})$ such that its projections on X and Y are μ and ν, respectively, and it minimizes the expression

$$K(\mu,\nu,\sigma) = K_h(\mu,\nu,\sigma) := \int_{X\times Y} h(x,y)\,\sigma(dx\,dy)$$

over all probability measures σ in the class $\Pi(\mu,\nu)$ of probability measures on the product $(X\times Y, \mathscr{A}\otimes\mathscr{B})$ that give μ and ν when projected on X and Y. It should be said that the paper [1] was part of a project that Kantorovich began in 1938 and presented in the booklet [26], which actually became the starting point of modern linear programming. The transport problem was one of the themes mentioned there.

Unlike the complicated non-linear Monge problem, the Kantorovich problem is linear: one is looking for the minimum of a linear functional on a convex set, and Kantorovich himself considered the case where X and Y are compact metric spaces, which makes $\Pi(\mu,\nu)$ compact in the weak topology, in which the convergence of measures ν_α to a measure ν means convergence of the integrals with respect to them of every bounded continuous function. Of course, the functional $\sigma \mapsto K(\mu,\nu,\sigma)$ turns out to be continuous in this topology if the function h is continuous. Thus, here there are solutions for every continuous function h, and all of them are extreme points of the compact set $\Pi(\mu,\nu)$ (that is, points that are not interior points of closed intervals with endpoints in $\Pi(\mu,\nu)$). In the general case a solution is a measure in $\Pi(\mu,\nu)$ at which the minimum is attained, and it is called *optimal* or an *optimal plan*. For non-compact spaces the situation is more complicated, but also here there are many sufficient conditions for the existence of solutions, while the Monge problem can be unsolvable even for rather simple functions and measures on compact sets in $\mathbb{R}^n$. It should be noted that in [2] Kantorovich called the Monge problem a special case of his problem and wrote that from a solution of the latter one can easily get a solution of the former. In the general case this is false, but if the Monge problem for a continuous cost function has a solution T, then the measure σ on the graph of T equal to the image of the measure μ under the map $x \mapsto (x, T(x))$ will be a solution to the Kantorovich problem. We note that the Kantorovich problem can be viewed as a 'Monge problem with a multivalued map' by representing a plan σ in the form $\sigma(dx\,dy) = \sigma^x(dy)\,\mu(dx)$ by means of conditional measures σ^x on Y (see below), which leads to the search for an optimal map $x \mapsto \sigma^x$ with values in the space of probability measures on Y. In this picture the existence of a pointwise optimal transportation means the possibility of choosing Dirac conditional measures σ^x. Here problems involving polymorphisms arise (see [27]). In some important special cases it is indeed possible to use a solution to the Kantorovich problem for constructing a solution to the Monge problem, but in general this is rather rare. Nevertheless, the connection between the problems of Monge and Kantorovich turns out to be surprisingly close, so the term 'Monge–Kantorovich problem' has become generally accepted (we saw the first use of these words in the title of Levin's paper [28]). For both problems it is useful to introduce the quantities

$$M(\mu,\nu) = M_h(\mu,\nu) := \inf\{M_h(\mu,\nu,T) \colon T \in T(\mu,\nu)\},$$

where $T(\mu,\nu)$ is the class of all measurable maps taking μ to ν, and

$$K(\mu,\nu) = K_h(\mu,\nu) := \inf\{K_h(\mu,\nu,\sigma) \colon \sigma \in \Pi(\mu,\nu)\}.$$

In dealing with a single cost function h we will omit it in the notation, but it will be important to use the index h when comparing cost functions. In the general case these infima are not minima, and one has the estimate

$$K(\mu,\nu) \leqslant M(\mu,\nu).$$

Finally, we denote by Φ_h the set of pairs of functions (φ, ψ), $\varphi\colon X \to \mathbb{R}^1$ and $\psi\colon Y \to \mathbb{R}^1$, that are measurable with respect to $\mathscr{A}$ and $\mathscr{B}$, respectively, and satisfy the inequality

$$\varphi(x) + \psi(y) \leqslant h(x, y), \qquad x \in X, \quad y \in Y.$$

For this set Kantorovich considered the so-called dual problem of finding the quantity (with values in $[0, +\infty]$)

$$J(\mu, \nu) = J_h(\mu, \nu) := \sup\biggl\{\int_X \varphi\, d\mu + \int_Y \psi\, d\nu,\ (\varphi, \psi) \in \Phi_h\biggr\}.$$

This problem, inspired by his investigations in linear programming involving dual problems, plays an important role in our circle of questions. We should say something about the case when all the quantities involved in some infimum are infinite: then the infimum is also infinite, of course. Thus, in the Kantorovich problem, the dual problem, and the Monge problem infinite values are allowed. For example, if the integral of h is infinite for all $\sigma \in \Pi(\mu, \nu)$, then we set $K(\mu, \nu) = +\infty$, and every measure in $\Pi(\mu, \nu)$ is optimal. Since obviously

$$\int_X \varphi\, d\mu + \int_Y \psi\, d\nu \leqslant \int_{X\times Y} h\, d\sigma$$

for all $(\varphi, \psi) \in \Phi_h$ and $\sigma \in \Pi(\mu, \nu)$, one has the estimate

$$J(\mu, \nu) \leqslant K(\mu, \nu) \leqslant M(\mu, \nu).$$

It turns out that under very broad assumptions these inequalities are actually equalities, and this will be the main topic of discussion in the first chapter. In the second chapter we turn to the Monge problem proper, presenting the principal results on the existence of its solutions and the properties (in particular, differential) of these solutions. An essential distinction of the Monge problem is that the requirements on the function h are considerably stronger here than in the Kantorovich problem, and these requirements are very restrictive even for nice measures on the space $\mathbb{R}^d$. So far one can state that only the case of the function $h(x, y) = |x - y|^2$ and some sufficiently regular functions with conditions of convexity/concavity type have been thoroughly studied. The last chapter is devoted to diverse applications and connections of the Monge–Kantorovich problem. There we discuss various non-linear inequalities and variational problems (isoperimetry, the logarithmic Sobolev inequality, and so on), geometric flows on finite-dimensional Riemannian manifolds and gradient flows on infinite-dimensional manifolds of probability measures, transport equations, infinite-dimensional variants of the Monge problem, and many other things; finally, we briefly mention some other interesting transformations of measures (certain competitors of optimal maps). Of course, the length of our survey does not allow us to go into details, so many directions in contemporary research are merely mentioned, with references (even in the very thorough monograph [17] by Villani, which is ten times longer than this paper, only certain selected problems are discussed in detail). But such a sketch of the general panorama is a goal of our survey, which is closely connected with another important goal: to attract the attention of our beginning mathematicians to this interesting and promising area created 50–70 years ago in the works of Kantorovich and also in closely related works of others of our outstanding researchers such as A. N. Kolmogorov, A. D. Aleksandrov, V. A. Rohlin, A. V. Pogorelov, R. L. Dobrushin, and Yu. V. Prokhorov.

Notation and terminology. Throughout, $\mathscr{B}(X)$ denotes the σ-algebra of Borel sets in a topological space X, that is, the smallest σ-algebra containing all open sets, and $C_b(X)$ denotes the set of all bounded continuous functions on X. We consider only Hausdorff (separated) spaces. Recall that a Borel measure is a real countably additive function μ on $\mathscr{B}(X)$. Such a function can be uniquely written in the form $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are non-negative Borel measures (called the positive and negative parts of μ) concentrated on disjoint sets; the measure $|\mu| = \mu^+ + \mu^-$ is called the total variation of the measure μ, and the quantity $\|\mu\| = |\mu|(X)$ is called the variation or the variation norm of μ. A Borel measure μ is called a Radon measure if, for every $B \in \mathscr{B}(X)$ and every $\varepsilon > 0$, there exists a compact set $K \subset B$ such that $|\mu|(B \setminus K) < \varepsilon$. A set $B \in \mathscr{B}(X)$ is called an atom of μ if $\mu(B) > 0$ and every set $A \in \mathscr{B}(X)$ contained in B has measure either 0 or $\mu(B)$. The smallest closed set of full measure is called the topological support of the measure μ and is denoted by $\mathrm{supp}(\mu)$. Such a support exists for every Radon measure on a topological space and also for every Borel measure on a separable metric space.

Let $a \wedge b = \min(a, b)$, and let δ_x be the Dirac measure at x.

In our discussion of certain questions it will be very useful to employ the notion of a Souslin space, which is a Hausdorff space that is a continuous image of a complete separable metric space (see [29], Chaps. 6 and 7). On such spaces all Borel measures are Radon measures. In a complete separable metric space all Borel sets are Souslin sets, but in uncountable spaces there always exist non-Borel Souslin sets. For the questions to be discussed an important advantage of Souslin sets compared to Borel sets is that a continuous and even a Borel image of a Souslin set is a Souslin set (the image of a Borel set need not be Borel even for an infinitely differentiable function on the real line).

Many aspects of the Monge–Kantorovich problem are connected with conditional measures, the construction and effective use of which goes back to the works of Kolmogorov, Rohlin, and other classics (see Chap. 10 in [29]). Here we only point out that for Souslin spaces X and Y every Radon probability measure μ on $X \times Y$ admits conditional measures μ^x, $x \in X$, on Y, that is, Radon probability measures such that for every Borel set $B \in X \times Y$ the function $x \mapsto \mu^x(B_x)$, where $B_x = \{y \in Y \colon (x, y) \in B\}$, is Borel measurable on X and

$$\mu(B) = \int_X \mu^x(B_x)\,\mu_X(dx),$$

where μ_X is the image of μ under projection on X. The measure $\mu^x(dy)$ is also denoted by $\mu(dy|x)$. In particular, letting $B = X \times C$ with $C \in \mathscr{B}(Y)$, we get that the function $x \mapsto \mu^x(C)$ is Borel measurable. For more general spaces the concept of conditional measures makes sense, but they do not always exist.

Let $\mathscr{M}_r(X)$ denote the space of all (including signed) Radon measures on X, let $\mathscr{M}_r^+(X)$ be the set of non-negative Radon measures, and let $\mathscr{P}_r(X)$ be its subset consisting of the probability measures. For most of the questions discussed below one can assume that our discussion concerns complete separable metric spaces, and on such spaces all Borel measures are Radon, so the notation $\mathscr{M}(X)$ and $\mathscr{P}(X)$ without the index r will be used. In the general case it is useful to consider also the Baire σ-algebra $\mathscr{B}a(X)$, which is generated by all the sets of the form $\{f > 0\}$, where $f \in C_b(X)$; measures defined on it are called Baire measures. It is clear that $\mathscr{B}a(X) \subset \mathscr{B}(X)$. For all metric spaces this inclusion is an equality, but in the general case it is strict even for compact spaces. The space of all Baire measures is denoted by $\mathscr{M}_\sigma(X)$, and its subset of probability measures is denoted by $\mathscr{P}_\sigma(X)$. This space (hence also $\mathscr{M}_r(X)$) is equipped with the weak topology generated by the seminorms

$$p_f(\mu) = \left|\int_X f\,d\mu\right|, \qquad f \in C_b(X).$$

Convergence of a sequence (or a net) of measures in the weak topology is convergence of the integrals of every fixed function in $C_b(X)$.

According to the classical Prokhorov theorem, if X is a complete separable metric space, then the following conditions are equivalent for a sequence of probability measures $\{\mu_n\}$ on X: 1) each subsequence of it contains a weakly convergent subsequence; 2) it is uniformly tight, that is, for every $\varepsilon > 0$ there is a compact set K_ε such that $\mu_n(X \setminus K_\varepsilon) < \varepsilon$ for all n.

The problems touched upon in this survey have been discussed with many colleagues. We are particularly indebted to L. Ambrosio, F. Barthe, S. G. Bobkov, F.-Yu. Wang, A. M. Vershik, A. V. Gasnikov, F. Götze, P. Catiaux, B. Klartag, Ph. Clement, A. Colesanti, M. Ledoux, V. L. Levin, R. MacCann, E. Milman, F. Morgan, M. K. von Renesse, M. Röckner, A. N. Sobolevskii, E. O. Stepanov, and V. N. Sudakov.

Chapter 1

The Kantorovich problem

1.1. Kantorovich metrics and Kantorovich–Rubinshtein norms

In this section X is a metric space with a metric ϱ. Thus, the classes of Borel and Baire measures coincide. For part of the results completeness and separability of the space will be important. It is known (see [29], Chap. 8) that, except for the case of finite X, the weak topology on the whole space $\mathscr{M}_\sigma(X)$ is not metrizable, hence it is not given by a norm. However, one can define a norm on $\mathscr{M}_\sigma(X)$ such that the generated topology coincides with the weak topology on the set of non-negative Radon measures (hence also on the set of probability measures). Let us consider the following Kantorovich–Rubinshtein norm on the space $\mathscr{M}_\sigma(X)$:

$$\|\mu\|_0 = \sup\left\{\int_X f\,d\mu\colon f \in \mathrm{Lip}_1(X),\ \sup_{x\in X}|f(x)| \leqslant 1\right\},$$

$$\mathrm{Lip}_1(X) := \{f\colon X \to \mathbb{R}^1,\ |f(x) - f(y)| \leqslant \varrho(x,y)\ \forall\, x,y \in X\}.$$

This norm generates the Kantorovich–Rubinshtein metric

$$d_0(\mu,\nu) := \|\mu - \nu\|_0.$$

Clearly, $\|\mu\|_0 \leqslant \|\mu\|$. If the space X contains an infinite convergent sequence, then the norm $\|\cdot\|_0$ is strictly weaker than the norm $\|\cdot\|$ (the variation norm). Indeed, if $x_n \to x$, then the measures δ_{x_n} converge in the norm $\|\cdot\|_0$ to the measure δ_x, since $|f(x_n) - f(x)| \leqslant \varrho(x_n, x)$ for $f \in \mathrm{Lip}_1(X)$, but $\|\delta_x - \delta_{x_n}\| = 2$ for $x_n \neq x$. In particular, under the indicated condition the space $\mathscr{M}_\sigma(X)$ cannot be complete in the norm $\|\cdot\|_0$, since it is complete in variation, and then by a theorem of Banach the two norms would be equivalent. If $\varrho(x,y) \geqslant \delta > 0$ whenever $x \neq y$, then the norms $\|\cdot\|_0$ and $\|\cdot\|$ are equivalent, since in this case $f \in \mathrm{Lip}_1(X)$ if $|f| \leqslant \delta/2$. According to Theorem 1.1.2 stated below, the topology generated by the norm $\|\cdot\|_0$ coincides with the weak topology on the set of non-negative τ-additive measures. Below we introduce a modified Kantorovich–Rubinshtein metric. The following equivalent norm is also frequently used:

$$\|\mu\|_{\mathrm{BL}}^* := \sup\biggl\{\int_X f\,d\mu\colon f \in \mathrm{BL}(X),\ \|f\|_{\mathrm{BL}} \leqslant 1\biggr\},$$

where $\mathrm{BL}(X)$ is the space of bounded Lipschitz functions on X with the norm

$$\|f\|_{\mathrm{BL}} := \sup_{x\in X} |f(x)| + \sup_{x\neq y} \frac{|f(x)-f(y)|}{\varrho(x,y)}\,.$$

It is readily seen that $\mathrm{BL}(X)$ with this norm is complete. Clearly,

$$\|\mu\|_{\mathrm{BL}}^* \leqslant \|\mu\|_0 \leqslant 2\|\mu\|_{\mathrm{BL}}^*,$$

since $\|f\|_{\mathrm{BL}} \leqslant 2$ if $f \in \mathrm{Lip}_1(X)$ and $\sup_X |f(x)| \leqslant 1$.

Remark 1.1.1. The weak convergence of a net $\{\mu_\alpha\}$ of *non-negative* measures to a measure μ is equivalent (see [29], Remark 8.3.1) to the equality

$$\lim_\alpha \int_X f(x)\,\mu_\alpha(dx) = \int_X f(x)\,\mu(dx)$$

for all bounded Lipschitz functions f on X. In particular, convergence of a net of non-negative measures in the Kantorovich–Rubinshtein metric implies weak convergence. However, if X is non-compact, then for some choice of a metric on X defining the original topology there necessarily exists a sequence of signed measures μ_n and a measure μ such that the integrals with respect to μ_n of every bounded uniformly continuous function f converge to the integral of f with respect to μ, but the measures μ_n do not converge weakly to μ (see Exercise 8.10.89 in [29]). The original metric does not always possess such a property (for example, in the case of $X = \mathbb{N}$ with the usual metric), but for $X = \mathbb{R}^1$ the standard metric is suitable: it suffices to have sequences $\{x_n\}$ and $\{y_n\}$ with $x_n \neq y_n$ and without limit points such that the distance between x_n and y_n tends to zero.

For any set $B \subset X$ let $B^\varepsilon = \{x\colon \operatorname{dist}(x,B) < \varepsilon\}$.

The following theorem employs yet another technical concept: a Borel measure μ is said to be τ-additive if $\mu(Z_\alpha) \to 0$ for every net of closed sets decreasing to the empty set. This property is shared by all Radon measures and also by all Borel measures on separable metric spaces.

Theorem 1.1.2. *The topology generated by the norm* $\|\cdot\|_0$ *coincides with the weak topology on the set* $\mathscr{M}_\tau^+(X)$ *of non-negative* τ*-additive measures. In addition, on the set* $\mathscr{P}_\tau(X)$ *of* τ*-additive probability measures the weak topology is given by the following Lévy–Prokhorov metric*:

$$d_P(\mu,\nu) = \inf\{\varepsilon > 0\colon \nu(B) \leqslant \mu(B^\varepsilon) + \varepsilon,\ \mu(B) \leqslant \nu(B^\varepsilon) + \varepsilon\ \forall\, B \in \mathscr{B}(X)\}.$$

In particular, if the space X *is separable, then the weak topology on the set* $\mathscr{M}_\sigma^+(X)$ *is generated by the metric* $d_0(\mu,\nu) = \|\mu-\nu\|_0$.

If $\mathscr{P}_\sigma(X) \neq \mathscr{P}_\tau(X)$, *then the weak topology on* $\mathscr{P}_\sigma(X)$ *is not metrizable.*

One should bear in mind that the non-coincidence of $\mathscr{P}_\sigma(X)$ and $\mathscr{P}_\tau(X)$ is rather exotic in applications. It is possible only for non-separable X, and in the case of a complete X it requires some additional set-theoretic assumptions (such as the existence of measurable cardinals; see, for example, [29], Proposition 7.2.10). Nevertheless, it is worthwhile to clarify the reason for non-metrizability of the weak topology on $\mathscr{P}_\sigma(X)$, since all the metrics considered are defined in any case. The point is that the set of measures with finite support is dense in $\mathscr{P}_\sigma(X)$ in the weak topology (this is true for all spaces). Hence the existence of a metric determining the weak topology on $\mathscr{P}_\sigma(X)$ gives countable weakly convergent sequences of discrete measures for every measure in $\mathscr{P}_\sigma(X)$, which thereby turns out to be a measure with separable support (the closure of the union of the atoms of this countable sequence), and this means τ-additivity (but not always the Radon property if the space is not complete).

Theorem 1.1.3. *For any two Borel probability measures* μ *and* ν *on a metric space* X *the following relations hold between the Lévy–Prokhorov metric and the Kantorovich–Rubinshtein metric*:

$$\frac{2d_P(\mu,\nu)^2}{2+d_P(\mu,\nu)} \leqslant \|\mu-\nu\|_{\mathrm{BL}}^* \leqslant \|\mu-\nu\|_0 \leqslant 3d_P(\mu,\nu).$$

In addition, $\|\mu-\nu\|_{\mathrm{BL}}^* \leqslant 2d_P(\mu,\nu)$. *If* X *is complete, then the space* $\mathscr{P}_r(X) = \mathscr{P}_\tau(X)$ *is also complete with either of these metrics.*

Therefore, for a complete separable metric space X *the space of Borel probability measures on* X *is complete with respect to these metrics and they determine the weak topology on it.*

It follows from the Prokhorov theorem that any weakly convergent sequence of signed Borel measures on a complete separable metric space converges in the norm $\|\cdot\|_0$, but the converse is not true: $\|\sqrt{n}(\delta_{1/n}-\delta_0)\|_0 \to 0$, although the measures $\sqrt{n}(\delta_{1/n}-\delta_0)$ do not converge weakly.

Let $\mathscr{P}^1(X)$ be the set of all Borel probability measures on X for which the function $x \mapsto \varrho(x,x_0)$ is integrable for some $x_0 \in X$ (and then for all x_0 by the triangle inequality). Let us equip $\mathscr{P}^1(X)$ with the following modified Kantorovich–Rubinshtein metric (introduced also by Fortet and Mourier [30] for a general separable metric space):

$$\|\mu-\nu\|_0^* := \sup\biggl\{\int_X f\,d(\mu-\nu)\colon f \in \mathrm{Lip}_1(X)\biggr\}.$$

We remark also that in [31] this norm was introduced on the linear span of Dirac measures for the study of isometric embeddings of X. Clearly,

$$\|\mu-\nu\|_0 \leqslant \|\mu-\nu\|_0^*,$$

since the restriction $\sup_x |f(x)| \leqslant 1$ is omitted when taking the supremum. In addition,

$$\|\mu-\nu\|_0^* \leqslant \int \varrho(x,a)\,(\mu+\nu)(dx) \qquad \forall\, a \in X,$$

since $f(x)$ can be replaced by $f(x)-f(a)$ in view of the equality $\mu(X)=\nu(X)$ and the inequality $|f(x)-f(a)| \leqslant \varrho(x,a)$. If the diameter of X does not exceed 1, then $\|\mu-\nu\|_0 = \|\mu-\nu\|_0^*$, because $|f(x)-f(a)| \leqslant 1$ for $f \in \mathrm{Lip}_1(X)$.

The quantity $\|\mu-\nu\|_0^*$ is indeed a norm of the measure $\mu-\nu$ if we introduce the linear space $\mathscr{M}_0(X)$ of all signed Borel measures σ on X such that $\sigma(X)=0$ and the function $x \mapsto \varrho(x,x_0)$ is integrable with respect to $|\sigma|$ (an equivalent condition: $\mathrm{Lip}_1(X) \in L^1(|\sigma|)$). On the space $\mathscr{M}_0(X)$ the indicated formula defines a norm $\sigma \mapsto \|\sigma\|_0^*$. We note that every measure $\sigma \in \mathscr{M}_0(X)$ has the form $\|\sigma^+\|\mu - \|\sigma^-\|\nu$, where $\mu,\nu \in \mathscr{P}^1(X)$, $\mu = \sigma^+/\|\sigma^+\|$, and $\nu = \sigma^-/\|\sigma^-\|$. The norm $\|\cdot\|_0^*$ can be extended to the linear space of all bounded Borel measures on X with respect to which all Lipschitz functions are integrable. To this end we set

$$\|\sigma\|_0^* = |\sigma(X)| + \sup\biggl\{\int_X f\,d\sigma\colon f \in \mathrm{Lip}_1(X),\ f(x_0)=0\biggr\}$$

for fixed $x_0 \in X$. If X is separable, then the dual of the space of measures with the norm $\|\cdot\|_0^*$ is the space $\mathrm{Lip}(X)$ of Lipschitz functions.

Proposition 1.1.4. *For all $\mu,\nu \in \mathscr{P}^1(X)$*

$$\|\mu-\nu\|_0^* = \sup\biggl\{\int_X f\,d\mu + \int_X g\,d\nu\colon \\ f \in C(X)\cap L^1(\mu),\ g \in C(X)\cap L^1(\nu),\ f(x)+g(y) \leqslant \varrho(x,y)\biggr\}.$$

Kantorovich himself introduced the following quantity $W(\mu,\nu)$, which is now called the Kantorovich metric (see [29], vol. 2 and [32] for a proof).

Theorem 1.1.5. *Let μ and ν be Radon measures in $\mathscr{P}^1(X)$. Then*

$$\|\mu-\nu\|_0^* = W(\mu,\nu) := \inf_{\lambda\in\Pi(\mu,\nu)} \int_{X\times X} \varrho(x,y)\,\lambda(dx,dy), \tag{1.1.1}$$

and there is a measure $\lambda_0 \in \Pi(\mu,\nu)$ at which the value $W(\mu,\nu)$ is attained.

If X is bounded, then $\mathscr{P}^1(X) = \mathscr{P}_\sigma(X)$. Note that for every metric space (X,ϱ) the metric $\varrho_0 := \varrho/(\varrho+1)$ (or $\varrho_1 := \min(\varrho,1)$) defines the original topology but is bounded, and hence the function W in (1.1.1) constructed from ϱ_0 is a metric for the weak topology on $\mathscr{P}_r(X)$. If the diameter of X does not exceed 1, then the embedding $X \to \mathscr{P}_r(X)$, $x \mapsto \delta_x$, preserves distances if $\mathscr{P}_r(X)$ is equipped with the metric d_0, since $d_0(\delta_a,\delta_b) = \varrho(a,b)$ (which is easy to get by taking $f(x)=\varrho(x,a)$).

A technical advantage of the modified Kantorovich–Rubinshtein metric $\|\mu-\nu\|_0^*$ is that the embedding $X\to\mathscr{P}^1(X)$ always preserves distances whatever the diameter of X: the equality

$$\|\delta_a-\delta_b\|_0^*=\varrho(a,b)$$

holds due to the estimate $|f(a)-f(b)|\leqslant\varrho(a,b)$ for all $f\in\mathrm{Lip}_1(X)$, and the supremum is attained at the function $f(x)=\varrho(x,a)$. Therefore, for any complete separable metric space X we obtain an increasing sequence of complete separable metric spaces

$$X\to\mathscr{P}^1(X)\to\mathscr{P}^1(\mathscr{P}^1(X))\to\cdots,$$

where the element V_{n+1} with index $n+1$ has the form $\mathscr{P}^1(V_n)$, the spaces of measures are equipped with the modified Kantorovich–Rubinshtein metric generated by the norm $\|\cdot\|_0^*$, and all the embeddings preserve distances. It is appropriate to call this chain *Vershik's tower*, in honour of Vershik, who introduced it more than 40 years ago (see [5] and [33], where there are additional references and discussions). If X is compact, then all levels of Vershik's tower are compact. In connection with the isometry of the embedding $X\to\mathscr{P}^1(X)$ we remark that the Kantorovich–Rubinshtein norm is maximal on the space V_0 of finite linear combinations σ of Dirac measures with $\sigma(X)=0$ among those norms q for which $q(\delta_x-\delta_y)=\varrho(x,y)$ (on related interesting problems and results, see [34]).

It is known that a general completely regular Hausdorff space X is homeomorphic to the subset of Dirac measures in the space $\mathscr{P}_r(X)$ of Radon probability measures on X with the weak topology (see [29], § 8.9). Compactness of X is equivalent to compactness of $\mathscr{P}_r(X)$.

Furthermore, metrizability of X is equivalent to metrizability of $\mathscr{P}_r(X)$, and metrizability of X by a complete metric is equivalent to metrizability of $\mathscr{P}_r(X)$ by a complete metric. Of course, on a metrizable space an incomplete metric can define the same topology as some complete metric (as in the case of the interval $(0,1)$). Unlike the weak topology, which is not connected with metrics, all the metrics considered by us on $\mathscr{P}_r(X)$ depend essentially on the original metric on X.

Similarly, one introduces the L^p-metric of Kantorovich (this was done explicitly in [35]; see also [36], [37], [15], and the references there). The quantity

$$W_p(\mu,\nu)=\inf_{\sigma\in\Pi(\mu,\nu)}\biggl(\int_{X\times X}\varrho(x,y)^p\,\sigma(dx\,dy)\biggr)^{1/p}$$

is a metric (for $p\geqslant 1$) on the space $\mathscr{P}_r^p(X)$ of Radon probability measures μ such that the function $x\mapsto\varrho(x,x_0)$ belongs to $L^p(\mu)$ for some (and then for all) $x_0\in X$ (if $0<p<1$, then W_p^p is a metric). The integrability of $\varrho(x,y)^p$ with respect to $\sigma\in\Pi(\mu,\nu)$ follows from the inequality $\varrho(x,y)^p\leqslant 2^p\varrho(x,x_0)^p+2^p\varrho(y,x_0)^p$ and the assumed integrability of the terms on the right-hand side with respect to μ and ν, respectively. The metric W_p is called the Kantorovich distance (or the p-distance). In foreign papers it is often called the Wasserstein distance under the influence of the celebrated paper [38] of Dobrushin, who efficiently used this metric and believed that it had been first introduced in the paper [39] of L. N. Vasershtein, who had indeed used it for $p=1$ (see p. 68 of the paper cited). Nevertheless, the historically wrong name has become firmly established in the foreign literature along with the distorted transliteration of the name of the author of [39] (by the way, translated in time into English with the spelling 'Vasershtein' for the author's name).

It is clear that $W_p(\mu,\nu)=W_p(\nu,\mu)$, and it is easily verified that $W_p(\mu,\nu)=0$ only when $\mu=\nu$ (the equality $W_p(\mu,\mu)=0$ is seen from the fact that one can take σ as the image of μ on the diagonal under the map $x\mapsto(x,x)$). Unlike for most metrics used, verification of the triangle inequality for W_p is not completely obvious. Given three measures $\mu_1,\mu_2,\mu_3\in\mathscr{P}_r^p(X)$, how can we estimate $W_p(\mu_1,\mu_3)$ in terms of $W_p(\mu_1,\mu_2)+W_p(\mu_2,\mu_3)$? If $\sigma_{1,2}$ and $\sigma_{2,3}$ are optimal plans for the pairs (μ_1,μ_2) and (μ_2,μ_3), then it suffices to have a probability measure η on $X\times X\times X$ whose projection on the product of the first two factors is $\sigma_{1,2}$ and whose projection on the product of the last two factors is the measure $\sigma_{2,3}$. Then the projections of η on the first and third factors are μ_1 and μ_3, whence we get that

$$\begin{aligned}W_p(\mu_1,\mu_3)&\leqslant\biggl(\int\varrho(x_1,x_3)^p\eta(dx_1\,dx_2\,dx_3)\biggr)^{1/p}\\&\leqslant\biggl(\int[\varrho(x_1,x_2)+\varrho(x_2,x_3)]^p\eta(dx_1\,dx_2\,dx_3)\biggr)^{1/p}\\&\leqslant\biggl(\int\varrho(x_1,x_2)^p\eta(dx_1\,dx_2\,dx_3)\biggr)^{1/p}+\biggl(\int\varrho(x_2,x_3)^p\eta(dx_1\,dx_2\,dx_3)\biggr)^{1/p},\end{aligned}$$

which coincides with $\|\varrho\|_{L^p(\sigma_{1,2})}+\|\varrho\|_{L^p(\sigma_{2,3})}=W_p(\mu_1,\mu_2)+W_p(\mu_2,\mu_3)$. However, the existence of such a measure η requires justification. This is done in the next lemma.

Lemma 1.1.6. *Let X_1, X_2, X_3 be completely regular spaces, and let $X_1\times X_2$ and $X_2\times X_3$ be equipped with Radon probability measures $\sigma_{1,2}$ and $\sigma_{2,3}$ with equal projections on X_2. Then on $X_1\times X_2\times X_3$ there is a Radon probability measure η with the projections $\sigma_{1,2}$ and $\sigma_{2,3}$ on $X_1\times X_2$ and $X_2\times X_3$, respectively.*

Proof. Let the given spaces be compact. Then on the linear subspace of $C(X_1\times X_2\times X_3)$ consisting of the functions of the form

$$\varphi(x_1,x_2,x_3)=f(x_1,x_2)+g(x_2,x_3),\qquad\text{where}\quad f\in C(X_1\times X_2),\quad g\in C(X_2\times X_3),$$

we can define the linear functional

$$l(\varphi)=\int f\,d\sigma_{1,2}+\int g\,d\sigma_{1,3}.$$

The condition of equality of the projections on X_2 guarantees that it is well-defined: if $f(x_1,x_2)+g(x_2,x_3)=f_0(x_1,x_2)+g(x_2,x_3)$, then $f(x_1,x_2)-f_0(x_1,x_2)$ does not depend on x_1 so has the form $\psi(x_2)$, and the measures $\sigma_{1,2}$ and $\sigma_{2,3}$ assign the same integral to $\psi(x_2)$. The norm of l equals 1, since $l(1)=1$ and $l(f+g)\leqslant1$ if $f+g\leqslant1$. The latter is seen from the representation

$$f(x_1,x_2)+g(x_2,x_3)=f(x_1,x_2)+\max_{x_3}g(x_2,x_3)+g(x_2,x_3)-\max_{x_3}g(x_2,x_3),$$

where $f(x_1,x_2)+\max_{x_3}g(x_2,x_3)\leqslant1$ and $g(x_2,x_3)-\max_{x_3}g(x_2,x_3)\leqslant0$. By the Hahn–Banach theorem l extends to a functional on $C(X_1\times X_2\times X_3)$ with unit norm, which by the Riesz theorem can be represented as the integral with respect to some Radon measure ν with $\|\nu\| = 1$. Since $l(1) = 1$, ν is a probability measure, and it has the required projections (it suffices to take $g = 0$, and then $f = 0$).

In the general case we may assume by the Radon property of the measures that all three spaces are countable unions of compact sets. Hence, they turn out to be Borel sets in the Stone-Čech compactifications $\overline{X}_1$, $\overline{X}_2$, and $\overline{X}_3$. Applying the fact proved above to the extensions of the measures $\sigma_{1,2}$ and $\sigma_{2,3}$ to $\overline{X}_1 \times \overline{X}_2$ and $\overline{X}_2 \times \overline{X}_3$, we obtain a Radon probability measure η on $\overline{X}_1 \times \overline{X}_2 \times \overline{X}_3$ with the required projections. This measure is concentrated on $X_1 \times X_2 \times X_3$, since $\eta(X_1 \times X_3 \times \overline{X}_3) = \eta(\overline{X}_1 \times X_2 \times X_3) = 1$ by our condition on the projections.

We remark that for many spaces (say, Souslin spaces) the required measure can easily be expressed explicitly via conditional measures:

$$\eta(dx_1\,dx_2\,dx_3) = \sigma_{1,2}(dx_1|x_2)\,\sigma_{2,3}(dx_3|x_2)\,\pi(dx_2),$$

where π is the common projection of $\sigma_{1,2}$ and $\sigma_{2,3}$ on X_2, and $\sigma_{1,2}(\,\cdot\,|x_2)$ and $\sigma_{2,3}(\,\cdot\,|x_2)$ are the corresponding conditional measures, that is, $\sigma_{1,2}(dx_1\,dx_2) = \sigma_{1,2}(dx_1|x_2)\,\pi(dx_2)$ and $\sigma_{2,3}(dx_2\,dx_3) = \sigma_{2,3}(dx_3|x_2)\,\pi(dx_2)$. In other words,

$$\int f\,d\eta = \int_{X_2}\int_{X_3}\int_{X_1} f(x_1,x_2,x_3)\,\sigma_{1,2}(dx_1|x_2)\,\sigma_{2,3}(dx_3|x_2)\,\pi(dx_2).$$

For example, if f depends only on x_2 and x_3, then the first integral gives $f(x_2,x_3)$, which gives the integral of f with respect to the measure $\sigma_{2,3}$. If f does not depend on x_3, then similarly we obtain the integral of f with respect to the measure $\sigma_{1,2}$, which means that the required conditions on the projections hold (on connections with multiplication of polymorphisms, see [27]). □

The metric W_p is bounded by the variation as follows (see [16], Proposition 7.10):

$$W_p(\mu,\nu)^p \leqslant 2^{p-1}\|\varrho(\,\cdot\,,x_0)(\mu-\nu)\|, \qquad \mu,\nu\in\mathscr{P}_r^p(X), \quad p\geqslant 1, \quad x_0\in X. \tag{1.1.2}$$

We note the following simple fact. Let a sequence of functions $\theta_j \in \mathrm{Lip}_1(X)$, $0 \leqslant \theta_j \leqslant 1$, be given such that $\theta_j(x) = 1$ for all x in the ball $B(x_0, j)$ of radius j with some common centre x_0 and $\theta_j(x) = 0$ if $x \notin B(x_0, j+1)$. It is clear that for every measure μ the measures $\theta_j \cdot \mu$ converge weakly to μ. An analogous fact is true for the metric W_p. Let $\xi_j(\mu)$ denote the probability measure $C_{n,j}\theta_j \cdot \mu$ with $C_{n,j} = (\theta_j \cdot \mu(X))^{-1}$ if $\theta_j \cdot \mu(X) > 0$, and let $\xi_j(\mu) = \mu$ if $\theta_j \cdot \mu(X) = 0$.

Lemma 1.1.7. *$W_p(\xi_j(\mu),\mu) \to 0$ as $j \to \infty$ for every measure $\mu \in \mathscr{P}_r^p(X)$. Moreover, $\lim_{j\to\infty}\sup_n W_p(\xi_j(\mu_n),\mu_n) = 0$ if $\mu_n,\mu \in \mathscr{P}_r^p(X)$ are such that the measures $(1+\varrho(\,\cdot\,,x_0)^p)\mu_n$ converge weakly to the measure $(1+\varrho(\,\cdot\,,x_0)^p)\mu$.*

Proof. The first assertion follows at once from (1.1.2). To prove the second one we have to verify that $\|\varrho(\,\cdot\,,x_0)(\xi_j(\mu_n)-\mu_n)\| \to 0$ as $j \to \infty$ uniformly with respect to n. The weak convergence of the indicated measures implies that for any given $\delta \in (0,1/2)$ there exists an $R > 0$ such that the integrals of $1+\varrho(x,x_0)^p$ over the exterior of the ball $B(R,x_0)$ with respect to all the measures μ_n and μ are smaller than δ. The integrals over the whole space are bounded by some number C. We can consider further only those n for which $\mu_n(B(R,x_0)) > 1-\delta$. If $j > R$, then we have $\theta_j(x) = 1$ for all $x \in B(R, x_0)$, so the restriction of the measure $\xi_j(\mu_n)$ to $B(R, x_0)$ equals the restriction of the measure $C_{n,j}\mu_n$, where $1 \leqslant C_{n,j} \leqslant 1 + \delta$. In addition, $\xi_j(\mu_n) \leqslant 2\mu_n$. Hence $\|\varrho(\,\cdot\,, x_0)(\xi_j(\mu_n) - \mu_n)\|$ is bounded by $C\delta + 3\delta$. □

Corollary 1.1.8. *The set of measures with bounded support, hence also the set of measures with finite support, is dense in* $(\mathscr{P}_r^p(X), W_p)$.

For $p > 1$ the metric W_p is connected with the weak topology in the following way.

Theorem 1.1.9. *A sequence of measures* $\mu_n \in \mathscr{P}_r^p(X)$ *converges to a measure* $\mu \in \mathscr{P}_r^p(X)$ *in the metric* W_p *precisely when* $\{\mu_n\}$ *converges to* μ *weakly and the equality*

$$\lim_{n\to\infty} \int_X \varrho(x, x_0)^p \, \mu_n(dx) = \int_X \varrho(x, x_0)^p \, \mu(dx) \tag{1.1.3}$$

is satisfied for some (and then for all) $x_0 \in X$. *This is also equivalent to the property that the measures* $(1 + \varrho(x, x_0)^p)\mu_n$ *converge weakly to* $(1 + \varrho(x, x_0)^p)\mu$.

Proof. The weak convergence of the sequence $\{\mu_n\}$ to μ along with (1.1.3) follows from the weak convergence of the measures $\nu_n = (1 + \varrho(x, x_0)^p)\mu_n$ to the measure $\nu = (1 + \varrho(x, x_0)^p)\mu$. It also implies the latter, since under the condition $\nu_n(X) \to \nu(X)$ weak convergence follows from convergence of the integrals of all the functions $f \in C_b(X)$ with bounded support. Let $W_p(\mu_n, \mu) \to 0$. Then $W(\mu_n, \mu) \to 0$, and hence we have weak convergence. Let us verify (1.1.3). The weak convergence implies, by use of the cut-off functions $\min(\varrho(x, x_0)^p, N)$, that the right-hand side does not exceed the $\liminf$ of the integrals of $\varrho(x, x_0)^p$ with respect to μ_n. On the other hand, the inequality $\varrho(x_0, x) \leqslant \varrho(x, y) + \varrho(x_0, y)$ implies that for every $q > 1$ there is a $C > 0$ such that $\varrho(x, x_0)^p \leqslant C\varrho(x, y)^p + q\varrho(x_0, y)$. Taking a measure $\sigma_n \in \Pi(\mu_n, \mu)$ with respect to which the integral of $\varrho(x, y)^p$ equals $W_p(\mu_n, \mu)^p$, and integrating the previous inequality with respect to it, we obtain the estimate

$$\int_X \varrho(x, x_0)^p \, \mu_n(dx) \leqslant C W_p(\mu_n, \mu)^p + q \int_X \varrho(x, x_0)^p \, \mu(dx).$$

Since $W_p(\mu_n, \mu) \to 0$, we see that the $\limsup$ of the left-hand side of this estimate is no greater than the product of q and the right-hand side of (1.1.3). Letting $q \to 1$, we obtain an estimate completing the proof of (1.1.3). Finally, we show that the weak convergence of the measures ν_n to ν implies the convergence $W_p(\mu, \mu_n) \to 0$. Lemma 1.1.7 reduces everything to the case of measures vanishing outside some ball, and hence to the case of a bounded metric. In the latter case our assertion is true by Theorems 1.1.2 and 1.1.5. □

For $p > 1$ the metric W_p on the simplex of probability measures cannot be the restriction of a norm on the space of measures due to the lack of convexity on intervals: already for $X = \{0, 1\}$ we have $W_p(\delta_0, (1 - t)\delta_0 + t\delta_1) = t^{1/p}$. However, the following is true.

Corollary 1.1.10. *The topology generated by the metric* W_p *on* $\mathscr{P}_r^p(X)$ *is also given by the norm*

$$K_p(\mu) := \|(1 + \varrho(\,\cdot\,, x_0)^p)\mu\|_0, \qquad \mu \in \mathscr{M}_r^p(X),$$

where the linear space $\mathscr{M}_r^p(X)$ *consists of all Radon measures* μ *such that the function* $\varrho(\,\cdot\,,x_0)^p$ *is integrable with respect to* $|\mu|$.

Proof. We know that convergence of sequences in the space $\mathscr{P}_r^p(X)$ with the indicated norm is equivalent to weak convergence after multiplication by $1+\varrho(\,\cdot\,,x_0)^p$, which by the previous theorem is equivalent to convergence in the metric W_p. It should be noted that the norm introduced does not necessarily generate the metric W_p on $\mathscr{P}_r^p(X)$, it just generates the same topology. $\square$

Already in the works of Dobrushin the metric W_1 was used for estimating the rate of convergence to an invariant distribution. An interesting geometric interpretation is given in [40]. We shall say that a discrete Markov chain M with transition probabilities P and metric d has Ricci curvature at least $k\geqslant 0$ if $W_1(P_x,P_y)\leqslant(1-k)d(x,y)$, $x,y\in M$, $P_x=P\delta_x, P_y=P\delta_y$. Many interesting properties can be derived from this. For example, for any probability measure μ on M one has $W_1(\mu P^N,\pi)\leqslant(1-k)^N W_1(\mu,\pi)$, where π is the invariant distribution, that is, there is an exponential rate of convergence to the invariant distribution. An analogy with the Ricci curvature for manifolds is confirmed by many examples (discrete versions of comparison theorems, Sobolev-type inequalities, and so on).

1.2. Existence of optimal plans

A number of very broad sufficient conditions are known for the existence of solutions of the Kantorovich problem, that is, the existence of measures minimizing the functional $\sigma\mapsto K(\mu,\nu,\sigma)$. We present the principal facts. Let us first observe that in the case of Radon measures μ and ν on completely regular spaces X and Y the set $\Pi(\mu,\nu)$ turns out to be uniformly tight: for every $\varepsilon>0$ there exists a compact set $K\subset X\times Y$ such that $\sigma((X\times Y)\backslash K)<\varepsilon$; for K one can take $K_1\times K_2$, where the compact sets K_1 and K_2 are such that $\mu(X\setminus K_1)+\nu(Y\setminus K_2)<\varepsilon$. In addition, $\Pi(\mu,\nu)$ is closed in the weak topology. Hence, $\Pi(\mu,\nu)$ turns out to be weakly compact (see [29], Theorem 8.6.7). By the linearity of the functional $K(\mu,\nu,\,\cdot\,)$, for proving the existence of its minimum on $\Pi(\mu,\nu)$ it suffices to have its continuity on $\Pi(\mu,\nu)$ in the weak topology or at least the weak closedness of the sets $\{\sigma\in\Pi(\mu,\nu)\colon K(\mu,\nu,\sigma)\leqslant c\}$, that is, its lower semicontinuity. We draw the reader's attention to the fact that these conditions are weaker than the corresponding requirements for the whole space of measures. For example, if the function h is bounded and lower semicontinuous, then we obtain lower semicontinuity for $K(\mu,\nu,\,\cdot\,)$. Indeed, it is known (see [29], Corollary 8.2.5) that for every weakly convergent net of measures $\sigma_\alpha\to\sigma$ in $\Pi(\mu,\nu)$

$$\liminf_{\alpha}\int h\,d\sigma_\alpha\geqslant\int h\,d\sigma.$$

If the function $h\geqslant 0$ is unbounded but lower semicontinuous, then we obtain the functional $K(\mu,\nu,\sigma)$ on $\Pi(\mu,\nu)$ in the form of an increasing sequence of functionals given by bounded lower semicontinuous functions $h_n=\min(h,n)$, whence the lower semicontinuity of the limit functional follows at once. This gives the next theorem (we recall that the existence of a solution does not mean that $K(\mu,\nu)<\infty$).

Theorem 1.2.1. *For any Radon measures μ and ν on completely regular spaces and any lower semicontinuous function $h \geqslant 0$ the Kantorovich problem has solutions.*

It is clear that the condition on h can be somewhat weakened. We actually need only the lower semicontinuity of the functionals given by the bounded functions $\min(h, n)$, and for this, as one can easily verify, it suffices to have the lower semicontinuity of the function h not on all of $X \times Y$, but on some sequence of compact sets of the form $A_j \times B_j$, where $\mu(A_j) \to 1$ and $\nu(B_j) \to 1$. The scope of applicability of this theorem is thereby considerably broadened.

It is well known and easily verified that a function h on a completely regular space is lower semicontinuous (that is, the sets $\{h \leqslant c\}$ are closed) precisely when $h = \sup_{\alpha \in A} h_\alpha$, where $h_\alpha \geqslant 0$ are continuous functions and A is an index set, and if $h \geqslant 0$, then one can find functions with $h_\alpha \geqslant 0$; if the space is metrizable, then one can take $A = \mathbb{N}$. These considerations along with weak compactness reduce finding $K_h(\mu, \nu)$ to the case of a bounded continuous cost function.

Proposition 1.2.2. *For any Radon measures μ and ν on completely regular spaces and any lower semicontinuous function $h \geqslant 0$*

$$K_h(\mu, \nu) = \sup_\alpha K_{h_\alpha}(\mu, \nu),$$

where $h_\alpha \geqslant 0$ are continuous functions such that $h = \sup_{\alpha \in A} h_\alpha$. In addition,

$$K_h(\mu, \nu) = \sup_N K_{h \wedge N}(\mu, \nu) = \sup_{N, \alpha} K_{h_\alpha \wedge N}(\mu, \nu).$$

Proof. Suppose that $\sup_\alpha K_{h_\alpha}(\mu, \nu) < K_h(\mu, \nu) - 2\delta$, where $\delta > 0$. Let us pass to a new collection of continuous functions h_α, indexed by finite subsets $\alpha \subset A$ partially ordered by inclusion, by setting $h_\alpha = \max(h_{\beta_1}, \dots, h_{\beta_k})$ for $\alpha = \{\beta_1, \dots, \beta_k\}$. Then h is the limit of the increasing net $\{h_\alpha\}$ (where $h_\alpha \geqslant h_\beta$ if $\alpha \geqslant \beta$). For every Radon measure $\sigma \geqslant 0$ on $X \times Y$ with respect to which the function h is integrable, the equality

$$\int_{X \times Y} h \, d\sigma = \lim_\alpha \int_{X \times Y} h_\alpha \, d\sigma$$

is satisfied. It is first verified for $h \wedge N$ (see [29], Lemma 7.2.6). As shown above, there is an optimal plan $\widehat{\sigma}$ for the lower semicontinuous function h, and for every function h_α there is an optimal plan $\widehat{\sigma}_\alpha$. Using the weak compactness of $\Pi(\mu, \nu)$, we pass to a subnet of $\{\widehat{\sigma}_\alpha\}$ that is weakly convergent to some measure $\sigma_0 \in \Pi(\mu, \nu)$; we can assume that this is the whole original net. We observe that $K_h(\mu, \nu, \sigma_0) = K_h(\mu, \nu)$. Indeed, if $K_h(\mu, \nu, \sigma_0) > K_h(\mu, \nu) + C$, where $C > 0$, then according to the foregoing there exists a β for which $K_{h_\beta}(\mu, \nu, \sigma_0) > K_h(\mu, \nu) + C$. Then by the weak convergence of $\widehat{\sigma}_\alpha$ to σ_0 and the continuity of h_β we obtain the existence of an $\alpha > 0$ such that $K_{h_\alpha}(\mu, \nu, \widehat{\sigma}_\alpha) \geqslant K_{h_\beta}(\mu, \nu, \widehat{\sigma}_\alpha) > K_h(\mu, \nu) + C$, which is impossible. Therefore, $K_h(\mu, \nu, \sigma_0) = K_h(\mu, \nu)$, so we may assume that $\widehat{\sigma} = \sigma_0$. Now take γ such that $K_{h_\gamma}(\mu, \nu, \sigma_0) > K_h(\mu, \nu, \sigma_0) - \delta$. Using the weak convergence once again, we find a $\beta > \gamma$ such that $K_{h_\gamma}(\mu, \nu, \widehat{\sigma}_\beta) > K_h(\mu, \nu, \sigma_0) - \delta$. Since $K_{h_\beta}(\mu, \nu, \widehat{\sigma}_\beta) \geqslant K_{h_\gamma}(\mu, \nu, \widehat{\sigma}_\beta)$, we arrive at a contradiction, which completes the proof of the first assertion. The remaining equalities are corollaries of it. □

Remark 1.2.3. For any bounded continuous cost function h a completely analogous argument enables us to obtain $K_h(\mu,\nu)$ as the limit of quantities $K_h(\mu_\alpha,\nu_\alpha)$, where the measures μ_α and ν_α have finite support. To this end we find a net of such measures σ_α on $X\times Y$ that are weakly convergent to an optimal plan $\widehat{\sigma}$. Their projections μ_α and ν_α on X and Y converge weakly to μ and ν, respectively. For them we take optimal plans $\widehat{\sigma}_\alpha$ and pass to a weakly convergent subnet $\{\widehat{\sigma}_\alpha\}$ whose weak limit σ_0 satisfies the estimate $K_h(\mu,\nu,\sigma_0)\leqslant K(\mu,\nu)$, since $K_h(\mu,\nu,\sigma_0) = \lim_\alpha K_h(\mu,\nu,\widehat{\sigma}_\alpha)\leqslant \lim_\alpha K_h(\mu,\nu,\sigma_\alpha)=K(\mu,\nu)$. Clearly, $\sigma_0\in\Pi(\mu,\nu)$, and hence $K_h(\mu,\nu,\sigma_0)=K(\mu,\nu)$. In the case of metrizable spaces (or spaces with metrizable compact sets) one can take countable sequences instead of nets.

Optimal plans are described by an interesting property of their supports. Let X and Y be sets and let c be a function on $X\times Y$.

Definition 1.2.4. A set $S\subset X\times Y$ with $\sum_{i=1}^n c(x_i,y_i)\leqslant\sum_{i=1}^n c(x_i,y_{i+1})$ for all $(x_1,y_1),\dots,(x_n,y_n)\in S$, where $y_{n+1}:=y_1$, is said to be c-cyclically monotone. If $\sum_{i=1}^n c(x_i,y_i)\leqslant\sum_{i=1}^n c(x_i,y_{\sigma(i)})$ for every permutation σ of the indices $\{1,\dots,n\}$, then S is said to be c-monotone.

As one can easily verify, these two notions are equivalent.

For functions ψ on Y and φ on X with values in $[-\infty,+\infty]$ we set

$$\psi^{c+}(x)=\inf_{y\in Y}[c(x,y)-\psi(y)],\quad \psi^{c-}(x)=\sup_{y\in Y}[-c(x,y)-\psi(y)],\qquad x\in X,$$
$$\varphi^{c+}(y)=\inf_{x\in X}[c(x,y)-\varphi(x)],\quad \varphi^{c-}(y)=\sup_{x\in X}[-c(x,y)-\varphi(x)],\qquad y\in Y.$$

A function $\varphi\colon X\to[-\infty,+\infty)$ is said to be c-concave if there is a function $\psi\colon Y\to[-\infty,+\infty)$ such that $\varphi=\psi^{c+}$. This is equivalent to the property that $\varphi=\varphi^{c+c+}$. A function φ is said to be c-convex if $-\varphi$ is c-concave. One defines c-concavity and c-convexity of functions on Y similarly.

For a c-concave function φ on X its c-superdifferential $\partial^{c+}\varphi\subset X\times Y$ is defined as the set $\partial^{c+}\varphi=\{(x,y)\colon \varphi(x)+\varphi^{c+}(y)=c(x,y)\}$. Let $\partial^{c+}\varphi(x):=\{y\in Y\colon (x,y)\in\partial^{c+}\varphi\}$. Similarly, the c-subdifferential of a convex function φ on X is the set $\partial^{c-}\varphi=\{(x,y)\colon \varphi(x)+\varphi^{c-}(y)=-c(x,y)\}$.

The inclusion $y\in\partial^{c+}\varphi(x)$ is equivalent to the estimate $\varphi(x)-c(x,y)\geqslant\varphi(z)-c(z,y)$ for all $z\in X$. It is verified directly that the c-superdifferential of any c-concave function is c-monotone. In [41] an important converse result was obtained.

Theorem 1.2.5. *Every c-cyclically monotone set S is contained in the c-superdifferential of some c-concave function.*

In the case of the function $c(x,y)=\langle x,y\rangle$ on $\mathbb{R}^d$ a set $C\subset\mathbb{R}^d\times\mathbb{R}^d$ is said to be cyclically monotone (without indication of a function) if the inequality $\sum_{i=1}^n\langle x_i,y_{i+1}-y_i\rangle\leqslant 0$ holds for every collection $(x_1,y_1),\dots,(x_n,y_n)\in C$, where $y_{n+1}=y_1$.

The subdifferential of a convex function φ on $\mathbb{R}^d$ at the point x is defined to be the set $\partial\varphi(x)=\{y\colon \forall\, z\in\mathbb{R}^d,\ \varphi(z)\geqslant\varphi(x)+\langle y,z-x\rangle\}$. The general definition above comes from this case (but the c-subdifferential is a set in $X\times Y$). A convex function φ is differentiable at x precisely when its subdifferential $\partial\varphi(x)$ consists of a single element. In this case $\partial\varphi(x)=\{\nabla\varphi(x)\}$. The following theorem of Rockafellar is a classical result in convex analysis, and is generalized by the previous theorem.

Theorem 1.2.6. *Every cyclically monotone set C in $\mathbb{R}^d\times\mathbb{R}^d$ is contained in the graph of the subdifferential of some convex function φ.*

The function φ can be explicitly defined as follows. Fix $(x_0,y_0)\in C$ and let $\varphi(x)=\sup\{\langle y_m,x-x_m\rangle+\cdots+\langle y_0,x_1-x_0\rangle;\ (x_0,y_0),\dots,(x_m,y_m)\in C\}$.

The following result goes back to [42], and its proof can be found in [10] and [17] (for Polish spaces, but the general case is similar).

Theorem 1.2.7. *Let X and Y be completely regular, let $\mu\in\mathscr{P}_r(X)$ and $\nu\in\mathscr{P}_r(Y)$, let the function $c\geqslant 0$ be continuous on $X\times Y$, and suppose that there are finite functions $a\in L^1(\mu)$ and $b\in L^1(\nu)$ such that $c(x,y)\leqslant a(x)+b(y)$. Then for any $\pi\in\Pi(\mu,\nu)$ the following conditions are equivalent*: (i) *the plan π is optimal*, (ii) *the topological support of π is c-cyclically monotone*, (iii) *there is a c-concave function φ on X such that $\varphi^+\in L^1(\mu)$ and the topological support of π is contained in $\partial^{c+}\varphi$.*

Corollary 1.2.8. *Suppose that in the situation of the previous theorem $\widehat{\sigma}$ is an optimal plan for μ and ν, and the support of a measure $\sigma_0\in\mathscr{P}_r(X\times Y)$ is contained in the support of $\widehat{\sigma}$. Then the measure σ_0 is an optimal plan for its projections.*

We also mention a result from [43], where the conditions on c were weakened.

Theorem 1.2.9. *Let X and Y be complete separable metric spaces and let a function $c\colon X\times Y\to[0,+\infty]$ be Borel measurable. Then each finite optimal plan is concentrated on a c-monotone set. Every plan with finite cost concentrated on a c-cyclically monotone set is optimal if there exist a closed set F and a $\mu\otimes\nu$-zero set N such that $c^{-1}(+\infty)=F\cup N$.*

The condition in Theorem 1.2.9 embraces functions c that are either finite or lower semicontinuous, but one cannot do without some kind of conditions on c (see [43]).

Remark 1.2.10. If X and Y are complete separable metric spaces, the finite function $c\geqslant 0$ is lower semicontinuous, and $K_c(\mu,\nu)<\infty$, then there exists a Borel c-cyclically monotone set $\Gamma\subset X\times Y$ with the following property: for *every* measure $\pi\in\Pi(\mu,\nu)$, optimality of π is equivalent to the equality $\pi(\Gamma)=1$.

We mention a convexity property of $K(\mu,\nu)$ (see [17], Theorem 4.8). Let X and Y be complete separable metric spaces, let $(T,\mathscr{B},\mathsf{P})$ be a probability space, and let $t\mapsto\mu_t$ and $t\mapsto\nu_t$ be measurable maps to the spaces $\mathscr{P}(X)$ and $\mathscr{P}(Y)$. We define $\mu=\int_T\mu_t\,\mathsf{P}(dt)$ and $\nu=\int_T\nu_t\,\mathsf{P}(dt)$.

Theorem 1.2.11. *Let the function c be lower semicontinuous and let $c(x,y)\leqslant a(x)+b(y)$, where $a\in L^1(\mu)$ and $b\in L^1(\nu)$. Then $K_c(\mu,\nu)\leqslant\int_T K_c(\mu_t,\nu_t)\,\mathsf{P}(dt)$.*

It is worth noting that already Kantorovich himself showed in the case of a compact space and cost function equal to the metric ϱ that an optimal plan π is characterized by the existence of a Lipschitz function U for which $U(x)-U(y)=\varrho(x,y)$ a.e. with respect to π.

Usually there are no explicit solutions to the Monge and Kantorovich problems (see [44] on the one-dimensional case, and also [45] on some examples on the plane).

1.3. The Kantorovich duality

For an arbitrary cost function h and the Kantorovich problem associated with it the following relationship, called the Kantorovich duality, holds under broad assumptions. For given probability measures μ and ν on the spaces $(X,\mathscr{A})$ and $(Y,\mathscr{B})$, respectively, consider the so-called dual functional and dual problem

$$J(\mu,\nu,\varphi,\psi)=\int_X \varphi\,d\mu+\int_Y \psi\,d\nu, \quad J(\mu,\nu)=\sup\{J(\mu,\nu,\varphi,\psi)\colon (\varphi,\psi)\in\Phi_h\}.$$

The duality theorem asserts that under certain conditions

$$J(\mu,\nu)=K(\mu,\nu).$$

The first case of such an equality goes back to Kantorovich's paper [1] and the joint papers [3] and [4] of Kantorovich and Rubinshtein for compact metric spaces. Later the class of spaces for which this is true was considerably enlarged. Before turning to precise formulations, we note that already contained in [3] and [4] is the close (but not equivalent) problem of finding the infimum (or minimum) $K_0(\varrho_0)$ of the functional

$$K_0(\varrho_0,\eta)=\int_{X\times X} h(x,y)\,\eta(dx\,dy)$$

over all non-negative measures η on the square $(X^2,\mathscr{A}\otimes\mathscr{A})$ of a measurable space $(X,\mathscr{A})$ such that the projections η_1 and η_2 on the first and second factors satisfy the equality $\eta_1-\eta_2=\varrho_0$ for a given (signed) measure ϱ_0 on X with $\varrho_0(X)=0$. For example, if $\varrho_0=0$, then the zero measure gives the zero solution, but if we are given two different probability measures μ and ν and $\varrho_0=\mu-\nu$, then the problem becomes non-trivial. It is known [46] that if h satisfies the triangle inequality (and only in this case), then $K_0(\mu-\nu)=K(\mu,\nu)$ under rather broad conditions, but in the general case there is no such connection. In some aspects this problem is more complicated, though the methods of investigation of the two problems have much in common. Let us confine ourselves to the technically simpler problem for $K(\mu,\nu)$.

We present the main results involving the equality $J(\mu,\nu)=K(\mu,\nu)$. In [47] the following theorem of a very general nature is proved, in which the notion of a perfect probability measure P on a measurable space $(\Omega,\mathscr{F})$ is used, that is, a measure such that for every $\mathscr{F}$-measurable function f the set $f(\Omega)$ contains a Borel set B for which $P(f^{-1}(B))=1$. In other words, the set $f(\Omega)$ is measurable with respect to $P\circ f^{-1}$ (on perfect measures see [29], § 7.5). Although there exist measures that are not perfect, all Radon measures are perfect.

Theorem 1.3.1. *If at least one of the measures μ and ν is perfect, then the equality $K(\mu,\nu)=J(\mu,\nu)$ holds for all bounded $\mathscr{A}\otimes\mathscr{B}$-measurable functions h, or more generally, for all $\mathscr{A}\otimes\mathscr{B}$-measurable functions that admit a pointwise estimate $h(x,y)\leqslant a(x)+b(y)$ with some functions $a\in L^1(\mu)$ and $b\in L^1(\nu)$. Hence, this equality holds if at least one of the two measures is Radon, and the latter is satisfied if one of the measures is Borel on a complete separable metric space.*

In [48] and [49] a probability space $(X,\mathscr{A},\mu)$ is called a space with the duality property if $K(\mu,\nu)=J(\mu,\nu)$ for every probability space $(Y,\mathscr{B},\nu)$ and every bounded measurable function h on $X\times Y$. It seems not to be known whether there are probability spaces without this property; as indicated by the authors of the paper [49], their counterexamples described in [47] and [48] are wrong. Hence, it follows from the facts proved in these papers that the perfectness of μ is necessary for the stronger property of strong duality, which in addition to the duality property requires that, for every probability space $(Y, \mathscr{B}, \nu)$, every sub-σ-algebra $\mathscr{B}_0 \subset \mathscr{B}$, and every bounded $\mathscr{A} \otimes \mathscr{B}_0$-measurable function h the quantity $J(\mu, \nu, h)$ is the same whether considering h on $\mathscr{A} \otimes \mathscr{B}$ or on $\mathscr{A} \otimes \mathscr{B}_0$. It is unclear whether this technical property is actually stronger.

In some important special cases a proof can be obtained from general results of convex analysis (the Fenchel–Rockafellar duality). It can be found in [16] along with a very interesting heuristic derivation of this relationship based on the minimax principle (see also [12] on connections between Kantorovich duality and linear programming). However, even for simple spaces with more complicated cost functions more special considerations are required. Among results of this sort we should mention the following recent achievement from [50].

Theorem 1.3.2. *Let X and Y be complete separable metric spaces with Borel probability measures μ and ν, respectively, and let $h\colon X \times Y \to [0, +\infty]$ be Borel measurable and finite $\mu \otimes \nu$-almost everywhere. Suppose also that there exists a measure in $\Pi(\mu, \nu)$ with respect to which h is integrable. Then $J(\mu, \nu) = K(\mu, \nu)$.*

It is clear from the formulation that instead of Borel measures on complete separable metric spaces one can take measures on abstract measurable spaces isomorphic (in the sense of the existence of bimeasurable and measure-preserving one-to-one maps defined almost everywhere) to such measurable spaces.

Of course, some restrictions on h are needed. For example, in Example 4.1 of [50] the following situation is considered: $X = Y = [0, 1]$, $\mu = \nu$ is Lebesgue measure, $h(x, y) = +\infty$ if $x < y$, $h(x, x) = 1$, and $h(x, y) = 0$ if $x > y$. Then $K(\mu, \nu) = 1$ (the transport plan is the normalized Lebesgue measure on the diagonal), but $J(\mu, \nu) = 0$. In [51] the following result is proved.

Theorem 1.3.3. *Let μ and ν be Radon measures. The equality $J(\mu, \nu) = K(\mu, \nu)$ holds if either the function h is lower semicontinuous or if h is Borel measurable and $h(x, y) \leqslant h_1(x) + h_2(y)$ for some $h_1 \in L^1(\mu)$ and $h_2 \in L^1(\nu)$. In the latter case the quantity $J(\mu, \nu)$ is attained on some pair $(\widehat{h}_1, \widehat{h}_2)$ with $\widehat{h}_1 \in L^1(\mu)$ and $\widehat{h}_2 \in L^1(\nu)$.*

In [52] (where in our terminology the function $-h$ is considered) a shorter proof of the equality $J(\mu, \nu) = K(\mu, \nu)$ is given in the case of Radon measures on completely regular spaces and non-negative lower semicontinuous functions h (since $-h$ is being considered, the paper deals with upper semicontinuity, which is not sufficient in our case, as Example 2.5 in [51] shows).

Remark 1.3.4. Even in the presence of the equality $J(\mu, \nu) = K(\mu, \nu)$ the infimum $J(\mu, \nu)$ need not be attained (need not be the minimum), and this does happen both in the case of a lower semicontinuous h on $[0, 1]$ and $\mu = \nu$ equal to Lebesgue measure, and in the case of absolutely continuous measures μ and ν on $\mathbb{R}$ and $h(x, y) = (x - y)^2$ (see [50], where $J(\mu, \nu)$ even equals the integral of $\varphi(x) + \psi(y)$ with respect to an optimal plan for some pair (φ, ψ), but these functions do not belong to $L^1(\mu)$ and $L^1(\nu)$, respectively).

Remark 1.3.5. Let X and Y be separable metric spaces, let the function $h \geqslant 0$ be lower semicontinuous, and let $K_h(\mu,\nu) < \infty$. For the existence of solutions to the direct and dual Kantorovich problems and for the equality of their values it suffices ([53], Theorem 3.2) that the integral of $h(x,y)$ with respect to μ not be infinite for ν-a.e. y and that the integral of $h(x,y)$ with respect to ν not be infinite for μ-a.e. x.

It is clear from the duality formula that for functions h of general form any solution $\widehat{\sigma}$ of the Kantorovich problem is concentrated on a small part of $X \times Y$. For example, if (φ,ψ) is a solution of the dual problem, then the integral of $\varphi(x)-\psi(y)$ with respect to the measure $\widehat{\sigma}$ coincides with the integral of $h(x,y)$, which implies that $\varphi(x) - \psi(y) = h(x,y)$ almost everywhere with respect to $\widehat{\sigma}$. This gives another explanation of the fact indicated above that any optimal plan is concentrated on a cyclically monotone set (see the previous section). In the situation of Remark 1.2.10 the set $\Gamma \subset X \times Y$ indicated there can be chosen so that the optimality of $\pi \in \Pi(\mu,\nu)$ is equivalent to the existence of an h-convex function ψ such that $h(x,y) = \psi^h(y) - \psi(x)$ π-a.e., and also equivalent to the existence of functions $\varphi\colon X \to [-\infty,+\infty)$ and $\psi\colon Y \to [-\infty,+\infty)$ such that $\varphi(x)+\psi(y) \leqslant h(x,y)$ and $\varphi(x)+\psi(y) = h(x,y)$ π-a.e.

Remark 1.3.6. There is another formulation of the dual Kantorovich problem in the case of completely regular topological spaces X and Y. Namely, instead of the class Φ_h of pairs of measurable functions φ and ψ one can consider a more narrow class Φ_h^C of pairs of bounded continuous functions φ and ψ for which $\varphi(x)-\psi(y) \leqslant h(x,y)$ for all $x \in X$ and $y \in Y$. Then instead of $J(\mu,\nu)$ one can consider the quantity $J^C(\mu,\nu)$, taking the supremum of $J(\mu,\nu,\varphi,\psi)$ over all pairs in Φ_h^C. It is clear that $\Phi_h^C \leqslant \Phi_h$. In the general case the inequality is strict, and it is shown in [46] that the equality $\Phi_h^C = \Phi_h$ is equivalent to the lower semicontinuity of the function h (see also [15], Theorem 4.6.8).

The total number of papers on the general Kantorovich problem in the topological setting and in the setting of general measure theory is too large to be adequately represented in the bibliography, but we particularly note a series of papers published over many years by Levin [28], [46], [54]–[58], where there are additional references to his works.

Let us turn to the case when $X = Y = \mathbb{R}^n$ and $h(x,y) = |x-y|^2/2$. We pass to the functions $\varphi(x) = |x|^2/2 - u(x)$ and $\psi(y) = |y|^2/2 - v(y)$. Then the problem dual to the Kantorovich problem takes the form of minimization of the functional

$$\widetilde{J}(\mu,\nu,\varphi,\psi) = \int_X \varphi\,d\mu + \int_Y \psi\,d\nu, \qquad \varphi(x)+\psi(y) \geqslant \langle x,y\rangle.$$

The Kantorovich problem itself can be reformulated as a search for a measure m with given projections at which the maximum is attained for the functional

$$\widetilde{K}(m) \mapsto \int_{X\times Y} \langle x,y\rangle\, m(dx\,dy).$$

Here the connection is seen between the Kantorovich problem and the Legendre transform, since it is clear from the formulation of the problem that from the very beginning we can confine ourselves to functions satisfying the equality $\psi = \varphi^*$, where $\varphi^*(y) = \sup_x(\langle x, y\rangle - \varphi(x))$. Similarly, $\varphi = \psi^* = \varphi^{**}$. Therefore, the functional $\widetilde{J}$ can be restricted to the set of pairs (φ, φ^*) with a convex function φ. In this case the following theorem holds.

Theorem 1.3.7. *There is a solution to the dual problem* (φ, ψ)*, where* φ *is convex,* $\psi = \varphi^*$*, and* $\varphi(x)+\varphi^*(y) \geqslant \langle x, y\rangle$*, and equality is attained on the topological support of any measure* m *giving a solution to the Kantorovich problem.*

In Theorem 2.1.3 below we present the principal facts about the dual problem in this case under the additional assumption that the measure μ vanishes on all sets of Hausdorff dimension at most $d-1$ (say, is absolutely continuous). Then it turns out that the map $T = \nabla\varphi$ is an optimal transportation of μ to ν, that is, a solution to the Kantorovich problem also gives a solution to the Monge problem. We see that here the two problems come together.

Chapter 2

The Monge problem

2.1. Existence and uniqueness of optimal maps

Let us turn to the Monge problem. Suppose that we are given a pair of probability measures μ and ν on measurable spaces $(X, \mathscr{A})$ and $(Y, \mathscr{B})$, respectively. A solution to the Monge problem is a map $T \in T(\mu, \nu)$ at which the minimum $M(\mu, \nu, T)$ is attained. Unlike the Kantorovich problem, a solution to the Monge problem need not exist even in the simplest cases.

Example 2.1.1. (i) Let $X = Y = [-1, 1]$, $\mu = \delta_0$, $\nu = 2^{-1}(\delta_{-1}+\delta_1)$, and $h(x, y) = |x - y|^2$. Then μ cannot be transformed into ν at all, but the half-sum of the Dirac measures at the points $(0, -1)$ and $(0, 1)$ serves as the unique solution to the Kantorovich problem.

(ii) Let $X = Y = [-1, 1]^2$, $\mu = \lambda \otimes \delta_0$, and $\nu = 2^{-1}(\lambda \otimes \delta_{-1} + \lambda \otimes \delta_1)$, let λ be the normalized Lebesgue measure on $[-1, 1]$, and let $h(x, y) = |x - y|^2$, where $|\cdot|$ is the usual norm in $\mathbb{R}^2$. Then both measures have no atoms, the Kantorovich problem has a solution and $K(\mu, \nu) = 1$, but the Monge problem has no solutions (although it has approximate solutions; see Theorem 2.3.1).

Proof. (i) If a measure σ on $[-1, 1]^2$ has projections μ and ν, then it is concentrated on the intersection of the interval $\{0\} \times [-1, 1]$ with the union of the intervals $[-1, 1] \times \{-1\}$ and $[-1, 1] \times \{1\}$, that is, is a combination of the Dirac measures at $(0, -1)$ and $(0, 1)$, whence it is seen that it must be their half-sum. Therefore, here $\Pi(\mu, \nu)$ consists of a single element.

(ii) Here $\Pi(\mu, \nu)$ contains many measures, but the quantity $K(\mu, \nu)$ is easily found, since by minimizing the integral of the function $|x - y|^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2$ with respect to measures $\sigma \in \Pi(\mu, \nu)$ we get that the integral of the first term is 1 due to the fact that $x_1 = 0$ σ-a.e. and $y_1^2 = 1$ σ-a.e., so that the minimum equal to 1 is attained precisely when $x_2 = y_2$ σ-a.e. An optimal plan is the distribution of the random vector $(0, \eta, \xi, \eta)$, where ξ and η are independent random variables such that ξ assumes the values -1 and 1 with probabilities 1/2 and η has the uniform distribution in $[-1, 1]$. If we suppose that the Monge problem has a solution $T = (T_1, T_2)$, then we get that $|T_1(x)| = 1$ μ-a.e., so that the integral of $|x - T(x)|^2 = (x_1 - T_1(x))^2 + (x_2 - T_2(x))^2$ with respect to the measure μ is equal to the integral of $1 + (x_2 - T_2(x))^2$ with respect to μ. In view of the equality $K(\mu, \nu) = 1$ and Theorem 2.3.1 this means that $T_2(x) = x_2$ μ-a.e., so that $T_1(x) = \xi(x_2)$ is either 1 or -1 for a.e. x_2. Then the measure $\mu \circ T^{-1}$ is concentrated on the union of sets $A \times \{-1\}$ and $B \times \{1\}$, where A and B are disjoint measurable sets with $A \cup B = [-1, 1]$, and therefore it cannot coincide with the measure ν. $\square$

We mention a general result on the unique solvability of the Monge problem (see [17], Theorem 5.28).

Theorem 2.1.2. *Let X and Y be complete separable metric spaces, let the function $h \geqslant 0$ be lower semicontinuous, and let $K_h(\mu, \nu) < \infty$. If for μ-a.e. x and for every h-convex function φ on X the set $\partial^h \varphi(x)$ contains at most one element, then the Monge problem has a unique solution T, $T(x) = \partial^h \psi(x)$ μ-a.e. for some h-convex function ψ, and the image of μ under the map $x \mapsto (x, T(x))$ gives a solution to the Kantorovich problem.*

In the rest of this subsection we shall discuss the important particular case when $X = Y = \mathbb{R}^d$ and h has the special form $h = |x - y|^p$, $p \geqslant 1$, where $|x|$ is the usual Euclidean norm. We shall assume that $\mu, \nu \in \mathscr{P}^p(\mathbb{R}^d)$. It follows that there exists a solution m_p of the corresponding Kantorovich problem. We recall that a solution to the Kantorovich problem gives a weaker result than a solution to the original Monge problem. Nevertheless, it turns out that under rather general conditions these problems are equivalent. In the case $p = 1$ the first existence result appeared in the paper [21] of Sudakov, according to which for any norm $\|\cdot\|$ on $\mathbb{R}^d$ (not necessarily generated by an inner product) and any pair of probability measures of the form $\mu = \varrho_\mu \, dx$, $\nu = \varrho_\nu \, dx$ on $\mathbb{R}^d$ there exists a map $T \colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\nu = \mu \circ T^{-1}$ and the image of the measure μ under the map $x \mapsto (x, T(x)) \in \mathbb{R}^d \times \mathbb{R}^d$ gives a solution to the corresponding Kantorovich problem for $h(x, y) = \|x - y\|$. In other words, the measure m_1 on $\mathbb{R}^d \times \mathbb{R}^d$ is concentrated on the graph of the map T: $m_1 = \mu \circ (x, T(x))^{-1}$. We shall call T an optimal map for the cost function $h(x, y) = \|x - y\|$, or an L^1-optimal map. Later a gap in Sudakov's proof was found which involved very subtle questions in the theory of conditional measures and non-linear versions of Fubini's theorem, and which by now has been completely filled (although a counterexample to one of the intermediate technical assertions has been constructed). However, it is impressive that this has taken 30 years, including a rather long period after detection of the gap. There is a whole series of papers in which different methods are employed to prove the existence of an L^1-optimal map for norms of different types (see [53], [59], [23], [60]–[62], and, finally, a very recent paper by Champion and De Pascale [24] in which there are no restrictions on the norm $\|\cdot\|$). All of these applications use very complicated constructions. The solution T itself has a non-trivial description.

It is simpler to prove the existence of a solution in the case of a so-called quadratic optimal map, or L^2-optimal map, that is, a map $T \colon \mathbb{R}^d \to \mathbb{R}^d$ transforming μ into ν for the cost function $|x - y|^2$. For such a T we have the equality

$$W_2(\mu, \nu) = \left(\int_{\mathbb{R}^d} |T(x) - x|^2 \, \mu(dx) \right)^{1/2}.$$

It turns out that such a T exists and has the form $T(x) = \nabla\varphi(x)$, where φ is some convex function.

It is surprising that the first general result on the existence of quadratic optimal maps appeared much later than Sudakov's work about L^1-optimal maps, even though quadratic optimal maps are considerably simpler and have wide applications in different areas of mathematics. Moreover, quadratic optimal maps actually arose long ago and were an object of study in geometry as solutions of a certain non-linear partial differential equation, the so-called Monge–Ampère equation (see [63], [64]). That equation and related geometric problems were considered in works of A. D. Aleksandrov, I. Ya. Bakel'man, Pogorelov, E. Calabi, Sh.-T. Yau, Nirenberg, and others. It arises, for example, in the classical Minkowski problem. The reason for the indicated delay is possibly that the quadratic Monge–Kantorovich problem had no clear physical meaning, while the Kantorovich functional with the cost function equal to the norm (and not the square of the norm) admitted a physical interpretation of 'work'. In any case, the existence of optimal maps under sufficiently general assumptions was first proved in Brenier's paper [65], which was motivated by an analysis of the equations of continuum mechanics. The main results of this paper are: 1) the existence and uniqueness of optimal maps for a broad class of measures, 2) a connection between the Monge–Kantorovich transport problem and the Monge–Ampère equation, 3) the polar factorization theorem, according to which a map $h\colon \Omega \to \mathbb{R}^d$ of class L^2 on a domain $\Omega \subset \mathbb{R}^d$ taking Lebesgue measure into an absolutely continuous measure has the form $h = \nabla\psi \circ U$, where $U\colon \Omega \to \Omega$ preserves Lebesgue measure and ψ is a convex function. We present a more general version of Brenier's result proved by McCann [66].

Theorem 2.1.3. *Let μ and ν be probability measures on $\mathbb{R}^d$ and let $\mu(A) = 0$ for every set A of Hausdorff dimension at most $d-1$. Then there exists a Borel map $T\colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\nu = \mu \circ T^{-1}$ and $T = \nabla\varphi$ for some convex function φ. This map is unique up to its definition on a set of μ-measure zero, that is, $T_1 = T_2$ holds μ-a.e. for any two such maps T_1 and T_2.*

If also $\nu(A) = 0$ for every set A of Hausdorff dimension at most $d-1$, then there exists an analogous map S for which $\mu = \nu \circ S^{-1}$ and $S = \nabla\varphi^*$, where $\varphi^*(y) = \sup_{x\in\mathbb{R}^d}\{\langle x, y\rangle - \varphi(x)\}$. The equalities $\nabla\varphi^* \circ \nabla\varphi(x) = x$ and $\nabla\varphi \circ \nabla\varphi^*(y) = y$ are satisfied for μ-a.e. x and ν-a.e. y.

The proof of this theorem is based on the Kantorovich duality theory and techniques of convex analysis. We indicate the main steps of a simplified proof from [66]. The problem of finding the minimum of the functional

$$m \mapsto \int |x-y|^2 \, m(dx\,dy)$$

is equivalent to the problem of finding the maximum of the functional $m \mapsto \int \langle x, y\rangle \, m(dx\,dy)$, since

$$\int |x-y|^2 \, m(dx\,dy) = \int |x|^2 \, \mu(dx) + \int |y|^2 \, \nu(dy) - 2\int \langle x, y\rangle \, m(dx\,dy).$$

Example 2.1.4. Let measures $\mu = N^{-1}\sum_{i=1}^N \delta_{x_i}$ and $\nu = N^{-1}\sum_{i=1}^N \delta_{y_i}$ be concentrated at N different points $x_1,\dots,x_N$ and $y_1,\dots,y_N$ in $\mathbb{R}^d$. Any transportation of the measure μ to ν is given by a permutation $\sigma\in S_N$: $\bigl(N^{-1}\sum_{i=1}^N \delta_{x_i}\bigr)\circ T_\sigma^{-1} = N^{-1}\sum_{i=1}^N \delta_{y_{\sigma(i)}}$. It is not difficult to show that T_σ is optimal (minimizes $\sum_{i=1}^N |x_i - y_{\sigma(i)}|^2$) precisely when $\sum_{k=1}^m \langle y_{\sigma(i_k)}, x_{i_{k+1}} - x_{i_k}\rangle \leqslant 0$, where $i_{m+1} = i_1$, for all $\{i_1,\dots,i_m\}\subset\{1,\dots,N\}$, that is, the graph of T_σ is cyclically monotone.

Corollary 2.1.5. *There exists a solution π_n of the Kantorovich problem for the pair of discrete measures $\mu = N^{-1}\sum_{i=1}^N \delta_{x_i}$, $\nu = N^{-1}\sum_{i=1}^N \delta_{y_i}$, and it is concentrated on a cyclically monotone discrete set of the form $\{(x_i, y_{\sigma(i)})\}$.*

Let μ and ν be probability measures on $\mathbb{R}^d$. Obviously, there exist weakly convergent sequences of discrete measures such that $\mu_n\to\mu$ and $\nu_n\to\nu$. Moreover, without loss of generality we may assume that each of the measures μ_n and ν_n is uniformly distributed at $N(n)$ atoms. It follows from the above that for every n there exists a measure π_n with corresponding projections μ_n and ν_n and concentrated on a discrete cyclically monotone set. We note that the family $\{\pi_n\}$ is weakly compact by the Prokhorov theorem. It is not difficult to deduce from the properties of the weak convergence that there is a measure π with projections μ and ν and having a cyclically monotone set as its topological support. By Theorem 1.2.6 it is contained in the graph of the subdifferential of some convex function φ. It is well known that any convex function is differentiable almost everywhere with respect to Lebesgue measure. It is less known that the set of points of non-differentiability of a convex function has Hausdorff dimension at most $d-1$. Suppose that $\mu(A)=0$ if the Hausdorff dimension of A is at most $d-1$. Then it follows that for μ-almost all x the intersection $\operatorname{supp}(\pi)\cap\{(x,y)\colon y\in\mathbb{R}^d\}\subset\mathbb{R}^d\times\mathbb{R}^d$ consists of the unique point $\nabla\varphi(x)$. It is not difficult to see that $\nabla\varphi$ is the required map. Thus, the existence of an optimal map follows from Theorem 1.2.6 and Example 2.1.4.

Remark 2.1.6. The uniqueness of the map $\nabla\varphi$ is a consequence of the following observation due to Aleksandrov: let φ_1 and φ_2 be two convex functions such that $\varphi_1(x_0)=\varphi_2(x_0)$ but $\nabla\varphi_1(x_0)\neq\nabla\varphi_2(x_0)$. Then $\nabla\varphi_1(D)$ is strictly contained in $\nabla\varphi_2(D)$, where $D=\{\varphi_1>\varphi_2\}$ and $\nabla\varphi_i(D)=\bigcup_{x\in D}\partial\varphi_i(x)$. The existence of two optimal maps $\nabla\varphi_1$ and $\nabla\varphi_2$ contradicts Aleksandrov's observation, since $\nabla\varphi_1(D)$ and $\nabla\varphi_2(D)$ have equal ν-measure.

Finally, we note that if the map $\nabla\varphi$ is smooth and $\mu=\varrho_\mu\,dx$ and $\nu=\varrho_\nu\,dx$, then the following change of variables formula must hold:

$$\varrho_\nu(\nabla\varphi)\det D^2\varphi = \varrho_\mu.$$

This equality can be regarded as an equation in φ. Equations of this form are called *Monge–Ampère equations*. It turns out surprisingly that the change of variables formula for measures with densities is always true, in a sense. As we have already noted, the first-order derivatives of a convex function exist almost everywhere. The second-order derivatives of a convex function can be understood in the generalized sense, that is, by definition $\partial_{e_i}\partial_{e_j}\varphi$ is a generalized function satisfying the equality

$$\langle\partial_{e_i}\partial_{e_j}\varphi,\psi\rangle = -\int_{\mathbb{R}^d}\partial_{e_i}\psi\,\partial_{e_j}\varphi\,dx$$

for all smooth compactly supported functions ψ. The convexity of φ implies that the functional $\partial_{e_i}\partial_{e_j}\varphi$ is non-negative, that is, is represented by a measure. By using convexity it is not difficult to show also that the generalized partial derivative $\partial_{e_i}\partial_{e_j}\varphi$ is also a measure (possibly signed). As is known, there exists a decomposition of the measure $\partial_{e_i}\partial_{e_j}\varphi = (\partial_{e_i}\partial_{e_j}\varphi)_{\mathrm{a}}\,dx + (\partial_{e_i}\partial_{e_j}\varphi)_{\mathrm{sing}}$ into an absolutely continuous part and a singular part. According to the Aleksandrov theorem (see [67], Theorem 2.3.2) the function $(\partial_{e_i}\partial_{e_j}\varphi)_a$ is the limit of the partial difference

$$\frac{1}{2t^2}[\varphi(x+te_i+te_j)+\varphi(x-te_i-te_j)-\varphi(x+te_i)-\varphi(x-te_i) \\ -\varphi(x+te_j)-\varphi(x-te_j)+2\varphi(x)]$$

for almost all x with respect to Lebesgue measure. This theorem has the following simplified formulation: *every convex function is twice differentiable almost everywhere*. Below, the symbol $D^2_a\varphi$ will denote the matrix made up of the functions $(\partial_{e_i}\partial_{e_j}\varphi)_a$. It turns out that only $D^2_a\varphi$ affects the change of variables formula. The next theorem was proved in [68].

Theorem 2.1.7. *The equality $\varrho_\nu(\nabla\varphi)\det D^2_{\mathrm{a}}\varphi = \varrho_\mu$ holds μ-almost everywhere.*

The Lagrange multiplier method. A non-rigorous but instructive proof of the fact that T is a gradient can be obtained by deriving the Euler–Lagrange equation (see [12]). Let T be an arbitrary map taking μ to ν. A solution of the Monge problem can be sought as a conditional extremum of the functional $\int_{\mathbb{R}^d}\langle T(x),x\rangle\,\mu(dx)$ under the condition that $\varrho_\nu(T)\det DT = \varrho_\mu$. Let us consider the Lagrange function

$$\int_{\mathbb{R}^d}\Bigl(\langle T(x),x\rangle\varrho_\mu(x)+\lambda(x)\bigl(\varrho_\nu(T(x))\det DT(x)-\varrho_\mu(x)\bigr)\Bigr)\,dx.$$

The function λ plays the role of the Lagrange multiplier (note that $\varrho_\nu(T)\det DT = \varrho_\mu$ by the change of variables formula). Since

$$\det(A+\varepsilon B)=\det A\det(I+\varepsilon A^{-1}B)\sim\det A\cdot(1+\varepsilon\operatorname{Tr}(A^{-1}B)),$$

the first variation of the Lagrange function equals

$$\int_{\mathbb{R}^d}\biggl(\langle\omega(x),x\rangle\varrho_\mu+\lambda\cdot\varrho_\mu\operatorname{Tr}[DT^{-1}D\omega]+\lambda\langle\nabla\varrho_\nu(T),\omega\rangle\frac{\varrho_\mu}{\varrho_\nu(T)}\biggr)\,dx,$$

where ω is a smooth vector field with compact support. Integrating by parts and using the change of variables formula, we readily see that

$$\int_{\mathbb{R}^d}\lambda\cdot\operatorname{Tr}[DT^{-1}D\omega]\varrho_\mu\,dx=\int_{\mathbb{R}^d}\lambda(T^{-1})\operatorname{div}(\omega(T^{-1}))\varrho_\nu\,dx \\ =-\int_{\mathbb{R}^d}\langle\nabla[\lambda(T^{-1})],\omega(T^{-1})\rangle\varrho_\nu\,dx-\int_{\mathbb{R}^d}\lambda(T^{-1})\Bigl\langle\omega(T^{-1}),\frac{\nabla\varrho_\nu}{\varrho_\nu}\Bigr\rangle\varrho_\nu\,dx.$$

Let $\lambda = u(T)$, where u is some function. Again using the change of variables formula, we get that for any smooth vector field ω

$$\int_{\mathbb{R}^d}\bigl(\langle\omega(x),x\rangle-\langle\nabla u(T(x)),\omega(x)\rangle\bigr)\varrho_\mu(x)\,dx=0.$$

Therefore, $\nabla u(T(x)) = x$, and thus $T^{-1} = \nabla u$. Due to the symmetry of the problem with respect to μ and ν the analogous assertion is true for the map T.

Non-quadratic cost functions. An interesting class of transformations of measures consists of optimal maps for non-quadratic cost functions. For a cost function $h(x,y) = h(x-y)$ with h strictly convex, they have the form $T(x) = x - \nabla h^*(\nabla\varphi(x))$, where φ is an h-convex function. The latter means that φ has the form $\varphi(x) = \inf_{y\in\mathbb{R}^d}\bigl(h(x,y) - \psi(y)\bigr)$ for some function ψ. For more details see [16] and [15]. About uniqueness without convexity conditions, see [69].

2.2. Regularity of solutions and a priori estimates

We discuss here the regularity properties of optimal transportations. This means properties like continuity, Hölder continuity, differentiability (ordinary and Sobolev), and growth estimates. The most complete picture exists so far for the quadratic cost function $|x-y|^2$ on $\mathbb{R}^d$, to which all the results presented in this section refer. Only at the end do we make a few remarks about other cost functions and about manifolds.

2.2.1. The maximum principle. We first discuss the non-linear maximum principle which was first obtained in the works of Aleksandrov on the geometry of convex surfaces. It generalizes the classical maximum principle and is an important tool in the theory of regularity of linear elliptic equations (in non-divergence form), and also of a broad class of non-linear elliptic equations. Below we give a classical proof, which, as one can easily see, is based on the idea of transportations of measures.

Definition 2.2.1. The convex envelope of a continuous function f is the function $f_* = \sup\{u \leqslant f,\ u \text{ is convex}\}$. The set $C_*(f) = \{x\colon f(x) = f_*(x)\}$ is called the set of contact points.

The concave envelope is defined similarly.

Definition 2.2.2. The Monge–Ampère measure associated with a convex function φ on a convex subset of $\mathbb{R}^d$ is the measure

$$\mu_\varphi(A) = \lambda_d\biggl(\bigcup_{x\in A}\partial\varphi(x)\biggr).$$

If the function φ is twice continuously differentiable, then

$$\mu_\varphi(A) = \int_A \det D^2\varphi(x)\,dx.$$

If $\det D^2\varphi > 0$, then μ_φ is the pre-image of Lebesgue measure under the map $\nabla\varphi$.

Theorem 2.2.3 (the maximum principle). *Let f be a continuous function on a convex set A. Then for some constant $C(d)$ depending only on the dimension,*

$$\sup_{x\in A} f(x) \leqslant \sup_{x\in\partial A} f(x) + C(d)\,\mathrm{diam}(A)[\mu_{(-f)_*}(C_*(-f))]^{1/d}.$$

In particular, if f is twice continuously differentiable, then

$$\sup_A f \leqslant \sup_{\partial A} f + C(d)\,\mathrm{diam}(A)\biggl[\int_{\{x\colon D^2 f(x)\leqslant 0\}} |\det D^2 f|\,dx\biggr]^{1/d}.$$

Proof. Passing to the function $g(x)=\sup_{y\in\partial A} f(y)-f(x)$, we obtain an equivalent inequality $\inf_{x\in A} g(x)\geqslant -C(d)\operatorname{diam}(A)[\mu_{g_*}(C_*(g))]^{1/d}$ for g with the condition $\inf_{x\in\partial A} g(x)\geqslant 0$. If $m=\inf_{x\in A} g(x)\geqslant 0$, then the inequality is trivial. Otherwise we fix a minimum point x_0 for the function g. Let us consider the cone in $\mathbb{R}^{d+1}$ with vertex at the point (x_0,m) and base $\partial A\subset\{x\colon x_{d+1}=0\}$ given by the equation $x_{d+1}=K(x)$. Let $v\in\partial K(x')$ for some point $x'\in A$. Consider the hyperplane L given by the equation $x_{d+1}=K(x')+\langle v,x-x'\rangle$ and tangent to K at the point x'. Since the graph of g lies above L on ∂A and coincides with L at x', by moving the graph of L continuously downwards along the x_{d+1}-axis we find a point x_v in the set $C_*(g)$ of contact points such that $v\in\partial g_*(x_v)$. Therefore, $\partial K(A)\subset\partial g_*(C_*(g))$ and $\mu_{\partial K}(A)\leqslant\mu_{g_*}(C_*(g))$. Since the function K is homogeneous, we have $\partial K(A)=\partial K(x_0)$. The measure of $\partial K(x_0)$ is easily estimated from below by the volume of the d-dimensional ball of radius $-m/\operatorname{diam}(A)$. Hence $\mu_{g_*}(A)\geqslant C(d)\bigl(-m/\operatorname{diam}(A)\bigr)^d$, which yields the desired assertion. □

Using the parabolic transportation of measures on $\mathbb{R}^{d+1}$ of the form $(t,x)\mapsto(\langle x,\nabla_x u\rangle-u,\nabla_x u)$, where $u(\,\cdot\,,\cdot\,)\colon\Omega=[0,T]\times Q\to\mathbb{R}$ and $Q\subset\mathbb{R}^d$ is a compact convex set, one can prove the parabolic maximum principle:

$$\sup_{\Omega} u\leqslant C(d,\operatorname{diam}(Q))\biggl(\int_{\Gamma_u}|\partial_t u\cdot\det D_x^2 u|\,dt\,dx\biggr)^{1/d},$$

where $\Gamma_u=\{\partial_t u\leqslant 0,\ D^2u\leqslant 0\}\subset\Omega$ and $u=0$ on $([0,T]\times\partial Q)\cup(T\times Q)$.

2.2.2. Regularity of solutions of the Monge–Ampère equation. The Monge–Ampère equation arising in the measure transportation problem is a particular case of equations of the form

$$\det D^2\varphi(x)=f(x,\nabla\varphi(x)) \tag{2.2.1}$$

in the class of so-called fully non-linear elliptic equations (see the surveys [70] and [71]). In [64] there is a description of the special case we are interested in (the Monge–Ampère equation for the transport problem). The regularity problem for the Monge–Ampère equation is highly non-trivial. A relatively complete picture has taken shape over more than 60 years in the works of Aleksandrov, Calabi, Pogorelov, N. V. Krylov, J. Spruck, L. Caffarelli, Nirenberg, and many others. In geometry there has been a parallel development in a closely related direction: the complex Monge–Ampère equation, where Yau, Calabi, and T. Aubin have been involved. The classical approach to the regularity theory of non-linear equations is based on differentiation of them and application of the linear theory to the resulting equations, which are linear with respect to the higher derivatives. An example of such techniques can be found in § 2.2.3, where we obtain some uniform estimates for the second derivatives of the potential in the transport problem. We quote Krylov from [70]: "To prove the existence of solutions of equations like (2.2.1) by the methods known before 1981 was no easy task. It involved finding a priori estimates for the solutions and their derivatives up to the *third* order. A large part of the work was based on differentiation of (2.2.1) three times and certain extremely cleverly organized manipulations invented by Calabi. After 1981 the approach to fully non-linear equations changed dramatically." In the 1980s, in the works of Krylov and Safonov and also (independently) of L. Evans, general regularity theorems were proved for elliptic equations defined by non-linear operators of the form $F(D^2\varphi)$, where F is a uniformly elliptic operator. Unfortunately, the Monge–Ampère operator $\varphi \mapsto \det D^2\varphi$ *is not* uniformly elliptic even on the space of convex functions, and the Krylov–Safonov–Evans theory is not directly applicable.

Below we use the concept of a solution of the Monge–Ampère equation in the sense of Aleksandrov. A solution of the equation

$$\det D^2\varphi = w$$

in the sense of Aleksandrov is a function φ for which the Monge–Ampère measure satisfies the equality $\mu_\varphi = w\,dx$. We observe that a solution in the sense of Aleksandrov with an absolutely continuous right-hand side automatically means the absence of a singular component of μ_φ, which, generally speaking, may not hold if we are concerned with an optimal transportation of $w\,dx$ to Lebesgue measure. In the case when φ is smooth the two concepts of a solution coincide.

2.2.3. The classical approach. Let $\nabla\Phi$ be an optimal transportation of a measure $e^{-V}dx$ to a measure $e^{-W}dx$. By the change of variables formula,

$$V = W(\nabla\Phi) - \log\det D^2\Phi.$$

Let us fix a unit vector e and differentiate this equality along e. To this end we use the relation

$$\partial_e \log\det D^2\Phi = \frac{\partial_e \det D^2\Phi}{\det D^2\Phi} = \operatorname{Tr}[(D^2\Phi)^{-1}\partial_e D^2\Phi].$$

Differentiating this along a vector v and using the fact that $\partial_v(D^2\Phi)(D^2\Phi)^{-1} + D^2\Phi\partial_v[(D^2\Phi)^{-1}] = 0$, we get that

$$\partial_v\partial_e \log\det D^2\Phi = \operatorname{Tr}[(D^2\Phi)^{-1}\partial_v\partial_e D^2\Phi] - \operatorname{Tr}[(D^2\Phi)^{-1}\partial_e(D^2\Phi)(D^2\Phi)^{-1}\partial_v D^2\Phi].$$

Using the change of variables formula again, we find that

$$\begin{aligned}\partial_e V &= \langle \nabla W(\nabla\Phi), D^2\Phi\cdot e\rangle - \operatorname{Tr}[(D^2\Phi)^{-1}\partial_e D^2\Phi],\\ \partial_e^2 V &= \langle D^2W(\nabla\Phi)D^2\Phi\cdot e, D^2\Phi\cdot e\rangle + \langle \nabla W(\nabla\Phi), \nabla\partial_e^2\Phi\rangle\\ &\quad - \operatorname{Tr}[(D^2\Phi)^{-1}\partial_e^2 D^2\Phi] + \operatorname{Tr}\bigl[(D^2\Phi)^{-1}\partial_e D^2\Phi\bigr]^2. \end{aligned} \tag{2.2.2}$$

Let us represent these relations in a more convenient form. We define a diffusion operator L_Φ by the formula $L_\Phi f = \operatorname{Tr}[D^2 f\cdot(D^2\Phi)^{-1}] - \langle\nabla f, \nabla W(\nabla\Phi)\rangle$.

Remark 2.2.4. If the measures μ and ν have locally Sobolev densities, then it is easily verified that

$$-\int \langle (D^2\Phi)^{-1}\nabla f, \nabla\eta\rangle\,d\mu = \int f\cdot L_\Phi\eta\,d\mu = \int \eta\cdot L_\Phi f\,d\mu, \qquad \eta\in C_0^\infty(\mathbb{R}^d).$$

Thus, $\partial_e V = -L_\Phi\partial_e\Phi$, where L_Φ is the generator of the Dirichlet form

$$\mathscr{E}_\Phi(f,\eta) = \int \langle (D^2\Phi)^{-1}\nabla f, \nabla\eta\rangle\,d\mu.$$

From (2.2.2),

$$\partial_e^2 V = \langle D^2W(\nabla\Phi)D^2\Phi\cdot e, D^2\Phi\cdot e\rangle - L_\Phi \partial_e^2\Phi + \mathrm{Tr}\big[(D^2\Phi)^{-1}\partial_e D^2\Phi\big]^2. \qquad (2.2.3)$$

Using (2.2.3) and the maximum principle, one can obtain a priori estimates for the supremum of $\partial_e^2\Phi$.

Example 2.2.5. Let the function $\Phi\in C^4$ be convex, let $\nabla\Phi$ be an optimal transportation of the measure $\mathrm{e}^{-V}dx$ with support in a convex set A to Lebesgue measure on the set B, and let e be a unit vector. Then

$$\sup_{x\in A}\partial_e^2\Phi(x) \leqslant \sup_{x\in\partial A}\partial_e^2\Phi(x) + C(d)\,\mathrm{diam}(A)\biggl(\int_A (\partial_e^2 V)_+^d \mathrm{e}^{-V}\,dx\biggr)^{1/d}.$$

Proof. By the change of variables formula, $\det D^2\Phi = 1/\lambda(B)$, and (2.2.3) implies that $\partial_e^2 V = -\,\mathrm{Tr}[(D^2\Phi)^{-1}\partial_e^2 D^2\Phi] + \mathrm{Tr}\big[(D^2\Phi)^{-1}\partial_e D^2\Phi\big]^2$. If $\partial_e^2 D^2\Phi(x)\leqslant 0$, then

$$\partial_e^2 V \geqslant \mathrm{Tr}[(D^2\Phi)^{-1}(-\partial_e^2 D^2\Phi)] \geqslant d\cdot\bigl(\det((D^2\Phi)^{-1}(-\partial_e^2 D^2\Phi))\bigr)^{1/d}.$$

By the non-linear maximum principle,

$$\begin{aligned}
\sup_{x\in A}\partial_e^2\Phi(x) - \sup_{x\in\partial A}\partial_e^2\Phi(x) &\leqslant C(d)\,\mathrm{diam}(A)\biggl(\int_{\{\partial_e^2 D^2\Phi\leqslant 0\}}\det(-\partial_e^2 D^2\Phi)\,dx\biggr)^{1/d}\\
&\leqslant C(d)\,\mathrm{diam}(A)\biggl(\int_{\{\partial_e^2 D^2\Phi\leqslant 0\}}\det D^2\Phi\cdot\det\big[(-\partial_e^2 D^2\Phi)\circ(D^2\Phi)^{-1}\big]\,dx\biggr)^{1/d}\\
&\leqslant C(d)\,\mathrm{diam}(A)\biggl(\int_A(\partial_e^2 V)_+^d\mathrm{e}^{-V}\,dx\biggr)^{1/d},
\end{aligned}$$

which completes the proof. □

With the aid of the classical maximum principle for elliptic equations one can prove the well-known Pogorelov lemma in which the maximum principle is applied to the function $(x,e)\to(C-\Phi)\partial_e^2\Phi(x)$ on the set $S^1\times\{\Phi\leqslant C\}$. As a result we obtain an estimate of $\sup_x\|D^2\Phi(x)\|$ on the sublevel set $\Omega=\{\Phi\leqslant C\}$ by a quantity depending on V and its derivatives up to the second order. The following formulation of Pogorelov's lemma is borrowed from [72].

Lemma 2.2.6. *Let $\Phi\in C^4(\Omega)$, where Ω is a bounded domain, $\Phi=0$ on $\partial\Omega$, and* $\det D^2\Phi=\mathrm{e}^{-V}$. *Then $-\Phi(x)\|D^2\Phi(x)\|\leqslant C(1+\sup_{y\in\Omega}|\nabla\Phi(y)|^2)$, where C depends on d,* $\sup\Phi$, *and* $\|V\|_{C^2(\Omega)}$.

Thus, by controlling the L^∞-norm of $\|D^2\Phi\|$ one can apply the classical techniques of regularity theory to the differential operator L_Φ, which will be (locally) uniformly elliptic. We briefly discuss Calabi's idea, which made it possible to obtain deep results about smoothness of solutions of the Monge–Ampère equation. For an arbitrary smooth convex function Φ we introduce a Riemannian metric on $\mathbb{R}^d$ by $g_{ij}=\Phi_{ij}$, where $\Phi_{ij}=\partial_{x_i}\partial_{x_j}\Phi$ (below, Φ_{ijk} is understood similarly). Let M denote the manifold obtained. It belongs to the class of the so-called Hessian manifolds, which are real analogues of Kähler manifolds, very popular in differential geometry and mathematical physics. In obtaining a priori estimates for the Monge–Ampère equation $\det D^2\Phi = 1$ it is useful to consider the Ricci tensor of this manifold. Direct calculations give the following result: $4\,\mathrm{Ric}_{ik} = g^{jl}g^{ms}(\Phi_{mil}\Phi_{sjk}-\Phi_{mik}\Phi_{sjl})$, where (g^{ij}) is the inverse matrix for (g_{ij}). Taking into account the Monge–Ampère equation, we get that $4\,\mathrm{Ric}_{ik} = g^{jl}g^{ms}\Phi_{mil}\Phi_{sjk}$, whence it follows that the Ricci tensor of the manifold M is non-negative. Let R denote the scalar curvature (the contraction of the Ricci tensor) and let Δ_M denote the Laplace–Beltrami operator. The basic estimate obtained by Calabi is as follows.

Theorem 2.2.7. $\Delta_M R \geqslant C(d)R^2$.

Using this inequality together with comparison theorems on a Riemannian manifold, one can obtain bounds on the growth of R (that is, on third-order derivatives of Φ). Some generalizations of these results to equations of the form $\mathrm{e}^{-V} = \mathrm{e}^{-W(\nabla\Phi)}\det D^2\Phi$ are obtained in a paper of the second author in preparation, where it is shown that the Dirichlet form $\mathscr{E}_\Phi$ considered above possesses a non-negative 'carré du champ' operator if V and W are convex.

2.2.4. Regularity in Hölder and Sobolev spaces. As already noted, the regularity theory for the Monge–Ampère equation has a long history and a very impressive list of publications. The first results in this direction were obtained by Aleksandrov, who introduced the concept of a solution generalized in the sense of Aleksandrov, and by Pogorelov, who gave a now well-known example (see [16]) of a non-smooth solution of the Monge–Ampère equation with an infinitely differentiable right-hand side. The reason for this phenomenon is the lack of uniform convexity of the solution. However, in the presence of uniform convexity it is possible to prove the regularity of the solution. Further development is connected with the works of Krylov, Nirenberg, Spruck, N.M. Ivochkina, J. Urbas, N. Trudinger, Yau, and others. The relative completeness of this theory was achieved in the 1990s in papers of Caffarelli. The survey [72] gives a concise exposition of many results that previously had long and cumbersome proofs. We mention two important regularity results of Caffarelli (see [72], [16], [64]).

Theorem 2.2.8. *Let $f\,dx$ and $g\,dy$ be probability measures on bounded connected open sets X and Y in $\mathbb{R}^d$ and let $\nabla\varphi$ be the corresponding optimal transportation. If Y is convex and if f and g are bounded and bounded away from zero, then $\nabla\varphi$ is Hölder of some order. If f and g are Hölder of some order, then the derivative of $\nabla\varphi$ is Hölder of some order.*

Constructive estimates of Hölder norms are obtained in [73]. It was recently shown in [74] that $\nabla\varphi \in W^{1,1}(X)$ in the first case in this theorem. The convexity of Y is important: examples are known of discontinuous optimal transportations of Lebesgue measure on a ball to Lebesgue measure on a non-convex connected set with discontinuous $\nabla\varphi$ (see Pogorelov's example in [16]). If Y is not connected, then $\nabla\varphi$ must be discontinuous. In [75] the continuity of optimal maps is proved for the cost function $|x-y|$ under the assumption that the measures μ and ν are given by strictly positive continuous densities on disjoint convex compact sets in $\mathbb{R}^2$.

Theorem 2.2.9. *Let Ω be a convex set and let φ be a solution of the equation* $\det D^2\varphi = f$, $\varphi|_{\partial\Omega} = 0$, *in the sense of Aleksandrov, where the function f is bounded and bounded away from zero. Then for every $p > 1$ there exists a number $\varepsilon > 0$ such that if $|f - 1| < \varepsilon$, then $\varphi \in W^{2,p}_{\mathrm{loc}}(\Omega)$.*

It is shown in [79] that under a general estimate $\lambda \leqslant f \leqslant \Lambda$ the exponent p cannot be taken arbitrarily large.

2.2.5. Global estimates and applications. Finite-dimensional techniques of regularity theory seldom yield a priori estimates independent of dimension or global estimates. It turns out that for these purposes the classical approach based on differentiating the equations is efficient. We present Caffarelli's contraction theorem.

Theorem 2.2.10. *Let $T = \nabla\Phi$ be an optimal transportation of a probability measure $\mu = \mathrm{e}^{-V}dx$ on $\mathbb{R}^d$ to a probability measure $\nu = \mathrm{e}^{-W}dx$, where V and W are twice continuously differentiable and $D^2W \geqslant K$. Then for every unit vector e*

$$\sup_{x\in\mathbb{R}^d} \Phi^2_{ee}(x) \leqslant \frac{1}{K} \sup_{x\in\mathbb{R}^d} V_{ee}(x).$$

In particular, if μ is the standard Gaussian measure and $K \geqslant 1$, then T is a 1-Lipschitz map.

The idea of the proof is that $L_\Phi \Phi_{ee}(x_0) \leqslant 0$ for a maximum point x_0 of Φ_{ee} (L_Φ is an elliptic operator), and then from (2.2.3) we get that $K\Phi^2_{ee}(x_0) \leqslant \sup_{x\in\mathbb{R}^d} V_{ee}(x)$. We remark that the original theorem in [77] is somewhat different from this formulation. Caffarelli's original result is stated below.

Theorem 2.2.11. *Let $\mu = \mathrm{e}^{-Q}dx$ be an arbitrary Gaussian measure. Then for every measure $\nu = \mathrm{e}^{-Q-P}dx$, where P is a convex function, the corresponding optimal transportation T is 1-Lipschitz.*

This theorem together with the Gaussian isoperimetric inequality (see Chap. 3) implies the following fact.

Corollary 2.2.12 (Bakry–Ledoux comparison theorem [78]). *Any probability measure $\mu = \mathrm{e}^{-W}dx$ with $D^2W \geqslant K$ satisfies the Gaussian-type isoperimetric inequality $\mu(A^h) \geqslant \Phi(\Phi^{-1}(\mu(A)) + Kh)$.*

The Bakry–Ledoux comparison theorem asserts that isoperimetric properties of measures of the form $\mu = \mathrm{e}^{-W}dx$ with $D^2W \geqslant K$ are no worse than the isoperimetric properties of the Gaussian measure $C_{d,K}\mathrm{e}^{-K|x|^2/2}$. The proof follows immediately from the existence of a 1-Lipschitz transportation of the measure μ to the measure $C_{d,K}\mathrm{e}^{-K|x|^2/2}$ and the Gaussian isoperimetric inequality.

Example 2.2.13. It is explained in [79] how the contraction result can imply the well-known particular case of the so-called correlation inequality $\gamma(A \cap B) \geqslant \gamma(A)\gamma(B)$, where γ is the standard Gaussian measure, if A is an absolutely convex set and B is an ellipsoid centred at the origin. It is not known whether this inequality is true for general absolutely convex sets B.

A proof of Caffarelli's theorem can also be obtained by applying the maximum principle to the difference

$$\Phi(x+te)+\Phi(x-te)-2\Phi(x)$$

rather than to second-order derivatives. This enables one to deal with non-smooth potentials. The next result is proved in [80].

Theorem 2.2.14. *Let* $\mu=\gamma=(2\pi)^{-d/2}\mathrm{e}^{-|x|^2/2}\,dx$ *and* $\nu=\mathrm{e}^{-W}dx$, *where* $W(x+y)+W(x-y)-W(x)\geqslant\delta(|y|)$ *for some increasing non-negative function* δ. *Then* $|\nabla\Phi(x)-\nabla\Phi(y)|\leqslant 8\delta^{-1}(4|x-y|^2)$.

From these estimates it is easy to get a proof of the concentration inequality for measures with uniformly convex potentials for the standard norm (see § 3.4). A question of obvious interest is whether it is possible to generalize Caffarelli's theorem to manifolds.

In connection with this problem we note that the Bakry–Ledoux comparison theorem is a 'flat' analogue of the Lévy–Gromov comparison theorem in Riemannian geometry. Moreover, both theorems are particular cases of comparison theorems for manifolds with measures. However, unlike in the 'flat' case, no transport proof is presently known for the Lévy–Gromov theorem. Another open problem is this: how can one estimate the contraction constant if the image-measure is not uniformly convex? In particular, let $\nabla\Phi$ be the optimal transportation of the standard Gaussian measure γ to the normalized Lebesgue measure on a convex set K. Is there an estimate

$$\int \|D^2\Phi(x)\|\,d\gamma\leqslant C,$$

where C is a constant that does not depend on the dimension (for example, $C=c\cdot\mathrm{diam}(K)$ or C is a universal constant for isotropic convex sets)?

This problem is motivated by the well-known Kannan–Lovasz–Simonovits conjecture (KLS-conjecture). We recall that Cheeger's constant $C_{\mathrm{Chig}}(K)$ for a convex body K is defined to be the smallest constant C for which

$$\int_K\biggl|f(x)-\frac{1}{\lambda(K)}\int_K f(y)\,dy\biggr|\,dx\leqslant C\int_K|\nabla f(x)|\,dx\qquad\forall\, f\in C_0^\infty(\mathbb{R}^d).$$

KLS-Conjecture. *There is a universal constant c such that $C_{\mathrm{Chig}}(K)\leqslant c$ for every convex set $K\subset\mathbb{R}^d$ satisfying the equalities*

$$\int_K x_i\,dx=0,\qquad \frac{1}{\lambda(K)}\int_K x_ix_j\,dx=\delta_i^j;$$

such bodies are said to be isotropic.

It is proved in [80] that $\sup_x\|D^2\Phi(x)\|\leqslant c\sqrt{d}\,\mathrm{diam}(K)$, but this does not give even an estimate $C_{\mathrm{Chig}}\leqslant c\cdot\mathrm{diam}(K)$. On the other hand, it is known that for estimating Cheeger's constant it suffices to estimate the L^1-norm of $\Lambda(x)=\|D^2\Phi(x)\|$. In a paper in preparation by the second author, Calabi's techniques is employed to prove that

$$\int\Lambda\,d\gamma-\biggl(\int\sqrt{\Lambda}\,d\gamma\biggr)^2\leqslant c\cdot\mathrm{diam}(A).$$

Some applications of Calabi's metric and techniques of Kähler manifolds to so-called thin-shell estimates for convex sets were obtained in [81].

Let us proceed to global Sobolev estimates. Integrating (2.2.3), one can obtain global estimates on the second and third derivatives of Φ. Indeed, since

$$\int V_{ee}\,d\mu=\int V_e^2\,d\mu,\qquad \int L_\Phi \Phi_{ee}\,d\mu=0,$$

we get that

$$\int V_e^2\,d\mu=\int\bigl\langle D^2W(\nabla\Phi)D^2\Phi\cdot e, D^2\Phi\cdot e\bigr\rangle\,d\mu+\int \mathrm{Tr}[(D^2\Phi)^{-1}D^2\Phi_e]^2\,d\mu.$$

The second term on the right-hand side is non-negative, and hence if $D^2W\geqslant K\cdot\mathrm{Id}$, then

$$\int|\partial_e V|^2\,d\mu\geqslant K\int|\nabla\partial_e\Phi|^2\,d\mu.$$

A more general result is obtained in [82].

Theorem 2.2.15. *Suppose that $D^2W\geqslant K\cdot\mathrm{Id}$ with $K>0$. Then for every unit vector e and for $p\geqslant 1$ the estimates*

$$K\|\Phi_{ee}^2\|_{L^p(\mu)}\leqslant\|(V_{ee})_+\|_{L^p(\mu)},\qquad K\|\Phi_{ee}^2\|_{L^p(\mu)}\leqslant\frac{p+1}{2}\|V_e^2\|_{L^p(\mu)}$$

are satisfied. Moreover, for any $r\geqslant 1$ the following inequality holds for the operator norm $\|D^2\Phi\|$:

$$K^r\int\|D^2\Phi\|^{2r}\,d\mu\leqslant\int\|(D^2V)_+\|^r\,d\mu,$$

where $(D^2V)_+$ is the positive part of the operator D^2V.

As $p\to\infty$ we again obtain the estimate $K\|\Phi_{ee}\|^2_{L^\infty(\mu)}\leqslant\|(V_{ee})_+\|_{L^\infty(\mu)}$ from Caffarelli's theorem. The estimates in this theorem can be generalized to the Wiener space (see § 3.8 and [83], [84], [82]).

About integral estimates, see [85]. Less is known about general cost functions (even smooth). In [86] there is an example of a smooth connected compact manifold in $\mathbb{R}^3$ and probability measures with smooth positive densities for which the optimal map (for the usual quadratic cost function on $\mathbb{R}^3$) is discontinuous (see also [10], Example 1.3.7). The situation is not saved even by the non-negative sectional curvature of the manifold; it turns out that here the major role is played by the so-called Ma–Trudinger–Wang tensor. About this see [87], [88], [72]. In [89]–[91] the problem of transportation of part of the mass is studied and an extensive bibliography is given.

2.3. Connections with the Kantorovich problem and approximate solutions

As already noted, the Monge problem does not always have a solution even for very simple spaces, measures, and functions. All the more surprising is the fact that under rather broad assumptions this problem has approximate solutions, that is, transformations $T \in T(\mu,\nu)$ for which $M(\mu,\nu,T)$ is as close to $K(\mu,\nu)$ as we wish. In other words, the Monge infimum coincides with the Kantorovich minimum. We give a precise formulation. Lipchius [92] strengthened a result of Pratelli [93] (who in turn strengthened results of W. Gangbo and L. Ambrosio) and established the following fact.

Theorem 2.3.1. *Let X and Y be completely regular topological spaces in which all compact sets are metrizable, let h be a continuous function on $X \times Y$, and let μ and ν be Radon probability measures on X and Y, respectively. Let μ have no atoms. Then there exists a Borel map $T\colon X \to Y$ taking μ to ν for which $M(\mu,\nu,T) \leqslant K(\mu,\nu,\mu\otimes\nu)$. Moreover,*

$$\min_{\sigma\in\Pi(\mu,\nu)} K(\mu,\nu,\sigma) = \inf_{T\in T(\mu,\nu)} M(\mu,\nu,T).$$

It is not known whether the metrizability of compact sets in this theorem can be omitted if we consider pairs of measures μ and ν such that μ can be transformed into ν. Of course, for general spaces the absence of atoms for the measure μ does not guarantee this. Since some details of the proof were skipped in [92], we give a detailed justification (borrowed from the Ph.D. dissertation of Lipchius). Let us introduce some auxiliary notation. If μ and ν are non-negative Radon measures on topological spaces X and Y and $\mu(X) = \nu(Y)$, then we set

$$\mu \boxtimes \nu := \frac{\mu\otimes\nu}{\mu(X)}\,.$$

Let $\pi_X\gamma$ and $\pi_Y\gamma$ denote the projections of a measure γ on $X\times Y$ onto the factors X and Y, respectively. The closure of a set A in a topological space will be denoted by $\overline{A}$. Let I be a finite or infinite binary (consisting of 0s and 1s) sequence. Then I_k denotes the first k symbols in I if I has at least k symbols, and $l(I)$ denotes the total number of symbols in a finite sequence of numbers.

The following lemma extends the result proved in [93] in the case of Polish spaces, to completely regular spaces with metrizable compact sets.

Lemma 2.3.2. *Let μ be a Radon probability measure on a completely regular space X in which all compact sets are metrizable. Then for every binary sequence I there exist a Borel set X_I in X and a Radon measure μ_I on X with the following properties:*

1) $\|\mu_I\| = 2^{-k}$ *for every sequence I of length k;*
2) $\mu_I = \mu_{(I,0)} + \mu_{(I,1)}$ *and* $\mu = \sum_{l(I)=k} \mu_I$;
3) X_I *has full μ_I-measure;*
4) *if $I \leqslant J$, then $X_J \subseteq X_I$;*
5) *the set of countable binary sequences for which the intersection $\bigcap_{k\in\mathbb{N}} \overline{X}_{I_k}$ contains more than one point is at most countable;*
6) 6) *if μ has no atoms, then for every natural number k the sets X_I with $l(I) = k$ are open and disjoint, and $\mu_I = \mu|_{X_I}$.*

Proof. Any Radon probability measure μ on a completely regular space X with metrizable compact sets is concentrated on a Souslin set (a countable union of metrizable compact sets). Hence we can assume without loss of generality that X is a completely regular Souslin space. We use the fact that on any completely regular Souslin space there exists a countable family of continuous functions separating the points. This means that any completely regular Souslin space can be regarded as a subset of $\mathbb{R}^\infty$ with a stronger topology, that is, we can assume that X is equipped with a metric ϱ such that ϱ-open sets are open in X. The boundary of any set in the X-topology belongs to its boundary with respect to the ϱ-topology. Since $\mathbb{R}^\infty$ is separable, X has a countable family of points x_n such that it can be covered by balls about these points of arbitrarily small radius with respect to the metric ϱ. By induction we construct open sets B_I, where I is a finite binary sequence. At the first step we construct sets B_n, $n \in \mathbb{N}$. For every x_n we choose $r_n \in [1/2, 1]$ such that the ball about x_n of radius r_n has ϱ- (hence also X-) boundary of μ-measure zero. Let

$$B_1 = B(z_1, r_1), \qquad B_n = B(z_n, r_n) \setminus (\overline{B}_1 \cup \dots \cup \overline{B}_{n-1}), \quad n \geqslant 2.$$

Then $\mu(\overline{B}_n \setminus B_n) = 0$ and $\mu\bigl(\bigcup B_n\bigr) = 1$. Next we argue by induction: if the B_I with $l(I) \leqslant n-1$ are already defined, then we continue our construction by replacing X by B_I, μ by $\mu|_{B_I}$, $\{x_n\}$ by an analogous sequence in B_I, and $1/2$ and 1 by 2^{-n} and 2^{1-n}, respectively, in the first step. Now we construct X_I and μ_I inductively. We use induction on the length of the binary sequence J. Let $B_I^J := B_I \cap X_J$. We assume that $X_\varnothing = X$ and $\mu_\varnothing = \mu$. Suppose that all the sets X_J and all the measures μ_J with $l(J) < n$ have been constructed. Let $l(J) = n$. We find a j_1 such that

$$\mu_J(B_1^J \cup \dots \cup B_{j_1-1}^J) \leqslant 2^{-n-1} < \mu_J(B_1^J \cup \dots \cup B_{j_1}^J).$$

Let $X_0^1 := B_1^J \cup \dots \cup B_{j_1-1}^J$ and $X_1^1 := B_{j_1+1}^J \cup B_{j_1+2}^J \cup \cdots$. Here we assume that $B_1^J \cup \dots \cup B_0^J = \varnothing$. Then X_0^1 and X_1^1 are open disjoint sets of measure at most 2^{-n-1}. We find a j_2 such that

$$\mu_J(X_0^1 \cup B_{j_1,1}^J \cup \dots \cup B_{j_1,j_2-1}^J) \leqslant 2^{-n-1} < \mu_J(X_0^1 \cup B_{j_1,1}^J \cup \dots \cup B_{j_1,j_2}^J).$$

Define

$$X_0^2 := X_0^1 \cup B_{j_1,1}^J \cup \dots \cup B_{j_1,j_2-1}^J, \quad X_1^2 := X_1^1 \cup B_{j_1,j_2+1}^j \cup B_{j_1,j_2+2}^j \cup \cdots.$$

Then X_0^2 and X_1^2 are open disjoint sets of measure at most 2^{-n-1}. Continuing this construction, we obtain sets X_0^i and X_1^i with the same properties. Let $\widehat{X}_0 = \bigcup_i X_0^i$ and $\widehat{X}_1 = \bigcup_i X_1^i$. We note that the set $X_J \setminus (\widehat{X}_0 \cup \widehat{X}_1) = \bigcap_k B_{j_1,\dots,j_k}^J$ contains at most one point. In addition, $\mu(\widehat{X}_0) = 2^{-n-1}$ and $\mu(\widehat{X}_1) = 2^{-n-1}$. If μ has no atoms or if the intersection is empty, then we let $X_{(J,0)} := \widehat{X}_0$ and $X_{(J,1)} := \widehat{X}_1$ and obtain open disjoint sets of measure 2^{-n-1}. In this case we let $\mu_{(J,0)} := \mu_J|_{X_{(J,0)}}$ and $\mu_{(J,1)} := \mu_J|_{X_{(J,1)}}$. Otherwise we let

$$X_{(J,i)} := \widehat{X}_i \cup \{z\}, \quad \mu_{(J,i)} := \mu_J|_{\widehat{X}_i} + (2^{-n-1} - \mu_J(\widehat{X}_i))\delta_z, \qquad i = 0, 1.$$

In both cases $\mu_J = \mu_{(J,0)} + \mu_{(J,1)}$. Since the new measures are obtained by restricting the previously constructed measures to measurable sets, and possibly by splitting up atoms of the old measures, the sets B_I preserve their properties also for the constructed measures. The inductive step is complete. It is seen from the construction that the constructed sets and measures possess the desired properties. Only the property 5) requires justification.

By construction, for any two finite binary sequences J and L of equal length there exists at most one countable binary sequence I with the property that $X_{I_k} \cap B_J \neq \varnothing$ and $X_{I_k} \cap B_L \neq \varnothing$ for all positive integers k. We denote this sequence by $I(J, L)$. Let $A = \{I(J, L)\}$, where J and L are finite sequences of equal length. Then A is at most countable. We show that for all other $I \in \{0,1\}^{\mathbb{N}}$ the set $\bigcap_{k\in\mathbb{N}} \overline{X}_{I_k}$ contains at most one point. To this end it suffices to show that $\bigcap_{k\in\mathbb{N}} X_{I_k}$ contains at most one point, since $\overline{X}_{I_k} \subset X_{I_{k-1}}$. For this, in turn, it suffices to show that $\bigcap_{k\in\mathbb{N}} X_{I_k} \subset B_{J^n}$, where J^n is a sequence of length $n \in \mathbb{N}$. Let us prove this by induction. The sequence of sets $T_k := \{m \in \mathbb{N} \colon X_{I_k} \cap B_m \neq \varnothing\}$ is decreasing. Its intersection cannot contain more than one point, because otherwise $I = I(j, l) \in A$, where j and l are elements in the intersection. Since $\bigcap_k T_k = \bigl\{m \in \mathbb{N} \colon \bigl(\bigcap_k X_{I_k}\bigr) \cap B_m \neq \varnothing\bigr\} = j$, we have obtained our assertion for $n = 1$. Passing to B_j and replacing the sets B_m in the definition of T_k by $B_{j,m}$, we complete the inductive step. □

Lemma 2.3.3. *Let X and Y be completely regular topological spaces with metrizable compact sets, let h be a continuous function on $X \times Y$, and let μ and ν be Radon probability measures on X and Y, respectively. Let γ be a Radon probability measure on $X \times Y$ with $\pi_X\gamma = \mu$ and $\pi_Y\gamma = \nu$. If μ has no atoms, then for every $\varepsilon > 0$ there exist sequences of measurable sets $A_n \subset X$, $B_n \subset Y$, and $X_n \subset A_n$ such that*

1) $\sup_{(x_1,y_1),(x_2,y_2)\in A_n\times B_n} |h(x_1, y_1) - h(x_2, y_2)| < \varepsilon$,
2) $(A_n \times B_n) \cap (A_k \times B_k) = \varnothing$ *for* $n \neq k$,
3) $\gamma\bigl(\bigcup(A_n \times B_n)\bigr) = 1$,
4) $\mu(X_n \cap X_k) = 0$ *for* $n \neq k$,
5) $\mu(X_n) = \gamma(A_n \times B_n)$.

Proof. Since μ and ν are Radon measures, there exist compact sets $K_n^1 \subset X$ and $K_n^2 \subset Y$ such that $\mu(X\backslash K_n^1) < n^{-1}$ and $\nu(Y\backslash K_n^2) < n^{-1}$. We may assume that $K_n^1 \subset K_{n+1}^1$ and $K_n^2 \subset K_{n+1}^2$. The compact sets K_n^1 and K_n^2 are metrizable. We denote their metrics by ϱ_n^1 and ϱ_n^2. The compact set $K_n^1 \times K_n^2$ is metrizable by the metric

$$\varrho_n(x_1 \times y_1, x_2 \times y_2) := \max(\varrho_n^1(x_1, x_2), \varrho_n^2(y_1, y_2)).$$

The continuous function h is uniformly continuous on the compact set $K_n^1 \times K_n^2$. Let $\delta_n > 0$ be such that $|h(x_1, y_1) - h(x_2, y_2)| < \varepsilon$ if $\varrho_n(x_1 \times y_1,\, x_2 \times y_2) < \delta_n$ and $x_1, x_2 \in K_n^1$, $y^1, y^2 \in K_n^2$. Let us cover K_n^1 by finitely many disjoint measurable sets $A_k^n \subset K_n^1$ of diameter less than δ_n. Passing to the intersection, we may assume that if $A_k^n \cap A_l^{n-1} \neq \varnothing$, then $A_k^n \subset A_l^{n-1}$. Now let us cover K_n^2 by finitely many disjoint measurable sets $B_k^n \subset K_n^2$ of diameter less than δ_n. Then $A_k^n \times B_l^n$ are disjoint rectangles (for fixed n) covering $K_n^1 \times K_n^2$, and the dispersion of the values of h on each of them is less than ε. Let

$$\widetilde{A}_k^n := (A_k^n \cap K_n^1) \setminus K_{n-1}^1, \qquad \widetilde{B}_k^n := (B_k^n \cap K_n^2) \setminus K_{n-1}^2,$$
$$\widehat{A}_k^n := A_k^n \cap K_{n-1}^1, \qquad \widehat{B}_k^n := B_k^n \cap K_{n-1}^2.$$

Then the disjoint rectangles $\widetilde{A}_k^n \times B_l^n$ and $\widehat{A}_k^n \times \widetilde{B}_l^n$ cover the union $\bigcup(K_n^1 \times K_n^2)$, and the dispersion of the values of h on each of them is less than ε. We number the countable collection of rectangles obtained in a consecutive order and denote them by $A_n \times B_n$. These rectangles possess the properties 1), 2), 3). By construction either $A_n \cap A_k = \varnothing$ or $A_n \subset A_k$ and $k \leqslant n$.

The Souslin set $S := \bigcup_n K_n^1$ has full μ-measure. Since μ has no atoms, by the isomorphism theorem the measure space (S, μ) is isomorphic to the unit interval with Lebesgue measure (see [29], Chap. 9). Hence, in every measurable set $A \subset S$ for every $t \leqslant \mu(A)$ one can find a measurable subset B with $\mu(B) = t$.

We construct an $X_n \subset A_n$ with the properties 4) and 5). Let X_1^1 be a set in A_1 such that $\mu(X_1^1) = \gamma(A_1 \times B_1)$. We find the smallest k_1 such that $\mu(A_{k_1} \cap A_1) > 0$. Let $X_2^1, \dots, X_{k_1-1}^1$ be any sets in $A_2, \dots, A_{k_1-1}$, respectively, for which $\mu(X_2^1) = \gamma(A_2 \times B_2), \dots, \mu(X_{k_1-1}^1) = \gamma(A_{k_1-1} \times B_{k_1-1})$. We find a set $X_{k_1}^1$ in A_{k_1} with $\mu(X_{k_1}^1) = \gamma(A_{k_1} \times B_{k_1})$. Let $R_1^1 = A_1 \setminus (X_1^1 \cup X_{k_1}^1)$, and let $X_1^2 = (X_1^1 \setminus X_{k_1}^1) \cup D_1^1$, where $D_1^1 \subset R_1^1$ and $\mu(D_1^1) = \mu(X_1^1 \cap X_{k_1}^1)$. Further, let

$$X_i^2 = X_i^1, \qquad i = 2, \dots, k_1.$$

We find the smallest $k_2 > k_1$ such that $\mu(A_{k_2} \cap A_i) > 0$ for some i. Let $X_{k_1+1}^2, \dots, X_{k_2-1}^2$ be any sets in $A_{k_1+1}, \dots, A_{k_2-1}$, respectively, for which $\mu(X_{k_1+1}^2) = \gamma(A_{k_1+1} \times B_{k_1+1}), \dots, \mu(X_{k_2-1}^2) = \gamma(A_{k_2-1} \times B_{k_2-1})$. There is a set $X_{k_2}^2$ in A_{k_2} with $\mu(X_{k_2}^2) = \gamma(A_{k_2} \times B_{k_2})$. For $j = k_1, k_1 - 1, \dots, 1$ we let

$$R_j^2 = A_j \setminus \biggl(\bigcup_{l:\ A_l \subset A_j} X_l^2 \cup \bigcup_{l>j} X_l^3 \cup X_{k_2}^2 \biggr),$$

$$D_j^2 \subset R_j^2, \quad \mu(D_j^2) = \mu(X_j^2 \cap X_{k_2}^2), \quad X_j^3 = (X_j^2 \setminus X_{k_2}^2) \cup D_j^2.$$

Define $C_j^2 = X_j^2 \cap X_{k_2}^2$, and let $X_i^3 = X_i^2$ for $i = k_1 + 1, \dots, k_2$. Continuing this construction, we obtain sets X_j^i, R_j^i, D_j^i, C_j^i. We have $R_j^k \subset R_j^l$ if $k > l$, and the sets D_j^i are pairwise disjoint. The sets C_j^i are pairwise disjoint for any fixed j. By construction $\sum_i \mu(D_j^i) = \sum_i \mu(C_j^i)$ for every j. For every n we can find the smallest m such that $n \leqslant k_m$. Let

$$X_n = \biggl(X_n^m \cup \bigcup_{l>m} D_n^l \biggr) \setminus \biggl(\bigcup_{l>m} C_n^l \biggr).$$

Then

$$\mu(X_n) = \mu(X_n^m) + \sum_i \mu(D_n^i) - \sum_i \mu(C_n^i) = \mu(X_n^m) = \gamma(A_n \times B_n)$$

since $\bigcup_{l>m} C_n^l \subset X_n^m \cup \bigcup_{l>m} D_n^l$. We now have $\mu(X_n \cap X_k) = 0$ for $n \neq k$. □

Proof of Theorem 2.3.1. As observed in the proof of Lemma 2.3.2, we may assume without loss of generality that X and Y are completely regular Souslin spaces. For the proof of the first assertion, (as in [93]) we use the indicated lemma to find the corresponding X_I, Y_I, μ_I, and ν_I. We construct a sequence

$$\gamma_k = \sum_{I\in\{0,1\}^k} \mu_I \boxtimes \nu_{\varphi_k(I)},$$

where $\varphi_k\colon \{0,1\}^k \to \{0,1\}^k$ is a bijection with the following properties:

1) the sequence $K(\gamma_k)$ is non-increasing;
2) φ_{k+1} extends φ_k, or more precisely, $\varphi_r(I_r) = (\varphi_s(I_s))_r$ for $0 \leqslant r \leqslant s$ and $I \in \{0,1\}^s$.

The sequence φ_k will be constructed by induction. At the first step we define φ_1 such that $K(\gamma_1) \leqslant K(\mu\otimes\nu)$. This can be done since

$$2\mu\boxtimes\nu = (\mu_0\boxtimes\nu_0 + \mu_1\boxtimes\nu_1) + (\mu_0\boxtimes\nu_1 + \mu_1\boxtimes\nu_0).$$

Therefore, $K(\mu\boxtimes\nu) = 2^{-1}(K(\mu_0\boxtimes\nu_0+\mu_1\boxtimes\nu_1)+K(\mu_0\boxtimes\nu_1+\mu_1\boxtimes\nu_0))$. In the last formula either the first or the second term does not exceed $K(\mu\boxtimes\nu)$. If this is the first term, then we let $\varphi_1(0)=0$ and $\varphi_1(1)=1$. If it is the second term, then we let $\varphi_1(0)=1$ and $\varphi_1(1)=0$. Suppose that $\varphi_1,\dots,\varphi_{k-1}$ are already defined. Let $l(I)=k$. Then $\varphi_k(I) = (\varphi_{k-1}(I_{k-1}),x)$, where x is chosen so that the property 1) holds. This can always be done. For the proof it suffices to replace μ and ν in the first step by $\mu_{I_{k-1}}$ and $\nu_{\varphi_{k-1}(I_{k-1})}$ and then to use the decompositions

$$\mu_{I_{k-1}} = \mu_{(I_{k-1},0)} + \mu_{(I_{k-1},1)}, \quad \nu_{\varphi_{k-1}(I_{k-1})} = \nu_{(\varphi_{k-1}(I_{k-1}),0)} + \nu_{(\varphi_{k-1}(I_{k-1}),1)}.$$

Thus, having defined φ_k, we obtain a sequence of Radon measures γ_k whose projections on X and Y are μ and ν. Therefore, these measures are uniformly tight. By the strengthened Prokhorov theorem the family $\{\gamma_k\}$ is sequentially weakly compact (see Theorem 8.6.7 in [29]). Hence we can pick a weakly convergent subsequence. Its limit will be denoted by γ. The projections of γ on X and Y are μ and ν, and $K(\gamma)\leqslant K(\mu\otimes\nu)$. We now prove that the support of γ is the graph of some map. Let us consider the set

$$Z := \bigcap_{k\in\mathbb{N}} \bigcup_{l(I)=k} X_I$$

and the function $\psi\colon Z\to\{0,1\}^{\mathbb{N}}$ which maps each x in Z into the unique countable binary sequence I such that $x = \bigcap_{k\in\mathbb{N}} X_{I_k}$. We note that Z has full μ-measure. Denote by A the set of countable sequences I such that $\bigcap_{k\in\mathbb{N}} \overline{Y}_{I_k}$ contains more than one point. Then by the property 5) in Lemma 2.3.3 we get that A is at most countable. Let $\varphi\colon\{0,1\}^{\mathbb{N}}\to\{0,1\}^{\mathbb{N}}$ be the extension of all the functions φ_k. By the definition of φ_k, φ is a bijection. Since μ has no atoms, the set $\widetilde{X} = \{x\in Z\colon \varphi(\psi(x))\notin A\}$ has full μ-measure. The set $S_k = \bigcup\{X_I\times\overline{Y}_{\varphi_k(I)}\colon I\in\{0,1\}^k\}$ has full γ_k-measure due to the fact that $\mu(X_I)=\mu(\overline{X}_I)$. This set S_k has full γ_n-measure for all $n\geqslant k$. We show that S_k has full γ-measure. Indeed,

$$\begin{aligned}\gamma(X_I\times\overline{Y}_{\varphi_k(I)}) &= \gamma(\overline{X}_I\times\overline{Y}_{\varphi_k(I)}) \geqslant \limsup_{n\to\infty}\gamma_n(\overline{X}_I\times\overline{Y}_{\varphi_k(I)})\\ &= \limsup_{n\to\infty}\gamma_n(\overline{X}_I\times Y) = \mu(\overline{X}_I) = \mu(X_I)\\ &= \gamma(X_I\times Y)\geqslant\gamma(X_I\times\overline{Y}_{\varphi_k(I)}).\end{aligned}$$

Let $S = \bigcap_k S_k$. Then S is a Borel set in $\widetilde{X} \times Y$ and $\gamma(S) = 1$. For every element $x \in \widetilde{X}$ the set $\{y\colon (x, y) \in S\} = \overline{Y}_{\varphi(\psi(x))}$ contains at most one point. The set $(\widetilde{X} \times Y) \cap S$ is Borel measurable. Its projection Π on $\widetilde{X}$ is a Souslin set and

$$\mu(X \setminus \Pi) = \gamma((X \setminus \Pi) \times Y) = 0,$$

because $\gamma(S) = 1$. For every $x \in \Pi$ there exists a unique element $y \in Y$ with $(x, y) \in S$, that is, the Souslin set $S \cap (\Pi \times Y)$ is the graph of some map $T\colon \Pi \to Y$. It is known (see Lemma 6.7.1 in [29]) that T is a Borel map from Π to Y. Extending it to have a constant value on the complement of Π, we obtain a Borel map from X to Y with the required properties. The first assertion of the theorem is proved.

Let us proceed to the second assertion. Let γ be a measure on $X \times Y$ on which the minimum is attained. Fix an $\varepsilon > 0$. We use Lemma 2.3.3 and take the sets A_n, B_n, and X_n indicated there. By the property 3) the rectangles $A_n \times B_n$ cover γ-almost all of $X \times Y$. On each of them the dispersion of the values of h is less than ε by the property 1). By the property 4), $\mu(X_n) = \gamma(A_n \times B_n) =: m_n$. We introduce the measures $\gamma_n := \gamma|_{A_n \times B_n}$ and $\widetilde{\gamma}_n := \mu|_{X_n} \boxtimes \pi_Y \gamma_n$. The projections of the measure $\widetilde{\gamma}_n$ are $\mu|_{X_n}$ and $\pi_Y \gamma_n$. Let us apply the first assertion of the theorem to them. We obtain a map $t_n\colon X_n \to B_n$ with the property that $M(t_n) \leqslant K(\widetilde{\gamma}_n)$. Let $T(x) = t_n(x)$ for $x \in X_n$. Since $\bigcup_n X_n$ has full μ-measure (this follows from the properties 3) and 5)), the map T takes μ to the measure $\sum_n \pi_Y \gamma_n = \pi_Y \sum_n \gamma_n = \pi_Y \gamma = \nu$. Since $|K(\widetilde{\gamma}_n) - K(\gamma_n)| \leqslant m_n \varepsilon$, we finally get that

$$M(T) \leqslant \sum_n M(t_n) \leqslant \sum_n K(\widetilde{\gamma}_n) \leqslant \sum_n (K(\gamma_n) + m_n \varepsilon) \leqslant K(\gamma) + \varepsilon \|\gamma\|.$$

The proof of the theorem is complete.

The continuity of the cost function is essential in the theorem proved and cannot even be replaced by lower semicontinuity. We give an example borrowed from [93].

Proposition 2.3.4. (i) *Let $X = Y = [0, 1] \times [-1, 1]$, let μ be the linear Lebesgue measure on the interval $I_0 = [0, 1] \times \{0\}$, let ν be half of the linear Lebesgue measure on the union of the intervals $I_1 = [0, 1] \times \{1\}$ and $I_2 = [0, 1] \times \{-1\}$, and let $h(x, y) = 0$ if $\|x - y\| = 1$ and $h(x, y) = 1$ otherwise, that is, h is the indicator function of the open set $U = \{(x, y) \in [0, 1]^4\colon \|x-y\| \neq 1\}$, which is the complement of the closed set $Z = \{(x, y) \in X \times Y\colon \|x - y\| = 1\}$. Then the Kantorovich problem and the Monge problem have solutions, but these solutions are different and $K(\mu, \nu) = 0$, $M(\mu, \nu) = 1$. In addition, the Kantorovich problem has a unique solution.*

(ii) *Let $T(x_1, x_2) = (2x_1, 1)$ for $0 \leqslant x_1 \leqslant 1/2$ and $T(x_1, x_2) = (2x_1 - 1, -1)$ for $1/2 < x_1 \leqslant 1$. Let h be redefined on the graph of T by setting $h = 1/2$ there. Then the assertion in* (i) *remains valid, but in addition also the Monge problem has a unique solution.*

Proof. (i) A solution to the Kantorovich problem is the measure $\widehat{\sigma}$ equal to half of the linear Lebesgue measure on the union of the diagonal intervals

$$D_1 = \{(x_1, 0, x_1, 1)\colon 0 \leqslant x_1 \leqslant 1\}, \qquad D_2 = \{(x_1, 0, x_1, -1)\colon 0 \leqslant x_1 \leqslant 1\}.$$

Its projections are μ and ν, and $K(\mu,\nu,\sigma)=0$, since D_1 and D_2 belong to Z. There are no other measures σ in $\Pi(\mu,\nu)$ with the same value, since we must have the equality $\sigma(Z)=1$ and also the equalities $\sigma(X\times(I_1\cup I_2))=1$ and $\sigma(I_0\times Y)=1$. In other words, the measure σ must be concentrated on the set of points of the form $(x_1,0,y_1,y_2)$, where $|y_2|=1$ and $(x_1-y_1)^2+y_2^2=1$, that is, $x_1=y_1$. Thus, σ is concentrated on $D_1\cup D_2$, which, along with the fact that the projection of σ on the first factor coincides with μ, implies that $\sigma=\widehat{\sigma}$.

Let us turn to the Monge problem. One of its solutions (but not the only one) is the map T indicated in (ii), for which $M(\mu,\nu,T)=1$, since there is no transportation S of μ to ν yielding a smaller value. To see this, we observe that if $\mu\circ S^{-1}=\nu$, then $\|x-S(x)\|\neq 1$ for μ-almost all x. Indeed, if $x=(x_1,0)\in I_0$ is such that $S(x)\in I_1\cup I_2$ (μ-almost each x is such a point), then the equality $\|x-S(x)\|=1$ is possible only if $S(x)=(x_1,y_2)$, where $|y_2|=1$. Thus, if the set $E=\{x\colon \|x-S(x)\|=1\}$ has positive measure, then μ-almost every point of it has the form $(x_1,0)$ and is moved either by 1 up or by 1 down. Hence, we may assume that $E\subset I_0$ and that every point of E is moved under S by 1 up or down. This contradicts the fact that ν is the image of μ: for example, if the set E_1 of points in E moved up has positive measure, then $\nu(E_1\times\{1\})=\mu(E_1)/2$, although $\nu(E_1)=\mu(S^{-1}(E_1\times\{1\}))\geqslant\mu(E_1)$.

(ii) The previous reasoning applies also to the new function h, but now the Monge problem also has a unique solution. Indeed, it is seen from these arguments that if $M(\mu,\nu,S)\leqslant M(\mu,\nu,T)$, then $S(x)=T(x)$ for μ-almost all x. $\square$

Using the isomorphism of spaces with atomless Borel measures, one can easily obtain from this example a case of a cost function h that is the indicator function of a Borel set in $[0,1]\times[0,1]$ such that the Kantorovich and Monge problems on $[0,1]$ with μ and ν equal to Lebesgue measure have different unique solutions.

We also give an example illustrating the difference between the cost functions $|x-y|$ and $|x-y|^2$ in the question of uniqueness.

Example 2.3.5. Let $X=Y=\mathbb{R}$, let $h(x,y)=|x-y|$, let μ be a probability measure on $(0,+\infty)$ with first moment, and let ν be the measure on $(-\infty,0)$ symmetric to it. Then every measure $\sigma\in\Pi(\mu,\nu)$ gives a solution to the Kantorovich problem, since

$$\int |x-y|\,\sigma(dx)=\int (x-y)\,\sigma(dx)=2\int_{[0,+\infty)} x\,\mu(dx),$$

because σ is concentrated on the lower right-hand quadrant. The same value is given by any map from μ to ν, and there are many such maps in the general case (one of them is the reflection).

On the connections between the Monge and Kantorovich problems see also [5], where there is a remark about reducing the Monge problem to the case of measure-preserving maps (which makes it a variational problem on the group of automorphisms). We remark that already in [18] Vershik posed the problem of minimizing the integral with respect to μ of the length of the curve $\{T_t(x)\}_{0\leqslant t\leqslant 1}$ in a suitable class of maps T_t for which $T_0(x)=x$ and T_1 takes μ to ν.

Chapter 3

Applications

We now briefly discuss applications of the transport problem to classical inequalities in analysis, probability theory and differential geometry, and also one of the most important ideas on which many applications are based: the connection between geodesics in spaces of measures and the Monge–Kantorovich problem. The following notation will be used below. If μ and ν are probability measures on $(X,\mathscr{A})$ and ν has a density ϱ with respect to μ, that is, $\nu=\varrho\cdot\mu$, then the entropy of ν (or ϱ) with respect to μ is defined by the formula

$$\operatorname{Ent}_\mu \nu = \operatorname{Ent}_\mu \varrho = \int_X \varrho\log\varrho\,d\mu$$

if $\varrho\log\varrho\in L^1(\mu)$; otherwise we set $\operatorname{Ent}_\mu \nu := +\infty$. By Jensen's inequality the entropy is non-negative.

3.1. Isoperimetric inequalities and the Brunn–Minkowski inequality

Here we give a proof proposed by Gromov for the classical isoperimetric inequality (see [94], Appendix). He found this proof by using triangular maps briefly discussed in § 3.9. The same reasoning remains valid in the case of optimal maps. It is known, though, that the isoperimetric inequality follows from the Brunn–Minkowski inequality, which is discussed below.

Theorem 3.1.1. *Let $A\subset\mathbb{R}^d$ be a Borel set. Then the following isoperimetric inequality holds*:

$$\lambda^{1-1/d}(A)\leqslant \kappa_d\mathscr{H}^{d-1}(\partial A), \qquad \kappa_d=\frac{\Gamma(1+d/2)^{1/d}}{d\sqrt{\pi}}\,.$$

Proof. Consider a ball $B_r=\{x\colon |x|\leqslant r\}$ satisfying the condition $\lambda(A)=\lambda(B_r)$. Let $T=\nabla W$ be the optimal transportation taking $\lambda|_A$ to $\lambda|_{B_r}$. We apply the change of variables formula $\det D^2_{\mathrm{a}}W=1$ on A. By the convexity of W we have $\Delta_{\mathrm{a}}W\,dx\leqslant\Delta W$, where ΔW is the Laplacian in the sense of distributions, that is, a measure. Let λ_i be the eigenvalues of $D^2_{\mathrm{a}}W$. By the inequality between the arithmetic mean and the geometric mean we have $\Delta_{\mathrm{a}}W=\sum_{i=1}^d\lambda_i\leqslant d(\lambda_1\cdots\lambda_d)^{1/d}=d$. Thus,

$$d\lambda(A)\leqslant\int_A \Delta_{\mathrm{a}}W\,dx\leqslant\Delta W(A).$$

Assuming that A has a sufficiently regular boundary and integrating by parts, we get that the latter quantity equals

$$\int_{\partial A}\langle\nabla W,n_A\rangle\,d\mathscr{H}^{d-1}\leqslant r\mathscr{H}^{d-1}(\partial A).$$

Here n_A is the unit normal to ∂A. The desired inequality follows from the equality $\lambda(A)=\lambda(B_r)=\pi^{d/2}r^d/\Gamma(1+d/2)$. $\square$

It is seen from the proof that the inequality becomes an equality when A is a shift of the ball B_r. Thus, the ball has the smallest surface area among sets of fixed Lebesgue measure.

Remark 3.1.2. Let H_d be the d-dimensional Lobachevskii space. We consider the Poincaré model $H_d = \mathbb{R}^{d-1}\times\mathbb{R}^+$ with metric $g = y_d^{-2}\,dy_1^2\cdots dy_d^2$. Using the optimal transportation (in the standard 'Euclidean' sense), one can prove the isoperimetric inequality

$$\nu^+(\partial A) \geqslant \max\biggl[\frac{\nu^{1-1/d}(A)}{\kappa_d}, (d-1)\nu(A)\biggr],$$

where $\nu = y_d^{-d} I_{\{y_d>0\}}\,dy_1\cdots dy_d$ is the Riemannian volume and $\nu^+ = y_d^{-d+1}\cdot\mathscr{H}^{d-1}$ is the corresponding surface measure.

We turn to the Brunn–Minkowski inequality. Relatively recently the following elegant form of this classical inequality was discovered.

Theorem 3.1.3. *Let f, g, and h be non-negative functions in $L^1(\mathbb{R}^d)$ such that $h(\lambda x + (1-\lambda)y) \geqslant f^\lambda(x)g^{1-\lambda}(y)$ for all $x, y \in \mathbb{R}^d$ and some $\lambda\in[0,1]$. Then*

$$\int_{\mathbb{R}^d} h\,dx \geqslant \biggl(\int_{\mathbb{R}^d} f\,dx\biggr)^{\lambda}\biggl(\int_{\mathbb{R}^d} g\,dx\biggr)^{1-\lambda}.$$

Proof. We may assume that $\|f\|_{L^1(\mathbb{R}^d)} = \|g\|_{L^1(\mathbb{R}^d)} = 1$. Let $\nabla\varphi_f$ and $\nabla\varphi_g$ be the optimal transportations of Lebesgue measure on $[0,1]^d$ to the measures $f\,dx$ and $g\,dx$, respectively. The change of variables formula takes the forms

$$f(\nabla\varphi_f)\det D_{\mathrm{a}}^2\varphi_f = 1, \qquad g(\nabla\varphi_g)\det D_{\mathrm{a}}^2\varphi_g = 1,$$

Let $\varphi = (1-\lambda)\varphi_f + \lambda\varphi_g$. By the change of variables formula and the known estimate for the determinant $\det((1-\lambda)M_1 + \lambda M_2) \geqslant (\det M_1)^{1-\lambda}(\det M_2)^\lambda$, which is true for any non-negative symmetric matrices M_1 and M_2, we have

$$\begin{aligned}\int_{\mathbb{R}^d} h\,dx &= \int_{[0,1]^d} h(\nabla\varphi)\det D_{\mathrm{a}}^2\varphi\,dx\\ &\geqslant \int_{[0,1]^d} h((1-\lambda)\varphi_f + \lambda\varphi_g)(\det D_{\mathrm{a}}^2\varphi_f)^{1-\lambda}(\det D_{\mathrm{a}}^2\varphi_g)^{\lambda}\,dx\\ &\geqslant \int_{[0,1]^d} f^{(1-\lambda)}(\nabla\varphi_f)g^{\lambda}(\nabla\varphi_g)(\det D_{\mathrm{a}}^2\varphi_f)^{1-\lambda}(\det D_{\mathrm{a}}^2\varphi_g)^{\lambda}\,dx = 1. \qquad \Box\end{aligned}$$

Applying this inequality to the indicator functions of sets A and B, we obtain the following form of the Brunn–Minkowski inequality:

$$\mathscr{H}^d((1-\lambda)A + \lambda B) \geqslant \bigl[\mathscr{H}^d(A)\bigr]^{\lambda}[\mathscr{H}^d(B)]^{1-\lambda}.$$

For an appropriate choice of λ this implies the classical Brunn–Minkowski inequality

$$[\mathscr{H}^d(A+B)]^{1/d} \geqslant [\mathscr{H}^d(A)]^{1/d} + [\mathscr{H}^d(B)]^{1/d}.$$

3.2. Sobolev inequalities and their generalizations

3.2.1. The classical Sobolev inequalities. It is well known that the isoperimetric inequality implies the Sobolev inequality. This was observed long ago independently in papers of V. G. Maz'ya and also of H. Federer and W. Flemming. For a linear combination $f=\sum_{i=1}^n \lambda_i I_{A_i}$ of indicator functions of Borel sets of bounded perimeter with disjoint boundaries the isoperimetric inequality implies that $\|f\|_{L^{d/(d-1)}}\leqslant\sum_{i=1}^n \lambda_i\|I_{A_i}\|_{L^{d/(d-1)}}\leqslant\kappa_d\sum_{i=1}^n \lambda_i\mathscr{H}^{d-1}(\partial A_i)=\kappa_d\|\nabla f\|_{L^1}$. This inequality can be extended to an arbitrary function $f\in W^{1,1}(\mathbb{R}^d)$ using approximation by simple functions in the norm of the space BV of functions of bounded variation. It implies the Sobolev embeddings $W^{p,1}\subset L^{dp/(d-p)}$. In [95] a direct transport proof was found for the classical Sobolev inequalities. The proof automatically yields formulae for the minimizing functions (the functions for which these inequalities become exact equalities). In addition, interesting duality relations have been obtained.

Theorem 3.2.1. *Let $f\in W^{p,1}(\mathbb{R}^d)$ and let $p^*=dp/(d-p)$. Then*

$$\|\nabla f\|_{L^p}\geqslant\|f\|_{L^{p^*}}\|\nabla h_p\|_{L^p},\qquad \textit{where}\quad h_p(x)=(\sigma_p+|x|^{p/(p-1)})^{1-d/p},$$

and σ_p satisfies the normalization condition $\|h_p\|_{L^{p^*}}=1$. Equality holds only for functions of the form $f(x)=Ch_p(x-x_0)$.

The proof can be found in [95] and [16]. In [96] the transport method was used to prove the following inequality for traces on half-spaces:

$$\|f\|_{L^{p^+}(\partial H)}\leqslant T_p(d)\|\nabla f\|_{L^p(H)}.$$

In [97] optimal inequalities were obtained for traces of Sobolev functions, in particular, inequalities of the form

$$\|f\|_{L^{p^*}(\Omega)}\leqslant S_p(d)\|\nabla f\|_{L^p(\Omega)}+C(p,\Omega)\|f\|_{L^{p^+}(\partial\Omega)},$$

where f is a Sobolev function on a domain $\Omega\subset\mathbb{R}^d$ with Lipschitz boundary, and $p^+=(d-1)p/(d-p)$. The maximal value of the constant $C(p,\Omega)$ is attained for balls. Generalizations of a number of classical inequalities (Gagliardo–Nirenberg, Faber–Krahn, Moser–Trudinger, the Euclidean logarithmic Sobolev inequality) to the case of traces were obtained in [97]. In [98] Gromov's method was used to find some strengthenings of the classical versions of the isoperimetric inequality and the Brunn–Minkowski inequality. It should be noted that transport considerations do not always work well in the manifold case. Nevertheless, in [99]–[101] there are interesting generalizations of the Brunn–Minkowski inequality as well as the isoperimetric and Sobolev inequalities for manifolds, obtained with the aid of the transport method.

There is an interesting connection between the Kantorovich distance and the dual Sobolev norm $H^{-1}(\mu)$ with respect to a measure μ, defined by

$$\|u\|_{H^{-1}(\mu)}=\sup\biggl\{\int u\varphi\,d\mu;\ \varphi\in C_0^\infty(\mathbb{R}^n);\ \int_{\mathbb{R}^n}|\nabla\varphi|^2\,d\mu\leqslant 1\biggr\}.$$

Let u be a bounded Borel function with zero integral with respect to μ, and let $\mu_\varepsilon = \mu + \varepsilon u\,dx$. Then

$$\|u\|_{H^{-1}(\mu)} = \liminf_{\varepsilon\to 0} \varepsilon^{-1} W_2(\mu,\mu_\varepsilon).$$

For interesting applications of this fact to convex geometry, see [81].

3.2.2. Logarithmic Sobolev inequalities. There are analogues of Sobolev inequalities for probability measures. The best known of them is the logarithmic Sobolev inequality proved by Gross in [102] for the standard Gaussian measure on $\mathbb{R}^d$ with density $(2\pi)^{-d/2}\mathrm{e}^{-|x|^2/2}$. Namely, for any smooth function f the following inequality is satisfied:

$$\operatorname{Ent}_\gamma f^2 := \int_{\mathbb{R}^d} f^2 \log f^2\,d\gamma - \int_{\mathbb{R}^d} f^2\,d\gamma\ \log \int_{\mathbb{R}^d} f^2\,d\gamma \leqslant 2\int_{\mathbb{R}^d} |\nabla f|^2\,d\gamma. \qquad (3.2.1)$$

The logarithmic Sobolev inequality is a natural analogue of the classical Sobolev inequalities. Along with the closely related hypercontractivity property of diffusion semigroups, it has many applications in mathematical physics, geometry, and probability theory. There is an extensive literature on this subject (see [103], [16], [17], [104]). A complete characterization of measures satisfying the logarithmic Sobolev inequality on the real line is given in [105]. In Perelman's paper [106] one of the key steps in the paper was proved with the aid of the logarithmic Sobolev inequality: monotonicity of the so-called $\mathscr{W}$-functional. At first sight the inequality (3.2.1) seems to be considerably weaker than the classical Sobolev inequalities, since the left-hand side contains a much more slowly growing function than in the analogous Sobolev inequality. However, its important advantage is that (3.2.1) does not depend on the dimension, and this enables one to apply it in the infinite-dimensional case.

A transport proof of (3.2.1) was first obtained in [79]. Before turning to it we note that this inequality does not change upon multiplication of f by a constant. Hence, it suffices to prove (3.2.1) for a function f with unit norm in $L^2(\gamma)$. Let us consider the optimal transportation $T=\nabla\varphi$ of the measure $f^2\cdot\gamma$ to the measure γ. We shall assume that the function f is sufficiently smooth, bounded, and bounded away from zero, which implies that T is smooth. By the change of variables formula, $f^2(x)\mathrm{e}^{-|x|^2/2} = \mathrm{e}^{-|\nabla\varphi(x)|^2/2}\det D^2\varphi(x)$, that is, $\log f^2(x) - |x|^2/2 = -|\nabla\varphi(x)|^2/2 + \log\det D^2\varphi(x)$. We rewrite the equality obtained in the form

$$\log f^2(x) = -\frac{|x-\nabla\varphi(x)|^2}{2} - \langle x, \nabla\varphi(x) - x\rangle + \log\det D^2\varphi(x)$$

and integrate it with respect to the measure $f^2\cdot\gamma$, obtaining

$$\begin{aligned}\int_{\mathbb{R}^d} f^2\log f^2\,d\gamma = &-\int_{\mathbb{R}^d} \frac{|x-\nabla\varphi(x)|^2}{2} f^2(x)\,\gamma(dx) \\ &-\int_{\mathbb{R}^d} \langle x,\nabla\varphi(x)-x\rangle f^2(x)\,\gamma(dx) + \int_{\mathbb{R}^d} \log\det D^2\varphi f^2\,d\gamma.\end{aligned}$$

Here we integrate by parts:

$$-\int_{\mathbb{R}^d} \langle x, \nabla\varphi(x) - x\rangle f^2(x)\,\gamma(dx)$$
$$= 2\int_{\mathbb{R}^d} \langle \nabla f(x), \nabla\varphi(x) - x\rangle f(x)\,\gamma(dx) - \int_{\mathbb{R}^d} (\Delta\varphi - d) f^2\,d\gamma.$$

For any symmetric $d \times d$ matrix $A \geqslant 0$ we have $\operatorname{Tr} A - d - \log\det A \geqslant 0$ (this inequality is proved by reducing the matrix to diagonal form and using the inequality $x\log x \geqslant x - 1$). Thus,

$$\int_{\mathbb{R}^d} f^2 \log f^2\,d\gamma \leqslant \int_{\mathbb{R}^d} \biggl[-\frac{|x - \nabla\varphi(x)|^2}{2} f^2(x) + 2\langle \nabla f(x), \nabla\varphi(x) - x\rangle f(x)\biggr]\,\gamma(dx).$$

The inequality (3.2.1) now follows from the obvious estimate

$$-\frac{|x - \nabla\varphi(x)|^2}{2} f^2(x) + 2\langle \nabla f(x), \nabla\varphi(x) - x\rangle f(x) \leqslant 2|\nabla f(x)|^2.$$

As in the case of the Sobolev inequality, there is a stronger isoperimetric inequality for Gaussian measures:

$$\gamma(A^h) \geqslant \Phi(\Phi^{-1}(\gamma(A)) + h), \qquad A^h = \{x \in \mathbb{R}^d \mid \exists\, a \in A : |a - x| < h\},$$
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\,dt.$$

In differential form it is

$$\gamma^+(\partial A) := \lim_{h\to 0} \frac{\gamma(A^h) - \gamma(A)}{h} \geqslant \frac{1}{[\Phi^{-1}]'(\gamma(A))}\,,$$

where $\gamma^+(\partial A)$ is the surface measure of A for γ. It is readily seen that this inequality is qualitatively equivalent to the inequality

$$\gamma^+(\partial A) \geqslant C\widetilde{\gamma}(A)\sqrt{-\log\widetilde{\gamma}(A)}, \qquad \widetilde{\gamma}(A) = \min(\gamma(A), 1 - \gamma(A)),$$

where C is a universal constant. The isoperimetric inequality was obtained by Sudakov and Tsirelson [107] and later by Borell [108]. There are several proofs of this inequality. The proof in [107] was obtained as a corollary of the Lévy–Gromov isoperimetric inequality on the sphere. Moreover, the Gaussian measure is one of the rare examples in which the exact solution of the isoperimetric problem is known. The sets of minimal surface measure are the half-spaces $\{x\colon \langle h, x\rangle \leqslant a\}$. No transport proof of the Gaussian isoperimetric inequality is known.

Gaussian measures belong to the important class of probability measures with interesting analytic and geometric properties that consists of convex (also called logarithmically concave) measures. Their role in analysis is analogous to the role played in geometry by manifolds with positive Ricci curvature. For more details on convex measures, see the book [67]; here we only briefly describe the basic properties that are important for what follows. A convex (or logarithmically concave) measure on $\mathbb{R}^d$ is a probability measure μ such that

$$\mu(\lambda A + (1 - \lambda)B) \geqslant \mu^\lambda(A)\mu(B)^{1-\lambda}$$

for all non-empty Borel sets A and B and numbers $\lambda \in [0,1]$. According to the well-known result of Borell, any convex measure is concentrated on some affine subspace, on which it possesses a density ϱ (with respect to the corresponding Lebesgue measure) of the form $\varrho = \mathrm{e}^{-V}$, where V is a convex function (and any such measure is convex). The class of convex measures is invariant with respect to affine maps. As we shall see below, a substantial number of known results connected with optimal maps and functional inequalities involve convex measures.

There are diverse conditions sufficient for a probability measure μ to satisfy the logarithmic Sobolev inequality. A necessary condition is the Poincaré inequality.

Definition 3.2.2. A probability measure μ is said to satisfy the Poincaré inequality if there exists a $C_p \geqslant 0$ such that for all smooth functions f

$$\operatorname{Var}_\mu f := \int_{\mathbb{R}^d} \biggl(f - \int_{\mathbb{R}^d} f\,d\mu\biggr)^2 d\mu \leqslant \frac{1}{C_p}\int_{\mathbb{R}^d} |\nabla f|^2\,d\mu.$$

One can verify that the logarithmic Sobolev inequality implies the Poincaré inequality by applying (3.2.1) to $1+\varepsilon f$ and passing to the limit as $\varepsilon \to 0$.

Another (much less obvious) necessary condition is the existence of a number $\varepsilon > 0$ such that

$$\int_{\mathbb{R}^d} \exp(\varepsilon x^2)\,\mu(dx) < +\infty. \tag{3.2.2}$$

The following generalization of the Gaussian Sobolev inequality was obtained in [109], and a transport proof was proposed in [110].

Theorem 3.2.3. *Let $\mu = \mathrm{e}^{-V}dx$ be a probability measure on $\mathbb{R}^d$ with a twice continuously differentiable function V. Suppose that there is a $\lambda \geqslant 0$ such that $D^2V \geqslant -\lambda\,\mathrm{Id}$, the inequality*

$$\int_{\mathbb{R}^{2d}} \exp\biggl(\frac{\lambda+\varepsilon}{2}|y-x|^2\biggr)\,\mu(dx)\,\mu(dy) < \infty$$

holds for some $\varepsilon > 0$, and μ satisfies the Poincaré inequality. Then there exists a $C > 0$ such that

$$\operatorname{Ent}_\mu f^2 \leqslant C\int_{\mathbb{R}^d} |\nabla f|^2\,d\mu.$$

Proof. We first prove that there are $C_1, C_2 > 0$ such that

$$\operatorname{Ent}_\mu f^2 \leqslant C_1\int_{\mathbb{R}^d} |\nabla f|^2\,d\mu + C_2\int_{\mathbb{R}^d} f^2\,d\mu. \tag{3.2.3}$$

Without loss of generality we may assume that f has unit norm in $L^2(\mu)$. Let $\nabla\varphi$ be the optimal transportation of the measure $f^2\cdot\mu$ to the measure μ. As in the case of Gaussian measures, we get that $\log f^2 = V - V(\nabla\varphi) + \log\det D^2_{\mathrm{a}}\varphi$. The inequality $V(x) - V(y) - \langle\nabla V(x), x-y\rangle \leqslant \lambda|x-y|^2/2$ implies the estimate

$$\log|f(x)^2| \leqslant \langle\nabla V(x), x - \nabla\varphi(x)\rangle + \log\det D^2_{\mathrm{a}}\varphi(x) + \frac{\lambda|x-\nabla\varphi(x)|^2}{2}\,.$$

Let us integrate the estimate obtained with respect to the measure $f^2\cdot\mu = f^2\cdot \mathrm{e}^{-V}\,dx$ and perform integration by parts. This gives

$$\int_{\mathbb{R}^d} f^2\log f^2\,d\mu \leqslant \int_{\mathbb{R}^d}(d-\mathrm{Tr}\,D_{\mathrm{a}}^2\varphi+\log\det D_{\mathrm{a}}^2\varphi)f^2\,d\mu$$
$$+\frac{\lambda}{2}\int_{\mathbb{R}^d}|x-\nabla\varphi(x)|^2 f(x)^2\,\mu(dx)+2\int_{\mathbb{R}^d}\langle\nabla f(x),x-\nabla\varphi(x)\rangle f(x)\,\gamma(dx).$$

Employing the inequality for determinants and traces used above along with the Cauchy inequality, we find that

$$\int_{\mathbb{R}^d} f^2\log f^2\,d\mu \leqslant \frac{\lambda+\varepsilon/2}{2}\int_{\mathbb{R}^d}|x-\nabla\varphi(x)|^2 f(x)^2\,\mu(dx)+2C_\varepsilon\int_{\mathbb{R}^d}|\nabla f|^2\,d\gamma. \quad (3.2.4)$$

Further, the definition of optimal maps implies the inequality

$$\frac{\lambda+\varepsilon/2}{2}\int_{\mathbb{R}^d}|x-\nabla\varphi(x)|^2\,f(x)^2\,\mu(dx)\leqslant\frac{\lambda+\varepsilon/2}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2 f(x)^2\,\pi(dx,dy),$$

where π is any probability measure on $\mathbb{R}^d\times\mathbb{R}^d$ with projections on both factors equal to μ (so that the projection of $f^2\cdot\pi$ on the first factor is $f^2\cdot\mu$). Let us take the measure $\pi=\mu\otimes\mu$. The elementary inequality $ab\leqslant \mathrm{e}^a-b+b\log b$ for $a,b>0$ implies that

$$\frac{\lambda+\varepsilon/2}{2}|x-y|^2 f(x)^2\leqslant\exp\biggl(\frac{\lambda+\varepsilon}{2}|x-y|^2\biggr)+\delta f(x)^2\log|f(x)^2|+C' f(x)^2,$$

where $\delta=\delta(\lambda,\varepsilon)<1$. The hypotheses of the theorem give the estimate

$$\frac{\lambda+\varepsilon/2}{2}\int_{\mathbb{R}^d}|x-\nabla\varphi(x)|^2 f(x)^2\,\mu(dx)\leqslant\delta\,\mathrm{Ent}_\mu f^2+C_0$$

for some $C_0=C_0(\varepsilon,\lambda)$. Then (3.2.4) gives us (3.2.3). According to the well-known inequality of Rothaus [111],

$$\mathrm{Ent}_\mu f^2\leqslant\mathrm{Ent}_\mu\biggl(f-\int_{\mathbb{R}^d}f\,d\mu\biggr)^2+2\int_{\mathbb{R}^d}\biggl(f-\int_{\mathbb{R}^d}f\,d\mu\biggr)^2\,d\mu.$$

Applying (3.2.3) to $f-\int_{\mathbb{R}^d}f\,d\mu$ and using the Rothaus inequality and the Poincaré inequality, we obtain the assertion of the theorem. □

Remark 3.2.4. (i) The condition that μ satisfies the Poincaré inequality can be omitted in the formulation of the theorem, since it follows from the remaining conditions. The proof of this fact is not very difficult but is rather lengthy (see [110]).

(ii) If $D^2V\geqslant\lambda>0$, then the logarithmic Sobolev inequality is automatically satisfied with the constant λ.

Corollary 3.2.5. *If the measure μ is convex, then a necessary and sufficient condition for* (3.2.1) *to hold is the existence of an $\varepsilon>0$ such that* (3.2.2) *holds.*

Below we consider some other examples in which analogues of the logarithmic Sobolev inequality are valid (see also [112]).

3.2.3. Uniformly convex potentials. Suppose that the potential V of the measure $\mu = \mathrm{e}^{-V}$ is uniformly convex in the sense that

$$V(x+y) + V(x) - \langle \nabla V(x), y \rangle \geqslant c(\|y\|), \tag{3.2.5}$$

where $\|\cdot\|$ is some norm (not necessarily Euclidean) and c is some increasing function. We remark that this condition holds, for example, for functions of the form $V(x) = |x|^\alpha$ with $\alpha > 2$. Using the reasoning from the proof of the logarithmic Sobolev inequality and Young's inequality, one can prove that the measure μ satisfies the inequality

$$\mathrm{Ent}_\mu f^2 \leqslant \int_{\mathbb{R}^d} c^*\left(\left\|2\frac{\nabla f}{f}\right\|_*\right) f^2\, d\mu, \tag{3.2.6}$$

where $\|\cdot\|_*$ is the dual norm and c^* is the corresponding convex conjugate function:

$$c^*(t) = \sup_{s>0}\{ts - c(s)\}.$$

Inequalities of the form (3.2.6) are called modified Sobolev logarithmic inequalities. They are generalizations of the classical logarithmic Sobolev inequality to the case of non-quadratic norms. It is important to single out the following features of this result: the inequality does not depend on the dimension, and it is preserved for product measures (tensorization). The latter means the following. Let μ_i be a collection of probability measures on $\mathbb{R}^d$ such that

$$\mathrm{Ent}_{\mu_i} g \leqslant \int c_i^*\left(\left\|\frac{\nabla g}{g}\right\|_*\right) g\, d\mu_i.$$

Let $\mu = \bigotimes_{i=1}^n \mu_i$. It is known (see, for example, [103], Proposition 5.6) that the entropy has the following remarkable property:

$$\mathrm{Ent}_\mu g = \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i} g_i(x_1, \dots, x_{i-1}, \cdot\,, x_{i+1}, \dots)\, \mu(dx),$$

where g_i is the function obtained from g by fixing all coordinates except x_i. Therefore,

$$\mathrm{Ent}_\mu g \leqslant \int \sum_{i=1}^n c_i^*\left(\left\|\frac{\nabla_{x_i} g_i}{g_i}\right\|_*\right) g\, d\mu.$$

Properties of this kind are important in infinite-dimensional analysis.

In [113] the modified logarithmic Sobolev inequality is extended to general cost functions by means of transport methods. In [114] the corresponding isoperimetric inequalities are proved for probability measures and convex bodies with uniform moduli of convexity.

Some properties of optimal transportations of measures satisfying (3.2.1) were described in § 2.2.5.

3.2.4. Boundedness from below for the second derivative and integrability. The condition of uniform convexity of the potential is rather restrictive and can be weakened at the expense of making the constant dependent on the dimension.

Definition 3.2.6. Let μ be a probability measure on a metric space (X,ϱ). The isoperimetric function of the measure μ is defined by

$$\mathscr{I}_\mu(t)=\inf\{\mu^+(\partial A);\ \mu(A)=t\},$$

where $\mu^+(\partial A)=\liminf_{h\to0^+}\mu(\{x\in X\setminus A;\ \varrho(x,A)\leqslant h\})/h$.

Remark 3.2.7. With the help of transportation of measures it is proved in [110] that the Bobkov isoperimetric inequality [115] holds for convex measures μ:

$$\mu(A)\log\frac{1}{\mu(A)}+\mu(A^c)\log\frac{1}{\mu(A^c)}+\log\mu(B_r)\leqslant 2r\mu^+(\partial A),$$

where $A\subset\mathbb{R}^d$, $B_r=\{x\colon |x|\leqslant r\}$, A^c is the complement of A, and $\mu(A)=\mu(B_r)$. We can deduce from this that if the measure satisfies the condition

$$\int_{\mathbb{R}^d}\exp\bigl(\varepsilon|x|^\alpha\bigr)\,\mu(dx)<+\infty,\quad\text{where}\quad \alpha>1, \tag{3.2.7}$$

then for sets of small measure we have the isoperimetric inequality

$$C\mu(A)|\log\mu(A)|^{1-1/\alpha}\leqslant\mu^+(\partial A).$$

To this end one has to consider transportations of the measures $\mu|_A/\mu(A)$ and $\mu|_{A^c}/\mu(A^c)$ to the measure $\mu|_{B_r}/\mu(B_r)$.

By the Gaussian isoperimetric inequality, the standard Gaussian measure γ and the standard norm satisfy the estimate

$$\mathscr{I}_\gamma(t)\geqslant Ct|\log t|^{1/2},\qquad t\in\left(0,\frac12\right],$$

where C does not depend on the dimension.

It follows from Remark 3.2.7 that the isoperimetric function of a convex measure μ with the condition (3.2.7) satisfies the inequality

$$\mathscr{I}_\mu(t)\geqslant Ct|\log t|^{1-1/\alpha}\quad\text{for}\quad t\leqslant\frac12\,.$$

More precisely, $\mathscr{I}_\mu(t)\geqslant C\min(t,1-t)|\log\min(t,1-t)|^{1-1/\alpha}$ for $t\in(0,1)$, where the constant C generally depends on the dimension.

A rather general form of relation between the isoperimetric function and Sobolev-type inequalities was obtained in [116] and [110], where the authors consider inequalities of the form

$$\int_{\mathbb{R}^d}f^2F\biggl(\frac{f^2}{\|f\|^2_{L^2(\mu)}}\biggr)\,d\mu\leqslant C\int_{\mathbb{R}^d}|\nabla f|^2\,d\mu,$$

which are called F-inequalities and which, along with the modified logarithmic Sobolev inequalities, generalize the classical logarithmic Sobolev inequality. They imply properties like hypercontractivity. There are also other functional inequalities naturally generalizing the logarithmic Sobolev inequality, for example, the Beckner–Latała–Oleszkiewicz inequality and the super-Poincaré inequality, both introduced by F.-Y. Wang. The abundance of inequalities is explained by the fact that the different kinds of them are differently adapted to various desired properties fully possessed by the logarithmic Sobolev inequality.

3.3. Transport inequalities

The transport inequality for the standard Gaussian measure γ was proved by Talagrand [117]. It is as follows:

$$\frac{1}{2}W_2^2(\gamma, g\cdot\gamma) \leqslant \operatorname{Ent}_\gamma g, \tag{3.3.1}$$

where $g\cdot\gamma$ is an arbitrary probability measure absolutely continuous with respect to γ for which the right-hand side is finite. This inequality is equivalent to the estimate

$$\frac{1}{2}\int_{\mathbb{R}^d} |x-\nabla\varphi(x)|^2\,\gamma(dx) \leqslant \int_{\mathbb{R}^d} g\log g\,d\gamma,$$

where $\nabla\varphi$ is the optimal transportation of the measure γ to $g\cdot\gamma$. A corollary of (3.3.1) and the triangle inequality for W_2 is the symmetric inequality

$$W_2(f\cdot\gamma, g\cdot\gamma) \leqslant \sqrt{2\operatorname{Ent}_\gamma f} + \sqrt{2\operatorname{Ent}_\gamma g}\,, \tag{3.3.2}$$

where f and g are probability densities with respect to γ.

For the proof of (3.3.1) we consider the change of variables formula

$$-\frac{x^2}{2} = \log g(\nabla\varphi(x)) - \frac{|\nabla\varphi(x)|^2}{2} + \log\det D_{\mathrm{a}}^2\varphi(x).$$

Writing it as $|x-\nabla\varphi(x)|^2/2 = \langle x, x-\nabla\varphi(x)\rangle + \log g(\nabla\varphi(x)) + \log\det D_{\mathrm{a}}^2\varphi(x)$, integrating with respect to the measure γ, and using the now familiar reasoning, we get that

$$\frac{1}{2}\int_{\mathbb{R}^d} |x-\nabla\varphi(x)|^2\,\gamma(dx) \leqslant \int_{\mathbb{R}^d} \log g(\nabla\varphi)\,d\gamma = \int_{\mathbb{R}^d} g\log g\,d\gamma.$$

The inequality (3.3.1) is proved.

One can prove (see [118]) that if $g\in L^p(\gamma)$ with $p>1$, then there exists a number $\varepsilon=\varepsilon(p)>0$ such that

$$\int_{\mathbb{R}^d} \exp\bigl(\varepsilon|x-\nabla\varphi(x)|^2\bigr)\,\gamma(dx) \leqslant C(p)\int_{\mathbb{R}^d} g^p\,d\gamma.$$

Below we need a more general version of the Talagrand inequality.

The Fredholm–Carleman determinant of an operator A on $\mathbb{R}^n$ is defined by the formula $\det_2 A := \mathrm{e}^{\operatorname{trace}(I-A)}\det A$.

Theorem 3.3.1. *For any two probability measures of the form* $\mu = \mathrm{e}^{-V}dx$ *and* $\nu = \mathrm{e}^{-W}dx$ *on* $\mathbb{R}^d$ *and for the corresponding optimal maps* $\nabla\Phi_\mu$ *and* $\nabla\Phi_\nu$ *taking* μ *and* ν *to a measure* $m = \mathrm{e}^{-P}dx$ *satisfying the condition* $D^2P \geqslant K\cdot\mathrm{Id}$ *with some number* $K > 0$, *the inequality*

$$\mathrm{Ent}_\nu\left(\frac{\mu}{\nu}\right) \geqslant \frac{K}{2}\int |\nabla\Phi_\mu - \nabla\Phi_\nu|^2\,d\mu - \int \log\det{}_2[D^2\Phi_\nu\cdot(D^2\Phi_\mu)^{-1}]\,d\mu$$

holds if the integrals exist.

The last term on the right-hand side is non-negative. This follows from the equality

$$-\log\det{}_2[D^2\Phi_\nu\cdot(D^2\Phi_\mu)^{-1}] = -\log\det{}_2[(D^2\Phi_\mu)^{-1/2}\cdot D^2\Phi_\nu\cdot(D^2\Phi_\mu)^{-1/2}]$$

and the fact that the matrix $(D^2\Phi_\mu)^{-1/2}\cdot D^2\Phi_\nu\cdot(D^2\Phi_\mu)^{-1/2}$ is non-negative.

We give an idea of the proof. By the change of variables formula,

$$\mathrm{e}^{-V} = \det D^2\Phi_\mu\cdot \mathrm{e}^{-P(\nabla\Phi_\mu)} \quad \mu\text{-a.e.}$$

Therefore, $-V((\nabla\Phi_\mu)^{-1}) = -P + \log\det D^2\Phi_\mu\circ((\nabla\Phi_\mu)^{-1})$. Similarly, using the change of variables formula for ν, we get that

$$-W((\nabla\Phi_\mu)^{-1}) = -P(\nabla\Phi_\nu\circ(\nabla\Phi_\mu)^{-1}) + \log\det D^2\Phi_\nu\circ((\nabla\Phi_\nu)^{-1}).$$

Letting $S = (\nabla\Phi_\mu)^{-1}$ and subtracting this equality from the previous one, we find that

$$(-V+W)\circ S = P(\nabla\Phi_\nu\circ S) - P - \log\det[D^2\Phi_\nu\cdot(D^2\Phi_\mu)^{-1}]\circ S.$$

Note that

$$P\bigl(\nabla\Phi_\nu\circ S(x)\bigr) - P(x) \geqslant \frac{K}{2}|\nabla\Phi_\nu(S(x)) - x|^2 + \langle\nabla P, \nabla\Phi_\nu(S(x)) - x\rangle.$$

Therefore,

$$\begin{aligned}(-V+W)\circ S(x) &\geqslant K|\nabla\Phi_\nu(S(x)) - x|^2 \\ &\qquad + \langle\nabla P, \nabla\Phi_\nu(S(x)) - x\rangle - \log\det[D^2\Phi_\nu\cdot(D^2\Phi_\mu)^{-1}]\circ S(x).\end{aligned}$$

Let us integrate the relation obtained with respect to the measure m. By the integration by parts formula and the equality $m = \mu\circ(\nabla\Phi_\mu)^{-1}$ we get that

$$\begin{aligned}&\int\langle\nabla P(x), \nabla\Phi_\nu(S(x)) - x\rangle\,m(dx) \\ &\qquad\geqslant \int \mathrm{Tr}[D^2\Phi_\nu\circ S\cdot(D^2\Phi_\mu)^{-1}\circ S]\,dm - d = \int\mathrm{Tr}[D^2\Phi_\nu\cdot(D^2\Phi_\mu)^{-1}]\,d\mu - d.\end{aligned}$$

It is easily seen that the proof of the transport inequality for Gaussian measures is similar to the proof of the logarithmic Sobolev inequality. In a sense the transport inequality is 'dual' to the logarithmic Sobolev inequality. We shall say that a probability measure μ satisfies the transport inequality if

$$\frac{1}{2}W_2^2(\mu, g\cdot\mu) \leqslant C\,\mathrm{Ent}_\mu\, g. \tag{3.3.3}$$

The logarithmic Sobolev inequality (LSI) and the transport inequality (TI) are connected in a non-trivial way. According to [119], the LSI implies the TI. We give a short proof from [120].

Theorem 3.3.2. *If a probability measure μ satisfies the* LSI, *that is,*

$$\operatorname{Ent}_\mu f^2 \leqslant 2C \int_{\mathbb{R}^d} |\nabla f|^2 \, d\mu, \qquad f \in C_0^\infty(\mathbb{R}^d),$$

then $W_2^2(\mu, g \cdot \mu) \leqslant 2C \operatorname{Ent}_\mu g$ for any probability measure $g \cdot \mu$.

Proof. For simplicity we suppose that $C = 1$. The proof is based on the properties of the Hopf–Lax semigroup

$$v(t,x) = Q_t f(x) = \inf_{y \in \mathbb{R}^d} \left[f(y) + \frac{1}{2t}|x-y|^2 \right].$$

It is known that v satisfies the Hamilton–Jacobi equation

$$\frac{\partial v}{\partial t} + \frac{1}{2}|\nabla v|^2 = 0.$$

In addition, $v(0,x) = f(x)$. Let $g(x,\lambda) = Q_1(\lambda f)(x)$. The Hamilton–Jacobi equation implies that $g = \lambda \dfrac{\partial g}{\partial \lambda} + \dfrac{|\nabla_x g|^2}{2}$. Let

$$M(\lambda) = \int_{\mathbb{R}^d} e^{\lambda g} \, d\mu.$$

Using the LSI, we get that $\lambda M'(\lambda) \leqslant M(\lambda) \log M(\lambda)$. This implies that $M(1) \leqslant e^{M'(0)}$. We note that $M'(0) = \int_{\mathbb{R}^d} f \, d\mu$. Thus,

$$\int_{\mathbb{R}^d} \exp(Q_1 f) \, d\mu \leqslant \exp\biggl(\int_{\mathbb{R}^d} f \, d\mu \biggr). \tag{3.3.4}$$

Now let $T = \nabla\varphi$ be the optimal transportation of the measure μ to the measure $g \cdot \mu$. Then

$$W_2^2(\mu, g \cdot \mu) = \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 \, \mu(dx).$$

As we have noted above, $\varphi(x) + \varphi^*(\nabla\varphi(x)) = \langle x, \nabla\varphi(x) \rangle$ for μ-almost all x. From this we have $|x - \nabla\varphi(x)|^2/2 = |x|^2/2 - \varphi(x) + \bigl(|x|^2/2 - \varphi^*(x)\bigr) \circ \nabla\varphi(x)$. By the change of variables formula,

$$\frac{1}{2} W_2^2(\mu, g \cdot \mu) = \int_{\mathbb{R}^d} \biggl(\frac{1}{2}|x|^2 - \varphi(x) \biggr) \mu(dx) + \int_{\mathbb{R}^d} \biggl(\frac{1}{2}|x|^2 - \varphi^*(x) \biggr) g(x) \, \mu(dx).$$

Without loss of generality φ can be chosen in such a way that

$$\int_{\mathbb{R}^d} \varphi \, d\mu = \frac{1}{2} \int_{\mathbb{R}^d} |x|^2 \, \mu(dx).$$

We observe that $Q_1(|x|^2/2-\varphi^*)(x)=\varphi(x)-|x|^2/2$. Using the known inequality $ab\leqslant \mathrm{e}^a-b+b\log b$ and (3.3.4), we find that

$$\begin{aligned}\frac{1}{2}W_2^2(\mu,g\cdot\mu)&\leqslant \exp\biggl(\int_{\mathbb{R}^d}\Bigl(\varphi(x)-\frac{|x|^2}{2}\Bigr)\,\mu(dx)\biggr)+\int_{\mathbb{R}^d}g\log g\,d\mu-\int_{\mathbb{R}^d}g\,d\mu\\&=\int_{\mathbb{R}^d}g\log g\,d\mu.\end{aligned}$$

The theorem is proved.

The converse assertion (the transport inequality implies the Sobolev inequality) is false in general; this was an open problem for some time. For some classes of measures (for example, convex) these inequalities are equivalent. The first counterexample was constructed in [121]. An example of a measure $\mathrm{e}^{-V}dx$ on the real line satisfying the transport inequality but not the logarithmic Sobolev inequality is given by the potential $V(x)=|x|^3+3x^2\sin^2 x+|x|^\beta$, $2<\beta<5/2$. The proof is based on certain criteria for both inequalities for measures on the real line. Gozlan [122], [123] gave a nice description of a broad class of measures satisfying the TI. Necessary and sufficient conditions for the logarithmic Sobolev inequality are known only in the one-dimensional case (see [105]).

An important property of the Gaussian transport inequality (3.3.1) (and also of the Gaussian logarithmic Sobolev inequality) is that it is independent of the dimension. A natural generalization to general convex cost functions c is the inequality

$$K_c(\mu,f\cdot\mu)\leqslant C\,\mathrm{Ent}_\mu f. \tag{3.3.5}$$

Inequalities of this type follow, for example, from the modified Sobolev inequalities (see [103], [120], [124]).

One can show that for any uniformly convex potential V, that is, for

$$V(x+y)+V(x)-\langle\nabla V(x),y\rangle\geqslant c(\|y\|),$$

the measure $\mu=\mathrm{e}^{-V}dx$ and the function $c(x,y)=c(\|x-y\|)$ satisfy (3.3.5).

In [125] a necessary and sufficient condition is found for the following analogue of the Talagrand inequality for the cost function $|x-y|$:

$$W_1(\mu,g\cdot\mu)\leqslant\sqrt{C\,\mathrm{Ent}_\mu g}\,.$$

It turns out that here it suffices to have an $\varepsilon>0$ such that

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathrm{e}^{\varepsilon|x-y|^2}\,\mu(dx)\,\mu(dy)<\infty.$$

The transport inequality is close in form to another classical inequality called the Pinsker–Kullbach–Csiszar inequality. Let μ and $\nu=f\cdot\mu$ be probability measures with $f>0$. Then

$$\frac{1}{2}\|\mu-\nu\|^2:=\frac{1}{2}\biggl(\int|f-1|\,d\mu\biggr)^2\leqslant\mathrm{Ent}_\mu f,$$

where $\|\mu-\nu\|$ is the variation distance between μ and ν. It turns out that this inequality can be regarded as the transport inequality for a suitable cost function (discrete). For more details, see [126].

Below we discuss the result of Gozlan on the equivalence of the transport inequality and the dimension-free concentration inequality together with connections with the theory of large deviations. A Sobolev-type inequality equivalent to the concentration inequality was recently obtained in [127]. With the aid of this result it was proved that, like the logarithmic Sobolev inequality, the transport inequality is preserved by bounded perturbations of the potential. It was also shown there that for a broad class of cost functions c the inequality

$$K_c(\mu,\nu) \leqslant C\,\mathrm{Ent}_\mu\,\nu = C\int \log\frac{\mu}{\nu}\,d\mu$$

is equivalent to the inequality

$$\mathrm{Ent}_\mu(\mathrm{e}^f) \leqslant \frac{f-Q_\lambda f}{1-\lambda C}\mathrm{e}^f\,d\mu, \qquad \lambda\in\left(0,\frac{1}{C}\right),$$

where $Q_\lambda f=\inf_y(f(y)+\lambda c(x-y))$.

On connections between transport inequalities and other functional inequalities, the theory of large deviations, information inequalities, and Lyapunov functions, see [128]–[130] and the survey [131]. Dynamical transport inequalities are considered in [132].

3.4. Concentration inequalities and large deviations

The following important observation was made by K. Marton. Let the measure μ satisfy (3.3.3), let $A\subset\mathbb{R}^d$, and let $\mu(A)\geqslant 1/2$. Consider the optimal map $\nabla\varphi$ of the measure $\mu_1=\dfrac{1}{\mu(A)}I_A$ to $\mu_2=\dfrac{1}{\mu((A^h)^c)}I_{(A^h)^c}$. Then $W_2(\mu_1,\mu_2)\geqslant h$. Moreover,

$$W_2(\mu_1,\mu_2)\leqslant W_2(\mu_1,\mu)+W_2(\mu_2,\mu)\leqslant\sqrt{2C\,\mathrm{Ent}_\mu\,\mu_1}+\sqrt{2C\,\mathrm{Ent}_\mu\,\mu_2}\,.$$

Since $\mathrm{Ent}_\mu\,\mu_1=\log\dfrac{1}{\mu(A)}$ and $\mathrm{Ent}_\mu\,\mu_2=\log\dfrac{1}{\mu((A^h)^c)}$, the inequality obtained implies that there exist $a,b>0$ such that

$$\mu(A^h)\geqslant 1-a\mathrm{e}^{-bh^2}. \tag{3.4.1}$$

This reasoning extends to more general cost functions. One can verify that (3.4.1) is equivalent to the following inequality for 1-Lipschitz functions: $\mu(x\colon f(x)-m_f\geqslant h)\leqslant a\mathrm{e}^{-bh^2}$ for $h>0$, where m_f is the median of f.

The inequality (3.4.1) is called the concentration inequality. Concentration inequalities arose in the asymptotic theory of convex bodies as a tool for investigating properties of convex bodies that are independent of the dimension. This direction has been especially actively developing in the works of V. Milman and his school (for interesting applications in probability theory, see [103]).

The following fact follows easily from the Chebyshev inequality: if a measure μ satisfies (3.4.1), then for some $\varepsilon > 0$ we have

$$\int_{\mathbb{R}^d} \mathrm{e}^{\varepsilon|x|^2}\,\mu(dx) < +\infty.$$

In particular, we obtain a necessary condition for the LSI and the TI.

In [133], [129], and [123] connections were investigated between optimal maps and the theory of large deviations. We explain this by the example of a nice result from [123].

Theorem 3.4.1. *A probability measure μ on $\mathbb{R}^d$ satisfies the transport inequality if and only if it has the dimension-free Gaussian concentration property, that is, there exist constants a, b, and r_0 such that for every power μ^n of μ one has the inequality*

$$\mu^n(A^r) \geqslant 1 - b\mathrm{e}^{-a(r-r_0)^2}, \quad \textit{where } r \geqslant r_0, \quad A \subset (\mathbb{R}^d)^n, \quad \mu^n(A) \geqslant \frac{1}{2}.$$

In one direction the proof is trivial and follows from the property that the transport inequality is preserved under products (tensorization) and the fact that the transport inequality implies concentration. Suppose now that the concentration property is satisfied. For every n consider the empirical average $L_n = n^{-1}\sum_{i=1}^n \delta_{x_i}$, where $\{x_i\}$ are independent random variables with distribution μ. It is known (Varadarajan's theorem) that the measures L_n converge weakly to μ. In addition, it is known (see § 1.1) that convergence in the Kantorovich metric W_p is equivalent to the weak convergence of measures with finite moments of order p and convergence of these moments. By Sanov's theorem, the quantity $\mathsf{P}(L_n \in \widetilde{A})$ for a set $\widetilde{A}$ in the space of measures behaves like $\mathrm{e}^{-nH(\widetilde{A}|\mu)}$ as n grows, where $H(\widetilde{A}|\mu) = \inf\{\mathrm{Ent}_\mu(d\nu/d\mu),\ \nu \in \widetilde{A}\}$. More precisely, using the techniques of the theory of large deviations, one can show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathsf{P}(W_p(L_n,\mu) > t) \geqslant -\inf\biggl\{\mathrm{Ent}_\mu\biggl(\frac{d\nu}{d\mu}\biggr) : W_p(\nu,\mu) > t\biggr\}$$

(see details in [123]). For any fixed $x \in (\mathbb{R}^d)^n$ let $L_n^x = n^{-1}\sum_{i=1}^n \delta_{x_i}$. By Theorem 1.2.11 we have

$$|W_2(L_n^x,\mu) - W_2(L_n^y,\mu)| \leqslant W_2(L_n^x, L_n^y) \leqslant \biggl(\frac{1}{n}\sum_{i=1}^n |x^i - y^i|^2\biggr)^{1/2} = \frac{1}{\sqrt{n}}|x-y|.$$

Let $A = \{x : W_2(L_n^x,\mu) \leqslant m_n\}$, where m_n is the median of $W_2(L_n,\mu)$. The estimate proved implies that $A^r \subset \{x\colon W_2(L_n^x,\mu) \leqslant m_n + r/\sqrt{n}\,\}$. For a random vector $x \in (\mathbb{R}^d)^n$ with independent components with distribution μ the concentration property (which we assume) implies that

$$\mathsf{P}\biggl(W_2(L_n^x,\mu) > m_n + \frac{r}{\sqrt{n}}\biggr) \leqslant b\mathrm{e}^{-a(r-r_0)^2}, \qquad r \geqslant r_0.$$

This is equivalent to the estimate $\mathsf{P}(W_2(L_n^x,\mu)>u)\leqslant b\exp\bigl(-a(\sqrt{n}(u-m_n)-r_0)^2\bigr)$ for $\sqrt{n}(u-m_n)\geqslant r_0$. The convergence of $W_2(L_n,\mu)$ to zero in probability implies the convergence of m_n to zero. Therefore if $u>0$, then

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathsf{P}(W_2(L_n^x,\mu)>u)\leqslant -au^2.$$

Then by the estimate obtained above we have $\inf\{\mathrm{Ent}_\mu(d\nu/d\mu):W_2(\nu,\mu)>u\}\geqslant au^2$. Thus, $aW_2^2(\mu,\nu)\leqslant\mathrm{Ent}_\mu(d\nu/d\mu)$, and the theorem is proved.

3.5. The hierarchy of inequalities

The examples considered above reveal a certain regularity. In the class of probability measures there is the following hierarchy of functional inequalities:

$$1)\ \text{isoperimetric inequalities}\ \ \mathscr{I}_\mu(t)\geqslant t\varphi(t),\ t\leqslant\frac{1}{2}$$

$$\Downarrow$$

$$2)\ \text{Sobolev-type inequalities}\ \ \mathrm{Ent}_\mu f^2\leqslant\lambda_S\int c^*\biggl(\biggl\|\frac{\nabla f}{f}\biggr\|\biggr)f^2\,d\mu$$

$$\Downarrow$$

$$3)\ \text{transport inequalities}\ \ W_c(\mu,f\cdot\mu)\leqslant\lambda_T\ \mathrm{Ent}_\mu f$$

$$\Downarrow$$

$$4)\ \text{concentration inequalities}\ \mu(A^r)\geqslant 1-\mathrm{e}^{-\lambda_C c(r)}$$

$$\Downarrow$$

$$5)\ \text{exponential integrability}\ \exists\varepsilon>0\colon\ \int\mathrm{e}^{\varepsilon c(|x|)}\,\mu(dx)<\infty.$$

In some situations the converse implications are also valid. For example, we have already seen that for convex measures the logarithmic Sobolev inequality and the corresponding isoperimetric inequality follow from the condition of exponential integrability 5) for $c(x)=|x|^2$. It turns out that in the convex case all these inequalities are equivalent. The most general result of this type was obtained in [134] (see also [135]). Moreover, the constants in these inequalities do not depend on the dimension.

Theorem 3.5.1. *Assume the following conditions for the convex measure* $\mu=\mathrm{e}^{-V}dx$:

(i) μ *satisfies the concentration inequality*

$$1-\mu(A^r)\leqslant\mathrm{e}^{-\mathscr{K}(r)},\quad\textit{where}\quad\mu(A)\geqslant\frac{1}{2},\quad\mathscr{K}\ \textit{is some function};$$

(ii) *the estimate* $\mathscr{K}(r)\geqslant\alpha(r)$ *holds for some increasing function* $\alpha(r)\colon\mathbb{R}^+\to\mathbb{R}^+$ *with* $\lim_{t\to\infty}\alpha(t)=\infty$.

Let $\widetilde{\mathscr{I}_\mu}(t)=\min\bigl(\mathscr{I}_\mu(t),\mathscr{I}_\mu(1-t)\bigr)$. *Then the isoperimetric inequality* $\widetilde{\mathscr{I}_\mu}(t)\geqslant\min\bigl(ct\gamma(\log(1/t)),c_\alpha\bigr)$ *holds with* $\gamma(t)=t/\alpha^{-1}(t)$, *where c is a universal constant and c_α depends only on α.*

This theorem is valid also for a manifold with a measure whose Bakry–Emery tensor is non-negative (§ 3.7.4).

3.6. Geodesics in the space of measures and gradient flows

3.6.1. Geodesics in the Kantorovich metric. Let us consider an optimal transportation T of a probability measure μ_0 to μ_1 on a metric space (X,ϱ) with the cost function $h=\varrho^p(x,y)$, $p>1$ (for brevity we call such maps p-optimal maps). The space of Radon probability measures with a finite p-moment, equipped with the metric W_p, will be denoted by $\mathscr{P}^p_r(X)$.

As we shall see, in physical terms one can imagine a mass transportation in $\mathbb{R}^d$ as a movement of particles with a constant speed $T(x)-x$ along straight lines:

$$T_t\colon x\mapsto (1-t)x+tT(x),\qquad x\in\mathbb{R}^d.$$

Thus, $T_0(x)=x$ and $T_1=T$. It should be said that already Monge himself regarded his transportations as processes in time and space, in particular, he considered the trajectories of mass transportations. So far this aspect has not been present in our discussions at all, since all transformations of measures are associated with an instantaneous action without any 'trajectories of displacement'; however, it is an aspect that can also be important. In spite of its simplicity, the family of interpolating maps $\{T_t\}_{t\in[0,1]}$ possesses very deep properties. We introduce an interpolating family of measures $\mu_t=\mu_0\circ T_t^{-1}$. It is fruitful and important for applications (first and foremost, to partial differential equations) to regard the family $\{\mu_t\}$ as a *geodesic* in the space of measures equipped with the Kantorovich metric. A closely related idea was advanced by V. I. Arnold in his papers on hydromechanics, where the Euler equations were represented as geodesic equations on the (infinite-dimensional) group of diffeomorphisms equipped with a certain Riemannian metric.

We have defined an interpolating family of measures using the notion of an optimal map (which may fail to exist). In the general case an interpolating family can be defined by means of a linear interpolation of an optimal plan and also by $\mu_t=\Pi\circ e_t^{-1}$, where Π is a dynamical optimal plan (see Definition 3.6.6). An interpolating curve is generally not unique. In an arbitrary metric space the length of a parametric curve $\gamma(t)$, $t\in[0,1]$, is defined as

$$L(\gamma)=\int_0^1|\dot\gamma(t)|\,dt,\qquad \dot\gamma(t)=\limsup_{\varepsilon\to 0}\frac{\varrho(\gamma(t+\varepsilon),\gamma(t))}{\varepsilon}\,.$$

Another (equivalent for absolutely continuous curves) definition is

$$L(\gamma)=\sup_n\sup_{0=t_0<t_1<\dots<t_n=1}\sum_{i=0}^{n-1}\varrho(\gamma(t_i),\gamma(t_{i+1})).$$

In §§ 3.6.1–3.6.3 we shall be concerned with the special class (LS) of metric spaces (called Aleksandrov–Busemann spaces or length spaces) in which the following condition is assumed: $\varrho(x,y)=\inf_{\Gamma^{x,y}}L(\gamma)$, where $\Gamma^{x,y}$ is the space of continuous curves with $\gamma(0)=x$ and $\gamma(1)=y$. The class of (LS)-spaces contains all smooth Riemannian manifolds and normed spaces. This class of spaces has proved to be very natural in problems where geodesics appear. In (LS)-spaces interpolation of optimal maps can be defined via shifts along geodesics.

Definition 3.6.1. A curve $\gamma\colon [0,1] \to X$ in an (LS)-space is called a geodesic of constant speed if

$$\varrho(\gamma(t), \gamma(s)) = |t - s| \varrho(\gamma(0), \gamma(1)).$$

It is easy to see that any geodesic of constant speed satisfies the equality $L(\gamma) = \varrho(\gamma(0), \gamma(1))$ and is the shortest curve joining γ_0 and γ_1. It turns out that the interpolating family $\{\mu_t\}$ is a geodesic of constant speed in the space $(\mathscr{P}_r^p(X), W^p)$.

Theorem 3.6.2. *Let μ_0 and μ_1 be probability measures on $X = \mathbb{R}^d$, let T be a p-optimal map of μ_0 to μ_1, and let $\mu_t = \mu_0 \circ T_t^{-1}$, with $T_t(x) = (1-t)x + tT(x)$. Then the family of measures $\{\mu_t\}$ satisfies the equality*

$$W_p(\mu_t, \mu_s) = (t - s) W_p(\mu_0, \mu_1).$$

Proof. The estimate $W_p(\mu_t, \mu_s) \leqslant (t - s) W_p(\mu_0, \mu_1)$ follows from the optimality of T and the inequality (valid for any measure μ and any maps r_1 and r_2)

$$W_p^p(\mu \circ r_1, \mu \circ r_2) \leqslant \int |r_1 - r_2|^p \, d\mu.$$

For the proof of the latter we note that $\mu \circ (r_1, r_2)^{-1}$ is a measure on $X \times X$ with marginals $\mu \circ r_1^{-1}$ and $\mu \circ r_2^{-1}$. Therefore,

$$W_p^p(\mu \circ r_1, \mu \circ r_2) \leqslant \int_{X^2} |x_1 - x_2|^p \, d\bigl(\mu \circ (r_1, r_2)^{-1}\bigr) = \int_X |r_1 - r_2|^p \, d\mu.$$

Now let $W_p(\mu_t, \mu_s) < (t - s) W_p(\mu_0, \mu_1)$ for some pair of points s and t. Since $W_p(\mu_0, \mu_s) \leqslant s W_p(\mu_0, \mu_1)$ and $W_p(\mu_t, \mu_1) \leqslant (1-t) W_p(\mu_0, \mu_1)$, the triangle inequality leads to a contradiction. $\square$

Remark 3.6.3. If X is an (LS)-space, then the space $\mathscr{P}_r^p(X)$ of probability measures is also an (LS)-space (see [136] and [11]).

Remark 3.6.4. The converse is also true: any geodesic of constant speed corresponds to some optimal plan π for the pair of measures (see Theorem 7.22 in [11] and Theorem 3.6.7 below).

On this subject see also [137], [138], and [139]. A new class of distances is considered in [140].

3.6.2. The Benamou–Brenier formula. The Benamou–Brenier formula interprets the metric $\mathscr{W}_p(\mathbb{R}^d)$ as the Riemannian length on the (infinite-dimensional) manifold of probability measures $\mathscr{P}_p(\mathbb{R}^d)$. Let $v_t\colon \mathbb{R}^d \to \mathbb{R}^d$, $t \in [0,1]$, be a family of smooth vector fields (a velocity field) and let $T_t\colon \mathbb{R}^d \to \mathbb{R}^d$ be the family of transformations generated by it according to the equation $dT_t/dt = v_t(T_t)$, $T_0(x) = x$. Let us consider the family of measures given by

$$\mu_t = \mu_0 \circ T_t^{-1} = \varrho_t \, dx, \qquad \mu_0 = \varrho_0 \, dx.$$

It is readily verified that μ_t satisfies the classical continuity equation (the transport equation)

$$\frac{\partial \varrho_t}{\partial t} + \operatorname{div}(\varrho_t \cdot v_t) = 0. \tag{3.6.1}$$

The following formula was obtained in papers of Benamou and Brenier (see [141]) for $p=2$:

$$W_p^p(\mu_0,\mu_1)=\inf_v \int_0^1 \int |v_t|^p \, d\mu_t \, dt,$$

where $\mu_t=\varrho_t \, dx$ satisfies (3.6.1) and the infimum is taken over the vector fields in some suitable class. The expression obtained resembles the formula for the length of a geodesic in a Riemannian manifold. We note that the inequality

$$W_p^p(\mu_0,\mu_1)\leqslant \int_0^1 \int |v_t|^p \, d\mu_t \, dt$$

follows from the relations

$$\begin{aligned} W_p^p(\mu_0,\mu_1) &\leqslant \int |T_t(x)-x|^p \varrho_0 \, dx \leqslant \int \int_0^1 \left|\frac{d}{dt}T_t(x)\right|^p dt \, \varrho_0 \, dx \\ &= \int \int_0^1 v_t^p(T_t) \, dt \, d\mu_0 = \int_0^1 \int v_t^p \, d\mu_t \, dt. \end{aligned}$$

Equality is formally attained for $T_t(x)=(1-t)x+tT_0$ (the case of geodesics of constant speed), $v_t=(T_0-x)\circ T_t^{-1}$. A rigorous justification (and formulation) can be found in [16] (Theorem 8.1) and in [11] (§ 8.3).

3.6.3. A dynamical mass transport plan. The theorem on geodesics has a broad generalization to the case of abstract 'Lagrangian actions'. This is explained in more detail in [17] (Chap. 7), but below we give a short formulation. A Lagrangian action is a family of lower semicontinuous functionals $\mathscr{A}^{s,t}\colon C\to\mathbb{R}$ on some class C of continuous curves $\gamma\colon [0,1]\to X$ and a family of cost functions $c^{s,t}\colon X\times X\to\mathbb{R}$ satisfying the conditions

1) $\mathscr{A}^{t_1,t_2}+\mathscr{A}^{t_2,t_3}=\mathscr{A}^{t_1,t_3}$, $0\leqslant t_1<t_2<t_3\leqslant 1$;
2) $c^{s,t}(x,y)=\inf\bigl\{\mathscr{A}^{s,t}(\gamma),\ \gamma\in C,\ \gamma(s)=x, \gamma(t)=y\bigr\}$ for all $x,y\in X$;
3) for all $\gamma\in C$

$$\mathscr{A}^{s,t}(\gamma)=\sup_{N\in\mathbb{N}} \sup_{0=t_0\leqslant t_1\leqslant t_N=1} \sum_{i=0}^{N-1} c^{t_i,t_{i+1}}(\gamma_{t_i},\gamma_{t_{i+1}}).$$

The class C of curves and the topology on it depend on the concrete problem. A typical example is the functional

$$\mathscr{A}^{s,t}(\gamma)=\int_s^t L(\gamma_\tau,\dot\gamma_\tau,\tau)\, d\tau, \quad \text{where } L(x,v,t) \text{ is some Lagrange function.}$$

Example 3.6.5. Let X be an (LS)-space and let c be a strictly convex function. We set

$$\mathscr{A}(\gamma)=\int_0^1 c(|\dot\gamma_t|)\, dt.$$

Jensen's inequality implies the estimate

$$c(d(\gamma_0,\gamma_1))\leqslant \int_0^1 c(|\dot\gamma_t|)\, dt.$$

Therefore, the minimum of $\mathscr{A}$ is attained on a geodesic of constant speed and equals $c(d(x,y))$.

It turns out that the theorem on geodesics has a broad generalization to the case of coercive actions. A Lagrangian action is said to be coercive if

1) $\inf_{s<t}\inf_\gamma \mathscr{A}^{s,t} > -\infty$,
2) for all $t<s$ and any two compact sets K_s and K_t the set of their minimizing curves $\Gamma^{s,t}_{K_s\to K_t}$ starting in K_s at the moment s and ending in K_t at the moment t is compact and non-empty.

Definition 3.6.6. Let Γ be the set of all curves minimizing the action $\mathscr{A}$. A dynamical optimal plan is a probability measure Π on Γ such that the measure $\Pi\circ(e_0,e_1)^{-1}$, where $e_t\colon\Gamma\to X$ is given by $e_t(\gamma)=\gamma(t)$, is an optimal plan for the pair μ_0, μ_1.

The following result ([17], Theorem 7.19) is a broad generalization of Theorem 3.6.2. As shown in [17], Theorem 3.6.2 follows from this result in the case of a Polish locally compact (LS)-space.

Theorem 3.6.7. *Let X be a Polish space and let $\mathscr{A}^{0,1}$ be a coercive Lagrangian action. Suppose that the cost functions $c^{s,t}$ are continuous. Denote by $C^{s,t}(\mu,\nu)$ the cost of the optimal transportation for the pair μ, ν. Then the following properties are equivalent for the curve $\{\mu_t\}_{0\leqslant t\leqslant 1}$:*

(i) *for $0\leqslant t\leqslant 1$ the measure μ_t equals the distribution of $\gamma(t)$, where $t\mapsto\gamma(t)$ is a random curve minimizing the action $\mathscr{A}$ and (γ_0,γ_1) is an optimal plan for the pair μ_0, μ_1;*

(ii) *for $t_1<t_2<t_3$,*

$$C^{t_1,t_2}(\mu_{t_1},\mu_{t_2})+C^{t_2,t_3}(\mu_{t_2},\mu_{t_3})=C^{t_1,t_3}(\mu_{t_1},\mu_{t_3});$$

(iii) *the curve $\{\mu_t\}$ in the space of measures minimizes the action of the coercive functional*

$$\mathscr{A}^{s,t}(\gamma)=\sup_{N\in\mathbb{N}}\ \sup_{0=t_0\leqslant t_1\leqslant t_N=1}\ \sum_{i=0}^{N-1}C^{t_i,t_{i+1}}(\mu_{t_i},\mu_{t_{i+1}})=\inf_\gamma \mathsf{E}\mathscr{A}^{s,t}(\gamma)$$

on the space $\mathscr{P}(X)$ of measures. The last infimum is taken in the space of random curves $\gamma\colon[s,t]\to X$ for which the distribution of γ_τ coincides with μ_τ, $s\leqslant\tau\leqslant t$.

Corollary 3.6.8. *Let X be a Polish (LS)-space and let Π be a dynamical optimal mass transportation plan for the cost function $d^p(x,y)$, $p>1$. Then the curve of measures $\mu_t=\Pi\circ e_t^{-1}$ is a geodesic of constant speed between μ_0 and μ_1. Moreover, each geodesic of constant speed $\{\mu_t\}$ has the form $\mu_t=\Pi\circ e_t^{-1}$ for some dynamical optimal mass transportation plan.*

3.6.4. Convex functionals. Functionals convex along geodesics on the space of measures were considered by McCann [68] in connection with the problems of uniqueness of solutions to certain variational problems. Convexity of this kind is called 'displacement convexity'. These techniques have found broad applications in both variational problems and probability problems (see Example 3.7.14). A function f on $\mathbb{R}^n$ is said to be λ-convex if the function $f(x)-\lambda|x|^2/2$ is convex.

Definition 3.6.9. A functional $\mathscr{F}\colon \mathscr{P}_r^p(X) \to (-\infty, \infty]$ is said to be λ-convex if for any pair of measures $\mu_0, \mu_1 \in W_p(X)$ the function $t \mapsto \mathscr{F}(\mu_t)$ is λ-convex, where $\{\mu_t\}$ is an arbitrary interpolating family of measures in $\mathscr{P}_r^p(X)$.

We give examples of convex functionals on the space $\mathscr{P}_r^p(X)$.

Example 3.6.10. A generalized entropy functional: for a given function F let

$$\mathscr{F}(\mu) = \int F(\varrho)\,dx$$

if $\mu = \varrho\,dx$ and $\mathscr{F}(\mu) = \infty$ otherwise. If the function $t \mapsto t^d F(t^{-d})$ is convex and decreasing, then $\mathscr{F}$ is a convex functional on $\mathscr{P}_r^p(\mathbb{R}^d)$.

The idea of the proof is to use the change of variables formula. Omitting regularity issues, we assume that the change of variables formula $\varrho_t(T_t) = \varrho_0 / \det DT_t$ holds for the interpolating map T_t. Then

$$\begin{aligned} \mathscr{F}(\mu_t) &= \int F(\varrho_t)\,d = \int F(\varrho_t(T_t)) \det DT_t\,dx \\ &= \int F\biggl(\frac{\varrho_0}{\det DT_t}\biggr) \det DT_t\,dx. \end{aligned} \tag{3.6.2}$$

We now observe that the matrix $(1-t)I+tDT_t$ is diagonalized and has non-negative values on the diagonal. It is not difficult to obtain from this that the function

$$t \mapsto \det{}^{1/d}\bigl((1-t)I + tDT_t\bigr)$$

is concave on $[0,1]$. Then by the conditions imposed on F the function $t \mapsto F(\varrho_0/\det DT_t)\det DT_t$ turns out to be convex. Therefore,

$$F\biggl(\frac{\varrho_0}{\det DT_t}\biggr) \det DT_t \leqslant (1-t)F(\varrho_0) + tF(\varrho_1).$$

Integrating the inequality obtained, we arrive at the required result.

The following functions satisfy the hypotheses in Example 3.6.10:

$$F(t) = t\log t, \quad F(t) = \frac{1}{m-1}s^m, \qquad m \geqslant 1 - \frac{1}{d}\,.$$

Example 3.6.11 (potential energy). For a function V satisfying the condition $V(x) \geqslant -a - b|x|^p$, take the functional

$$\mathscr{V}(\mu) = \int V\,d\mu.$$

If V is a λ-convex function, then $\mathscr{V}$ is a λ-convex functional on $\mathscr{P}^2(\mathbb{R}^d)$. This follows from the relations

$$\begin{aligned}\mathscr{V}(\mu_t) &= \int V((1-t)x + tT_t(x))\,\mu_0(dx)\\ &\leqslant \int \Bigl((1-t)V(x) + tV(T_t(x)) - \frac{\lambda}{2}t(1-t)|x-T_t(x)|^2\Bigr)\,\mu_0(dx)\\ &\leqslant (1-t)\int V\,d\mu_0 + t\int V\,d\mu_1 - \frac{\lambda}{2}t(1-t)W_2^2(\mu_0,\mu_1)\\ &= (1-t)\mathscr{V}(\mu_0) + t\mathscr{V}(\mu_1) - \frac{\lambda}{2}t(1-t)W_2^2(\mu_0,\mu_1).\end{aligned}$$

This result is valid for $p \leqslant 2$, $\lambda \geqslant 0$ and for $p \geqslant 2$, $\lambda \leqslant 0$ (see [11], § 9.3).

Corollary 3.6.12. *Let $\nu = \mathrm{e}^{-V}dx$ be a convex measure. Then the relative entropy $\mathrm{Ent}_\nu(\mu)$ is a convex functional on $\mathscr{P}^p_r(\mathbb{R}^d)$.*

Proof. This follows from the representation of the relative entropy $\mathrm{Ent}_\nu(\varrho\,dx)$ as the sum of two convex functionals: the integral of $\varrho\log\varrho$ with respect to Lebesgue measure and the integral of V with respect to the measure μ. □

Example 3.6.13 (interaction energy). Any lower semicontinuous function $W(x_1, \dots, x_k)\colon (\mathbb{R}^d)^k \to \mathbb{R}$ defines the following functional on $\mathscr{P}^p_r(\mathbb{R}^d)$:

$$\mathscr{W}(\mu) = \int_{(\mathbb{R}^d)^k} W\,d\mu^k.$$

If the function W is convex, then $\mathscr{W}$ is a convex functional on $\mathscr{P}^p_r(\mathbb{R}^d)$. This is seen from the fact that $\mathscr{W}$ is the restriction of the potential interaction functional

$$\widetilde{\mu} \mapsto \int W\,d\widetilde{\mu}, \qquad \widetilde{\mu} \in \mathscr{P}^p_r((\mathbb{R}^d)^k),$$

to the subset of product measures.

The following result is proved in [68].

Theorem 3.6.14. *For any strictly convex functions V and W the functional*

$$F(\mu) = \int U(\varrho)\,dx + \int V\,d\mu + \frac{1}{2}\int_{(\mathbb{R}^d)^2} W(x-y)\,\mu(dx)\,\mu(dy)$$

possesses a unique minimum point.

The 'displacement convexity' of sets is defined naturally. It was recently shown in [142] that this property does not hold for the set of measures in $\mathscr{P}^2(\mathbb{R}^d)$ vanishing on all sets of non-differentiability of convex functions on $\mathbb{R}^d$.

3.6.5. Additional properties of the space $\mathscr{P}^p$. We mention some properties of the space $\mathscr{P}^p_r(X)$ important for what follows, with references to some papers where they are discussed.

1) A formula for differentiating W_2 (see [11], Theorem 8.4.7):

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t,\sigma)=\int_{X^2}\langle v_t(x_2),x_1-x_2\rangle\,\gamma_t(dx_1\,dx_2),$$

where σ is a fixed measure, μ_t is an absolutely continuous curve in $\mathscr{P}_2(X)$,

$$\frac{\partial}{\partial t}\mu_t+\operatorname{div}(v\cdot\mu_t)=0,$$

and γ_t is an optimal transport plan between μ_t and σ. The functional W_2^2 is not convex on $\mathscr{P}^2(\mathbb{R}^d)$.

2) $\mathscr{P}_r^p(X)$ is an (LS)-space if X is (Theorem 3.6.7).

3) 'The parallelogram inequality' (see [11]) for $\mathscr{P}^2(\mathbb{R}^n)$:

$$W_2(\mu_t^{1\to 2},\mu^3)\geqslant(1-t)W_2^2(\mu^1,\mu^3)+tW_2^2(\mu^2,\mu^3)-t(1-t)W_2^2(\mu^1,\mu^2),$$

where $\mu_t^{1\to 2}$ is the interpolation of the measures μ_1 and μ_2. This inequality implies that $\mathscr{P}^2(\mathbb{R}^d)$ is a space of non-negative curvature. More generally, the space $\mathscr{P}_r^p(M)$, where M is a smooth manifold, has non-negative Aleksandrov curvature if (and only if) M has non-negative sectional curvature (see [143], [136]).

4) In [144] some geometric quantities (the connection, the Riemannian tensor) were calculated on the space $\mathscr{P}_r^p$.

3.6.6. Subdifferentials. Subdifferentials for convex functions $\Phi\colon\mathscr{P}^2(X)\to\mathbb{R}$ are studied in the book [11]. Below we present the basic concepts and results. Let $\mathscr{D}(\Phi)$ denote the set of points x where $\Phi(x)<\infty$. For $v\in\mathscr{D}(\Phi)$ the absolute value of the metric gradient is defined by

$$|\partial\Phi|(v)=\limsup_{\omega\to v}\frac{(\Phi(\omega)-\Phi(v))_+}{W_2(\omega,v)}\,.$$

If a measure μ is regular (for example, has a density), then for every measure ν there exists an optimal transportation T of μ to ν. The following definition of a subdifferential for a regular measure agrees well with the classical notion of a subdifferential for functions on a Hilbert space.

Definition 3.6.15. Let $\mu\in\mathscr{D}(|\partial\Phi|)$. A vector field $\xi\in L^2(\mu;\mathbb{R}^d)$ belongs to the subdifferential $\partial\Phi$ if

$$\Phi(\nu)-\Phi(\mu)\geqslant\int_{\mathbb{R}^d}\langle\xi(x),T(x)-x\rangle\,\mu(dx)+o(W_2(\mu,\nu))$$

for the optimal transportation T of μ to ν.

A minimal selection of the subdifferential $\partial\Phi$ is a vector $\xi\in\partial\Phi$ with minimal norm $\|\cdot\|_{L^2(\mu;\mathbb{R}^d)}$. The concept of a subdifferential can be extended to the case of a non-regular measure if in place of optimal transportations we use optimal transport plans (see [11], § 10.3).

Example 3.6.16. For the formal calculation of subdifferentials we employ (3.6.2), assuming that all the calculated objects are smooth:

$$\mathscr{F}(\mu_t) = \int F\biggl(\frac{\varrho_0}{\det((1-t)I + tD^2\varphi)}\biggr) \det((1-t)I + tD^2\varphi)\, dx.$$

Taking into account that $\dfrac{d}{dt}\det\bigl((1-t)I + tD^2\varphi\bigr)\big|_{t=0} = \Delta\varphi - d$, we easily get that

$$\frac{d}{dt}\mathscr{F}(\mu_t)\bigg|_{t=0} = \int \bigl(F(\varrho_0) - \varrho_0 F'(\varrho_0)\bigr)(\Delta\varphi - d)\, dx.$$

Let $L_F(\varrho_0) = \varrho_0 F'(\varrho_0) - F(\varrho_0)$. Integrating by parts, we find that

$$\frac{d}{dt}\mathscr{F}(\mu_t)\bigg|_{t=0} = \int \biggl\langle \frac{\nabla L_F(\varrho(x))}{\varrho(x)}\, , \nabla\varphi(x) - x \biggr\rangle \varrho(x)\, dx.$$

In the same way one also obtains a formula for subdifferentials of the potential energy and interaction energy functionals:

$$\frac{d}{dt}\mathscr{V}(\mu)\bigg|_{t=0} = \int \langle \nabla V(x), \nabla\varphi(x) - x\rangle \varrho_0(x)\, dx,$$
$$\frac{d}{dt}\mathscr{W}(\mu)\bigg|_{t=0} = \int \langle (\nabla W * \mu(x)), \nabla\varphi(x) - x\rangle \varrho_0(x)\, dx.$$

A rigorous justification of these formulae is rather involved technically. The following result is proved in [11], Theorem 10.4.13.

Theorem 3.6.17. *Consider the functional*

$$\Phi(\mu) = \int F(\varrho)\, dx + \int V\, d\mu + \frac{1}{2}\int W\, d\mu \times d\mu, \qquad \mu = \varrho\, dx,$$

on $\mathscr{P}_r^p(\mathbb{R}^d)$, where $\Phi(\mu) = \infty$ if μ has no density. Suppose that the function $F\colon [0,\infty) \to \mathbb{R}$ is convex, differentiable, and superlinear, the function $t \mapsto t^d F(t^{-d})$ is convex and decreasing, there is an $\alpha > d/(d+p)$ such that

$$F(0) = 0, \qquad \liminf_{s\to 0} F(s)s^{-\alpha} > -\infty,$$

and also the doubling condition is satisfied: there exists a number $C > 0$ such that

$$F(a+b) \leqslant C(1 + F(a) + F(b)).$$

In addition, let $V\colon \mathbb{R}^d \to (-\infty, +\infty)$ be a lower semicontinuous λ-convex function, let $\mathscr{D}(V)$ have a non-empty interior Ω, and let $W\colon \mathbb{R}^d \to [0,+\infty)$ be a convex differentiable even function satisfying the doubling condition. Under these conditions the measure $\mu = \varrho\, dx$ belongs to $\mathscr{D}(|\partial\varphi|)$ precisely when $L_F(\varrho) \in W^{1,1}_{\mathrm{loc}}(\Omega)$, $L_F = sF'(s) - F(s)$, and

$$\varrho\omega = \nabla L_F(\varrho) + \varrho\nabla V + \varrho(\nabla W) * \varrho$$

for some $\omega \in L^q(\mu;\mathbb{R}^d)$. In this case the vector ω is the minimal selection in $\partial\Phi(\mu)$.

3.6.7. Gradient flows. Otto's calculus. A gradient flow in $\mathbb{R}^d$ is a solution $X(t)$ of the equation $dX(t)/dt = -\nabla E(X(t))$, where E is some potential. We note that $dE(X(t))/dt = -|\nabla E(X(t))|^2$. The first part of the book [11] is devoted to the concept of a gradient flow on a metric space. A gradient flow can be defined in some sense as a 'curve of minimal descent'. A gradient flow on a metric space X satisfies the equation

$$-\frac{d}{dt}E(X(t)) = \frac{1}{2}|X'(t)|^2 + \frac{1}{2}|\partial E(X(t))|^2,$$

where $|X'(t)|$ is the metric derivative of the curve $X(t)$. It turns out that many equations of mathematical physics admit descriptions in the form of gradient flows of some functionals on the space $\mathscr{P}^2(X)$.

Example 3.6.18. Consider the equation $\partial\varrho/\partial t - \operatorname{div}\bigl(\nabla L_F(\varrho)\bigr) = 0$ with respect to the function $\varrho\colon (t,x) \mapsto \varrho_t(x)$. For $F(s) = s\log s$ we obtain the heat equation $\partial\varrho/\partial t = \Delta\varrho$. If $v = -\nabla L_F(\varrho)/\varrho$, then we obtain the transport equation $\partial\varrho/\partial t + \operatorname{div}(\varrho\cdot v) = 0$. It is natural to interpret v as the velocity of the curve $t \mapsto \varrho_t(\cdot)$. By Theorem 3.6.17 the function $\nabla L_F(\varrho)/\varrho$ serves as the (sub)differential of the functional $\mathscr{F}$. Therefore, the original equation is a gradient flow and (formally) $v = -\nabla\mathscr{F}$ or $v \in -\partial\mathscr{F}$.

Gradient flows connected with functionals on the space $\mathscr{P}^2(X)$ have an interesting geometric interpretation proposed by Otto [145] (see also [146]). He introduced a formal Riemannian structure on $\mathscr{P}^2(X)$ such that W_2 becomes a Riemannian metric on the 'manifold' of measures. We remark, however, that in many cases the Otto calculus so far remains a heuristic approach requiring a rigorous justification.

We recall that by the Benamou–Brenier formula,

$$W_2(\mu_0,\mu_1) = \inf \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \varrho_t(x)\,dx\,dt,$$

where the infimum is taken over all ϱ_t and v_t with $\varrho_t\big|_{t=0} = \varrho_0$ and $\varrho_t\big|_{t=1} = \varrho_1$ satisfying the transport equation $\partial\varrho_t/\partial t + \operatorname{div}(\varrho_t v_t) = 0$. The minimum is attained on the curve $\{\varrho_t\}$ satisfying the equations

$$\varrho_t(x) = \varrho_t\bigl((1-t)x + t\nabla\varphi(x)\bigr)\det\bigl((1-t)I + tD^2\varphi(x)\bigr), \qquad \frac{\partial\varrho_t}{\partial t} + \operatorname{div}(\varrho_t v_t) = 0,$$

where $v_t(x) = \nabla\varphi(S_t(x)) - S_t(x)$, $S_t(x) = ((1-t)I + t\nabla\varphi)^{-1}(x)$, and $\nabla\varphi$ is the optimal transportation of μ_0 to μ_1. Otto introduced the following differential structure on $\mathscr{P}^2(X)$, which we consider for $X = \mathbb{R}^d$. The tangent space $T\mathscr{P}^2(\mathbb{R}^d)$ at the point $\varrho\,dx$ consists of all measures with densities of the form $-\operatorname{div}(\varrho v)$, where v is a vector field. We define the inner product (a Riemannian metric) on $T\mathscr{P}^2(\mathbb{R}^d)$ as follows:

$$\int_X \langle v_1, v_2\rangle \varrho\,dx.$$

Let us consider the entropy functional $\operatorname{Ent}\colon \varrho\,dx \mapsto \int_{\mathbb{R}^d}\varrho\log\varrho\,dx$ from this point of view. Differentiating it along a geodesic $t \mapsto \mu_0\circ T_t^{-1}$, we get that

$$\frac{\partial\operatorname{Ent}\varrho_t}{\partial t}\bigg|_{t=0} = -\int_X \varrho(\Delta\varphi - d)\,dx = \int_X \langle\nabla\varrho, \nabla\psi\rangle\,dx,$$

where $\psi(x) = \varphi(x) - |x|^2/2$. We note that $\partial\varrho_t/\partial t\big|_{t=0} + \operatorname{div}(\nabla\psi\cdot\varrho_0) = 0$. Therefore, the 'gradient' of the functional Ent satisfies the equalities

$$\left\langle \nabla \operatorname{Ent}(\varrho), \frac{\partial \varrho_t}{\partial t} \right\rangle\bigg|_{t=0} := \frac{\partial \operatorname{Ent} \varrho_t}{\partial t}\bigg|_{t=0} = \int_X \langle \nabla \varrho, \nabla\psi\rangle\, dx = \int_X \left\langle \frac{\nabla\varrho}{\varrho}, \nabla\psi \right\rangle \varrho\, dx,$$

and $\nabla \operatorname{Ent}(\varrho)$ is identified with the function $-\operatorname{div}\left(\varrho\dfrac{\nabla\varrho}{\varrho}\right) = -\Delta\varrho$. The corresponding *gradient flow* is given by the heat equation $\dfrac{\partial\varrho}{\partial t} = \Delta\varrho$. Let us fix a probability measure $\nu = e^{-V}\,dx$. We set $\mu = f\cdot\nu = \varrho\, dx$. Considering in the same manner the relative entropy

$$\operatorname{Ent}_\nu \mu = \int_X f \log f\, d\nu = \int_X \varrho \log \varrho\, dx + \int_X V\varrho\, dx,$$

we get that the corresponding gradient flow satisfies the equation

$$\frac{\partial\varrho}{\partial t} = \Delta\varrho + \operatorname{div}(\varrho\cdot\nabla V).$$

Below we give some other examples of gradient flows.

Example 3.6.19. The porous media equation:

$$E(\mu) = \frac{1}{m-1}\int \varrho^m\, dx, \qquad \frac{\partial\varrho}{\partial t} = \Delta\varrho^m.$$

The non-linear diffusion equation: $\dfrac{\partial\varrho}{\partial t} - \Delta L_F(\varrho) = 0$.

The McKean–Vlasov equation:

$$E(\mu) = \frac{1}{2}\int W(x-y)\varrho(x)\varrho(y)\, dx\, dy, \qquad \frac{\partial\varrho}{\partial t} = \operatorname{div}(\varrho\nabla(\varrho * \nabla W)).$$

The quantum drift-diffusion equation (see [147]):

$$E(\mu) = \int \frac{|\nabla\varrho|^2}{\varrho}\, dx, \qquad \frac{\partial\varrho}{\partial t} + 4\operatorname{div}\left(\varrho\nabla\frac{\Delta\sqrt{\varrho}}{\sqrt{\varrho}}\right) = 0.$$

The Kantorovich metric (under the name 'the Vasershtein metric') was applied to such equations already by Dobrushin [148], and among more recent papers we mention [149], [132], [150]. Estimates of solutions of the heat equation connected with the metric W_p can be found in [151].

3.6.8. Applications to equations of mathematical physics. A rigorous justification of the heuristic manipulations above can be given in different ways. One possible way is a formalization of the notion of a Riemannian structure and the Otto differential calculus. The situation is greatly complicated by the fact that the objects considered are not smooth. Below we present some results from the book [11], where the main emphasis is on work with curves in the space of measures and a formalization of the notion of a subdifferential. It turns out that in many cases (including the λ-convex case) to show the existence/uniqueness/regularity of the gradient flow it suffices to consider the inclusion $v \in -\partial\mathscr{F}$ instead of the equality $v = -\nabla\mathscr{F}$.

Let X be a Hilbert space.

Theorem 3.6.20 ([11], Theorem 8.3.1). *Any absolutely continuous curve $\{\mu_t\}$ has a tangent vector v_t: $\partial\mu_t/\partial_t + \operatorname{div}(v_t \cdot \mu_t) = 0$, where the equality is understood in the weak sense.*

Definition 3.6.21. A curve $\{\mu_t\} \in AC^p_{\text{loc}}(\mathbb{R}^+; \mathscr{P}^p_r(X))$ is called a gradient flow for the functional Φ if $j_p(v_t) \in -\partial\Phi(\mu_t)$ for $t > 0$, where $j_p(v) = |v|^{p-2}v$ is the dual map.

The next theorem is of considerable importance since it enables one to apply the 'metric' theory of gradient flows to gradient flows in $\mathscr{P}^p_r(X)$.

Theorem 3.6.22 [11]. *For λ-convex functionals the notion of a gradient flow in $\mathscr{P}^p_r(X)$ coincides with the notion of a gradient flow in the metric space $\mathscr{P}^p_r(X)$.*

Approximations by minimizing motions are used to prove the existence of a gradient flow. This is based on the following simple observation. Suppose for simplicity that ψ is a lower semicontinuous functional on a Hilbert space and v minimizes the functional

$$\omega \mapsto \psi(\omega) + \frac{1}{2t}|\omega - a|^2.$$

Then $-(v-a)/t \in \partial\psi(v)$. An analogue of this for $\mathscr{P}^2$ is the following fact.

Lemma 3.6.23. *Let ν minimize the functional $\nu \mapsto \Phi(\nu) + W_2^2(\nu,\mu)/(2t)$ and let T be the optimal transportation of ν to μ. Then $(T-x)/\tau \in \partial\Phi(\nu)$.*

Let us now split the real line into intervals of length $\tau > 0$. Let $M^0_\tau = \mu_0$ and let M^{n+1}_τ be the minimum point of the functional

$$\mu \mapsto \Phi(\mu) + \frac{1}{2\tau}W_2^2(\mu, M^{n-1}_\tau).$$

If M^{n+1}_τ is a sufficiently nice measure, then for the optimal map T^n_τ of M^{n+1}_τ to M^n_τ the field $-(x - T^n_\tau(x))/\tau$ belongs to $\partial\Phi(M^{n+1}_\tau)$. Intuitively this means that $(x - T^n_\tau(x))/\tau$ gives an approximation of the velocity v of the limiting (as $\tau \to 0$) curve $\{\mu_t\}$ to which the discrete broken lines $\mu^\tau_t : t \mapsto M^\tau_n$ must converge, where $n\tau \leqslant t < (n+1)\tau$. The velocity vector v must satisfy the inclusion $v \in -\partial\Phi$. The scheme described does work in the case of λ-convex functionals. The following theorem is a corollary of Theorem 11.2.1 in [11], where among other things important estimates are proved for the approximations themselves, and diverse variational formulations of the problem are given.

Theorem 3.6.24. *If X is a Hilbert space and Φ is a λ-convex functional, then as $\tau \to 0$ the discrete broken lines $\{\mu^\tau_t\}$ converge locally uniformly in the metric W_2 to the unique gradient flow $\{\mu_t\}$ of the functional Φ.*

Uniqueness of the gradient flow $\{\mu_t\}$ is proved in the following way. Since $-v_t \in \partial\Phi(\mu_t)$, by the definition of a differential

$$\Phi(\sigma) \geqslant \Phi(\mu_t) + \int_{X^2} \langle v_t(x_2), x_1 - x_2\rangle\, \gamma_t(dx_1\, dx_2) + \frac{\lambda}{2} W_2^2(\mu_t, \sigma),$$

where γ_t is an optimal plan for the pair μ_t, σ. Using the equality

$$\frac{1}{2}\,\frac{d}{dt} W_2^2(\mu_t, \sigma) = \int_{X^2} \langle v_t(x_2), x_1 - x_2\rangle\, \gamma_t(dx_1\, dx_2),$$

we get that

$$\frac{1}{2}\,\frac{d}{dt} W_2^2(\mu_t, \sigma) + \frac{\lambda}{2} W_2^2(\mu_t, \sigma) \leqslant \Phi(\sigma) - \Phi(\mu_t).$$

In particular, for two different gradient flows μ_t^1 and μ_t^2,

$$\frac{d}{dt} W_2^2(\mu_t^1, \mu_t^2) \leqslant -\lambda W_2^2(\mu_t^1, \mu_t^2).$$

Therefore, $W_2^2(\mu_t^1, \mu_t^2) \leqslant \mathrm{e}^{-\lambda t} W_2^2(\mu_0^1, \mu_0^2)$, whence uniqueness follows.

3.6.9. The Riemannian volume on $\mathscr{P}^2$. In the paper [152] of von Renesse and Sturm the concept of the Riemannian volume on $\mathscr{P}^2(X)$ was introduced, where $X = [0, 1]$. Heuristically, the measure

$$P_\beta = \mathrm{e}^{-\beta\, \mathrm{Ent}_\nu \mu} \cdot P_0$$

was defined, where ν is some fixed measure and P_0 is the 'uniform distribution' on $\mathscr{P}^2(X)$. We employ the natural isomorphism between measures and monotone functions on $[0, 1]$. For every probability measure μ on $[0, 1]$ we find a non-decreasing function ξ_μ mapping Lebesgue measure on $[0, 1]$ to μ. Then

$$\mathrm{Ent}_\lambda \mu = \int_0^1 \varrho \log \varrho\, dx = \int_0^1 \log \varrho(\xi_\mu(t))\, dt = -\int_0^1 \log \xi_\mu'(t)\, dt.$$

By using this, one can construct a measure Q_β on the space of monotone functions. This measure Q_β is determined by its finite-dimensional projections

$$\begin{aligned} &Q_\beta(\xi_1 \in dx_1, \dots, \xi_N \in dx_N) \\ &\quad = \frac{1}{Z_{\beta,N}} \exp\biggl(\beta \sum_{i=1}^{N+1} \log \frac{x_i - x_{i-1}}{t_i - t_{i-1}} \cdot (t_i - t_{i-1})\biggr) q_N(x_1, \dots, x_N). \end{aligned} \tag{3.6.3}$$

Here $x_0 = 0, x_{N+1} = 1$, and $q_N(x_1, \dots, x_N)$ is a distribution on $\Sigma_N = \{(x_1, \dots, x_N) \in [0, 1]^N : 0 < x_1 < \dots < x_N < 1\}$. Under the requirement of certain invariance properties and the consistency of the conditional distributions, the measure q_N is uniquely determined by the formula

$$q_N(x_1, \dots, x_N) = C^N \frac{dx_1 \cdots dx_N}{x_1(x_2 - x_1) \cdots (x_N - x_{N-1})(1 - x_N)}\,.$$

We note that q_N defines a σ-finite measure. Nevertheless, for a suitably chosen constant the equality (3.6.3) defines a probability measure. The existence of Q_β follows from Kolmogorov's theorem. Taking into account the correspondence between measures and functions, we obtain a measure P_β on $\mathscr{P}^2(X)$. This measure has the following two properties.

1) Its support consists of purely singular measures.

2) It is quasi-invariant with respect to transformations of the following form: let h be a C^2-diffeomorphism of $[0,1]$ with $h(0)=0$, and define an action of it on measures by $\mu \mapsto \mu \circ h^{-1}$; then the Radon–Nikodym density is

$$\frac{dP_\beta(\mu \circ h^{-1})}{dP_\beta(\mu)} = \exp\biggl(\beta \int_0^1 \log h'(s)\, d\mu\biggr) \frac{1}{\sqrt{h'(0)\cdot h'(1)}} \prod_{I\in \mathrm{gaps}(\mu)} \frac{\sqrt{h'(I_-)h'(I_+)}}{|h(I)|/|I|},$$

where $\mathrm{gaps}(\mu)$ denotes the set of closed intervals of maximal length $I=[I_-,I_+]$ with $\mu(I)=0$.

3.7. Geometric applications

3.7.1. Classical problems of differential geometry. The Monge–Ampère equation has been actively studied in connection with a number of problems in differential geometry. This includes the problem of the existence of a surface of given Gaussian curvature. A function $\varphi(x)$ is to be found such that the Gaussian curvature of its graph at the point x coincides with a given function $K(x)$. The problem reduces to solving the Monge–Ampère equation

$$\det D^2\varphi(x) = K(x)(1+|\nabla\varphi(x)|^2)^{(d+2)/2}.$$

The classical Minkowski problem consists in constructing a convex polyhedron having given normals $\mathrm{n}_1,\dots,\mathrm{n}_k$, and given areas of the corresponding faces $S_1,\dots,S_k$. In the continuous case we have to find a convex surface whose curvature at a point with normal n equals a given function $K(\mathrm{n})$. It turns out that a solution exists under the condition

$$\int_{S^{d-1}} \frac{\mathrm{n}}{K(\mathrm{n})}\, d\mathscr{H}^{d-1} = 0.$$

Substantial contributions to the solution of the Minkowski problem are due to Minkowski, Aleksandrov, Lewy, Calabi, Nirenberg, Pogorelov, Sh.-Yu. Cheng, and Yau (see [72]). As an example of applications of transport techniques we briefly discuss the problem of Aleksandrov and its transport solution. For an arbitrary convex surface $F\subset\mathbb{R}^d$ we consider its generalized normal map to the sphere S^{d-1}: $F\ni x\mapsto N(x)$, where $N(x)$ is the set of all normals to support planes of F at the point x. Suppose that the origin of coordinates is interior to F. Then F can be parameterized by means of a radial function: $F\ni r(x)=\varrho(x)x$, $x\in S^{d-1}$. We define a multivalued map $\alpha_F\colon S^{d-1}\to S^{d-1}$ by $\alpha_F(x)=N(r(x))$. The measure μ_F is defined by

$$\mu_F(A) = \mathscr{H}^{d-1}\biggl(\bigcup_{x\in A}\alpha_F(x)\biggr),$$

and is called the integral Gaussian curvature of F. Aleksandrov's problem (posed and solved by him [153]) is this: for a given measure μ find a convex surface F such that $\mu = \mu_F$. The following conditions are necessary and sufficient for the existence of a solution to Aleksandrov's problem (necessity is easily verified):

1) $\mu(S^{d-1}) = \mathscr{H}^{d-1}(S^{d-1})$,
2) $\mu(S^{d-1} \setminus A) \geqslant \mathscr{H}^{d-1}(A^*)$ for every spherically convex set $A \subset S^{d-1}$, where $A^* = \{y \in S^{d-1} \colon \langle x, y\rangle \leqslant 0,\ x \in A\}$.

An interesting transport solution of this problem was found in [154]. We define the support function h by the formula

$$h(y) = \sup_{x \in S^{d-1}} \varrho(x)\langle x, y\rangle, \qquad y \in S^{d-1}.$$

The radial function and the support function are connected by duality relations in terms of the Legendre transform (in this form also called the Young transform): $1/\varrho(x) = \sup\{\langle x, y\rangle / h(y) \mid y \in S^{d-1}\}$. Thus, $h(y)/\varrho(x) \geqslant \langle x, y\rangle$, or

$$\log h(y) - \log \varrho(x) \geqslant \log\langle x, y\rangle. \tag{3.7.1}$$

Exact equality is attained for $y \in N(r(x))$. The duality theorem suggests that the pair $(\log h, -\log \varrho)$ is a solution to the following optimal problem (dual to the Monge–Kantorovich problem): find the minimum of the functional

$$\int_{S^{d-1}} \log h \, d\mathscr{H}^{d-1} - \int_{S^{d-1}} \log \varrho \, d\mu$$

among functions $(\log h, -\log \varrho)$ satisfying (3.7.1), Indeed, it is proved in [154] that a solution to this problem gives a solution to Aleksandrov's problem (the required surface can be uniquely reconstructed from either of the given functions h, ϱ). The corresponding Monge–Kantorovich problem has the following form: for $c(x, y) = \log\langle x, y\rangle$ find the maximum of the functional

$$\int_{S^{d-1} \times S^{d-1}} c(x, y) \, dm(x, y) \quad \text{with} \quad m_x = \mathscr{H}^{d-1}, \quad m_y = \mu,$$

where $c(x, y) = -\infty$ if $\langle x, y\rangle < 0$. Analytically the problem reduces to a Monge–Ampère type equation on the sphere.

3.7.2. The Monge–Kantorovich problem on a Riemannian manifold. Existence and uniqueness. The Monge–Kantorovich problem generalizes naturally to Riemannian manifolds. A solution to this problem was obtained by McCann in [155]. Below we present the basic elements of his construction. On this subject, see [10], [156]–[158], [86], [88], and [72]. Let (M, g) be a smooth (of class C^3) compact connected Riemannian manifold. Let $T_x M$ denote the tangent space at the point $x \in M$. The distance $d(x, y)$ between points $x, y \in M$ is defined as

$$d(x, y) = \inf_{\gamma_{x,y}} \int \sqrt{g_{ij}(\gamma)\dot\gamma^i \dot\gamma^j} \, dt,$$

where the infimum is taken over all possible curves $\gamma_{x,y}$ joining x and y. A geodesic is a curve $t \mapsto \gamma(t)$ satisfying the condition $\nabla_{\dot\gamma}\dot\gamma = 0$, where $\dot\gamma \in TM_{\gamma(t)}$ is the velocity vector of the curve and ∇ is the covariant derivative. The exponential map $\exp_x \colon TM_x \to M$ is defined as the map $TM_x \ni v \mapsto y$, where y is the endpoint of the geodesic with starting point at x and initial tangent vector v. The Monge problem for a quadratic functional has the following natural setting. Given two probability measures μ and ν having densities ϱ_μ and ϱ_ν with respect to the Riemannian volume dx, one has to find the minimum of the functional

$$M(\mu, \nu, T) = \int_M d^2(x, T(x))\,\mu(dx), \qquad T \in T(\mu, \nu).$$

In the rest of this subsection $c(x,y) = d^2(x,y)$. As in the flat case, let us consider the operation of convex conjugation $\varphi^{c+}(y) = \inf_{x\in M}\bigl(c(x,y) - \varphi(x)\bigr)$. It is natural to expect that the optimal transportation will take the form $T(x) = \exp_x(-\nabla\varphi(x))$, where $\exp_x(v)$ is the exponential map at a point x and the potential φ satisfies the condition $\varphi^{cc} = \varphi$. We note that for $\mathbb{R}^d$ this would give $T(x) = x - \nabla\varphi(x)$, and not $T(x) = x + \nabla\varphi(x)$ as we had above. As in the Euclidean case, the corresponding Kantorovich problem on the space of measures has the dual formulation $J(\varphi, \psi) \to \sup$, where the functional J is considered on pairs of continuous functions φ, ψ satisfying the condition $\varphi(x) + \psi(y) \leqslant d^2(x,y)$. The results of the duality theory fully extend to the case of a Riemannian manifold. In particular, we obtain the existence of the potential φ. Then it is proved that $\varphi^{cc} = \varphi$ and $\psi = \varphi^c$. The main problem now is to prove that the corresponding map gives the desired mass transport. The following lemma (Rademacher's theorem on manifolds) is proved in [155].

Lemma 3.7.1. *Any Lipschitz function φ on a manifold M is differentiable at all points in M except for a set of measure zero* (*with respect to the Riemannian volume*).

Remark 3.7.2. By representing the potential φ as the result of the operation of conjugation one can prove that φ is a Lipschitz function.

The principal complication is the non-differentiability of the Riemannian metric $d(x,y)$ (for example, at conjugate points where uniqueness of a connecting geodesic is lost). It is intuitively clear (although it requires a non-trivial justification) that $d(x,y)$ is a smooth function for y in a small neighbourhood of x. In the general case (see details in [155]) the squared distance $f(y) = d^2(x,y)$ is *superdifferentiable*, that is, there exists a vector $p \in TM_x$ such that for all $v \in TM_x$

$$f(\exp_x(v)) \leqslant f(x) + g^{ij}(x)p^i v^j + o(|v|).$$

Subdifferentiability is defined similarly. Using the local differentiability of $d(x,y)$ and the triangle inequality, one can show the superdifferentiability of the function $y \mapsto d(x,y)$ and the superdifferentiability of $d^2(x,y)$. It follows from Remark 3.7.2 and a version of Rademacher's theorem that $\nabla\varphi$ exists μ-almost everywhere. Moreover, the following important fact follows from the definitions of the operation of conjugation and the superdifferentiability of the function $c(x,y)$ (see details in [155]).

Lemma 3.7.3. *The inequality $\varphi(x) + \varphi^c(y) \leqslant c(x,y)$ holds for all $x, y \in M$. If x is a point of differentiability of the function φ, then the equality*

$$\varphi(x) + \varphi^c(y) = c(x,y)$$

holds precisely when $y = \exp_x(-\nabla\varphi)$. *In this case* $|\nabla\varphi(x)| = d(x, y)$.

Proof. The inequality follows directly from the definition. Let x be a point of differentiability of φ. If $\varphi(x) + \varphi^c(y) = c(x, y)$ for some point $y \in M$, then for all $z \in M$

$$c(z, y) - \varphi(z) - \varphi^c(y) \geqslant c(x, y) - \varphi(x) - \varphi^c(y).$$

Let $z = \exp_x(v)$, $v \in TM_x$, and $f(z) = c(z, y)$. Then

$$f(z) \geqslant f(x) - \varphi(x) + \varphi(z) = f(x) - \varphi(x) + \varphi(x) + g(\nabla\varphi(x), v) + o(|v|).$$

Therefore, $\nabla\varphi(x)$ belongs to the subgradient of f at the point x. On the other hand, by classical results in differential geometry (the Hopf–Rinow theorem) there is a minimal geodesic joining y and x and having initial tangent vector v. One can deduce from this that f is superdifferentiable at x and v belongs to its superdifferential at x. It follows that f is differentiable at x and $y = \exp_x(-\nabla\varphi(x))$, $v = \nabla\varphi(x)$. The first part of the assertion is proved.

Now for the proof of the second part it suffices to show that the equality $\varphi(x) + \varphi^c(y) = c(x, y)$ is attained at least at one point. Since $\varphi = \varphi^{cc}$, Remark 3.7.2 implies that the function φ^c is Lipschitz. Hence in the equality $\varphi^{cc}(x) = \inf_{y\in M}\bigl(c(x, y) - \varphi^c(y)\bigr)$ the infimum is attained at some point (since M is compact and f is continuous). □

Theorem 3.7.4. *The image of* μ *with respect to* $T(x) = \exp_x(-\nabla\varphi(x))$ *is* ν.

Proof. Let $f \in C_b(M)$. We prove that the integral of f with respect to ν equals the integral of $f \circ T$ with respect to μ. For $\varepsilon \in (-1, 1)$ we consider the perturbations $\psi_\varepsilon(y) := \psi(y) + \varepsilon h(y)$ and $\varphi_\varepsilon(x) = (\psi_\varepsilon)^c(x) = \inf_{y\in M}\bigl(c(x, y) - \psi(y) - \varepsilon h(y)\bigr)$. Let the function φ be differentiable at the point x. Then the minimum of $c(x, y) - \psi(y)$ is attained at a unique point $t(x)$ by Lemma 3.7.3. For small ε the minimum of $c(x, y) - \psi(y) - \varepsilon h(y)$ is attained at a nearby point $y_\varepsilon = t(x) + o(1)$. Thus, for all $y \in M$

$$c(x, t(x)) - \varphi(t(x)) - \varepsilon f(t_\varepsilon(x)) \leqslant \varphi_\varepsilon(x) \leqslant c(x, y) - \varphi(y) - \varepsilon f(y).$$

Take $y = t(x)$. Then $\varphi_\varepsilon(x) = \varphi(x) - \varepsilon f(t(x)) + o(\varepsilon)$, and the estimate $o(\varepsilon) \leqslant \varepsilon \sup_{x\in M}|f(x)|$ holds. Since $\varepsilon \mapsto J(\varphi_\varepsilon, \psi_\varepsilon)$ has a maximum at $\varepsilon = 0$, we have

$$\lim_{\varepsilon\to 0}\frac{J(\varphi_\varepsilon, \psi_\varepsilon) - J(\varphi, \psi)}{\varepsilon} = \int f\,d\nu + \lim_{\varepsilon\to 0}\int\frac{\varphi_\varepsilon - \varphi}{\varepsilon}\,d\mu = \int f\,d\nu + \int f(t(x))\,d\mu = 0$$

by the dominated convergence theorem. □

The following uniqueness theorem holds [155].

Theorem 3.7.5. *The map* $T = \exp_x(\nabla\varphi)$ *minimizes the functional* $M(\mu, \nu, S)$ *among all maps* $S \in T(\mu, \nu)$, *and up to equivalence it is the unique minimizing map.*

3.7.3. The Monge–Kantorovich problem on a Riemannian manifold. Change of variables, the Ricci tensor, and geometric inequalities. As in the flat case, any optimal map T on a Riemannian manifold is almost everywhere differentiable. Here the situation is more complicated due to the fact that T is the composition of two maps, namely, the gradient of a function and the exponential map $TM_x \ni v \mapsto \exp_x(v)$, and the latter map may not be differentiable for all $v \in TM_x$ in general. For any $x \in M$ we define the set $\mathrm{cut}(x) \subset M$ (the cut locus) as the set of points y for which there exists the shortest geodesic $t \mapsto \exp(tv)$ for $t \in [0,1]$ joining x and y, but any such geodesic is not the shortest path for $t \in [0, 1+\varepsilon)$. It is well known that the exponential map $v \mapsto \exp_x(v)$ is differentiable for all v satisfying the condition $\exp(v) \notin \mathrm{cut}(x)$. Obviously, $|v| = d(x, \exp_x(v))$ for all v such that $\exp(tv)$ does not intersect $\mathrm{cut}(x)$ for $0 \leqslant t \leqslant 1$. If $x \notin \mathrm{cut}(x)$, then $y = \exp_x(-\nabla_x[d^2(y,x)]/2)$. The following change of variables formula was obtained in [99]. As in the flat case, a Hessian means a Hessian in the sense of Aleksandrov (more precisely, its generalization to manifolds).

Theorem 3.7.6. *Let $\mu = f\,dx$ and $\nu = g\,dx$ be probability measures with compact support, where dx is the Riemannian volume. Let $T(x) = \exp_x(-\nabla\varphi(x))$ be the optimal map of μ to ν for the 'quadratic' cost function $d^2(x,y)$. Then there exists a Borel set K of full measure μ such that* (i) *the function φ has a Hessian* $\mathrm{Hess}_x\,\varphi$ *at every point $x \in K$ and $T(x) \notin \mathrm{cut}(x)$,* (ii) *$f(x) = g(T(x))\det[Y(H - \mathrm{Hess}_x\,\varphi)]$ for $x \in K$, where $Y = d(\exp_x)_{-\nabla\varphi(x)}$ and $H = (1/2)\,\mathrm{Hess}\,d^2(x,y)\big|_{y=T(x)}$.*

In applications of optimal transportation, an important role is played by the interpolating family of maps T_t which in the flat case is defined by $T_t(x) = (1-t)x + tT(x)$. By analogy, for manifolds we set

$$T_t(x) = \exp_x(-t\nabla\varphi(x)).$$

This family of maps is closely connected with the important concept of Jacobi fields. We recall that a Jacobi field $J(t)$ is a vector field obtained by variation of a family of geodesics: $J(t) := \dfrac{d}{ds}\gamma_s(t)\Big|_{s=0}$, where $[0,1] \ni t \mapsto \gamma_s(t)$ is a geodesic for each s. As is known from Riemannian geometry, an arbitrary Jacobi field satisfies the equation

$$\nabla^2_{\dot\gamma(t)} J(t) + R_{\gamma(t)}(\dot\gamma(t), J(t))\dot\gamma(t) = 0, \tag{3.7.2}$$

where $R_x\colon TM_x \times TM_x \times TM_x \mapsto TM_x$ is the Riemann tensor at the point x, $\dot\gamma(t) = \dfrac{d\gamma}{dt}(t)$, and $\nabla^2_{\dot\gamma(t)}$ is the second covariant derivative along γ.

It is convenient to write differential equations along geodesics in a moving orthonormal basis. Let us fix an orthonormal basis $\{e_1(0), \ldots, e_n(0)\}$ in $T_{\gamma(0)}M$ and consider its parallel displacement along γ. Let $Y(t)$ be the coordinates of $J(t)$ in the basis $\{e_1(t), \ldots, e_n(t)\}$. The equation (3.7.2) will be written as $Y'' + RY = 0$, where $R = R(t)$ is some symmetric matrix whose trace gives the value of the Ricci tensor on the vector $\dot\gamma$. We observe that for any vector $v_0 \in TM_{x_0}$ the field $t \mapsto A(t)(x_0)v_0 := d_x(T_t)\big|_{x=x_0} v_0 = d_x(\exp_x(-t\varphi(x)))|_{x=x_0} v_0$, where $d_x(T_t)\big|_{x=x_0}$ is the differential of the map $x \mapsto T_t(x)$ at x_0, is a Jacobi field. The change of variables formula for the family of interpolating measures takes the form $\varrho_0(x) = \varrho_t(T_t x)\det A(t)$, where $\varrho_0 = f$, $\varrho_1 = g$, and A_t is the unique Jacobi tensor field determined by $A(0) = I$, $A'(0) = -\operatorname{Hess}\varphi$ along the geodesic $\exp(-t\nabla\varphi)$.

Let us see how one can employ this technique to obtain a generalization of the Brunn–Minkowski inequality to manifolds. For two points x and y we set

$$Z_t(x,y) = \{z \in M;\ d(x,z) = td(x,y),\ d(z,y) = (1-t)d(x,y)\}.$$

If these two points are joined by a unique geodesic, then $Z_t(x,y)$ is a single point.

The Ricci curvature tensor will be denoted by Ric.

Theorem 3.7.7. *Let $\mu = \mathrm{e}^{-V}dx$ be a probability measure on M. Suppose that for some $\lambda \in \mathbb{R}$ the Bakry–Emery tensor of the manifold M satisfies the inequality* $\mathrm{Ric} + \operatorname{Hess} V \geqslant \lambda$. *Let f, g, and h be non-negative functions such that*

$$h(z) \geqslant \mathrm{e}^{-\lambda s(1-s)d^2(x,y)/2} f^{1-s}(x) g^s(y)$$

for all $x, y \in M$, $s \in [0,1]$, and $z \in Z_s(x,y)$. Then

$$\int_M h\,d\mu \geqslant \biggl(\int_M f\,d\mu\biggr)^{1-s} \biggl(\int_M g\,d\mu\biggr)^{s}.$$

We explain the idea of the proof. Let ϱ_s be the density of the measure $\mu\circ T_s^{-1}$, where T_t was defined before (3.7.2). Let $\gamma(s)$ be the geodesic starting at the point x (which is omitted in the notation) with the velocity $-\nabla\varphi(x)$. As in Theorem 3.1.3, it suffices to prove the inequality $h(\gamma(s))\exp\bigl(-V(\gamma(s))\bigr) \geqslant \varrho_s(\gamma(s))$ (because the integral of the right-hand side equals 1 by the change of variables formula). Let $\psi(t) = -\log\det d(T_t))$. By the change of variables formula,

$$f(\gamma(0))\mathrm{e}^{-V(\gamma(0))} = \varrho_s(\gamma(s))\mathrm{e}^{-\psi(s)} = g(\gamma(1))\mathrm{e}^{-V(\gamma(1))-\psi(1)}.$$

Our assumption about the function h implies that

$$h(\gamma(s))\mathrm{e}^{-V(\gamma(s))} \geqslant f^{1-s}(\gamma(0))g^s(\gamma(1))\exp\biggl(-V(\gamma(s)) - \frac{1}{2}\lambda s(1-s)d^2(\gamma(0),\gamma(1))\biggr).$$

By the above equations the logarithm of the right-hand side divided by $\varrho_s(\gamma(s))$ equals

$$(1-s)V(\gamma(0)) + sV(\gamma(1)) - V(\gamma(s)) + s\psi(1) - \psi(s) - \frac{1}{2}\lambda s(1-s)d^2(\gamma(0),\gamma(1)).$$

Let $\alpha(t) := V(\gamma(t)) + \psi(t)$. Taking into account that $\psi(0) = 0$, we get that it suffices to prove the estimate $(1-s)\alpha(0) + s\alpha(1) - \alpha(s) \geqslant \lambda s(1-s)d^2(\gamma(0),\gamma(1))/2$. Note that $\alpha''(t) = \operatorname{Hess} V(\gamma(t))(\dot\gamma(t),\dot\gamma(t)) + \psi''(t)$. If we prove the estimate $\psi''(t) \geqslant \mathrm{Ric}_{\gamma(t)}(\dot\gamma(t),\dot\gamma(t))$, then by the condition on the Bakry–Emery tensor we obtain the inequality $\alpha''(t) \geqslant \lambda|\dot\gamma(t)|^2$, which implies the desired relation. The required estimate is indeed true. Its proof is based on methods developed in Riemannian geometry for proving comparison theorems. Let us pass to a moving orthogonal basis and use the fact that $A(t) = d_x(T_t)$ is a solution to the equation $A''(t) + R(t)A(t) = 0$. By the symmetry of the initial conditions the symmetry of A is preserved. Let $B = A'A^{-1}$. Differentiating the determinant, we find that $\psi' = -\operatorname{Tr} B$ and $\psi'' = -\operatorname{Tr} B'$. By differentiating $B^* - B$ and using the symmetry of R it is not difficult to show that $(B^* - B)' = 0$, and therefore B is symmetric. Hence

$$\begin{aligned}\psi'' &= -\operatorname{Tr} B' = \operatorname{Tr} B^2 + \operatorname{Tr} R \geqslant \frac{(\operatorname{Tr} B)^2}{d} + \operatorname{Tr} R \geqslant \frac{(\operatorname{Tr} B)^2}{d} + \operatorname{Tr} R \\ &= \frac{(\psi)^2}{d} + \operatorname{Ric}(\gamma(t))(\dot\gamma(t), \dot\gamma(t)),\end{aligned}$$

as required.

We also mention some results from [99]. The volume distortion coefficient is defined as follows: $v_t(x, y) = \lim_{r\to 0} Z_t(x, B_r(y))/\operatorname{vol}(B_{tr}(y))$. The volume distortion coefficient has to do with how ‘curved’ the manifold is. For a Euclidean space, $v_t = 0$. Now let

$$S_k(t) = \frac{\sin(\sqrt{k}\,t)}{\sqrt{k}\,t} \quad \text{if} \quad k > 0, \qquad S_k(t) = \frac{\sinh(\sqrt{-k}\,t)}{\sqrt{-k}\,t} \quad \text{if} \quad k < 0.$$

Theorem 3.7.8. *If* $\operatorname{Ric} \geqslant (d-1)k$ *on a* d*-dimensional manifold* M, *then*

$$v_t(x, y) \geqslant \left(\frac{S_k(td(x, y))}{S_k(d(x, y))}\right)^{d-1}.$$

Equality is attained for the model spaces of constant sectional curvature (*the sphere* S^d, $\mathbb{R}^d$, *and the Lobachevskii space* H^d).

On this subject, see also [159].

3.7.4. Spaces of non-negative curvature. Many analytic and geometric properties of manifolds can be expressed in terms of the Ricci tensor. Particularly many interesting results are known about manifolds for which the Ricci tensor is bounded from below by a constant tensor: $\operatorname{Ric} \geqslant K$ (another form of this estimate is $\operatorname{Ric} \geqslant K \cdot g$, where g is the metric tensor on M). We mention the best known results.

1. *The Bishop–Gromov comparison theorem.* If $x \in M$, then the volume of the ball about x with radius r is growing no faster than the volume of the ball in the model space of constant sectional curvature K. The model spaces are the sphere of curvature K if $K > 0$, $\mathbb{R}^d$ if $K = 0$, and the Lobachevskii space of curvature K if $K < 0$.

2. *The Bonnet–Myers compactness theorem.* If $K > 0$, then the diameter of M does not exceed $\pi\sqrt{(d-1)/K}$.

3. *The Lévy–Gromov isoperimetric inequality.* If $K>0$, then

$$\frac{\nu^+(\partial A)}{\nu(A)^{(d-1)/d}} \geqslant \frac{\mu^+(\partial B)}{\mu(B)^{(d-1)/d}},$$

where $A\subset M$, ν is the Riemannian volume on M, ν^+ is the surface measure on M, $B\subset S$ is a ball in the Riemannian metric on the model sphere S, and μ and μ^+ are analogous measures on S. It is assumed that $\nu(A)/\nu(M)=\mu(B)/\mu(S)$. The Lévy–Gromov inequality is closely connected with the Gaussian isoperimetric inequality, which can be regarded as an 'infinite-dimensional version' of the latter.

This list of interesting results in this area is far from being complete. One could also mention sharp Sobolev inequalities, Lee–Yau type inequalities, Harnack inequalities, gradient estimates and estimates of heat kernels, and various topological results for spaces of non-negative curvature. Also closely related is the Bakry–Emery theorem, which implies the logarithmic Sobolev inequality. The key analytic object in this result is the Bakry–Emery tensor $\operatorname{Hess} V+\operatorname{Ric}$. It turns out that this tensor is responsible for many analytic properties of manifolds. The idea of studying the geometry of manifolds by equipping them with probability measures and considering measure-preserving isometries was proposed by Gromov [160], who called this object a 'metric measure space'. A collection consisting of a space, a metric and a measure on it was called 'a metric triple' by Vershik [161], [162]. This topic is presented in detail in the book [17] (see also [163]–[165], [40]), so we confine ourselves to a brief exposition of the principle concepts and formulations of the main results. These results are based on the concept of convexity of suitable energy functionals on the space of measures with the Kantorovich metric. The term 'displacement convexity' is common in the literature. Remembering that the interpolation of measures by means of optimal transportations defines geodesics in the space of measures, one can also speak of the geodesic convexity. The Otto calculus is used as a formal technique for working with convex functionals. We have already discussed calculating gradients of such functionals. The next formula (see [17]) is understood heuristically (in the framework of the same Otto calculus). Let dx be the Riemannian volume, let the functional $\mathscr{F}$ be given by the formula

$$\mathscr{F}(\mu)=\int_M F(\varrho)\,d\nu, \qquad \mu=\varrho\,d\nu, \quad \nu=\mathrm{e}^{-V}dx,$$

and let $v=\nabla f$ be the vector field defining the curve $\dot\varrho+\operatorname{div}(\nabla f\cdot\varrho)=0$. Then by direct calculations one verifies the following formula.

Theorem 3.7.9. *The Hessian on the space $\mathscr{P}^2(\mathbb{R}^d)$ has the form*

$$\begin{gathered}\operatorname{Hess}_\mu\mathscr{F}(\nabla f,\nabla f)=\int\bigl[\|D^2f\|^2_{\mathscr{H}\mathscr{S}}+(\operatorname{Hess}V+\operatorname{Ric})(\nabla f,\nabla f)\bigr]p(\varrho)\,d\nu \\ +\int(Lf)^2p_2(\varrho)\,d\nu=\int\Gamma_2(f)p(\varrho)\,d\nu+\int(Lf)^2p_2(\varrho)\,d\nu, \\ Lf=\Delta f-\langle\nabla V,\nabla f\rangle, \quad p(\varrho)=\varrho F'(\varrho)-F(\varrho), \quad p_2(\varrho)=\varrho p'(\varrho)-p(\varrho).\end{gathered}$$

The expression (the iteration of the diffusion operator L)

$$\Gamma_2(f)=L\biggl(\frac{|\nabla f|^2}{2}\biggr)-\langle\nabla f,\nabla Lf\rangle$$

is called the 'carré du champ itéré' operator. The non-trivial and very useful equality

$$\Gamma_2(f) = \|D^2 f\|^2_{\mathscr{H}\mathscr{S}} + (\mathrm{Hess}\, V + \mathrm{Ric})(\nabla f, \nabla f)$$

is called the (generalized) Bochner formula. For $V = 0$ it takes the form

$$\frac{1}{2}\Delta|\nabla f|^2 = \langle \nabla f, \nabla \Delta f\rangle + \|D^2 f\|^2_{\mathscr{H}\mathscr{S}} + \mathrm{Ric}(\nabla f, \nabla f). \qquad (3.7.3)$$

The Bakry–Emery tensor is defined by $R_{\infty,\mu} = \mathrm{Ric} + \mathrm{Hess}\, V$. Let us now consider the modified (depending on the dimension) Bakry–Emery tensor

$$R_{N,\mu} = \mathrm{Ric} + \mathrm{Hess}\, V - \frac{\nabla V \oplus \nabla V}{N - d}, \qquad N > d.$$

It is straightforward to verify the inequality

$$\Gamma_2(f) \geqslant \frac{(Lf)^2}{N} + R_{N,\mu}(\nabla f, \nabla f).$$

Finally, together with the formula for a Hessian it implies the important estimate

$$\mathrm{Hess}_\mu \mathscr{F} \geqslant \int R_{N,\mu}(\nabla f, \nabla f) p(\varrho)\, d\nu + \int (Lf)^2 \Big[p_2(\varrho) + \frac{p(\varrho)}{N}\Big]\, d\nu.$$

In particular, the convexity property of the corresponding functionals can be formally deduced from this inequality.

Example 3.7.10. If $R_{\infty,\mu} \geqslant 0$, then $\mathrm{Hess}_\mu \mathscr{F}(\nabla f, \nabla f) \geqslant 0$ for $F(\varrho) = \varrho \log \varrho$. If $R_{N,\mu} \geqslant 0$, then $\mathrm{Hess}_\mu \mathscr{F}(\nabla f, \nabla f) \geqslant 0$ for $F(\varrho) = -N(\varrho^{1-1/N} - \varrho)$.

Definition 3.7.11. Let V be a twice continuously differentiable function on a smooth Riemannian manifold M with metric ϱ, and let $\nu = \mathrm{e}^{-V} dx$. We shall say that the space (M, ϱ, ν) belongs to the class $CD(K, N)$, where $K \in R$ and $N \in [n, \infty]$, if $R_{N,\nu} \geqslant K$.

The case $N = n$ makes sense for a constant V. In this case we set $R_{N,\mu} = \mathrm{Ric}$ by definition.

Example 3.7.12. The following examples of measures on $\mathbb{R}$ with the standard metric give model examples of spaces in $CD(K, N)$:

1) $M = (-\pi\sqrt{N-1}/(2K), \pi\sqrt{N-1}/(2K))$, $K > 0$, $1 < N < \infty$, $\nu = \cos^{N-1}\big(\sqrt{K/(N-1)}\, x\big)\, dx$,
2) $M = \mathbb{R}$, $K < 0$, $1 < N < \infty$, $\nu = \cosh^{N-1}\big(\sqrt{-K/(N-1)}\, x\big)\, dx$,
3) $M = (0, +\infty)$, $K = 0$, $1 < N < \infty$, $\nu = x^{N-1}\, dx$,
4) $K > 0$, $N = \infty$, $\nu = \mathrm{e}^{-Kx^2/2}\, dx$.

The principle lying at the basis of analytic theorems for metric measure spaces and comparison theorems in Riemannian geometry says: the analytic properties of the spaces in $CD(K, N)$ are no worse than those of the corresponding model spaces.

Remark 3.7.13 (the tangent formalism). The techniques for proving analytic relations based on the concept of geodesic convexity of functionals give far-reaching generalizations of many of the above results for $\mathbb{R}^d$ to the case of Riemannian manifolds. The following elementary relation is used here: if $D^2F \geqslant K$, then

$$F(tx+(1-t)y) \leqslant tF(x)+(1-t)F(y) - \frac{Kt(1-t)}{2}|x-y|^2, \qquad 0\leqslant t\leqslant 1,$$
$$F(y) \geqslant F(x) + \langle F(x), y-x\rangle - \frac{K}{2}|y-x|^2, \tag{3.7.4}$$

generalized to functionals on $\mathscr{P}^2$ in the proper way.

Example 3.7.14 (The Sobolev inequality and the tangent formalism). Let $K>0$, $N=\infty$, and let $T=I+\nabla\varphi$ be the optimal transportation of μ to ν on $\mathbb{R}^d$. Consider the standard interpolation by means of a geodesic between ν and $\mu=\varrho\,d\nu$ in the space of measures. We make use of the convexity of the entropy functional, where $F(\varrho)=\varrho\log\varrho$ and $p(\varrho)=\varrho$ (see Theorem 3.7.9). Using (3.7.4) and the formula

$$\nabla\mathscr{F}(\nabla\varphi) = \int \biggl\langle \frac{\nabla p(\varrho(x))}{\varrho(x)}, \nabla\varphi(x)\biggr\rangle \varrho(x)\,dx,$$

we have

$$\int \varrho\log\varrho\,d\nu \leqslant -\int \langle\nabla\varrho,\nabla\varphi\rangle\,d\nu - \frac{K}{2}W_2^2(\mu,\nu),$$

which we know in the flat case, and which has the logarithmic Sobolev inequality as a corollary. In the same manner the standard Sobolev inequalities on $\mathbb{R}^d$ follow from the convexity of the functional given by the integral of $-\varrho^{1-1/d}$.

In particular, in this way we obtain the Bakry–Emery theorem [166].

Theorem 3.7.15. *Let M be a Riemannian manifold with a probability measure $\mu=\mathrm{e}^{-V}dx$, where dx is the Riemannian volume. If $R_{\infty,\mu}\geqslant K$ with $K>0$, then the logarithmic Sobolev inequality* $\operatorname{Ent}_\mu f^2 \leqslant \dfrac{2}{K}\displaystyle\int_M |\nabla f|^2\,d\mu$ *holds.*

There are different characterizations of $(CD(K,N))$-spaces (see [143], [136]). We give the principal results, following [17].

Theorem 3.7.16. *A smooth manifold M with a measure $\nu=\mathrm{e}^{-V}dx$, where $V\in C^2(M)$, belongs to the class $CD(K,N)$ precisely when the functional*

$$\mathscr{F}(\mu) = \int_M F(\varrho)\,d\nu$$

is convex for $F(\varrho)=-N(\varrho^{1-1/N}-\varrho)$, or for $F(\varrho)=\varrho\log\varrho$ if $N=+\infty$.

This theorem lets us define the class of $CD(K,N)$-spaces in the weak sense as the limits of $CD(K,N)$-spaces in the topology of Gromov–Hausdorff convergence. Such a limit is a metric measure space, but may fail to be a smooth manifold. In particular, one can define spaces with Ricci curvature bounded from below as the (LS)-spaces for which the corresponding functional is convex on the space of measures with the Kantorovich metric. There are also various 'local' characteristics of $CD(K,N)$-spaces. We have already seen in Theorems 3.7.7 and 3.7.8 that in work with volumes, geometric inequalities, and the change of variables formula the distortion coefficients for shifts along geodesics arise in a natural way. Due to their variational nature these coefficients are solutions to the Jacobi equation. Its solutions can be estimated by standard methods. For the proof of the following fact, see [17].

Theorem 3.7.17. *Let $K > 0$ and $N < \infty$. Any weak $CD(K,N)$-space satisfies the inequality* $\operatorname{diam}(\operatorname{supp}\nu) \leqslant \pi\sqrt{(N-1)/K}$.

3.7.5. Geometric flows: flows of Gaussian curvature and parabolic equations. In this subsection we consider the so-called Gauss transportations of measures, which is a class of close-to-optimal maps of measures. The general form of a Gauss transportation is $T = \varphi\nabla\varphi/|\nabla\varphi|$, where φ is a function (a potential) with convex sets $A_t = \{\varphi \leqslant t\}$. For a broad class of measures the existence of such maps is established in [167]. The name 'Gauss' is due to the fact that this object is closely connected with Gauss (spherical or normal) maps of surfaces, and also with Gaussian curvature flows. Let us consider a simple example. Let $\gamma\colon S^1 \to \mathbb{R}^2$ be a diffeomorphic embedding of the circle in the plane, that is, $\gamma(S^1)$ is a smooth closed contour without self-intersections bounding a simply connected area. Let $\gamma_t\colon S^1 \to \mathbb{R}^2$ be the family of curves given by the equations

$$\gamma_0 = \gamma, \qquad \frac{d}{dt}\gamma_t(s) = -K(\gamma_t(s)) \cdot n(\gamma_t(s)),$$

where $K(\gamma_t(s))$ and $n(\gamma_t(s))$ are the curvature and the outer normal of the curve γ_t at the corresponding point. In other words, γ_t is a flow of curves starting at γ and moving with velocity K in the direction of the normal (that is, inwards if $K > 0$ and outwards if $K < 0$). In the literature this object is called a 'curve shortening flow' and is the simplest example of geometric flows, which include mean curvature flows and Gaussian curvature flows and also Ricci flows. It is readily seen that the area of the domain A_t enclosed by γ_t is changing at the constant rate

$$\frac{d}{dt}\mathscr{H}^2(A_t) = -\int_{\gamma_t} K\, d\mathscr{H}^1 = -2\pi.$$

The latter equality holds by the Gauss–Bonnet theorem. It is not difficult to prove that the length of the contour ∂A_t is changing according to the formula

$$\frac{d}{dt}\mathscr{H}^1(\partial A_t) = -\int_{\gamma_t} K^2\, d\mathscr{H}^1.$$

The formula for the variation of the area shows that as time goes on, the curves must shrink to some set of measure zero. This is indeed so. Moreover, it is known that even if the initial curve γ is non-convex (that is, the set A_0 is not convex), the curves become convex in finite time. Convexity is then further preserved, and the sets A_t are embedded one in another: $A_{t_2} \subset A_{t_1}$ if $t_2 \geqslant t_1$. In the course of time the set A_t becomes more and more 'round', which means that the curvature of γ_t is distributed more and more uniformly on the curve. Finally, the curves are shrinking to a point. Flows of manifolds constitute a natural generalization of this construction. Let $M \subset \mathbb{R}^d$ be a $d-1$-dimensional surface in $\mathbb{R}^d$. A geometric flow is a family of embeddings $M_t \subset \mathbb{R}^d$ with $M_0 = M$ that are moving in the direction of the normal n with some velocity F: $\dot{x}(s) = -F(x)\cdot n(x)$. If F is the mean curvature H, then one speaks of a mean curvature flow. If F is the Gaussian curvature K, then one speaks of a Gaussian curvature flow. Unlike in the one-dimensional case, both these and other flows can be much less regular. In the multidimensional case a discontinuity may appear in finite time and a connected manifold may split into parts. However, if the initial surface is convex, then both types of flows preserve convexity, and the surface will shrink to a point in finite time. In the case of mean curvature flows this was proved in [168] and in the case of Gaussian curvature flows it was proved in [169].

We explain a connection between curvature flows and transportations of measures. Let $\{\gamma_t\}$ be a flow of one-dimensional curves in the plane and let $\gamma = \gamma_0$ be a smooth convex curve. As we know, all other curves γ_t remain convex and the flow exists for some finite time T_0. Let φ be the function equal to $T_0 - t$ on γ_t. What is the image of Lebesgue measure under the map $T = \varphi\nabla\varphi/|\nabla\varphi|$? We observe that $T = \varphi \cdot n$, where n is the normal to γ_t. Let v be a tangent vector to γ_t, so that (n, v) is an orthonormal basis. We differentiate T in this basis. Since φ is constant on γ_t, we find that $\partial_n T = \partial_n\varphi \cdot n + \varphi \cdot \partial_n n$ and $\partial_v T = \varphi \cdot \partial_v n$. We note that $\partial_v n = Kv$ (the Frenet formula) and $\partial_n\varphi = |\nabla\varphi|$. Therefore, we get that $\det DT = K\varphi|\nabla\varphi|$. By the change of variables formula, Lebesgue measure is taken to the measure $\varrho\,dx$ under the action of T, where $\varrho(\varphi\nabla\varphi/|\nabla\varphi|)K\varphi|\nabla\varphi| = 1$. The level sets $\{x\colon \varphi(x) = T_0 - t\}$ of the function φ change with the velocity $1/|\nabla\varphi|$. Since we are dealing with the curvature flow, $K|\nabla\varphi| = 1$. Therefore, $\varrho(\varphi\nabla\varphi/|\nabla\varphi|) = 1/\varphi$, that is, the image of Lebesgue measure with respect to the map T is the measure $\nu = dx/|x|$. In the general case the following theorem is true.

Theorem 3.7.18. *Let $A \subset \mathbb{R}^d$ be a compact convex set and let $\mu = \varrho_0\,dx$ be a probability measure on A equivalent to the restriction of Lebesgue measure. Let $\nu = \varrho_1\,dx$ be a probability measure on $B_r = \{x\colon |x| \leqslant r\}$ equivalent to the restriction of Lebesgue measure. Then there exist a Borel map $T\colon A \to B_r$ and a continuous function $\varphi\colon A \to [0, r]$ with convex sets $A_s = \{\varphi \leqslant s\}$ such that $\nu = \mu\circ T^{-1}$ and $T = \varphi \cdot \mathrm{n}$ almost everywhere with respect to $\mathscr{H}^d$, where $\mathrm{n} = \mathrm{n}(x)$ is the unit outer normal vector to the boundary of the sublevel set $\{y\colon \varphi(y) \leqslant \varphi(x)\}$.*

If the function φ is smooth, then the level sets of φ move according to the following equation of the Gaussian curvature flow:

$$\dot{x}(s) = -s^{d-1}\frac{\varrho_1(s\mathrm{n})}{\varrho_0(x)}K(x)\cdot \mathrm{n}(x),$$

where $x(0) \in \partial A$ is an arbitrary initial point. If the set A is strictly convex, then $\varphi\big|_{\partial A} = r$.

The proof is based on optimal transportation techniques. As a corollary we obtain a transport proof of the existence of the Gauss flow for strictly convex surfaces (see details in [167]). We note that by a change of coordinates the equation of the Gaussian curvature flow reduces (at least locally) to the Monge–Ampère parabolic equation

$$\partial_t u(t,x)\det D_x^2 u(t,x) = f(t,x), \qquad u(0,x) = u_0, \quad (t,x) \in \mathbb{R}^+ \times \mathbb{R}^{d-1}.$$

The function u is assumed to be convex in x and increasing in t. This equation can be rewritten in a 'transport' form. The map $T\colon (t,x) \mapsto (\langle x, \nabla_x u\rangle - u, \nabla_x u)$ serves as a 'parabolic' transportation. In addition, $\det DT = \partial_t u \det D_x^2 u$. For a development of this theme (the change of variables formula, connections with the parabolic Monge–Ampère equation, and the parabolic maximum principle), see [170]. On parabolic optimal equations, see also [171].

3.7.6. Geometric flows: Ricci flows. Let M be a smooth n-dimensional Riemannian manifold with metric g. The evolution of g defined by the equation $\partial g/\partial t = -2\,\mathrm{Ric}(g)$ is called the Ricci flow. Ricci flows became a main tool in the proof by G. Ya. Perelman of the Poincaré conjecture. In the first approximation the Ricci flow is a non-linear analogue of the heat semigroup and acts on the manifold as an 'averaging'. Nevertheless, the hope that under the action of the Ricci flow the manifold will shrink to a space of constant curvature is not realized. It turns out that singularities of the manifold may arise. In order to prevent this, one has to complicate the procedure and control various analytic characteristics, for example, constants in Sobolev-type inequalities on the manifold. Among other things this is done by the method of introducing various functionals such as the entropy or energy functionals, which turn out to be monotone under the action of the Ricci flow. As we have seen, for two gradient flows μ_t^1 and μ_t^2 of a convex functional the Kantorovich distance $t \mapsto W_2(\mu_t^1, \mu_t^2)$ is a non-increasing function. In particular, monotonicity holds for the diffusion with generator $\Delta - \nabla V \cdot \nabla$ for a convex function V. It turns out that a suitable generalization of this property to Ricci flows implies the monotonicity of certain functionals as proved by Perelman. Let us consider an example from [172]. We introduce the 'reverse' time $\tau = b - t$ and consider the non-linear 'adjoint' heat equation

$$\frac{\partial u}{\partial \tau} = \Delta_{g(\tau)} u - \frac{1}{2}\mathrm{R}u,$$

where $\mathrm{R} = \mathrm{Tr}(\mathrm{Ric})$ is the scalar curvature, $\Delta_{g(\tau)}$ is the Laplace–Beltrami operator in the metric $g(\tau)$, and $u(x,0)$ us a non-negative function with $\int_M u(x,0)\,d\,\mathrm{vol}_0 = 1$, where vol_t is the Riemannian volume corresponding to the metric $g(t)$. Differentiating the corresponding quantities with respect to t and using the above equations, one can see that $\nu_t = u(\tau, x)\ \mathrm{vol}_\tau$ is a probability measure for all τ. Here the formula $d\,\mathrm{vol}_\tau/d\tau = 2^{-1}\,\mathrm{Tr}(dg(\tau)/d\tau)\,\mathrm{vol}_{g(\tau)}$ for the volume evolution is used. The flow of measures constructed in this way will be called a 'diffusion'. Moreover, the integral

$$\int f(x,t)\,\nu_t(dx)$$

is constant if f is a solution of the heat equation $\partial_t f(x,t) = \Delta_{g(t)} f(x,t)$.

Let us now consider two diffusions ν_τ^1 and ν_τ^2. It is proved in [172] (see also [173]) that the functions $\tau \mapsto W_1^{g(\tau)}(\nu_\tau^1, \nu_\tau^2)$ and $\tau \mapsto W_2^{g(\tau)}(\nu_\tau^1, \nu_\tau^2)$ are decreasing (this generalizes the monotonicity property of the standard Kantorovich distance with respect to the Brownian diffusion in the flat case). The proof for W_1 is based on the following observations.

1) The Lipschitz constant of a function f satisfying the heat equation $\partial_t f = \Delta_{g(t)} f$ is non-increasing along t. This can be seen by applying the Bochner formula (3.7.3) to $g(t)$ and f and computing $\partial |\nabla f|^2_{g(t)}/\partial t$. Taking into account that

$\frac{d}{dt}A(t)^{-1} = -A(t)^{-1}\frac{d}{dt}A(t)A(t)^{-1}$ for any differentiable matrix function, we get that

$$\frac{\partial}{\partial t}|\nabla f|_g^2 = \frac{\partial}{\partial t}(g^{ij} f_{x_i} f_{x_j}) = -|\nabla f|^2_{\partial_t g} + 2\langle \nabla f, \nabla \Delta f\rangle$$
$$= |\nabla f|^2_{\partial_\tau g} + \Delta|\nabla f|^2 - 2|\operatorname{Hess}(f)|^2 - 2\operatorname{Ric}(\nabla f, \nabla f).$$

Therefore (in the metric $g(t)$), $\partial|\nabla f|^2/\partial t = \Delta|\nabla f|^2 - 2|\operatorname{Hess}(f)|^2 \leqslant \Delta|\nabla f|^2$. The desired property follows by the maximum principle.

2) We have the duality relation

$$W_1^{g(\tau)}(\nu^1_\tau, \nu^2_\tau) = \max\biggl\{\int \varphi\, d\nu^1_\tau - \int \varphi\, d\nu^2_\tau : \varphi \text{ is a 1-Lipschitz function}\biggr\}.$$

If now φ is a solution of the dual Monge–Kantorovich problem for the metric g_{τ_0}, then, letting $\varphi_\tau = -\Delta_{g(\tau)}\varphi$, we obtain a family of 1-Lipschitz functions for which the quantity

$$\int \varphi\, d\nu^1_\tau - \int \varphi\, d\nu^2_\tau$$

is preserved. This immediately implies the monotonicity of W_1. A more general result was obtained in [174], [173]. A natural object connected with Ricci flows is the so-called Perelman $\mathscr{L}$-length

$$\mathscr{L}(\gamma) = \int_{\tau_1}^{\tau_2} \sqrt{\tau}(\mathrm{R}(\gamma(\tau),\tau) + |\gamma'(\tau)|^2_{g(\tau)})\, d\tau,$$

where γ is some curve in M (for each moment τ its own Riemannian metric is involved); see [175]. The indicated $\mathscr{L}$-functional determines a Riemannian 'metric' on $M \times [\tau_1, \tau_2]$ (this functional does not completely correspond to the classical concept of length, since $\mathscr{L}$ may be negative). P. Topping considered an $\mathscr{L}$-optimal transportation of probability measures and the corresponding Kantorovich 'distance'. It turns out that on the space of measures there is a functional Θ constructed as a certain 'renormalized' Kantorovich $\mathscr{L}$-distance and possessing monotonicity along the flow of diffusions. From the monotonicity of this functional one can derive (by a suitable limit procedure) both the result of McCann and Topping in [172] and the monotonicity of the so-called $\mathscr{W}$-entropy introduced by Perelman in his proof of the Poincaré conjecture.

3.7.7. Transport networks. The theory of optimal transport networks is an interesting and in a sense alternative branch of the theory of optimal transportation. Unlike the familiar regular solutions to the Monge–Kantorovich problem, the behaviour of solutions to problems in the theory of transport networks is completely different. Visually these solutions can be represented as discrete collections of one-dimensional branching objects similar to trees or blood vessels. The cost of transportation can be easily defined on discrete objects. For example, suppose that for a discrete set of vertices (terminals) we are given flows between φ terminals. The total cost of transportation equals $\sum_i \varphi_i^\alpha \lambda_i$ for some constants $0 < \alpha < 1$ (λ_i is the edge length). For formulating the problem in the continuous case we recall that the Monge–Kantorovich problem admits a formulation in the language of dynamical optimal plans, that is, measures on the space of curves. Similarly, an optimal transport network is defined as a measure P on the space Ω of Lipschitz curves $\gamma\colon [0,+\infty) \to \mathbb{R}^d$ with given projections $P\circ\pi_0^{-1}$ ('the irrigating measure') and $P\circ\pi_\infty^{-1}$ ('the irrigated measure') that minimizes Gilbert's functional

$$\mathscr{E}^\alpha(P) = \int_\Omega \int_{\mathbb{R}^+} |\gamma(\omega,t)|_P^{\alpha-1} |\dot\gamma(t)|\,dt\,dP(\gamma).$$

Here $0 \leqslant \alpha \leqslant 1$ and the quantity $|\gamma(\omega,t)|_P$, called the multiplicity, is defined as $P(\{\gamma\colon \text{ there is a } t \text{ for which } \gamma(t)=x\})$. When $\alpha=1$ we obtain the classical Monge problem. For more details on this, see [176], [177].

3.8. Infinite-dimensional Monge–Kantorovich problems

3.8.1. Existence of optimal maps on the Wiener space. In this subsection we discuss analysis on an infinite-dimensional locally convex space X equipped with a Radon Gaussian measure γ.

Definition 3.8.1. A Radon probability measure γ on a locally convex space X is said to be Gaussian if its one-dimensional image $\gamma\circ f^{-1}$ is a Gaussian measure for all $f\in X^*$. If all the measures $\gamma\circ f^{-1}$ are symmetric, then γ is said to be symmetric or centred.

Below we shall assume for simplicity (if is not explicitly stated otherwise) that $X=\mathbb{R}^\infty$ (the countable power of the real line), and by an infinite-dimensional Gaussian measure we shall mean the countable product $\gamma=\bigotimes_{i=1}^\infty \gamma_i$ of the standard Gaussian measures on the real line. According to the well-known Tsirelson theorem, every centred Radon Gaussian measure not concentrated on a finite-dimensional subspace is isomorphic to γ via a measurable linear map, so the results discussed below are valid for the whole class of these measures. The general theory of Gaussian measures is presented in detail in the book [104]. Below we briefly discuss some particular features of analysis on a space with a Gaussian measure. Unlike in the finite-dimensional case, the measure γ_h obtained from the measure γ by a shift by some vector $h\in H$, that is, $\gamma_h(A)=\gamma(A-h)$, need not be absolutely continuous with respect to γ. Let us consider this question for the countable product $\gamma = \bigotimes_{i=1}^\infty \gamma_i$. Let $h=(h_i)\in\mathbb{R}^\infty$. A formal computation of the Radon–Nikodym density gives the expression

$$\frac{d\gamma_h}{d\gamma} = \mathsf{E}\biggl(\sum_{i=1}^\infty h_i x_i - \frac{1}{2}\sum_{i=1}^\infty h_i^2\biggr). \tag{3.8.1}$$

This makes sense if $\sum_{i=1}^\infty h_i^2<\infty$, that is, $h\in l^2$. Indeed, the coordinate functions x_i are independent standard Gaussian random variables on the space $\mathbb{R}^\infty$ with the measure γ. Since $\|S_{m+n}-S_m\|_{L^2(\gamma)}^2 = \sum_{i=m+1}^{m+n} h_i^2$, where $S_n=\sum_{i=1}^n\langle x,h_i\rangle$, the series $\sum_{i=1}^\infty h_i x_i$ converges in $L^2(\gamma)$ for $h\in l^2$. Moreover, one has convergence γ-a.e., since $\{S_n\}$ is a martingale with $\sup_n \mathsf{E}S_n^2<\infty$. This implies the equivalence of the measures γ_h and γ for $h\in l^2$. But if $h\notin l^2$, then the measures γ and γ_h are mutually singular, since one can find an element $k\in l^2$ for which

$\sum_{n=1}^{\infty} k_n h_n = +\infty$, and then the set K of sequences (x_n) for which the series of terms $k_n x_n$ converges has γ-measure 1, but $(K+h)\cap K = \varnothing$.

Definition 3.8.2. The space $H = \{h \in X : \gamma_h \sim \gamma\}$ is called the Cameron–Martin space of the measure γ.

It is clear from what has been said that $\gamma_h \sim \gamma$ (that is, the measures are equivalent: each of them is absolutely continuous with respect to the other) for all $h \in H$, and for all other h the measures γ and γ_h are mutually singular. The space H is linear and has measure zero with respect to γ if it is infinite-dimensional.

We recall that the Wiener process $\{w_t\}_{t\in[0,1]}$, $w_0 = 0$, gives a measure on the space of paths on $[0,1]$. In a special way it can be restricted to the space of continuous paths, which gives the Wiener measure P^W on the space $C[0,1]$ of continuous functions.

Example 3.8.3. For the countable product γ of standard Gaussian measures we have $H = l^2$. For P^W the Cameron–Martin space is the Sobolev class

$$W_0^{2,1} = \{h\colon h \text{ is absolutely continuous, } h(0)=0,\ h' \in L^2[0,1]\}.$$

Every vector h in the Cameron–Martin space defines a so-called measurable linear functional $\widehat{h}$ on X. For $\gamma = \prod_{i=1}^{\infty} \gamma_i$ this is $\widehat{h}(x) = \sum_{i=1}^{\infty} h_i x_i$. In the case of the measure P^W this is the stochastic integral

$$w \mapsto \int_0^1 h'(t)\, dw_t$$

of the deterministic function h'. Thus, to the vector h there corresponds a random variable (a measurable linear functional) $\widehat{h}$ on the space with a Gaussian measure. The space H is equipped with a natural norm by the equality

$$|h|_H^2 = \|\widehat{h}\|_{L^2(\gamma)}^2 = \int_X \widehat{h}^2\, d\gamma,$$

with which H becomes a separable Hilbert space.

Analysis on the space X with the Gaussian measure γ is closely related to the Cameron–Martin space. In a certain sense, naturally differentiable functions are differentiable only along the Cameron–Martin space. With this understanding of differentiation, it is possible to construct analogues of Sobolev spaces on the Wiener space. For defining these Sobolev spaces let us introduce some notation. As before, it is convenient to assume that $X = \mathbb{R}^\infty$ and $x = (x_i)$. Let $L^2(\gamma, H)$ denote the Hilbert space of H-valued γ-measurable maps v with finite norm

$$\|v\|_{L^2(\gamma,H)} = \left(\int_X |v(x)|_H^2\, \gamma(dx)\right)^{1/2}.$$

The Hilbert–Schmidt norm of a symmetric operator A on H is defined by the formula $\|A\|_{\mathscr{H}} = \bigl(\sum_{i=1}^{\infty} (Ae_i, Ae_i)_H\bigr)^{1/2}$, where $\{e_i\}$ is an orthogonal basis in H. The σ-algebra generated by the coordinate functions $x_1, \dots, x_n$ is denoted by $\mathscr{F}_n$. The space $\mathscr{F}C_b^\infty$ of smooth cylindrical functions consists of all functions of the form

$\zeta(x_1, \dots, x_n)$, where $\zeta \in C_b^\infty(\mathbb{R}^n)$ for some n. The Sobolev class $W^{2,1}(\gamma)$ consists of all functions $f \in L^2(\gamma)$ having a generalized gradient $\nabla_H f \in L^2(\gamma, H)$ along H determined by the equality

$$\int_X \partial_h \varphi\, f\, d\gamma = -\int_X \varphi \langle \nabla_H f, h\rangle_H\, d\gamma + \int_X \varphi f \widehat{h}\, d\gamma$$

for all $h \in H$ and $\varphi \in \mathscr{F}\mathscr{C}_b^\infty$, where $\partial_h \varphi(x) = \lim_{t\to 0}(\varphi(x+th) - \varphi(x))/t$. The Sobolev class $W^{2,2}(\gamma)$ consists of the functions that are twice H-differentiable in the generalized sense and have finite norm

$$\|f\|_{W^{2,2}(\gamma)} = \left(\int_X f^2\, d\gamma + \int_X |\nabla_H f|_H^2\, d\gamma + \int_X \|D_H^2 f\|_{\mathscr{H}}^2\, d\gamma\right)^{1/2}.$$

The matrix elements $\partial_{e_i}\partial_{e_j} f$ of the operator $D_H^2 f$ are defined via the integration by parts formula.

Finite-dimensional results independent of the dimension (the logarithmic Sobolev inequality and so on) are automatically true on the infinite-dimensional space if in place of the usual Euclidean norm we take the norm on the Cameron–Martin space. For example, the logarithmic Sobolev inequality takes the form

$$\int f^2 \log \frac{f^2}{\|f\|_{L^2(\gamma)}^2}\, d\gamma \leqslant 2 \int |\nabla_H f|_H^2\, d\gamma.$$

The Radon–Nikodym density $d\gamma_h/d\gamma$ in the case of an abstract Gaussian measure has the form

$$\frac{d\gamma_h}{d\gamma} = \mathsf{E}\left(\widehat{h}(x) - \frac{1}{2}|h|_H^2\right).$$

In the case $\gamma = \bigotimes_{i=1}^\infty \gamma_i$ this is the formula (3.8.1). In the case of P^W we arrive at the classical Cameron–Martin formula

$$\frac{dP_h^W}{dP^W} = \mathsf{E}\left(\int_0^1 h'(t)\, dw_t - \frac{1}{2}\int_0^1 |h'(t)|^2\, dt\right).$$

For Gaussian measures it is useful to consider non-linear shifts along the Cameron–Martin space of the form

$$T(x) = x + F(x), \qquad (3.8.2)$$

where F is a map with values in H. Under rather broad assumptions it is possible to prove for maps of this form that the measure $\gamma\circ T^{-1}$ is absolutely continuous with respect to γ and to write explicitly the change of variables formula. For example, such maps can be defined by solutions to stochastic differential equations, and the corresponding 'change of variables formula' is called Girsanov's formula. For quite a long time it was an open question (discussed in particular in [178]) as to which measures of the form $\nu = g\cdot\gamma$ are images of γ under maps of the form (3.8.2). An answer was obtained in [179].

Theorem 3.8.4. *Every probability measure of the form $\nu = g\cdot\gamma$ is representable in the form $\nu = \gamma\circ T^{-1}$, where $T(x) = x + F(x)$, $F\colon X \to H$.*

The map T was constructed with the aid of triangular maps, and an analogue of the transport inequality for triangular maps was essentially used; such maps are discussed in the last section. However, the question is open as to whether a map T of the form (3.8.2) can itself always be chosen to be triangular. As we shall see, non-linear shifts along the Cameron–Martin space arise also in the Monge–Kantorovich problem for $|\cdot|_H^2$, which was first studied by Feyel and Üstünel [180], [181]. In this situation, the problem is to minimize the functional

$$K_H(\mu,\nu,m) = \int_{X\times X} |x-y|_H^2\,dm, \qquad m\in\Pi(\mu,\nu).$$

Let

$$W_2(\mu,\nu) = \sqrt{K_H(\mu,\nu)}\,.$$

Letting $K_H(\mu,\nu) = \inf\{K_H(\mu,\nu,m) : m\in\Pi(\mu,\nu)\}$ and introducing the finite-dimensional continuous functions $h_n(x,y) = \sum_{i=1}^n |x_i-y_i|^2$, we note that

$$K_H(\mu,\nu) = \sup_n K_{H,n}(\mu_n,\nu_n), \tag{3.8.3}$$

where $K_{H,n}$ corresponds to the finite-dimensional cost function h_n, and μ_n and ν_n are the projections of μ and ν on $\mathbb{R}^n$. This follows from the monotone convergence theorem and the facts that $h_n\uparrow h$ and the integral of the function h_n with respect to the measure σ equals the integral with respect to the projection of σ on $\mathbb{R}^n\times\mathbb{R}^n$.

A substantial novelty in this situation is that here we have a much more narrow class of measures m on which $K_H(\mu,\nu,m)$ assumes finite values. Partly this is connected with the fact that $h(x,y)=+\infty$ almost everywhere with respect to $\gamma\otimes\gamma$. We have not encountered this phenomenon in the previous sections. Of course, if we deal with measures concentrated on H, then all these troubles disappear, but the whole substantive part of the case attached to the measure γ will also disappear with them. We stress that this concerns both measures and not just μ. For example, if $\mu=\gamma$ and the measure ν is concentrated on H (even if it is the Dirac measure at the origin), then we have $K_H(\mu,\nu,m)=+\infty$ for all $m\in\Pi(\mu,\nu)$ since otherwise writing m in the form $m(dxdy) = m^y(dx)\nu(dy)$, we get for ν-a.e. $y\in H$ that the measure m^y must also be concentrated on H (if $|x-y|_H<\infty$, then $x\in H$), and hence $\gamma(H)=1$, which is false. If the quantity $K_H(\mu,\nu)$ is finite, then the general reasoning about the existence of the corresponding solution to the Kantorovich problem remains valid due to the fact that the lower semicontinuity of the functional $m\mapsto K_H(\mu,\nu,m)$ is preserved here.

We consider another example: ν is the image of γ under the homothety $x\mapsto 2x$. Then $K_{H,n}(\gamma_n,\nu_n)=n$, since the map $Tx=2x$ taking γ_n to ν_n is a gradient, and hence by the uniqueness theorem it serves as the optimal transportation of γ_n to ν_n for the cost function h_n, and then $K_{H,n}(\gamma_n,\nu_n) = M(\gamma_n,\nu_n,T)$ is the integral of $|x-2x|_H^2 = |x|_H^2$ with respect to the measure γ_n and equals n.

Thus, we have to impose certain additional conditions on the measure ν. A common (but not the only possible) situation arises if both measures μ and ν are absolutely continuous with respect to γ (but even this still does not guarantee that $K_H(\mu,\nu)<\infty$). An effectively verifiable sufficient (but by no means necessary) condition is the following.

Proposition 3.8.5. *Let* $\mu = \varrho_\mu \cdot \gamma$ *and* $\nu = \varrho_\nu \cdot \gamma$ *be probability measures such that* $\mathrm{Ent}_\gamma\, \mu < \infty$ *and* $\mathrm{Ent}_\gamma\, \nu < \infty$. *Then*

$$K_H(\mu,\nu) = W_2^2(\mu,\nu) \leqslant 4\,\mathrm{Ent}_\gamma\, \mu + 4\,\mathrm{Ent}_\gamma\, \nu. \tag{3.8.4}$$

Proof. This follows from (3.8.3) and the finite-dimensional Talagrand inequality (3.3.2), since $\mathrm{Ent}_\gamma\, \mu = \sup_n \mathrm{Ent}_{\gamma_n}\, \mu_n$ and similarly for ν. The last facts follow from the fact that the Radon–Nikodym densities $\varrho_n = d\mu_n/d\gamma_n$, considered on the space $\mathbb{R}^\infty$ with the measure γ, form a martingale with respect to the σ-algebras generated by the projections on $\mathbb{R}^n$, and it converges γ-a.e. to $\varrho = d\mu/d\gamma$. □

We note that by Jensen's inequality for the conditional expectations the quantities $\mathrm{Ent}_{\gamma_n}\, \mu_n$ are increasing and do not exceed $\mathrm{Ent}_\gamma\, \mu$.

Without Talagrand's inequality it is not even obvious that $K_H(\gamma,\nu) < \infty$ for the measure $\nu = \gamma/\gamma(A)$, where A is a set with $\gamma(A) > 0$. Thus, here it is rare to find explicitly at least one measure $\sigma \in \Pi(\mu,\nu)$ with $K(\mu,\nu,\sigma) < \infty$.

Of course, in addition to being an analogue of the Kantorovich problem the problem here is also an analogue of the Monge problem of minimizing the functional

$$\int_X |x - T(x)|_H^2\, \mu(dx)$$

over maps T taking the measure μ to the measure ν, where T must necessarily be of the form (3.8.2). Consider a pair of probability measures $\mu = \varrho_\mu \cdot \gamma$, $\nu = \varrho_\nu \cdot \gamma$. We are interested in the question of whether one can map μ to ν by a map of the form (3.8.2) minimizing the indicated functional. It turns out that this is possible, and under broad assumptions F must in a certain sense be a gradient along H. We remark that even for the identity map $T(x) = x$ the equality $x = \nabla_H |x|_H^2/2$ is not meaningful in the infinite-dimensional space, since the Cameron–Martin space has measure zero. Therefore, one cannot expect that T itself will be a gradient along H. A situation arises that is similar to the situation on manifolds. One should look for a solution of the form

$$T(x) = x + \nabla_H f(x),$$

where f is a so-called 1-convex function (see [182]). A function f is said to be 1-convex if the map $h \mapsto F_x(h) := f(x+h) + |h|^2/2$ from H to $L^0(\gamma)$ (with its natural ordering) is convex, that is, given $h, k \in H$ and $\alpha \in [0,1]$, one has

$$F_x(\alpha h + (1-\alpha)k) \leqslant \alpha F_x(h) + (1-\alpha) F_x(k) \quad \text{for } \gamma\text{-a.e. } x,$$

where the corresponding measure-zero set may depend on h, k, and α. One can show that, given an orthonormal basis $\{e_i\}$ in H, for every fixed i there is a version of f such that the functions $t \mapsto f(x + te_i) + t^2/2$ are convex. Hence, the partial derivatives $\partial_{e_i} f$ exist almost everywhere. We define $\nabla_H f(x)$ as $\sum_{i=1}^\infty \partial_{e_i} f \cdot e_i$ if this series converges in H. Finite-dimensional 1-convex functions are (up to modifications) functions of the form $\varphi(x) - |x|^2/2$, where φ is convex. It should be noted that typically in such problems partial derivatives are defined in the Sobolev sense (see [104]).

We now give a result of Feyel and Üstünel [180].

Theorem 3.8.6. *Let $\mu = \varrho_\mu \cdot \gamma$ and $\nu = \varrho_\nu \cdot \gamma$ be probability measures such that* $\operatorname{Ent}_\gamma \mu < \infty$ *and* $\operatorname{Ent}_\gamma \nu < \infty$. *Suppose additionally that μ satisfies the Poincaré inequality (for instance, $\mu = \gamma$), that is, there exists a $C > 0$ such that*

$$\int \Bigl(f - \int f\, d\mu\Bigr)^2 d\mu \leqslant C^{-1} \int |\nabla_H f|^2\, d\mu \quad \textit{for all} \quad f \in \mathscr{F}\mathscr{C}_b^\infty.$$

Then there exists a map $T(x) = x + F(x)$ such that $\nu = \mu \circ T^{-1}$ and $F = \nabla_H f \in H$, where f is a 1-convex function. The map T is unique up to μ-equivalence, is a unique solution to the Monge problem, and generates a unique solution to the Kantorovich problem.

Proof. As indicated above, the Kantorovich problem described for the functional K_H and the pair μ, ν has a solution. Moreover, $W_2(\mu,\nu) < \infty$ by (3.8.4). We assume further that $X = \mathbb{R}^\infty$ and γ is the countable power of the standard Gaussian measure. Then $H = l^2$, so in the notation for the norm and gradient we omit the index H. Denote by π and π_n some solutions to the Kantorovich problem for the pairs μ, ν and μ_n, ν_n, respectively. Since the projection of the measure π on $\mathbb{R}^n \times \mathbb{R}^n$ has the marginals μ_n and ν_n, we obtain the obvious inequality

$$W_2(\mu_n,\nu_n) \leqslant \int \sum_{i=1}^n |x_i - y_i|^2\, \pi(dx\, dy) \leqslant W_2(\mu,\nu).$$

Similarly, $W_2(\mu_m,\nu_m) \leqslant W_2(\mu_n,\nu_n)$ if $m \leqslant n$. It follows from this and the above inequalities that $\lim_{n\to\infty} W_2(\mu_n,\nu_n) = W_2(\mu,\nu)$. Let $\Phi_n(x)$, $\Psi_n(y)$ be a solution to the dual Kantorovich problem for μ_n, ν_n, and let $\varphi_n(x) = \Phi_n(x) - |x|^2/2$ and $\psi_n(y) = \Psi_n(y) - |y|^2/2$. We observe that

$$F_n(x,y) = \varphi_n(x) + \psi_n(y) + \frac{1}{2}|x-y|^2 \geqslant 0,$$

and exact equality is attained π_n-a.e. It is clear that

$$\int F_n(x,y)\, \pi(dx\, dy) = W_2(\mu,\nu) - W_2(\mu_n,\nu_n) \to 0.$$

Therefore, $F_n \to 0$ in $L^1(\pi)$. Hence the sequence $\{F_n\}$ is uniformly integrable. Since μ satisfies the Poincaré inequality, this inequality holds for the projections μ_n (with the same constant). By the Poincaré inequality,

$$W_2(\mu,\gamma) \geqslant W_2(\mu_n,\gamma_n) = \frac{1}{2}\int |\nabla\varphi_n|^2\, d\mu_n \geqslant \frac{C}{2}\int \Bigl(\varphi_n - \int \varphi_n\, d\mu_n\Bigr)^2 d\mu_n.$$

Subtracting constants, we pass to the case $\int \varphi_n\, d\mu_n = 0$. The estimate obtained gives the uniform integrability of the sequence $\{\varphi_n\}$ with respect to the measure μ. Hence, the sequence $\{\psi_n\}$ is uniformly integrable with respect to the measure ν. One can thus assume that the sequences $\{\varphi_n\}$ and $\{\psi_n\}$ are uniformly integrable with respect to the measure π. By a well-known result, for some subsequence of indices $\{n_k\}$ sequences $\{\varphi_n'\}$ and $\{\psi_n'\}$ of convex combinations of the functions φ_{n_k}

and ψ_{n_k} converge π-almost everywhere and in $L^1(\pi)$ to some functions φ and ψ, respectively. Let us set

$$\varphi = \limsup \varphi'_n, \qquad \psi = \limsup \psi'_n$$

everywhere (and not just where they converge). We have $\varphi(x)+\psi(y)+|x-y|^2/2 \geqslant 0$, and the equality is satisfied π-a.e. Fix an $h \in H$. The equality $|x+h-y|^2 = |x-y|^2 + 2\langle h, x-y\rangle + |h|^2$ implies that $\varphi(x+h) - \varphi(x) \geqslant -\langle h, x-y\rangle - |h|^2/2$ π-a.e. The latter means that $y = x + \nabla\varphi(x)$ for μ-a.e. x, and $T(x) = x + \nabla\varphi(x)$ is the desired map. We now observe that every $L^1(\pi)$-limit $\widetilde{\varphi}$ of convex combinations of the functions φ_n will satisfy the equality $\nabla\widetilde{\varphi}(x) = y - x$ π-a.e. It follows that $\nabla\varphi(x) = \nabla\widetilde{\varphi}(x)$ for μ-a.e. x. If now $\widetilde{\gamma}$ is some other solution to the Kantorovich problem, then it is also concentrated on the graph of $\nabla\varphi$, where φ is any limit point of our finite-dimensional approximations. It follows that we have obtained a unique solution to both the Kantorovich problem and the Monge problem. □

Remark 3.8.7. It is not difficult to prove that the map $T(x) = x+\nabla\varphi(x)$ is invertible and $T^{-1}(y) = S(y)$ for ν-a.e. y and that $S(y) = y + \nabla\psi(x)$ is also an optimal map.

The paper [180] contains a formulation (with a proof that is not completely clear to us) of a more general result on the existence of an optimal map between a pair of measures with a finite Kantorovich distance. In [183] an alternative proof of the existence of a transportation of a measure $f\cdot\gamma$ to the measure γ was proposed, based on the inequality (3.3.1). Using this inequality in the case when $m = \gamma$, $\mu = f\cdot\gamma$, and $\nu = g\cdot\gamma$, we obtain the inequality

$$\operatorname{Ent}_\mu \nu = \int g \log\frac{f}{g}\, d\gamma \geqslant \frac{1}{2}\int |\nabla\Phi_f - \nabla\Phi_g|^2 f\, d\gamma. \tag{3.8.5}$$

From (3.8.5) one can readily deduce the existence of an optimal map T of a measure $f\cdot\gamma$ with $\operatorname{Ent}_\gamma f < \infty$ to the measure γ. Indeed, the finite-dimensional optimal transportations T_n mapping $f_n\cdot\gamma$ to γ, where $f_n = \mathsf{E}_\gamma^{\mathscr{F}_n} f$, satisfy the inequality

$$\int f_n \log\frac{f_m}{f_n}\, d\gamma \geqslant \frac{1}{2}\int |\nabla\Phi_{f_m} - \nabla\Phi_{f_n}|^2 f_m\, d\gamma.$$

From the fact that

$$\lim_{n,m\to\infty}\int f_n \log\frac{f_m}{f_n}\, d\gamma = 0,$$

it is easy to show that the sequence $\{\nabla\Phi_{f_n}\}$ converges $f\cdot\gamma$-a.e. This implies the existence of the desired map. Moreover, this proof can be extended to broader classes of measures; in particular, it can be used to construct examples of optimal transportations of mutually singular infinite-dimensional measures (see Theorem 3.8.8).

In the recent paper [184] the existence of an optimal map was established between any two probability measures absolutely continuous with respect to γ.

The question of which pairs of measures (mutually singular in general) possess optimal transportations is undoubtedly a central issue in the infinite-dimensional Monge–Kantorovich theory. A weaker version of a solution to this problem is given in the next theorem, which we state without proof (it is based on the estimate (3.8.5)). In this theorem T is obtained as a limit (almost everywhere) of finite-dimensional optimal maps.

Theorem 3.8.8. *Let μ be a probability measure on $\mathbb{R}^\infty$. Suppose that the finite-dimensional projections $\mu_n = \mu \circ P_n^{-1} = g_n\,dx$ have the property that*

$$\lim_{m\to\infty} \sup_{n>m} e(n,m) = 0,$$

where $e(n,m) = \inf_{\tau \in M_{(n,m)}} \mathrm{Ent}_{\mu_m\times\tau}\bigl(d\mu_n/d(\mu_m\times\tau)\bigr)$, and $M_{(n,m)}$ is the space of probability measures on the subspace $L_{m,n}$ generated by $\{e_{m+1},\dots,e_n\}$. Then there is a map T of μ to γ almost everywhere equal to the limit of the finite-dimensional optimal maps.

The hypothesis of the theorem is satisfied for any measure μ having finite entropy with respect to some product measure.

3.8.2. The change of variables formula. Under some restrictions on the measures, it becomes possible to prove an infinite-dimensional change of variables formula for optimal maps. Let γ be the standard Gaussian measure on $\mathbb{R}^\infty$ as above. Let $T(x) = x + \nabla\varphi(x)$ be the optimal transportation of the measure $g\cdot\gamma$ to γ, and let $S(x) = T^{-1}(x) = x + \nabla\psi(x)$ be the inverse map. A formal expression for the change of variables formula takes the form

$$g = \det{}_2(\mathrm{I} + D^2\varphi)\mathsf{E}\biggl(\mathscr{L}\varphi - \frac{1}{2}|\nabla\varphi|^2\biggr),$$

where D^2 is the second derivative,

$$\mathscr{L}\varphi(x) = \Delta\varphi(x) - \langle x, \nabla\varphi(x)\rangle = \mathrm{div}_\gamma(\nabla\varphi)(x)$$

is the Ornstein–Uhlenbeck operator, and $\det_2$ is the Fredholm–Carleman determinant defined by the formula $\det_2(\mathrm{I}+K) = \prod_{i=1}^\infty (1+k_i)\mathrm{e}^{-k_i}$, where K is a symmetric Hilbert–Schmidt operator with eigenvalues k_i. In the class $W^{p,2}(\gamma)$ one can introduce the operator $\mathscr{L}$ as the generator of the Ornstein–Uhlenbeck semigroup in $L^p(\gamma)$ or one can define $\mathscr{L}f$ as a limit in $L^p(\gamma)$ of the functions $\mathscr{L}f_n$ for some functions $f_n \in \mathscr{FC}_b^\infty$ convergent to f in $W^{p,2}(\gamma)$.

For the inverse map $T^{-1}(x) = x + \nabla\psi(x)$ the change of variables formula takes the form

$$g(x+\nabla\psi(x))\det{}_2(\mathrm{I} + D^2\psi(x))\mathsf{E}\biggl(\mathscr{L}\psi - \frac{1}{2}|\nabla\psi|^2\biggr) = 1. \tag{3.8.6}$$

In the general case the question as to which maps satisfy the change of variables formula is non-trivial (see [104], [67], and [178]). In our case one can make use of a special form of T. The equality (3.8.6) was obtained in [180] under the assumption that $g\cdot\gamma$ is uniformly convex, that is, $-D^2\log g + \mathrm{I} \geqslant \varepsilon\,\mathrm{Id}$ for some $\varepsilon > 0$. As shown in [185], under the assumption that $\mathrm{Ent}_\gamma g < \infty$ and $g > c > 0$ one has

$$g = \det{}_2(\mathrm{I} + D_a^2\varphi)\mathsf{E}\biggl(\mathscr{L}_a\varphi - \frac{1}{2}|\nabla\varphi|^2\biggr),$$

where $D_a^2\varphi$ and $\mathscr{L}_a\varphi$ are the absolutely continuous parts of $D^2\varphi$ and $\mathscr{L}\varphi$, respectively. Similarly, under the assumptions that $\mathrm{Ent}_{g\cdot\gamma}\frac{1}{g} < \infty$ and $g < c$ one has

$$g(x+\nabla\psi(x))\det{}_2(\mathrm{I} + D_a^2\psi(x))\mathsf{E}\biggl(\mathscr{L}_a\psi - \frac{|\nabla\psi|^2}{2}\biggr) = 1.$$

The next theorem is proved in [84].

Theorem 3.8.9. *Suppose that* $\mathrm{I}_\gamma g := \int \frac{|\nabla g|^2}{g}\, d\gamma < \infty$. *Then* $D^2\varphi$ *belongs to* $\mathscr{H}\mathscr{S}$ $g\cdot\gamma$*-a.e.,* $\int \|D^2\varphi\|^2_{\mathscr{H}\mathscr{S}} g\, d\gamma < \infty$, $\mathscr{L}\varphi \in L^1(g\cdot\gamma)$, *and the change of variables formula* $g = \det(\mathrm{I} + D^2\varphi)\mathsf{E}(\mathscr{L}\varphi - |\nabla\varphi|^2/2)$ *holds* $g\cdot\gamma$*-a.e.*

In this theorem $\mathscr{L}\varphi$ is understood in the sense of integration by parts, but under the additional condition that $1/g \in L^r(\gamma)$ for some $r > 1$ one has the inclusion $\varphi \in W^{2r/(1+r),2}(\gamma)$, so that $\mathscr{L}\varphi$ exists in the sense of $W^{2r/(1+r),2}(\gamma)$. In [84] there are some other results on the regularity of φ (see also § 2.2.5). A natural problem in this area is the regularity of the optimal map (solved partially in Theorem 3.8.9). Let γ be a Gaussian measure on an infinite-dimensional space, and let $T(x) = x + \nabla\varphi(x)$ be the optimal transportation of a probability measure $f\cdot\gamma$ to a probability measure $g\cdot\gamma$. It is of interest to find a priori Sobolev estimates for φ in the general case.

The Kantorovich metric connected with the Wiener space is studied also in [186]. Optimal transportations on configuration spaces are considered in [187].

3.9. Some other types of transformations of measures

3.9.1. Triangular maps. Suppose that we are given a finite or countable number of spaces X_k and let $X = \prod_k X_k$. A map $T = (T_1, T_2, \dots)\colon X \to X$ is said to be *triangular* if $T_k(x) = T_k(x_1, \dots, x_k)$ for all k. If $X_k = \mathbb{R}^1$ for all k, then a triangular map is said to be *increasing* if all the functions $x_k \mapsto T_k(x_1, \dots, x_k)$ are increasing. Similarly, one defines increasing triangular maps on subsets of $\mathbb{R}^\infty$. We note that no monotonicity in other variables is required. Triangular maps appeared in the 1950s (M. Rosenblatt, H. Knothe; see references in [17]). Later, Gromov used triangular maps for deriving the classical isoperimetric inequality (see § 3.1). Basic properties of triangular maps are presented in [29], § 10.10 (vii) and [67], § 10.7. The word 'triangular' in this name is explained by the fact that the derivative of a differentiable triangular map on $\mathbb{R}^n$ is given by a triangular matrix. In spite of their rather special form, triangular maps possess rich possibilities for transforming measures. In the one-dimensional case Lebesgue measure λ on $(0,1)$ can be taken to an arbitrary probability measure ν on $(0,1)$ by an increasing function T by using the distribution function F_ν of the measure ν (if F_ν is strictly increasing, then $T = F_\nu^{-1}$). It follows from this that any atomless probability measure μ on the real line can be taken to any probability measure on the real line (if μ has a positive density, then this transformation can be found as a non-decreasing function). We show how to construct a triangular map on the plane transforming an absolutely continuous probability measure μ to a probability measure ν. Let μ_1 and ν_1 denote the projections of the measures μ and ν on the first coordinate line, and consider the one-dimensional map T_1 taking μ_1 to ν_1. Let μ^{x_1} and ν^{x_1} denote the conditional probability measures on the second coordinate line, $x_1 \in \mathbb{R}^1$. For μ_1-almost every x_1 the measure μ^{x_1} has a density. Hence, it can be taken to the measure $\nu^{T_1(x_1)}$ by a function $x_2 \mapsto T_2(x_1, x_2)$, which can be made to be increasing in the case of a positive density. It is clear that $T := (T_1, T_2)$ is a triangular map which is increasing if the density of μ is positive. One can prove directly that it takes μ to ν. The construction is continued inductively by using one-dimensional conditional measures on the last coordinate line. Therefore, for any pair of absolutely continuous probability measures on $\mathbb{R}^n$ one can construct a triangular map T transforming one of the measures into the other. A more general fact can be proved in a completely analogous way.

Theorem 3.9.1. *Let $X = \prod_{n=1}^{\infty} X_n$, where X_n are Souslin spaces. Let μ be a Borel probability measure on X such that for all n its projections on $\prod_{j=1}^{n} X_j$ and its conditional measures on X_n have no atoms. Then for every Borel probability measure ν on X there is a triangular Borel map $T\colon X \to X$ such that $\mu\circ T^{-1} = \nu$.*

If $X_n = \mathbb{R}^1$ for all n and the projections of μ on all the spaces $\mathbb{R}^n$ have positive densities, then this triangular map can be made to be increasing. The latter property uniquely defines it up to μ-equivalence, and it is called the canonical triangular map.

Triangular maps are almost never optimal (except for the one-dimensional case); this is already seen from the fact that the Jacobi matrix of a triangular map is triangular and for an optimal map it is symmetric, which can occur simultaneously only for diagonal matrices. Nevertheless, many properties of increasing triangular maps are close to properties of optimal transportations. For example, it is proved in [188] that the following transport inequality holds. A measure $\mathrm{e}^{-V}dx$ on $\mathbb{R}^n$ is said to be convex with constant $C > 0$ if $V'' \geqslant C\cdot \mathrm{Id}$.

Theorem 3.9.2. *Suppose that a Borel probability measure μ on $X = \mathbb{R}^\infty$ is uniformly convex with constant $C > 0$, that is, its projections on all the spaces $\mathbb{R}^n$ are convex with a constant $C > 0$, and let $H = l^2$. If $\nu \ll \mu$ is a probability measure such that $f = d\nu/d\mu$ satisfies $f\log f \in L^1(\mu)$, then the canonical triangular map $T_{\mu,\nu}$ has the property that*

$$\int_X |T_{\mu,\nu}(x) - x|_H^2\,\mu(dx) \leqslant \frac{2}{C}\int_X f\log f\,d\mu.$$

About probabilistic applications of triangular maps see [188], [189]. In [190] triangular maps are represented in the form of limits of solutions of the Monge problem with specially chosen cost functions.

3.9.2. Moser transformations, semigroups, and transport equations. Some maps transforming a given pre-image measure to a given image measure can be constructed by means of the classical Liouville theorem on change of volume. The idea of such constructions goes back to Moser's paper [191] on diffeomorphisms of manifolds with measures. For example, in [192] the following construction was realized. Let us consider the semigroup $P_t = \mathrm{e}^{tL}$ generated by the elliptic operator $L = \Delta - \langle\nabla V, \nabla\rangle = \mathrm{e}^V \operatorname{div}(\mathrm{e}^{-V}\cdot\nabla)$, and the flow of probability measures $\nu_t = P_t(\mathrm{e}^{-W+V})\cdot\mu$. It is obvious that μ is an invariant measure for P_t, $\nu_0 = \nu$, and $\nu_\infty = \mu$. We write an equation for ν_t:

$$\frac{d}{dt}\nu_t = LP_t(\mathrm{e}^{-W+V})\cdot\mu = \operatorname{div}\bigl[\nabla P_t(\mathrm{e}^{-W+V})\cdot \mathrm{e}^{-V}\bigr] = \operatorname{div}\bigl[\nabla \log P_t(\mathrm{e}^{-W+V})\cdot\nu_t\bigr].$$

By Liouville's theorem, if $X(t)$ is the solution of the equation $dX(t)/dt = b(X(t))$ with $X(0) = I$, where b is smooth, then the image $\varrho_t\,dx$ of Lebesgue measure with respect to the diffeomorphism $X(t)$ satisfies the equation $\partial\varrho_t/\partial t = \operatorname{div}(\varrho_t\cdot b)$.

The corresponding flow of diffeomorphisms is given by the equation $dS_t/dt = -\nabla \log P_t(\mathrm{e}^{-W+V}) \circ S_t$ with $S_0 = \mathrm{Id}$, where ν_t and S_t are related by the equality $\nu_t = \nu \circ S_t^{-1}$. In particular, the limiting map $S_\infty = \lim_{t\to\infty} S_t$ takes ν to μ.

For results on evolution equations and results related to optimal transportations, see [193]. On connections with Young measures, see [194]. On the Monge and Kantorovich problems with several measures, see [195] and [15]. Differential forms on the space of probability measures are considered in [196]. The flow of publications on the questions considered above is very intensive; the bibliography below covers considerably less than half of it.

Bibliography

[1] Л. В. Канторович, "О перемещении масс", *Докл. АН СССР* **37** (1942), 227–229; English transl., L. V. Kantorovitch, "On the translocation of masses", *C. R. (Doklady) Acad. Sci. URSS (N. S.)* **37** (1942), 199–201.

[2] Л. В. Канторович, "Об одной проблеме Монжа", *УМН* **3**:2(24) (1948), 225–226. [L. V. Kantorovich, "A problem of Monge", *Uspekhi Mat. Nauk* **3**:2(24) (1948), 225–226.]

[3] Л. В. Канторович, Г. Ш. Рубинштейн, "Об одном функциональном пространстве и некоторых экстремальных задачах", *Докл. АН СССР* **115**:6 (1957), 1058–1061. [L. V. Kantorovich and G. Sh. Rubinshtein, "One function space and some extremal problems", *Dokl. Akad. Nauk SSSR* **115**:6 (1957), 1058–1061.]

[4] Л. В. Канторович, Г. Ш. Рубинштейн, "Об одном пространстве вполне аддитивных функций множества", *Вестн. ЛГУ* **13**:7 (1958), 52–59. [L. V. Kantorovich and G. Sh. Rubinshtein, "A space of completely additive set functions", *Vestn. Leningrad. Univ.* **13**:7 (1958), 52–59.]

[5] А. М. Вершик, "Метрика Канторовича: начальная история и малоизвестные применения", *Теория представлений, динамические системы*. XI, Специальный выпуск, Зап. научн. сем. ПОМИ, **312**, ПОМИ, СПб. 2004, с. 69–85; English transl., A. M. Vershik, "The Kantorovich metric: initial history and little-known applications", *J. Math. Sci. (N. Y.)* **133**:4 (2006), 1410–1417.

[6] G. Monge, *Mémoire sur la théorie des déblais et de remblais*, Histoire de l'Académie Royale des Sciences avec les Mémoires de Mathématique & de Physique, Paris 1781.

[7] В. Л. Канторович, С. С. Кутателадзе, Я. И. Фет (ред.), *Леонид Витальевич Канторович: человек и ученый*, т. 1, 2, Наука Сибири в лицах, Изд-во СО РАН, филиал "Гео", Новосибирск 2002, 543, 614 с. [V. L. Kantorovich, S. S. Kutateladze, and Ya. I. Fet (eds.), *Leonid Vital'evich Kantorovich: person and scientist*, vols. 1, 2, Siberian Science in Persons, Academic Publishing House of the Siberian Branch of the RAS, 'Geo' subsidiary, Novosibirsk 2002, 543, 614 pp.]

[8] А. М. Вершик, С. С. Кутателадзе, С. П. Новиков, "Леонид Витальевич Канторович (к 100-летию со дня рождения)", *УМН* **67**:3(405) (2012), 185–191; English transl., A. M. Vershik, S. S. Kutateladze, and S. P. Novikov, "Leonid Vital'evich Kantorovich (on the 100th anniversary of his birth)", *Russian Math. Surveys* **67**:3 (2012), 589–597.

[9] L. Ambrosio, "Lecture notes on optimal transport problems", *Mathematical aspects of evolving interfaces* (Funchal 2000), Lecture Notes in Math., vol. 1812, Springer, Berlin 2003, pp. 1–52.

[10] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, Scuola Normale Superiore di Pisa, Pisa 2011.

[11] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures Math. ETH Zurich, Birkhäuser Verlag, Basel 2005, viii+333 pp.

[12] L. C. Evans, "Partial differential equations and Monge–Kantorovich mass transfer", *Current developments in mathematics*, 1997 (Cambridge, MA), Int. Press, Boston, MA 1999, pp. 65–126.

[13] W. Gangbo and R. J. McCann, "The geometry of optimal transportation", *Acta Math.* **177**:2 (1996), 113–161.

[14] R. J. McCann and N. Guillen, *Five lectures on optimal transportation: geometry, regularity and applications*, 2010, arXiv: 1011.2911v1.

[15] S. T. Rachev and L. Rüschendorf, *Mass transportation problems*, vols. I, II, Probab. Appl. (N. Y.), Springer-Verlag, New York 1998, xxvi+508, xxvi+430 pp.

[16] C. Villani, *Topics in optimal transportation*, Grad. Stud. Math., vol. 58, Amer. Math. Soc., Providence, RI 2003, xvi+370 pp.

[17] C. Villani, *Optimal transport. Old and new*, Grundlehren Math. Wiss., vol. 338, Springer-Verlag, Berlin 2009, xxii+973 pp.

[18] А. М. Вершик, "Несколько замечаний о бесконечномерных задачах линейного программирования", *УМН* **25**:5(155) (1970), 117–124; English transl., A. M. Vershik, "Some remarks on the infinite-dimensional problems of linear programming", *Russian Math. Surveys* **25**:5 (1970), 117–124.

[19] А. Н. Боголюбов, *Гаспар Монж*, Наука, М. 1978, 184 с. [A. N. Bogolyubov, *Gaspard Monge*, Nauka, Moscow 1978, 184 pp.]

[20] P. Appell, "Mémoire sur les déblais et les remblais des systèmes continus ou discontinus", Mém. Acad. Sci. Inst. France, vol. 29, no. 3, Paris 1887, 208 pp.

[21] В. Н. Судаков, "Геометрические проблемы теории бесконечномерных вероятностных распределений", Тр. МИАН СССР, **141**, 1976, с. 3–191; English transl., V. N. Sudakov, "Geometric problems of the theory of infinite-dimensional probability distributions", *Proc. Steklov Inst. Math.* **141** (1979), 1–178.

[22] L. Ambrosio, B. Kirchheim, and A. Pratelli, "Existence of optimal transport maps for chrystalline norms", *Duke Math. J.* **125**:2 (2004), 207–241.

[23] L. C. Evans and W. Gangbo, *Differential equations methods for the Monge–Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., vol. 137, no. 653, 1999, viii+66 pp.

[24] T. Champion and L. De Pascale, "The Monge problem in $\mathbb{R}^d$", *Duke Math. J.* **157**:3 (2011), 551–572.

[25] В. И. Богачев, "О работах Г. М. Фихтенгольца по теории интеграла", Истор.-матем. исслед. Сер. 2, no. 9(44), 2005, с. 252–264. [V. I. Bogachev, "On G. M. Fichtenholz's work on the theory of the integral", Istor.-Mat. Issled. (2), no. 9(44), 2005, pp. 252–264.]

[26] Л. В. Канторович, *Математические методы организации и планирования производства*, Изд-во ЛГУ, Л. 1939, 67 с.; репринтное изд.:, Изд. дом СПбГУ, СПб. 2012. [L. V. Kantorovich, *Mathematical methods for manufacturing, organizing and planning*, Leningrad State University Publishing House, Leningrad 1939, 67 pp.]; Reprint of the 1939 edition, St. Petersburg University Publishing House, St. Petersburg 2012.

[27] А. М. Вершик, "Многозначные отображения с инвариантной мерой (полиморфизмы) и марковские операторы", *Проблемы теории*

вероятностных распределений. IV, Зап. научн. сем. ЛОМИ, **72**, Наука, Л. 1977, с. 26–61; English transl., A. M. Vershik, "Many-valued measure-preserving mappings (polymorphisms) and Markovian operators", *J. Soviet Math.* **23**:3 (1983), 2243–2266.

[28] В. Л. Левин, "О теоремах двойственности в задаче Монжа–Канторовича", *УМН* **32**:3(195) (1977), 171–172. [V. L. Levin, "Duality theorems in the Monge–Kantorovich problem", *Uspekhi Mat. Nauk* **32**:3(195) (1977), 171–172.]

[29] В. И. Богачев, *Основы теории меры*, т. 1, 2, 2-е изд., НИЦ "Регулярная и хаотическая динамика", М.–Ижевск 2006, 584, 680 с.; English transl., V. I. Bogachev, *Measure theory*, vols. I, II, Springer-Verlag, Berlin 2007, xviii+500, xiv+575 pp.

[30] R. Fortet and E. Mourier, "Convergence de la répartition empirique vers la répartition théorique", *Ann. Sci. Ecole Norm. Sup.* (3) **70**:3 (1953), 267–285.

[31] R. F. Arens and J. Eells, Jr., "On embedding uniform and topological spaces", *Pacif. J. Math.* **6** (1956), 397–403.

[32] D. A. Edwards, "On the Kantorovich–Rubinshtein theorem", *Expos. Math.* **29**:4 (2011), 387–398.

[33] A. M. Vershik, "Dynamics of metrics in measure spaces and their asymptotic invariants", *Markov Process. Related Fields* **16**:1 (2010), 169–184.

[34] J. Melleray, F. V. Petrov, and A. M. Vershik, "Linearly rigid metric spaces and the embedding problem", *Fund. Math.* **199**:2 (2008), 177–194.

[35] В. М. Золотарев, "О непрерывности стохастических последовательностей, порождаемых рекуррентными процедурами", *Теория вероятн. и ее примен.* **20**:4 (1975), 834–847; English transl., V. M. Zolotarev, "On the continuity of stochastic sequences generated by recurrent processes", *Theory Probab. Appl.* **20**:4 (1976), 819–832.

[36] В. М. Золотарев, "Метрические расстояния в пространствах случайных величин и их распределений", *Матем. сб.* **101(143)**:3(11) (1976), 416–454; English transl., V. M. Zolotarev, "Metric distances in spaces of random variables and their distributions", *Math. USSR-Sb.* **30**:3 (1976), 373–401.

[37] С. Т. Рачев, "Задача Монжа–Канторовича о перемещении масс и ее применения в стохастике", *Теория вероятн. и ее примен.* **29**:4 (1984), 625–653; English transl., S. T. Rachev, "The Monge–Kantorovich mass transference problem and its stochastic applications", *Theory Probab. Appl.* **29**:4 (1985), 647–676.

[38] Р. Л. Добрушин, "Задание системы случайных величин при помощи условных распределений", *Теория вероятн. и ее примен.* **15**:3 (1970), 469–497; English transl., R. L. Dobrushin, "Prescribing a system of random variables by conditional distributions", *Theory Probab. Appl.* **15**:3 (1970), 458–486.

[39] Л. Н. Васерштейн, "Марковские процессы на счетном произведении пространства, описывающие большие системы автоматов", *Пробл. передачи информ.* **5**:3 (1969), 64–72; English transl., L. N. Vasershtein, "Markov processes over denumerable products of spaces describing large system of automata", *Problems Inform. Transmission* **5**:3 (1969), 47–52.

[40] Y. Olivier, "A survey of Ricci curvature for metric spaces and Markov chains", *Probabilistic approach to geometry*, Adv. Stud. Pure Math., vol. 57, Math. Soc. Japan, Tokyo 2010, pp. 343–381.

[41] L. Rüschendorf, "On c-optimal random variables", *Statist. Probab. Lett.* **27**:3 (1996), 267–270.

[42] M. Knott and C. S. Smith, "On the optimal mapping of distributions", *J. Optim. Theory Appl.* **43**:1 (1984), 39–49.

[43] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer, "Optimal and better transport plans", *J. Funct. Anal.* **256**:6 (2009), 1907–1927.

[44] А. Ю. Плахов, "Точные решения одномерной задачи Монжа–Канторовича", *Матем. сб.* **195**:9 (2004), 57–74; English transl., A. Yu. Plakhov, "Precise solutions of the one-dimensional Monge–Kantorovich problem", *Sb. Math.* **195**:9 (2004), 1291–1307.

[45] В. Л. Левин, "Условия оптимальности и точные решения двумерной задачи Монжа–Канторовича", *Теория представлений, динамические системы.* XI, Специальный выпуск, Зап. научн. сем. ПОМИ, **312**, ПОМИ, СПб. 2004, с. 150–164; English transl., V. L. Levin, "Optimality conditions and exact solutions to the two-dimensional Monge–Kantorovich problem", *J. Math. Sci. (N. Y.)* **133**:4 (2006), 1456–1463.

[46] V. L. Levin, "General Monge–Kantorovich problem and its applications in measure theory and mathematical economics", *Functional analysis, optimization, and mathematical economics*, A collection of papers dedicated to the memory of L. V. Kantorovich (L. J. Leifman, ed.), Oxford Univ. Press, New York 1990, pp. 141–176.

[47] D. Ramachandran and L. Rüschendorf, "A general duality theorem for marginal problems", *Probab. Theory Relat. Fields* **101**:3 (1995), 311–319.

[48] D. Ramachandran and L. Rüschendorf, "Duality and perfect probability spaces", *Proc. Amer. Math. Soc.* **124**:7 (1996), 2223–2228.

[49] D. Ramachandran, L. Rüschendorf, "On the Monge–Kantorovich duality theorem", *Теория вероятн. и ее примен.* **45**:2 (2000), 403–409; also published as, D. Ramachandran and L. Rüschendorf, "On the Monge–Kantorovich duality theorem", *Theory Probab. Appl.* **45**:2 (2001), 350–356.

[50] M. Beiglböck and W. Schachermayer, "Duality for Borel measurable cost functions", *Trans. Amer. Math. Soc.* **363**:8 (2011), 4203–4224.

[51] H. G. Kellerer, "Duality theorems for marginal problems", *Z. Wahrscheinlichkeitstheor. verw. Geb.* **67**:4 (1984), 399–432.

[52] D. A. Edwards, "A simple proof in Monge–Kantorovich duality theory", *Studia Math.* **200**:1 (2010), 67–77.

[53] L. Ambrosio and A. Pratelli, "Existence and stability results in the L^1 theory of optimal transportation", *Optimal transportation and applications* (Martina Franca 2001), Lecture Notes in Math., vol. 1813, Springer, Berlin 2003, pp. 123–160.

[54] В. Л. Левин, "Формула для оптимального значения задачи Монжа–Канторовича с гладкой функцией стоимости и характеризация циклически монотонных отображений", *Матем. сб.* **181**:12 (1990), 1694–1709; English transl., V. L. Levin, "A formula for the optimal value in the Monge–Kantorovich problem with a smooth cost function, and a characterization of cyclically monotone mappings", *Math. USSR-Sb.* **71**:2 (1992), 533–548.

[55] В. Л. Левин, "К теории двойственности для нетопологических вариантов задачи о перемещении масс", *Матем. сб.* **188**:4 (1997), 95–126; English transl., V. L. Levin, "On duality theory for non-topological variants of the mass transfer problem", *Sb. Math.* **188**:4 (1997), 571–602.

[56] В. Л. Левин, "Двойственность Монжа–Канторовича и ее применение в теории полезности", *Экономика и матем. методы* **47**:4 (2011), 143–165. [V. L. Levin, "Monge–Kantorovich duality theory and its application in utility theory", *Economics and Mathematical Methods* **47**:4 (2011), 143–165.]

[57] V. Levin, "Abstract cyclical monotonicity and Monge solutions for the general Monge–Kantorovich problem", *Set-Valued Anal.* **7**:1 (1999), 7–32.

[58] В. Л. Левин, А. А. Милютин, “Задача о перемещении масс с разрывной функцией стоимости и массовая постановка проблемы двойственности выпуклых экстремальных задач”, *УМН* **34**:3(207) (1979), 3–68; English transl., V. L. Levin and A. A. Milyutin, “The problem of mass transfer with a discontinuous cost function and a mass statement of the duality problem for convex extremal problems”, *Russian Math. Surveys* **34**:3 (1979), 1–78.

[59] L. A. Caffarelli, M. Feldman, and R. J. McCann, “Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs”, *J. Amer. Math. Soc.* **15**:1 (2002), 1–26.

[60] N. S. Trudinger and X.-J. Wang, “On the Monge mass transfer problem”, *Calc. Var. Partial Differential Equations* **13**:1 (2001), 19–31.

[61] T. Champion and L. De Pascale, “The Monge problem for strictly convex norms in $\mathbb{R}^d$”, *J. Eur. Math. Soc.* **12**:6 (2010), 1355–1369.

[62] L. Caravenna, “A proof of Sudakov theorem with strictly convex norms”, *Math. Z.* **268**:1-2 (2011), 371–407.

[63] А. В. Погорелов, *Многомерное уравнение Монжа–Ампера* $\det\|z_{ij}\| = \phi(z_1,\dots,z_n,z,x_1,\dots,x_n)$, Наука, М. 1988, 94 с. [A. V. Pogorelov, *Multidimensional Monge–Ampère equation* $\det\|z_{ij}\| = \phi(z_1,\dots,z_n,z,x_1,\dots,x_n)$, Nauka, Moscow 1988, 94 pp.]

[64] C. E. Gutiérrez, *The Monge–Ampère equation*, Progr. Nonlinear Differential Equations Appl., vol. 44, Birkhäuser Boston, Inc., Boston, MA 2001, xii+127 pp.

[65] Y. Brenier, “Polar factorization and monotone rearrangement of vector valued functions”, *Comm. Pure Appl. Math.* **44**:4 (1991), 375–417.

[66] R. J. McCann, “Existence and uniqueness of monotone measure-preserving maps”, *Duke Math. J.* **80**:2 (1995), 309–323.

[67] В. И. Богачев, *Дифференцируемые меры и исчисление Маллявэна*, НИЦ “Регулярная и хаотическая динамика”, М.–Ижевск 2008, 544 с.; English transl., V. I. Bogachev, *Differentiable measures and the Malliavin calculus*, Math. Surveys Monogr., vol. 164, Amer. Math. Soc., Providence, RI 2010, xvi+488 pp.

[68] R. J. McCann, “A convexity principle for interacting gases”, *Adv. Math.* **128**:1 (1997), 153–179.

[69] R. van der Putten, “Uniqueness and nonuniqueness of optimal maps in mass transport problem with not strictly convex cost”, *Comment. Math. Univ. Carolin.* **51**:1 (2010), 67–83.

[70] N. V. Krylov, “Fully nonlinear second order elliptic equations: recent development”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci* (4) **25**:3-4 (1997), 569–595.

[71] X. Cabré, “Elliptic PDE's in probability and geometry: symmetry and regularity of solutions”, *Discrete Contin. Dyn. Syst.* **20**:3 (2008), 425–457.

[72] N. S. Trudinger and X.-J. Wang, “The Monge–Ampère equation and its geometric applications”, *Handbook of geometric analysis*, № 1, Adv. Lect. Math. (ALM), vol. 7, Int. Press, Somerville, MA 2008, pp. 467–524.

[73] L. Forzani and D. Maldonaldo, “Properties of the solutions to the Monge–Ampère equation”, *Nonlinear Anal.* **57**:5-6 (2004), 815–829.

[74] G. De Philippis and A. Figalli, *$W^{2,1}$-regularity for solutions of the Monge–Ampère equation*, 2011, arXiv: 1111.7207v2.

[75] I. Fragalà, M. S. Gelli, and A. Pratelli, “Continuity of an optimal transport in Monge problem”, *J. Math. Pures Appl.* (9) **84**:9 (2005), 1261–1294.

[76] X.-J. Wang, “Some counterexamples to the regularity of Monge–Ampère equations”, *Proc. Amer. Math. Soc.* **123**:3 (1995), 841–845.

[77] L. A. Caffarelli, "Monotonicity properties of optimal transportation and the FKG and related inequalities", *Comm. Math. Phys.* **214**:3 (2000), 547–563.

[78] D. Bakry and M. Ledoux, "Lévy–Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator", *Invent. Math.* **123**:2 (1996), 259–281.

[79] D. Cordero-Erausquin, "Some applications of mass transport to Gaussian-type inequalities", *Arch. Ration. Mech. Anal.* **161**:3 (2002), 257–269.

[80] А. В. Колесников, "Глобальные гёльдеровы оценки для оптимальной транспортировки", *Матем. заметки* **88**:5 (2010), 708–728; English transl., A. V. Kolesnikov, "Global Hölder estimates for optimal transportation", *Math. Notes* **88**:5-6 (2010), 678–695.

[81] B. Klartag, *Poincaré inequalities and moment maps*, 2011, arXiv: 1104.2791.

[82] А. В. Колесников, "Соболевская регулярность транспортировки вероятностных мер и транспортные неравенства", *Теория вероятн. и ее примен.* **57**:2 (2012), 296–321. [A. V. Kolesnikov, "Sobolev regularity of the transportation of probability measures and transport inequalities", *Teor. Veroyatn. i Primenen.* **57**:2 (2012), 296–321.]

[83] В. И. Богачев, А. В. Колесников, "Соболевская регулярность для бесконечномерного уравнения Монжа–Ампера", *Докл. РАН* **444**:2 (2012), 131–136; English transl., V. I. Bogachev and A. V. Kolesnikov, "Sobolev regularity for infinite-dimensional Monge–Ampère equation", *Dokl. Math.* **85** (2012), 331–335.

[84] V. I. Bogachev and A. V. Kolesnikov, "Sobolev regularity for the Monge–Ampere equation in the Wiener space", *Kyoto J. Math.* (to appear); 2011, arXiv: 1110.1822.

[85] L. De Pascale, L. C. Evans, and A. Pratelli, "Integral estimates for transport densities", *Bull. London Math. Soc.* **36**:3 (2004), 383–395.

[86] G. Loeper, "On the regularity of solutions of optimal transportation problems", *Acta Math.* **202**:2 (2009), 241–283.

[87] J. Liu, N. S. Trudinger, and X.-J. Wang, "Interior $C^{2,\alpha}$ regularity for potential functions in optimal transportation", *Comm. Partial Differential Equations* **35**:1 (2010), 165–184.

[88] G. Loeper and C. Villani, "Regularity of optimal transport in curved geometry: the nonfocal case", *Duke Math. J.* **151**:3 (2010), 431–485.

[89] L. Caffarelli and R. J. McCann, "Free boundaries in optimal transport and Monge–Ampère obstacle problems", *Ann. of Math.* (2) **171**:2 (2010), 673–730.

[90] A. Figalli, "The optimal partial transport problem", *Arch. Ration. Mech. Anal.* **195**:2 (2010), 533–560.

[91] A. Figalli and Y.-H. Kim, "Partial regularity of Brenier solutions of the Monge–Ampère equation", *Discrete Contin. Dyn. Syst.* **28**:2 (2010), 559–565.

[92] А. А. Липчюс, "Замечание о равенстве в задачах Монжа и Канторовича", *Теория вероятн. и ее примен.* **50**:4 (2005), 779–782; English transl., A. A. Lipchius, "A note on the equality in the Monge and Kantorovich problems", *Theory Probab. Appl.* **50**:4 (2006), 689–693.

[93] A. Pratelli, "On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation", *Ann. Inst. H. Poincaré Probab. Statist.* **43**:1 (2007), 1–13.

[94] V. D. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, with an appendix by M. Gromov, Lecture Notes in Math., vol. 1200, Springer-Verlag, Berlin 1986, viii+156 pp.

[95] D. Cordero-Erausquin, B. Nazaret, and C. Villani, “A mass-transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities”, *Adv. Math.* **182**:2 (2004), 307–332.

[96] B. Nazaret, “Best constant in Sobolev trace inequalities on the half space”, *Nonlinear Anal.* **65**:10 (2006), 1977–1985.

[97] F. Maggi and C. Villani, “Balls have the worst best Sobolev inequalities”, *J. Geom. Anal.* **15**:1 (2005), 83–121; “Balls have the worst best Sobolev inequalities. II: Variants and extensions”, *Calc. Var. Partial Differential Equations* **31**:1 (2008), 47–74.

[98] A. Figalli, F. Maggi, and A. Pratelli, “A mass transportation approach to quantitative isoperimetric inequalities”, *Invent. Math.* **182**:1 (2010), 167–211.

[99] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, “A Riemannian interpolation inequality à la Borell, Brascamp and Lieb”, *Invent. Math.* **146**:2 (2001), 219–257.

[100] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, “Prékopa–Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport”, *Ann. Fac. Sci. Toulouse Math.* (6) **15**:4 (2006), 613–635.

[101] P. Castillon, “Submanifolds, isoperimetric inequalities and optimal transportation”, *J. Funct. Anal.* **259**:1 (2010), 79–103; 2009, arXiv: 0908.2711.

[102] L. Gross, “Logarithmic Sobolev inequalities”, *Amer. J. Math.* **97**:4 (1975), 1061–1083.

[103] M. Ledoux, *The concentration of measure phenomenon*, Math. Surveys Monogr., vol. 89, Amer. Math. Soc., Providence, RI 2001, x+181 pp.

[104] В. И. Богачев, *Гауссовские меры*, Наука, М. 1997, 352 с.; English transl., V. I. Bogachev, *Gaussian measures*, Amer. Math. Soc., Providence, RI 1998, xii+433 pp.

[105] S. G. Bobkov and F. Götze, “Exponential integrability and transportation cost related to logarithmic Sobolev inequality”, *J. Funct. Anal.* **163** (1999), 1–28.

[106] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, 2002, arXiv: math/0211159.

[107] В. Н. Судаков, Б. С. Цирельсон, “Экстремальные свойства полупространств для сферически инвариантных мер”, *Проблемы теории вероятностных распределений*. II, Зап. научн. сем. ЛОМИ, **41**, Наука, Л. 1974, с. 14–24; V. N. Sudakov and B. S. Tsirelson, “Extremal properties of half-spaces for spherically invariant measures”, *J. Math. Sci.* **9**:1 (1978), 9-18.

[108] C. Borell, “The Brunn–Minkowski inequality in Gauss space”, *Invent. Math.* **30**:2 (1975), 207–216.

[109] F.-Y. Wang, “Logarithmic Sobolev inequalities on non-compact Riemannian manifolds”, *Probab. Theory Related Fields* **109**:3 (1997), 417–424.

[110] F. Barthe and A. V. Kolesnikov, “Mass transport and variants of the logarithmic Sobolev inequality”, *J. Geom. Anal.* **18**:4 (2008), 921–979.

[111] O. S. Rothaus, “Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities”, *J. Funct. Anal.* **64**:2 (1985), 296–313.

[112] F. Barthe, P. Cattiaux, and C. Roberto, “Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry”, *Rev. Mat. Iberoam.* **22**:3 (2006), 993–1067.

[113] D. Cordero-Erausquin, W. Gangbo, and C. Houdré, “Inequalities for generalized entropies and optimal transportation”, *Recent advances in the theory and applications of mass transport*, Contemp. Math., vol. 353, Amer. Math. Soc., Providence, RI 2004, pp. 73–94.

[114] E. Milman and S. Sodin, “An isoperimetric inequality for uniformly log-concave measures and uniformly convex bodies”, *J. Funct. Anal.* **254**:5 (2008), 1235–1268.

[115] S. G. Bobkov, “Isoperimetric and analytic inequalities for log-concave probability measures”, *Ann. Probab.* **27**:4 (1999), 1903–1921.

[116] A. V. Kolesnikov, “Modified log-Sobolev inequalities and isoperimetry”, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei* (9) *Mat. Appl.* **18**:2 (2007), 179–208.

[117] M. Talagrand, “Transportation cost for Gaussian and other product measures”, *Geom. Funct. Anal.* **6**:3 (1996), 587–600.

[118] В. И. Богачев, А. В. Колесников, “Интегрируемость абсолютно непрерывных преобразований мер и применения к оптимальному переносу масс”, *Теория вероятн. и ее примен.* **50**:3 (2005), 433–456; English transl., V. I. Bogachev and A. V. Kolesnikov, “Integrability of absolutely continuous transformations of measures and applications to optimal mass transportation”, *Theory Probab. Appl.* **50**:3 (2006), 367–385.

[119] F. Otto and C. Villani, “Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality”, *J. Funct. Anal.* **173**:2 (2000), 361–400.

[120] S. G. Bobkov, I. Gentil, and M. Ledoux, “Hypercontractivity of Hamilton–Jacobi equations”, *J. Math. Pures Appl.* (9) **80**:7 (2001), 669–696.

[121] P. Cattiaux and A. Guillin, “On quadratic transportation cost inequalities”, *J. Math. Pures Appl.* (9) **86**:4 (2006), 342–361.

[122] N. Gozlan, “Characterization of Talagrand’s like transportation-cost inequalities on the real line”, *J. Funct. Anal.* **250**:2 (2007), 400–425.

[123] N. Gozlan, “A characterization of dimension free concentration in terms of transportation inequalities”, *Ann. Probab.* **37**:6 (2009), 2480–2498.

[124] I. Gentil, A. Guillin, and L. Miclo, “Modified logarithmic Sobolev inequalities and transportation inequalities”, *Probab. Theory Related Fields* **133**:3 (2005), 409–436.

[125] H. Djellout, A. Guillin, and L. Wu, “Transportation cost-information inequalities and applications to random dynamical systems and diffusions”, *Ann. Probab.* **32**:3B (2004), 2702–2732.

[126] F. Bolley and C. Villani, “Weighted Csiszár–Kullback–Pinsker inequalities and application to transportation inequalities”, *Ann. Fac. Sci. Toulouse Math.* (6) **14**:3 (2005), 331–352.

[127] N. Gozlan, C. Roberto, and P. M. Samson, “A new characterization of Talagrand’s transport-entropy inequalities and applications”, *Ann. Probab.* **39**:3 (2011), 857–880.

[128] P. Cattiaux, A. Guillin, and L.-M. Wu, “A note on Talagrand’s transportation inequality and logarithmic Sobolev inequality”, *Probab. Theory Related Fields* **148**:1-2 (2010), 285–304.

[129] A. Guillin, C. Léonard, L. Wu, and N. Yao, “Transportation-information inequalities for Markov processes”, *Probab. Theory Related Fields* **144**:3-4 (2009), 669–695.

[130] A. Guillin, C. Leonard, F.-Y. Wang, and L. Wu, *Transportation-information inequalities for Markov processes* (II): *relations with other functional inequalities*, 2009, arXiv: 0902.2101.

[131] N. Gozlan and C. Léonard, “Transport inequalities. A survey”, *Markov Process. Related Fields* **16**:4 (2010), 635–736.

[132] J. Dolbeault, B. Nazaret, and G. Savaré, “A new class of transport distances between measures”, *Calc. Var. Partial Differential Equations* **34**:2 (2009), 193–231.

[133] L. Gozlan and C. Léonard, "A large deviation approach to some transportation cost inequalities", *Probab. Theory Related Fields* **139**:1-2 (2007), 235–283.

[134] E. Milman, "Isoperimetric and concentration inequalities: equivalence under curvature lower bound", *Duke Math. J.* **154**:2 (2010), 207–239.

[135] E. Milman, "Properties of isoperimetric, functional and transport-entropy inequalities via concentration", *Probab. Theory Related Fields* **152**:3-4 (2012), 475–507.

[136] J. Lott and C. Villani, "Ricci curvature for metric-measure spaces via optimal transport", *Ann. Math.* (2) **169**:3 (2009), 903–991.

[137] L. Ambrosio and A. Figalli, "Geodesics in the space of measure-preserving maps and plans", *Arch. Ration. Mech. Anal.* **194**:2 (2009), 421–462.

[138] S. Angenent, S. Haker, and A. Tannenbaum, "Minimizing flows for the Monge–Kantorovich problem", *SIAM J. Math. Anal.* **35**:1 (2003), 61–97.

[139] L. De Pascale, M. S. Gelli, and L. Granieri, "Minimal measures, one-dimensional currents and the Monge–Kantorovich problem", *Calc. Var. Partial Differential Equations* **27**:1 (2006), 1–23.

[140] A. Figalli and N. Gigli, "A new transportation distance between non-negative measures, with applications to gradient flows with Dirichlet boundary conditions", *J. Math. Pures Appl.* (9) **94**:2 (2010), 107–130.

[141] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem", *Numer. Math.* **84**:3 (2000), 375–393.

[142] N. Juillet, "On displacement interpolation of measures involved in Brenier's theorem", *Proc. Amer. Math. Soc.* **139**:10 (2011), 3623–3632.

[143] K.-Th. Sturm, "On the geometry of metric measure spaces, I, II", *Acta Math.* **196** (2006), 65–131, 133–177.

[144] J. Lott, "Some geometric calculations on Wasserstein spaces", *Comm. Math. Phys.* **277**:2 (2008), 423–437.

[145] F. Otto, "The geometry of dissipative evolution equations: the porous medium equation", *Comm. Partial Differential Equations* **26**:1-2 (2001), 101–174.

[146] F. Otto and M. Westdickenberg, "Eulerian calculus for the contraction in the Wasserstein distance", *SIAM J. Math. Anal.* **37**:4 (2005), 1227–1255 (electronic).

[147] U. Gianazza, G. Savaré, and G. Toscani, "The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation", *Arch. Ration. Mech. Anal.* **194**:1 (2009), 133–220.

[148] Р. Л. Добрушин, "Уравнения Власова", *Функц. анализ и его прил.* **13**:2 (1979), 48–58; English transl., R. L. Dobrushin, "Vlasov equations", *Funct. Anal. Appl.* **13**:2 (1979), 115–123.

[149] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation", *SIAM J. Math. Anal.* **29**:1 (1998), 1–17.

[150] L. Natile, M. A. Peletier, and G. Savaré, "Contraction of general transportation costs along solutions to Fokker–Planck equations with monotone drifts", *J. Math. Pures Appl.* (9) **95**:1 (2011), 18–35.

[151] K. Kuwada, "Duality on gradient estimates and Wasserstein controls", *J. Funct. Anal.* **258**:11 (2010), 3758–3774.

[152] M.-K. von Renesse and K.-Th. Sturm, "Entropic measure and Wasserstein diffusion", *Ann. Probab.* **37**:3 (2009), 1114–1191.

[153] А. Д. Александров, *Избранные труды*, т. 1, 2, Наука, Новосибирск 2006, 2007, lii+748, iv+491 с. [A. D. Aleksandrov, *Selected works*, vols. 1, 2, Nauka, Novosibirsk 2006, 2007, lii+748, iv+491 pp.]

[154] V. Oliker, “Embeddings of S^n into $\mathbb{R}^{n+1}$ with given integral Gauss curvature and optimal mass transport on S^n”, *Adv. Math.* **213**:2 (2007), 600–620.

[155] R. J. McCann, “Polar factorization of maps on Riemannian manifolds”, *Geom. Funct. Anal.* **11**:3 (2001), 589–608.

[156] A. Fathi and A. Figalli, “Optimal transportation on non-compact manifolds”, *Israel J. Math.* **175** (2010), 1–59.

[157] A. Figalli and L. Rifford, “Mass transportation on sub-Riemannian manifolds”, *Geom. Funct. Anal.* **20**:1 (2010), 124–159.

[158] Y.-H. Kim and R. J. McCann, “Continuity, curvature, and the general covariance of optimal transportation”, *J. Eur. Math. Soc.* **12**:4 (2010), 1009–1040.

[159] J.-P. Bourguignon, “Ricci curvature and measures”, *Japan. J. Math.* **4**:1 (2009), 27–45.

[160] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progr. Math., vol. 152, Birkhäuser Boston, Inc., Boston, MA 1999, xx+585 pp.; translated from the French original, M. Gromov, *Structures métriques pour les variétés riemanniennes*, Textes Math., vol. 1 (J. Lafontaine and P. Pansu, eds.), CEDIC, Paris 1981, iv+152 pp.

[161] А. М. Вершик, “Универсальное пространство Урысона, метрические тройки Громова и случайные метрики на натуральном ряде”, *УМН* **53**:5(323) (1998), 57–64; English transl., A. M. Vershik, “The universal Urysohn space, Gromov metric triples and random metrics on the natural numbers”, *Russian Math. Surveys* **53**:5 (1998), 921–928.

[162] А. М. Вершик, “Случайные метрические пространства и универсальность”, *УМН* **59**:2(356) (2004), 65–104; English transl., A. M. Vershik, “Random metric spaces and universality”, *Russian Math. Surveys* **59**:2 (2004), 259–295.

[163] K. Bacher, “On Borell–Brascamp–Lieb inequalities on metric measure spaces”, *Potential Anal.* **33**:1 (2010), 1–15.

[164] K. Bacher and K.-Th. Sturm, “Localization and tensorization properties of the curvature-dimension condition for metric measure spaces”, *J. Funct. Anal.* **259**:1 (2010), 28–56.

[165] Q. Deng and K.-Th. Sturm, “Localization and tensorization properties of the curvature-dimension condition for metric measure spaces, II”, *J. Funct. Anal.* **260**:12 (2011), 3718–3725.

[166] D. Bakry and M. Émery, “Diffusions hypercontractives”, *Séminaire de probabilités*, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin 1985, pp. 177–206.

[167] V. I. Bogachev and A. V. Kolesnikov, “Mass transport generated by a flow of Gauss maps”, *J. Funct. Anal.* **256**:3 (2009), 940–957.

[168] G. Huisken, “Flow by mean curvature of convex surfaces into spheres”, *J. Differential Geom.* **20**:1 (1984), 237–266.

[169] K. Tso, “Deforming a hypersurfaces by its Gauss–Kronecker curvature”, *Comm. Pure Appl. Math.* **38**:6 (1985), 867–882.

[170] A. V. Kolesnikov, *Weak regularity of Gauss mass transport*, 2009, arXiv: 0904.1852.

[171] Y.-H. Kim, J. Streets, and M. Warren, *Parabolic optimal transport equations on manifolds*, 2010, arXiv: 1008.3824v1.

[172] R. J. McCann and P. M. Topping, “Ricci flow, entropy and optimal transportation”, *Amer. J. Math.* **132**:3 (2010), 711–730.

[173] P. Topping, *Ricci flows: the foundations via optimal transportation*, http://homepages.warwick.ac.uk/~maseq/grenoble_20100324.pdf.

[174] P. Topping, "$\mathscr{L}$-optimal transportation for Ricci flows", *J. Reine Angew. Math.* **636** (2009), 93–122.

[175] J. Lott, "Optimal transport and Perelman's reduced volume", *Calc. Var. Partial Differential Equations* **36**:1 (2009), 49–84.

[176] G. Buttazzo, A. Pratelli, S. Solimini, and E. Stepanov, *Optimal urban networks via mass transportation*, Lecture Notes in Math., vol. 1961, Springer-Verlag, Berlin 2009, x+150 pp.

[177] M. Bernot, V. Caselles, and J.-M. Morel, *Optimal transportation networks. Models and theory*, Lecture Notes in Math., vol. 1955, Springer-Verlag, Berlin 2009, x+200 pp.

[178] A. S. Üstünel and M. Zakai, *Transformation of measure on Wiener space*, Springer Monogr. Math., Springer-Verlag, Berlin 2000, xiv+296 pp.

[179] В. И. Богачев, А. В. Колесников, К. В. Медведев, "Треугольные преобразования мер", *Матем. сб.* **196**:3 (2005), 3–30; English transl., V. I. Bogachev, A. V. Kolesnikov, and K. V. Medvedev, "Triangular transformations of measures", *Sb. Math.* **196**:3-4 (2005), 309–335.

[180] D. Feyel and A. S. Üstünel, "Monge–Kantorovitch measure transportation and Monge–Ampère equation on Wiener space", *Probab. Theory Related Fields* **128**:3 (2004), 347–385.

[181] D. Feyel and A. S. Üstünel, "Solution of the Monge–Ampère equation on Wiener space for general log-concave measures", *J. Funct. Anal.* **232**:1 (2006), 29–55.

[182] D. Feyel and A. S. Üstünel, "The notions of convexity and concavity on Wiener space", *J. Funct. Anal.* **176**:2 (2000), 400–428.

[183] A. V. Kolesnikov, "Convexity inequalities and optimal transport of infinite-dimensional measures", *J. Math. Pures Appl.* (9) **83**:11 (2004), 1373–1404.

[184] F. Cavalletti, "The Monge problem in Wiener space", *Calc. Var. Partial Differential Equations* **45**:1-2 (2012), 101–124; 2011, arXiv: 1103.2798v2.

[185] V. I. Bogachev and A. V. Kolesnikov, "On the Monge–Ampère equation in infinite dimensions", *Infin. Dimen. Anal. Quantum Probab. Related Top.* **8**:4 (2005), 547–572.

[186] S. Fang, J. Shao, and K.-Th. Sturm, "Wasserstein space over the Wiener space", *Probab. Theory Related Fields* **146**:3-4 (2010), 535–565.

[187] L. Decreusefond, "Wasserstein distance on configuration space", *Potential Anal.* **28**:3 (2008), 283–300.

[188] В. И. Богачев, А. В. Колесников, "Нелинейные преобразования выпуклых мер", *Теория вероятн. и ее примен.* **50**:1 (2005), 27–51; English transl., V. I. Bogachev and A. V. Kolesnikov, "Nonlinear transformations of convex measures", *Theory Probab. Appl.* **50**:1 (2006), 34–52.

[189] S. G. Bobkov, "Large deviations via transference plans", *Advances in mathematics research*, vol. 2, Adv. Math. Res., vol. 2, Nova Sci. Publ., Hauppauge, NY 2003, pp. 151–175.

[190] G. Carlier, A. Galichon, and F. Santambrogio, "From Knothe's transport to Brenier's map and a continuation method for optimal transport", *SIAM J. Math. Anal.* **41**:6 (2009/10), 2554–2576.

[191] J. Moser, "On the volume elements on a manifold", *Trans. Amer. Math. Soc.* **120** (1965), 286–294.

[192] Y.-H. Kim and E. Milman, "A generalization of Caffarelli's contraction theorem via (reverse) heat flow", *Math. Ann.* **354**:3 (2012), 827–862.

[193] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev, "Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations", *Duke Math. J.* **156**:2 (2011), 229–271.

[194] P. Bernard, “Young measures, superposition and transport”, *Indiana Univ. Math. J.* **57**:1 (2008), 247–275.

[195] W. Gangbo and A. Święch, “Optimal maps for the multidimensional Monge–Kantorovich problem”, *Comm. Pure Appl. Math.* **51**:1 (1998), 23–45.

[196] W. Gangbo, H. K. Kim, and T. Pacini, *Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems*, Mem. Amer. Math. Soc., vol. 211, no. 993, 2011, vi+77 pp.

V. I. Bogachev
Moscow State University;
St. Tikhon’s Orthodox University
E-mail: vibogach@mail.ru

Received 20/JUN/12
Translated by V. BOGACHEV

A. V. Kolesnikov
Higher School of Economics
E-mail: sascha77@mail.ru