Monge–Kantorovich transportation problem and optimal couplings

Ludger Rüschendorf
Mathematical Stochastics, University of Freiburg, Germany

Abstract

- Mathematics Subject Classification: 60E15, 90C40, 49N15
- Keywords and Phrases: Optimal coupling, minimal metric, duality theory, polar factorization, c-convexity, stochastic equation, recursive algorithm.

The Monge–Kantorovich mass-transportation problem has been shown in recent years to be fundamental for various basic problems in analysis and geometry. In this paper we describe some of the historical developments of this problem and some of the basic results. In particular we emphasize the probabilistic aspects and contributions to this subject and its relevance for various classical and recent developments in probability theory ranging from probability metrics and functional inequalities over estimates for risk measures to the analysis of algorithms. The paper is based on a lecture of the author delivered at the MSRI meeting on mass transportation problems in November 2005 in Berkeley.

1 Introduction

The classical mass transportation problem of Monge and its version of Kantorovich has found a lot of recent interest because of its importance for several problems in nonlinear PDEs, Riemannian geometry, variational problems and for several interesting inequalities and concentration results, see in particular the recent excellent presentations of Ambrosio (2003), Villani (2003, 2006) and Ambrosio, Gigli, and Savaré (2005). In this paper we survey some of the probabilistic developments of the transportation problem and the related optimal coupling problem and its applications. The probabilistic development of the subject was mainly concentrated on the Kantorovich formulation of the problem which turned out to be also instrumental for the Monge formulation and its applications in analysis.

The probabilistic interest in this topic was essentially connected with some naturally defined minimal metrics on the space of probability measures which are defined via optimal coupling properties. In particular to mention are the minimal ℓ_p-metrics, the Kantorovich–Rubinstein theorem and others. Much of this development and many probabilistic applications are discussed in Rachev and Rüschendorf (1998a,b).

After the introduction of the connections between optimal couplings and mass transportation we discuss in section 3 the development of the basic duality theory which gives the clue to many of the optimal coupling results. We then present in section 4 as consequences some of the main results for optimal L^2-couplings (the classical L^2-distance) and also for general coupling functions. This includes in particular the important characterizations of optimal transportation plans based on generalized convexity notions (c-convexity, c-subgradients, c-cyclical monotonicity). At this point roughly around 1990 the development of this subject in analysis began. Here this survey tries to relate the various historical sources and to describe the probabilistic contributions. We also describe some of the more concrete probabilistic applications and developments as e.g. to the optimal coupling of normal or discrete distributions or to obtain bounds for the risk of portfolios arising from positive dependence. In the final part of this paper we discuss a recently introduced modification of the minimal ℓ_s-metric and its application in the analysis of recursive algorithms of divide-and-conquer type.

2 The mass transportation problem

In 1942 Kantorovich introduced the problem of *optimal mass transport* in the following form:

$$\int c(x,y)d\mu(x,y) = \inf_{\mu\in M(P_1,P_2)} =: \widehat{\mu}_c(P_1,P_2), \tag{2.1}$$

where $c : U_1 \times U_2 \to \mathbb{R}$ is a measurable real cost function, $P_i \in M^1(U_i)$ are probability measures on U_i and

$$M(P_1,P_2) = \{\mu \in M^1(U_1 \times U_2); \mu^{\pi_i} = P_i, \quad i = 1,2\} \tag{2.2}$$

is the class of all probability measures on $U_1 \times U_2$ with marginals P_1, P_2. Here π_i are the projections on the i-th components and μ^{π_i} is the image of μ under π_i. $\widehat{\mu}_c$ is called the *Monge–Kantorovich functional.* In terms of random variables on a non-atomic probability space $(\Omega, \mathfrak{A}, P)$ problem (2.1) is equivalent with the problem to find an *optimal coupling* of P_1, P_2 w.r.t. the coupling function c, i.e.

$$Ec(X_1,X_2) = \inf \tag{2.3}$$

over all couplings X_1, X_2 of P_1, P_2, i.e. such that $P^{X_i} = P_i$, $i = 1, 2$. Any $\mu \in M(P_1, P_2)$ describes a transference (transportation) plan for the mass distribution P_1 to P_2 or equivalently the joint distribution of a pair of (X_1, X_2) of couplings of P_1, P_2. Using conditional distributions we obtain for $\mu \in M(P_1, P_2)$

$$\int c(x,y)d\mu(x,y) = \int \left(\int c(x,y)\mu(dy|x) \right) P_1(dx). \tag{2.4}$$

Any mass at point x is transported to y according to $\mu(dy|x)$ and thus $\int c(x,y)d\mu(x,y)$ denotes the total cost of transportation using this plan. In the optimal coupling problem (2.3) c is understood as a distance (dissimilarity) and it is a natural problem to find an coupling (X_1, X_2) of P_1, P_2 with minimal expected dissimilarity.

A subclass of all transport plans are deterministic transport plans of the form $\mu(\cdot \mid x) = \varepsilon_{\phi(x)}$, where ϕ is a function which transports P_1 to P_2, i.e. $P_1^\phi = P_2$. The additional restriction is that no mass is allowed to be split. Denoting by

$$S(P_1, P_2) = \{\phi : U_1 \to U_2, \phi \text{ measurable}, P_1^\phi = P_2\} \tag{2.5}$$

the set of all deterministic transport plans one obtains the corresponding *Monge transportation problem*

$$\int c(x, \phi(x))dP_1(x) = \inf_{\phi \in S(P_1, P_2)} \tag{2.6}$$

resp. the deterministic coupling problem

$$Ec(X_1, \phi(X_1)) = \inf_{\phi \in S(P_1, P_2)}. \tag{2.7}$$

This problem was introduced in 1781 by Monge for the special case that $U_i \subset \mathbb{R}^3$ are two bounded domains with volume measures P_i and $c(x,y) = \|x - y\|$ is the Euclidean distance. Monge detected that optimal transport should go along straight lines which are orthogonal to a family of surfaces (formally worked out by Appel (1887)). Also he found the no-crossing rule of optimal transport rays.

From the probabilistic point of view the Kantorovich formulation of the transport problem is more ‘natural’ than the Monge formulation. Similar extensions of deterministic optimization problems are quite often to find in probability and statistics as e.g. the transition from deterministic decision rules (like deterministic tests and estimators) to randomized decision rules (like randomized tests and estimators) is a classical example from the early period of statistics. Kantorovich obviously was not aware of the Monge problem when he formulated his transport problem in 1942. In 1948 he wrote a short note of three pages where he made the connection to the Monge problem and stated that in case (2.1) has a deterministic solution ϕ, then ϕ is also a solution of the Monge problem, i.e., the Kantorovich problem is a relaxation of the Monge problem. In fact Kantorovich’s problem was one of the earliest infinite dimensional linear programming problems considered. In 1975 Kantorovich got together with Koopman the Nobel price in economics for his development of linear programming and the application to mathematical economics.

The Kantorovich problem and some variants have been developed in the probabilistic literature since the mid seventies. For various kinds of optimization problems they have been established as a basic and natural tool. A detailed exposition with many applications of these developments is given in Rachev and Rüschendorf (1998a,b). Starting with the late eighties, early nineties, important connections of the transportation problem with problems from analysis and geometry, partial differential equations, fluid mechanics, general curvature theory, variational problems, geometric and functional inequalities like isoperimetric and concentration inequalities, gradient flows in metric spaces, and many others have been detected. This lead to a very active and wide ranged research area. This line of research is excellently described and developed in the books of Ambrosio (2003), Villani (2003, 2006), and Ambrosio, Gigli, and Savaré (2005).

In the following we review some of the history of the probabilistic development of the transport problem, put it into line with the developments described above in analysis. In this way we obtain e.g. a new extension of Brenier's polar factorization result. Finally, we point out to some of the more recent applications in various areas of probabilistic analysis.

3 Duality theory and optimal couplings

In this section we describe developments of the duality theory which is the main tool and the basis for determining optimal couplings and transport plans. In some more recent work starting with McCann (1995) and Gangbo and McCann (1996) more direct methods have been developed to determine optimal transport plans. We begin this section with stating some of the classical results on minimal probability metrics which stand at the beginning of optimal transportation problems.

3.1 Minimal probability metrics

a) Minimal ℓ_1-metric. Let (U,d) be a separable metric space and $P_1, P_2 \in M^1(U)$ be probability measures on U with its Borel σ-Algebra. We denote the *minimal ℓ_1-metric* on $M^1(U)$ by

$$\ell_1(P_1,P_2) := \inf\left\{ \int d(x,y)d\mu(x,y); \quad \mu \in M(P_1,P_2)\right\} \tag{3.1}$$

i.e. ℓ_1 is the minimal version of the usual L_1-metric

$$L_1(X,Y) = Ed(X,Y) \tag{3.2}$$

of random variables X, Y in U and is identical to the solution of the transportation problem with cost function $c = d$.

The *Lipschitz metric* μ_L is defined by

$$\mu_L(P_1,P_2) = \sup\left\{ \int f d(P_1 - P_2); \ \operatorname{Lip} f \le 1\right\}. \tag{3.3}$$

Kantorovich–Rubinstein Theorem: The minimal ℓ_1-metric is identical to the Lipschitz metric, i.e., for all $P_1, P_2 \in M^1(U)$ holds

$$\ell_1(P_1,P_2) = \mu_L(P_1,P_2). \tag{3.4}$$

This result was proved by Kantorovich and Rubinstein (1957) in the case of compact metric spaces and then extended by de Acosta (1982), Dudley (1976), Fernique (1981), Levin (1975), and Kellerer (1984a).

In the case of the real line $U = \mathbb{R}^1$ and $d(x,y) = |x-y|$ one gets the explicit expression

$$\begin{aligned}\ell_1(P_1,P_2) &= \int_0^1 |F_1^{-1}(u) - F_2^{-1}(u)|du \\ &= \int |F_1(x) - F_2(x)|dx\end{aligned} \tag{3.5}$$

where F_i are the distribution functions of P_i. In this case the results go back to early work of Gini (1914), Salvemini (1949), and Dall'Aglio (1956) (even for the case $c(x,y) = |x-y|^\alpha$) Vallander (1973) and Szulga (1978).

Fréchet (1940) was the first to note formally the metric properties of ℓ_1 in general metric spaces, Hoeffding (1940) gave a formula for ℓ_2 in the real case and Vasershtein (1969) 'introduced' ℓ_1 again in his paper on Markov processes. Dobrushin (1970) was the first to call ℓ_1 Wasserstein metric (Wasserstein the English transcripted version of Vasershtein).

b) Total variation metric. Let V denote the total variation metric on U and let

$$\sigma(X,Y) := P(X \neq Y) \tag{3.6}$$

denote the compound probability metric on the space of random variables, then

$$V(P_1,P_2) = \widehat{\sigma}(P_1,P_2) = \inf\{\sigma(X,Y); X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2\}. \tag{3.7}$$

This result is due to Dobrushin (1970). It is a basic result to many of the optimal coupling results in probability theory which extend the classical paper of Doeblin (1938) giving a coupling proof of the limit theorem for Markov chains.

c) Prohorov metric. A similar result holds true for the Prohorov metric π on $M^1(U)$ which is the classical metric for the weak convergence topology. Strassen (1965) proved that π is the minimal metric of the Ky Fan metric K, i.e.

$$\pi = \widehat{K}. \tag{3.8}$$

In all three cases the transition to the minimal metric yields a change of the topology. This has important applications, e.g. to Skorohod type results (relation between weak and strong convergence) in the proof of central limit theorems, in matching theory and in robustness results and many others. More generally for any compound probability metric $\mu(X,Y)$ the corresponding minimal metric $\widehat{\mu}$ is defined by

$$\widehat{\mu}(P_1,P_2) = \inf\{\mu(X,Y); X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2\}. \tag{3.9}$$

Zolotarev (1976) used this principle for the construction of several examples of ideal metrics. The interesting question of characterization of the minimal metrics with ζ-structure, i.e. which have a sup-representation similar to (3.3), is so far only partially answered.

3.2 Monge–Kantorovich duality theory

The duality theory for the transportation problem began with Kantorovich's 1942 result which stated equivalence of (2.1) with a dual problem for the case of compact metric spaces and continuous cost functions $c(x,y)$. The proof however worked only for the case where

$c(x,y) = d(x,y)$ is a metric on $U = U_1 = U_2$. For the metric case the MK-transportation problem is equivalent to the mass transfer problem where for $P_1, P_2 \in M^1(U)$ the class of transport plans $M(P_1, P_2)$ is replaced by the class of *mass transference plans*

$$M = M(P_1 - P_2) = \{\gamma \in M(U \times U);\ \ \gamma^{\pi_1} - \gamma^{\pi_2} = P_1 - P_2\}, \tag{3.10}$$

i.e. all transport plans with fixed difference of the marginals. With respect to this class it is allowed to transfer a mass point of x to y via some route $x = x_1, x_2, \ldots, x_n = y$ such that the cost $c(x,y)$ is replaced by the cost $\sum_{i=1}^n c(x_i, x_{i+1})$. The basic result is an extension of the Kantorovich–Rubinstein theorem of the form

$$\begin{aligned} \mathring{\mu}_c(P_1, P_2) &= \inf\left\{ \int c(x,y) d\mu(x,y); \mu \in M(P_1 - P_2)\right\} \\ &= \sup\left\{ \int_U f d(P_1 - P_2); f(x) - f(y) \le c(x,y)\right\}. \end{aligned} \tag{3.11}$$

After the Kantorovich–Rubinstein (1957) paper this kind of duality theorems for the mass transfer problem was intensively discussed in the Russian probability school in particular by Levin (1975) and Levin and Milyutin (1979). Also the papers of de Acosta (1982), Dudley (1976), Fernique (1981), and Rachev and Shortt (1990) concerned the Kantorovich–Rubinstein functional $\mathring{\mu}_c$. It coincides with the MK-functional $\widehat{\mu}_c$ only if c is a metric (see Neveu and Dudley (1980)). An important role in this development is played by the Lipschitz norm in (3.11) (see Fortet and Mourier (1953)) and by related approximation arguments. A very complete duality theory of the KR-functional $\mathring{\mu}_c$ has been developed by Levin (see corresponding references and presentation in Rachev and Rüschendorf (1998a,b)).

The MK-problem with fixed marginals can be considered also on n-fold products of probability spaces $(U_i, \mathfrak{A}_i, P_i)$. Let $h : \prod_{i=1}^n U_i \to \mathbb{R}$ and $M = M(P_1, \ldots, P_n)$ be the set of all transport plans, i.e. measures with marginals P_i then we define

$$\begin{aligned} S(h) &= \sup\left\{ \int h d\mu; \mu \in M\right\} \\ I(h) &= \inf\left\{ \sum_{i=1}^n \int f_i dP_i; h \le \oplus f_i, f_i \in L^2(P_i)\right\}, \end{aligned} \tag{3.12}$$

where $\oplus f_i(x) = \sum_{i=1}^n f_i(x_i)$. We say that *duality holds* if

$$S(h) = I(h). \tag{3.13}$$

Here the equivalent problem of maximizing the gain (profit) is considered which transfers to the problem of minimizing the cost by switching to $c = -h$. For the proof of duality theorems of MK-type several strategies have been developed. One approach is to establish via Hahn–Banach and Riesz-type results in the first step the equality

$$\widetilde{S}(h) = I(h) \tag{3.14}$$

where $\widetilde{S}(h)$ is the supremum problem where the measures with fixed marginals are relaxed to the finite additive measures $\mathrm{ba}(P_1, \ldots, P_n)$. In the second step conditions on h are identified (Riesz-type results) which ensure that $S(h) = \widetilde{S}(h)$. This approach was followed in Rüschendorf (1979–1981) (in the following abbreviated by Rü) and Gaffke and Rü (1981) (without being aware at that time of the MK-problem in the Russian school). Motivated by this development Kellerer (1984b) followed a different route to obtain more general results. Starting from the duality for simple cases he investigated in detail continuity properties of the functionals S, I which allowed him by Choquet's capacity theorem to obtain very general duality results. Rachev (1985) extended approximation arguments as used in the KR-case to some instances of the MK-problem. Levin (since 1984) established some techniques which allowed him to prove reduction results from the MK-problem to the KR-problem.

Here is a list of some of the basic duality results for the MK-duality problem. The spaces U_i are assumed to be Hausdorff and the measures are restricted to the class of tight measures.

Theorem 3.1 (Duality Theorem.) *a) Duality holds on the class of all lower-majorized product-measurable functions:*

$$\mathcal{L}_m(U) = \{h \in \mathcal{L}(\mathfrak{A}_1 \otimes \cdots \otimes \mathfrak{A}_n); \exists f_i \in L^1(P_i); h \geq \oplus_{i=1}^n f_i\}. \tag{3.15}$$

b) Duality holds on $\overline{F}(U)$, (3.16)

where $F(U)$ are the upper semicontinuous functions and closure is w.r.t. $I(|f-g|)$.

Duality also holds on $\overline{G}_m(U)$, (3.17)

the closure of the lower majorized lower semicontinuous functions.

c) Existence of an optimal measure on $\overline{F}$ for the S-functional. (3.18)

d) Existence of minimal functions (f_i) on $\mathcal{L}_m(U)$ for the I-functional. (3.19)

e) For $P \in M$, $f_i \in \mathcal{L}^1(P_i)$, $\oplus f_i \geq h$ holds:

$$P, (f_i) \text{ are solutions for } S, I \iff h = \oplus_{i=1}^n f_i \ \ [\mathrm{P}] \tag{3.20}$$

Remarks 3.2 a) The duality and existence results were proved in Rü (1981) essentially for the bounded product measurable case and in Kellerer (1984b) for the general case. The conditions in the results are sharp, i.e. there exist counter-examples of the duality and existence results, e.g. on $G(U)$ or without lower boundedness in d).

b) Condition (3.20) characterizes optimal transport plans P under the existence condition e.g. for $h \in \mathcal{L}_m(U)$. The sufficiency part does not need any conditions, i.e. the r.h.s. of (3.20) implies optimality of P and (f_i).

c) In Kellerer (1984b) a simple example is given where $h \in C(U \times U)$, $S(h) = I(h)$, $P_1 = P_2$, but h does not allow a representation of the form $h = f_1 \oplus f_2\,[P]$ with $f_i \in L^1(P_i)$, $P \in M(P_1, P_2)$.

d) The existence of solutions of the dual problem is closely connected with the following *closedness problem:* Let $P \in M(P_1, P_2)$, $s \geq 0$ and consider

$$\begin{aligned} F_s &= L^s(P_1) \oplus L^s(P_2) \\ &= \{f = f(x,y) = g(x) + h(y); g \in L^s(P_1), h \in L^s(P_2)\} \end{aligned} \tag{3.21}$$

When is F_s closed in $L^s(P)$? In general closedness does not hold true (see Rü and Thomsen (1993)). Several partial results are known, e.g. in case $s = 0$ and $P \ll P_1 \otimes P_2$ any element $\Phi \in \overline{F}_0$, the closure w.r.t. $L^0(P)$, has a representation of the form

$$\Phi(x,y) = f(x) + g(y) \ \ [P] \tag{3.22}$$

but in general f, g cannot be chosen measurable. Several positive results are established. (3.22) is sufficient for proving the existence of a general version of Schrödinger bridges and the positive results also allow to give an extension of the Kolmogorov representation result for continuous functions of n variables by a superposition of functions of one variable to the case of locally bounded measurable functions with equality holding a.s. (see Rü and Thomsen (1997)).

Immediately after establishing the duality results (3.15)–(3.20) some interesting consequences were established in the early eighties in particular sharpening some classical bounds. Here are some examples:

a) **Sharpness of Fréchet-bounds:** For $A_i \in \mathfrak{A}_i$ holds

$$\sup\{P(A_1 \times \cdots \times A_n); P \in M(P_1, \ldots, P_n)\} = \min\{P_i(A_i); 1 \le i \le n\} \tag{3.23}$$

$$\inf\{P(A_1 \times \cdots \times A_n); P \in M(P_1, \ldots, P_n)\} = \left(\sum_{i=1}^n P_i(A_i) - (n-1)\right) \tag{3.24}$$

These are classical bounds in probability theory. They could be shown to be sharp by calculating the dual problem explicitly (see Rü (1981)).

b) **Hölder and Jensen inequality:** For $\alpha_i > 0$, $\sum 1/\alpha_i = 1$, $X_i \ge 0$ the Hölder inequality

$$E \prod_{i=1}^n X_i \le \prod_{i=1}^n \|X_i\|_{\alpha_i} \tag{3.25}$$

is an optimal upper bound in the class of distributions with given α_i-th moments of X_i. One can improve this bound by

$$E \prod_{i=1}^n X_i \le E \prod_{i=1}^n F_i^{-1}(U), \tag{3.26}$$

where U is uniform on $(0,1)$, F_i the distribution functions of P_i. This bound is sharp in the class of all distributions with marginals P_i (with distribution functions F_i).

Similarly the Jensen inequality

$$E\varphi(X) \underset{(\le)}{\ge} \varphi(EX), \qquad \varphi \text{ convex (concave)} \tag{3.27}$$

is sharp in the class of all distributions with given expectation. For large classes of convex functions one can improve the bounds. E.g. for $\varphi(x) = \max x_i - \min x_i$, the span of x, holds

$$E \operatorname{span}(X_i) \ge E \operatorname{span}(F_i^{-1}(U)) \tag{3.28}$$

which is sharp in the class of distributions with given marginals.

c) **Sharp bounds for the sum:** For $P_i \in M^1(\mathbb{R}^1)$ with distribution functions F_i holds

$$\begin{aligned} \sup\{P(X_1 + X_2 \le t); X_i \sim P_i\} &= F_1 \wedge F_2(t) \\ &= \inf\{F_1(u) + F_2(t-u); u \in \mathbb{R}^1\}, \end{aligned} \tag{3.29}$$

$F_1 \wedge F_2$ is the infimal convolution of the distribution functions F_i of P_i. Similarly,

$$\inf\{P(X_1 + X_2 \le t); X_i \sim P_i\} = (F_1 \vee F_2(t) - 1)_+, \tag{3.30}$$

where $F_1 \vee F_2$ is the supremal convolution.

This problem of sharp bounds for the distribution of the sum was solved independently by Makarov (1981) and Rü (1982). In Rü (1982) the proof was based on the duality theorem. This result has found recently great interest in risk theory since it allows to derive sharp bounds for the *'value at risk' measure* in a portfolio caused by dependence of the components. There are a sequence of recent papers using the duality result in order to establish extensions of the bounds for more than two random variables and thus to obtain effective bounds for the risk in greater portfolios.

d) For Δ-monotone and for quasi monotone functions $h : \mathbb{R}^n \to \mathbb{R}^1$ as e.g. $h(x) = \phi(x_1 + \cdots + x_n)$, ϕ convex, sharp bounds were established

$$\sup\{Eh(X); X_i \sim P_i\} = Eh(F_1^{-1}(U), \dots, F_n^{-1}(U)) \tag{3.31}$$

(see Tchen (1980), Rü (1980, 1983)).

The duality theorem (3.15)–(3.20) was established for tight measures on a Hausdorff space which corresponds roughly to the case of complete separable metric spaces. A natural question is in what generality does the duality theorem hold true? We consider the case $n = 2$ and probability spaces $(U_i, \mathfrak{A}_i, P_i)$ and define:

(D) holds if $S(h) = I(h)$ for all $h \in B(U_1 \times U_2, \mathfrak{A}_1 \otimes \mathfrak{A}_2)$ (or $h \in \mathcal{L}_m(U_1 \times U_2)$).

Remind that $P \in M^1(\Omega, \mathfrak{A})$ is called *perfect* if for all $f \in \mathcal{L}(\mathfrak{A})$ there exists a Borel set $B \subset f(\Omega)$ such that

$$P(f^{-1}(B)) = 1. \tag{3.32}$$

This notion was introduced by Kolmogorov and is instrumental for various measure theoretic constructions like conditional probability measures (see Ramachandran (1979a,b)). The following general duality theorem holds if one of the underlying marginal measures is perfect.

Theorem 3.3 (Perfectness and duality.) *(Ramachandran and Rü (1995)) If P_2 is perfect, then $(U_2, \mathfrak{A}_2, P_2)$ is a* duality space*, i.e. (D) holds for any further probability space $(U_1, \mathfrak{A}_1, P_1)$.*

The proof starts with the case $U_i = [0, 1]$, $i = 1, 2$, where (D) holds by the Duality Theorem 3.1. It then uses various measure theoretic properties as the outer measure property of Pachl, the Marczewski imbedding theorem, and a measure extension property.

In the following we deal with the problem whether perfectness is also a necessary condition. To study this question we introduce the notion of a strong duality space.

Definition 3.4 *$(U_1, \mathfrak{A}_1, P_1)$ is a 'strong duality space' if it is a duality space and the functional I is stable under extensions, i.e. for any $(U_2, \mathfrak{A}_2, P_2)$ and any sub σ-algebra $\mathfrak{C}_2 \subset \mathfrak{A}_2$ and $h \in B(\mathfrak{A}_1 \otimes \mathfrak{C}_2)$ holds*

$$I_{\mathfrak{A}_1 \otimes \mathfrak{C}_2}(h) = I_{\mathfrak{A}_1 \otimes \mathfrak{A}_2}(h). \tag{3.33}$$

Equivalently one could also postulate stability of S. As consequence of the general duality theorem one obtains that *perfectness* implies *strong duality space*. We need further two measure theoretic properties.

Definition 3.5 $(U_1, \mathfrak{A}_1, P_1)$ *has the 'projection property' if for all* $(U_2, \mathfrak{A}_2, P_2)$ *and* $C \in \mathfrak{A}_1 \otimes \mathfrak{A}_2$ *there exists* $A_1 \in \mathfrak{A}_1$ *with* $P_1(A_1) = 1$ *and*

$$\pi_2(C \cap (A_1 \times U_2)) \in \overline{\mathfrak{A}_2}^{P_2}, \tag{3.34}$$

$\overline{\mathfrak{A}_2}^{P_2}$ *the* P_2*-completion of* $\mathfrak{A}_2$.

This notion is a measure theoretic analog of the projection property in descriptive set theory. The classical result in this area says that the projection of a Borel set in a product of two standard Borel spaces is analytic and thus universally measurable. The second property is the measure extension property.

Definition 3.6 $(U_1, \mathfrak{A}_1, P_1)$ *has the 'measure extension property' if for all* $(U_2, \mathfrak{A}_2, P_2)$ *for all* $\mathcal{D}_2 \subset \mathfrak{A}_2$ *and all* $P \in M(P_1, P_2/\mathcal{D}_2)$ *there exists an extension* $\overline{P} \in M(P_1, P_2)$ *such that* $\overline{P}/\mathfrak{A}_1 \otimes \mathcal{D}_2 = P$.

We say that $(U_1, \mathfrak{A}_1, P_1)$ *has the 'charge extension property' if the extension can be found in the set* $\mathrm{ba}(P_1, P_2)$ *of charges, i.e. non-negative finitely additive measures.*

It now turns out that the strong duality spaces are exactly the perfect spaces. Thus a general duality theorem in the strong sense implies perfectness. The following theorem states that the strong duality property is even equivalent with any of the measure theoretic notions introduced above.

Theorem 3.7 (Characterization theorem.) *(Ramachandran and Rü (2000)) For a probability space* $(U_1, \mathfrak{A}_1, P_1)$ *the following statements a)–e) are equivalent:*

a) $(U_1, \mathfrak{A}_1, P_1)$ *is a strong duality space.*

b) $(U_1, \mathfrak{A}_1, P_1)$ *is perfect.*

c) $(U_1, \mathfrak{A}_1, P_1)$ *has the measure extension property.*

d) $(U_1, \mathfrak{A}_1, P_1)$ *has the projection property.*

e) $(U_1, \mathfrak{A}_1, P_1)$ *has the charge extension property.*

As consequence all structure theorems for perfect spaces are also valid for strong duality spaces. On the other hand this result says that one cannot expect 'good' duality results on 'general' infinite dimensional spaces. There remain the following important

Open problems:

a) Is any measure space a duality space?

b) Is any duality space a strong duality space?

c) Is $M(P_1, P_2) \subset \mathrm{ba}(P_1, P_2)$ dense in weak $*$-topology from $B(U_1 \times U_2)$?

4 Optimal multivariate couplings

The duality results of section 3 were developed into more concrete optimal coupling results in the early nineties. Here also started the development on the subject by several researchers from analysis since the strong and fruitful connections to several problems in analysis soon became clear. In particular to mention is the work of Brenier, Gangbo and McCann and later on Ambrosio, Villani, Otto, Caffarelli, Evans, Trudinger, Lott, and Sturm.

4.1 The squared norm cost

For the squared norm cost $c(x,y) = \|x-y\|^2$, $x, y \in \mathbb{R}^k$ the problem of optimal transport or optimal couplings is given by

$$E\|X-Y\|^2 = \inf_{X \sim P_1, Y \sim P_2} \tag{4.1}$$

where $P_i \in M^1(\mathbb{R}^k, \mathfrak{B}^k)$ have covariance matrices $\sum_i = \operatorname{Cov}(P_i)$. This problem is equivalent to maximizing the trace $\operatorname{tr} \Psi$

$$\operatorname{tr} \Psi = \max! \tag{4.2}$$

over all

$$\Psi \in C(P_1, P_2) = \left\{ \Psi : \exists P \in M(P_1, P_2) \text{ such that } \begin{pmatrix} \Sigma_1 & \Psi \\ \Psi^T & \Sigma_2 \end{pmatrix} \in \operatorname{Cov}(P) \right\}. \tag{4.3}$$

In general $C(P_1, P_2)$ is a complicated set but for normal distributions $P_i = N(a_i, \sum_i)$ one gets the maximal possible class $C(P_1, P_2)$ with covariance matrices $\sum_i$ of P_i. The covariance condition (4.3) is in this case equivalent to

$$\begin{pmatrix} \Sigma_1 & \Psi \\ \Psi^T & \Sigma_2 \end{pmatrix} \geq 0 \text{ in the sense of positive semidefiniteness.} \tag{4.4}$$

The corresponding optimization problem (4.2) was analytically solved in Olkin and Pukelsheim (1982) and Dowson and Landau (1982), leading in particular to an universal lower bound of $\ell_2(P_1, P_2)$ depending only on first and second moments a_i, Σ_i for any pair P_1, $P_2 \in M^1(\mathbb{R}^k, \mathfrak{B}^k)$.

For general distributions P_i the following is the basic optimal coupling result which is due to Knott and Smith (1984, 1987), Rü and Rachev (1990), and Brenier (1987, 1991).

Theorem 4.1 (Optimal L^2-couplings.) *Let $P_i \in M^1(\mathbb{R}^k, \mathfrak{B}^k)$ with $\int \|x\|^2 dP_i(x) < \infty$, then*

a) *There exists an optimal L^2-coupling, i.e. a solution of* (4.1).

b) $X \stackrel{d}{=} P_1$, $Y \stackrel{d}{=} P_2$ *is an optimal L^2-coupling*

$$\Longleftrightarrow \exists \text{ convex, lsc } f \in L^1(P_1) \text{ such that } Y \in \partial f(X) \text{ a.s.} \tag{4.5}$$

c) *If $P_1 \ll \lambda^k$, then for f as in a)*

$$\partial f(X) = \nabla f(X) \text{ a.s. and } (X, \nabla f(X)) \tag{4.6}$$

is a solution of the Monge problem.

d) *If $P_1 \ll \lambda\!\!\lambda^k$, then there exists a P_1 a.s. unique gradient ∇f of a convex function f, such that*

$$P_1^{\nabla f} = P_2. \tag{4.7}$$

Remarks 4.2 a) Part b) of this theorem was given in this form first in Rü and Rachev (1990). The proof was based on the duality theorem. The sufficiency part for b) is contained already in Knott and Smith (1984, 1987). Brenier (1991) established the uniqueness result in d) as well as b) while a special version of c) is already in his 1987 paper. Note that the existence of a Monge solution in c) is an immediate consequence of b) and a.s. differentiability of convex functions (see Rockafellar (1970)). By this history it seems appropriate to describe this important theorem to the authors from probability and analysis as mentioned above.

b) **Cyclically monotone support.** By convex analysis condition (4.5) is equivalent to *cyclically monotone support* Γ of the optimal transportation measure $\mu = P^{(X,Y)}$, i.e. $\forall (x_1, y_1), \dots, (x_m, y_m) \in \Gamma$ holds

$$\sum_{i=1}^m y_i x_{i+1} \le \sum_{i=1}^m y_i x_i \tag{4.8}$$

with $x_{m+1} := x_1$. This equivalence lead Gangbo and McCann (1995) to a new strategy of proof. If uniqueness holds (as in the case $P_1 \ll \lambda\!\!\lambda^k$) then *cyclical monotonicity* of the support of μ implies optimality. In this way they were able to replace the moment assumptions on P_1, P_2 by the uniqueness condition.

c) For $P_1 = f\lambda\!\!\lambda^k$, $P_2 = g\lambda\!\!\lambda^k$ absolutely continuous w.r.t. Lebesgue measure Caffarelli (1992, 1996) established regularity estimates of the optimal Monge solution Φ:
If $f, g \in C^{k,\alpha}$ (i.e. the partial derivatives up to order k are of Hölder type α) and $g > 0$ then $\Phi \in C^{k+2,\alpha}$. In particular if $f, g \in C^\infty$ and locally bounded from below, their supports and $\operatorname{supp} g$ is convex, then $\Phi \in C^\infty$ and Φ is a classical solution of the Monge–Ampère equation

$$\det D^2\Phi(x) = \frac{f(x)}{g(\nabla\Phi(x))} \quad [P_1] \tag{4.9}$$

(see Villani (2006) for a more detailed exposition).

A corollary of the optimal L^2-coupling theorem is the polar factorization theorem due to Brenier (1987).

Corollary 4.3 (Polar factorization theorem.) *Let $E \subset \mathbb{R}^d$ be a bounded subset with positive Lebesgue measure, $h : E \to \mathbb{R}^d$ a measurable map with $P^h \ll \lambda\!\!\lambda^d$, where $P = \lambda_E$ is the normalized Lebesgue measure on E. Then there exists a unique gradient ∇f of a convex lsc function f and a measure preserving map s on (E, P) such that*

$$h = \nabla f \circ s \;\; [P]. \tag{4.10}$$

Remarks 4.4 a) The nondegeneracy condition of the polar factorization theorem has been weakened by Burton and Douglas (1998, 2003). Also a counter-example is given there to show that the theorem is false without any further assumption.

b) The nondegeneracy condition of h in the polar factorization theorem can also be replaced by the following independence assumption $(\mathrm{I_h})$:

($\mathbf{I_h}$) There exists a random variable V on (E, P) such that V is independent of h and $P^V = U(0,1)$ is the uniform distribution on $(0,1)$.

For the proof let (X, Y) be an optimal coupling of (P^h, P), $X \stackrel{d}{=} P^h$, $Y \stackrel{d}{=} P$. Since $P \ll \lambda\!\!\!\lambda^d$, by the optimal coupling result Theorem 4.1 there exists a unique gradient ∇f of a convex function f such that

$$X = \nabla f \circ Y \ \text{ a.s..} \tag{4.11}$$

Now one can apply the following result of Rü (1985) on the solutions of *stochastic equations* (see also Rachenv and Rü (1990)):

Let $(E, \mu) \stackrel{h}{\to} B$, B a Borel space, let $(F, R) \stackrel{f}{\to} B$, h, f measurable, μ, R probability measures, and let (E, μ) be rich enough (i.e. it allows a uniformly distributed r.v. V on (E, μ) independent of h). If the distributional equation

$$\mu^h = R^f \tag{4.12}$$

holds, then there exists an r.v. $U : E \to F$ with $\mu^U = R$ such that the stochastic equation

$$h = f \circ U \ \ [\mu] \tag{4.13}$$

holds.

Applying this general factorization theorem with $\mu = \lambda^d_E = P$, $R = P = \mathcal{L}(Y)$ we obtain the existence of a measurable factorization

$$h = \nabla f \circ U \ \ [P] \tag{4.14}$$

with some measure preserving map U on (E, P), i.e. the polar factorization result.

Corollary 4.5 *For any measurable function h the independence hypothesis* $(\mathrm{I_h})$ *implies the existence of a polar factorization.*

In general the independence hypothesis does not hold. If e.g. $d = 1$ and $h(u) = u$, $u \in [0,1] = E$ then I_h does not hold. If $V = V(u)$ would be independent of h, then $P^{V|h=u} = \varepsilon_{V(u)}$, a contradiction. But by enlarging E to e.g. $E' := E \times [0,1]$ and considering $P' = P \otimes \lambda\!\!\!\lambda_{[0,1]}$ we can consider h formally as function on E' by $h(x,u) := h(x)$. The independence hypothesis holds in this extended framework and thus there exists an r.v. U on E' such that $P'^U = P = \lambda_E$ and

$$h = \nabla f \circ U \ \ [P'], \tag{4.15}$$

i.e. $h(x) = \nabla f \circ U(x,u) \ \ [P']$.

Thus, we obtain a polar factorization theorem in the 'weak sense' without any nondegeneration condition on h.

Corollary 4.6 *The polar factorization theorem holds in the extended sense* (4.15) *without any further nondegeneration assumption.*

This extension also holds for McCann's (2001) version of the polar factorization theorem in Riemannian manifolds.

4.2 General coupling function

For the case of general coupling (resp. cost) functions $c = c(x, y)$ and probability measures P, Q we consider the corresponding optimal coupling (transport) problem

$$S(c) = \sup\left\{ \int c d\mu; \mu \in M(P, Q) \right\} \tag{4.16}$$

with dual problem

$$I(c) = \inf\left\{ \int h_1 dP + \int h_2 dQ; c \leq h_1 \oplus h_2, h_i \in L^1 \right\} \tag{4.17}$$

The following notions from nonconvex optimization theory as discussed in Elster and Nehse (1974) and Dietrich (1988) are useful and were introduced in the context of the transportation problem in Rü (1991b). A proper function $f : \mathbb{R}^k \to \mathbb{R} \cup \{\infty\}$ is called c-convex if it has a representation of the form

$$f(x) = \sup_y (c(x, y) + a(y)) \tag{4.18}$$

for some function a. The *c-conjugate* f^c of f is defined by

$$f^c(y) = \sup_x (c(x, y) - f(x)), \tag{4.19}$$

the sup being over the domain of f. Defining further the *double c-conjugate* f^{cc} by

$$f^{cc}(x) = \sup_y (c(x, y) - f^c(y)) \tag{4.20}$$

then f^c, f^{cc} are c-convex, f^{cc} is the largest c-convex function majorized by f and $f = f^{cc}$ if and only if f is c-convex. The pair f^c, f^{cc} is an admissible pair in the sense that

$$f^c(y) + f^{cc}(x) \geq c(x, y) \text{ for all } x, y. \tag{4.21}$$

Obviously this construction is similarly possible on a general pair U_1, U_2 of spaces replacing $\mathbb{R}^k$ and $c : U_1 \times U_2 \to \mathbb{R} \cup \{\infty\}$. The (double) c-conjugate functions are basic for the theory of inequalities as in (4.21). The generalized *c-subgradient* of a function f at a point x is defined by

$$\partial_c f(x) = \{y; f(z) - f(x) \geq c(z, y) - c(x, y) \ \ \forall z \in \operatorname{dom} f\} \tag{4.22}$$

further

$$\partial_c f = \{(x, y) \in U_1 \times U_2; y \in \partial_c f(x)\}. \tag{4.23}$$

Denoting by $\mathfrak{C} := \{\Psi_{y,a}; y \in U_2, a \in \mathbb{R}\}$ the class of all shifts and translates of c, $\Psi_{y,a}(x) := c(x, y) + a$, c-convexity of f is equivalent to a representation of the form $f(x) = \sup_{\Psi \in \mathfrak{C}'} \Psi(x)$ for some $\mathfrak{C}' \subset \mathfrak{C}$ and further (with $a := f(x) - c(x, y_0)$)
$y_0 \in \partial_c f(x)$ if and only if $\exists\, a \in \mathbb{R}$ such that

$$\Psi_{y_0,a}(x) \ = \ f(x) \tag{4.24}$$

$$\Psi_{y_0,a}(x') \ \leq \ f(x'), \quad \forall x' \in \operatorname{dom} f. \tag{4.25}$$

This geometric description of c-subgradients generalizes the corresponding description in the case of convex functions and is very useful and intuitive. The following is an extension of the optimal coupling (transportation) result in Theorem 4.1 to general cost functions and general measure spaces U_i as in the basic Duality Theorem 4.1.

Theorem 4.7 (Optimal c-couplings.) *(Rü (1991b)) Let c be a lower majorized function (i.e. $c(x,y) \geq f_1(x) + f_2(y)$ for some $f_1 \in L^1(P)$, $f_2 \in L^1(Q)$) and assume that $I(c) < \infty$. Then a pair (X,Y) with $X \stackrel{d}{=} P$, $Y \stackrel{d}{=} Q$ is an optimal c-coupling between P and Q if and only if*

$$(X,Y) \in \partial_c f \quad a.s. \tag{4.26}$$

for some c-convex function f, equivalently, $Y \in \partial_c f(X)$ a.s.

The characterization in (4.26) is equivalent to the condition that the support Γ of the joint distribution of X, Y is *c-cyclically monotone*, i.e. for all $(x_i, y_i) \in \Gamma$, $1 \leq i \leq n$, $x_{n+1} := x_1$ holds

$$\sum_{i=1}^{n} \big(c(x_{i+1}, y_i) - c(x_i, y_i)\big) \leq 0. \tag{4.27}$$

This notion was introduced in Smith and Knott (1992), who reformulated Theorem 4.7 in terms of c-cyclical monotonicity. For the equivalence see also Rü (1995, 1996b) and Gangbo and McCann (1996). Note that for the corresponding inf problem (transportation problem) the inequality sign in (4.27) has to be chosen in the opposite direction.

Without the moment assumptions in Theorem 4.7 c-cyclically monotone support is in general not a sufficient condition for optimality (see Ambrosio and Pratelli (2003)). Gangbo and McCann (1996) have developed a characterization of c-optimality based on c-cyclically monotone support plus a uniqueness property. They also have studied some regularity properties of c-convex functions. The moment assumptions of the duality theorem have been weakened in Ambrosio and Pratelli (2003) and Schachermayer and Teichmann (2006) for lower semicontinuous cost functions c. The notion of 'strongly c-monotone' support is introduced and shown as a sufficient condition in their paper.

The characterization of solutions of the optimal coupling problem and the results on existence of solutions (X,Y) resp. f imply the following necessary condition for differentiable cost functions: If (X,Y), f are solutions of the optimal c-coupling problem on $\mathbb{R}^k$ and if $P \ll \lambda^k$ and f is differentiable almost everywhere, then

$$\nabla f(X) = \nabla_x c(X,Y) \quad \text{a.s.} \tag{4.28}$$

(see Rü (1991b)).

In this direction Gangbo and McCann (1996) have shown that if $c = c(x-y)$ is locally Lipschitz, then c-convex functions are differentiable almost everywhere. For an extension see (Villani, 2006, p. 125). As consequence one obtains:
If (4.28) has a unique solution in $Y = \Phi(X)$, then

$$(X, \Phi(X)) \quad \text{is a Monge solution.} \tag{4.29}$$

In the case where $c(x,y) = h(x-y)$, h strictly convex, c-convex functions f are convex and thus ∇f exists almost everywhere and then (4.28) implies

$$Y = X - \nabla h^*(\nabla f(X)) =: \Phi(X) \tag{4.30}$$

where h^* is the convex conjugate of h (see Gangbo and McCann (1996)). A similar example for the concave functions $h(|x-y|)$ is in Rü and Uckelmann (2000). Note that Monge solutions for (4.16) are solutions in weak sense for generalized PDE's of Monge–Ampére type.

Remarks 4.8 a) Call a function Φ c-cyclically monotone if $\Gamma = \text{graph}(\Phi)$ is c-cyclically monotone. Several sufficient and several necessary conditions for Φ to be c-cyclically monotone have been given in Smith and Knott (1992) and in Rü (1995). For the optimal ℓ_p-coupling with $c(x,y) = \|x-y\|^p$, $\|\cdot\|$ the Euclidean norm, one obtains e.g. for $p > 1$ that for h

$$\Phi(x) := \|h(x)\|^{-\frac{p-2}{p-1}} h(x) + x \tag{4.31}$$

is c-cyclically monotone. Similar extensions hold for p-norms $\|\cdot\|_p$. In particular it is shown that radial transformations

$$\Phi(x) = \alpha(\|x\|)\frac{x}{\|x\|}, \quad \alpha(t) \geq t \tag{4.32}$$

are optimal. This allows e.g. to determine optimal couplings between uniform distributions on two ellipsoids. For the Euclidean norm $\|\cdot\| = \|\cdot\|_2$ and $p = 1$, the classical Monge case, one obtains that the optimal transport is concentrated on lines $Y \in \{X + t\nabla f(X); 0 \leq t \leq T\}$. It is however not uniquely determined by this property. Existence of Monge-solutions for the classical Monge case where c is the Euclidean norm has a long history, starting with early work of Sudakov (1979) (for details see Villani (2006)).

b) **One-dimensional case.** In the one-dimensional case the optimal coupling result in Theorem 4.7 has been applied to determine optimal couplings for some classes of nonconvex functions, see Uckelmann (1997), Rü and Uckelmann (2000), and, based on a direct analysis of monotonicity properties, in McCann (1999). In the case where $P = Q = U(0,1)$ and $c(x,y) = \Phi(x+y)$ and for coupling functions Φ such that

$$\Phi''(t) \begin{cases} > 0 & t \in [0,k_1) \cup (k_2, 2] \\ < 0 & t \in (k_1, k_2), \end{cases} \tag{4.33}$$

i.e. Φ is of the type: convex-concave-convex, the following result was proved in Uckelmann (1997):

Proposition 4.9 *If α, β are solutions of*

$$\begin{cases} \Phi(2\alpha) - \Phi(\alpha+\beta) + (\beta-\alpha)\Phi'(\alpha+\beta) = 0, \\ \Phi(2\beta) - \Phi(\alpha+\beta) + (\alpha-\beta)\Phi'(\alpha+\beta) = 0, \end{cases} \tag{4.34}$$

$0 < \alpha < \beta < 1$, *then $(X, T(X))$ is an optimal c-coupling where*

$$T(x) = \begin{cases} x, & x \in [0,\alpha] \cup [\beta, 1], \\ \alpha + \beta - x, & x \in (\alpha, \beta). \end{cases} \tag{4.35}$$

Similar results have been shown for $c(x,y) = \Phi(|x-y|)$, Φ convex-concave for the case of uniform marginals. Extensions of these results are in Rü and Uckelmann (2000). Also some results for nonuniform marginals and numerical results are given there. These results confirm some general conjectures on the solutions in the one dimensional case. The case $c(x,y) = \Phi(|x-y|)$, Φ concave was studied in McCann (1999) in detail.

c) **Discrete distributions, Voronoi cells.** In the case where one distribution is discrete $Q = \sum_{i=1}^n \alpha_j \varepsilon_{x_j}$ the relevant c-convex functions for the optimal couplings are of the form

$$f(x) = \max(c(x, x_j) + a_j). \tag{4.36}$$

The subgradients are to be the support points x_j of Q and we have to determine only the shifts a_j. Let

$$\begin{aligned} A_j &:= \{x : f(x) = c(x, x_j) + a_j\} \\ &= \{x : x_j \in \partial_c f(x)\} \end{aligned} \tag{4.37}$$

denote the corresponding Voronoi cells, then the subgradients are unique except at the boundaries of A_j. This observation implies a uniqueness result for discrete distributions (see Cuesta-Albertos and Tuero-Diaz (1993)). The optimal transportation problem between P and Q reduces to finding shifts a_j such that $P(A_j) = \alpha_j$, $1 \le j \le n$. For the case of $c(x, y) = \|x - y\|^a$ one gets for $a = 2$ linear boundaries. Some cases for $a = 1$, 2, 4 are dealt with explicitly in Rü (1997, 2000) in the case where P is uniform on a square in $\mathbb{R}^2$ or a cube in $\mathbb{R}^3$. For the solution in not too large discrete cases one can apply sophisticated linear programming techniques or algorithms developed for Voronoi cells (see Rü and Uckelmann (2000)). For an alternative continuous time algorithm see Benamou and Brenier (1999).

4.3 The n-coupling problem

The coupling problem between two distributions on $\mathbb{R}^k$ is naturally extended to the optimal coupling (transportation) between n probability measures $P_1, \dots, P_n$ on $\mathbb{R}^k$. This might be used as discrete time approximation of a continuous time flow along some time interval $[0, T]$ where P_i are intermediate distributions on the transport from P_1 to P_n. A new problem arises only if in the formulation of the coupling problem there is a cycle or a back coupling e.g. between P_n and P_1. For the L^2-cost one version of this problem is

$$E\Big\| \sum_{i=1}^n X_i \Big\|^2 = \max \tag{4.38}$$

over all $X_i \overset{d}{=} P_i$, $1 \le i \le n$.

For the case $n = 3$ and $P_i = N(0, \Sigma_i)$ this problem was posed in Olkin and Rachev (1993). It is for $n = 3$ equivalent to

$$E(\langle X, Y\rangle + \langle Y, Z\rangle + \langle X, Z\rangle) = \max! \tag{4.39}$$

and thus includes cycles: some mass should be cycled around three places from 1) to 2) from 2) to 3) but then also back from 3) to 1) in an optimal way. Knott and Smith (1994) (in the case $n = 3$) proposed the fruitful idea that “optimal coupling to the sum $T = \sum_{i=1}^n X_i$ should imply multivariate optimal coupling in the sense of (4.38) ”.

The reason for this conjecture is the equivalence of (4.38) with the problem

$$E \sum_{i=1}^n \|X_i - T\|^2 = \min_{X_i \overset{d}{=} P_i} ! \tag{4.40}$$

In the case of multivariate normal distributions $P_i = N(0, \Sigma_i)$, $1 \le i \le 3$, this idea leads then to the following algorithm: Let $T \stackrel{d}{=} N(0, \Sigma_0)$, Σ_0 positive definite, and define

$$X = S_1 T, \;\; Y = S_2 T, \;\; Z = S_3 T \tag{4.41}$$

the optimal 2-couplings with

$$S_i = \Sigma_i^{1/2} \big(\Sigma_i^{1/2} \Sigma_0 \Sigma_i^{1/2} \big)^{-1/2} \Sigma_i^{1/2}.$$

Then one needs that $T = X + Y + Z = (S_1 + S_2 + S_2) T$ and thus one needs the identity $S_1 + S_2 + S_2 = I$. This leads to the following nonlinear equation for Σ_0:

$$\sum_{i=1}^{3} \big(\Sigma_0^{1/2} \Sigma_i \Sigma_0^{1/2} \big)^{1/2} = \Sigma_0. \tag{4.42}$$

As consequence of this idea one gets the following result:

If (4.42) has a positive definite solution Σ_0, then X, Y, Z as defined in (4.41) are optimal solutions of the 3-coupling problem (see Knott and Smith (1994)).

The natural iterative algorithm for the solution of (4.42) converges rapidly in $d = 2$ (as reported in Knott and Smith (1994)) but in $d = 3$ it turns out that convergence depends very sensitive on the initial conditions.

It was shown only in Rü and Uckelmann (2002) based on the uniqueness result for optimal couplings that this idea of Knott and Smith is valid in the normal case (in particular (4.42) has a solution) for any $n \ge 3$.

For general distributions P_i however optimal coupling to the sum is not sufficient for optimal n-coupling. There are some simple counter-examples. But a positive result is given in Rü and Uckelmann (2002) saying that optimal coupling to the sum T, continuity of $P^T \ll \lambda^d$ and 'maximality' of the domain of P^T imply optimal n-coupling.

The paper contains also a simple proof of the existence result of Gangbo and Święch (1998) on Monge solutions (in the case without cycles) of the form

$$X = (X_1, \Phi_2(X_1), \dots, \Phi_n(X_1)) \tag{4.43}$$

if $P_i \ll \lambda^d$ and gives a one-to-one equivalence of the n-coupling problem with several modified 2-coupling problems.

5 A variant of the minimal ℓ_s-metric – application to the analysis of algorithms

Many algorithms of recursive structure (divide and conquer type algorithms) allow to derive limit theorems for their important parameters by the *contraction method* (see Rösler and Rü (2001) and Neininger and Rü (2004)). The limits then are characterized by stochastic fix point equations typically of the form

$$X \stackrel{d}{=} \sum_{i=1}^{K} (A_i X_i + b_i) \tag{5.1}$$

where X_i are independent copies of X and where (X_i) and the random coefficients (A_i, b_i) are also independent.

For example: The path length Y_n of the Quicksort algorithm satisfies the recursive equation

$$Y_n \stackrel{d}{=} I_n Y_{I_n} + (n - I_n - 1)\overline{Y}_{n-I_n-1} + (n-1), \tag{5.2}$$

where I_n is uniform on $\{0, \dots, n-1\}$ distributed and $(\overline{Y}_n)$ are independent copies of (Y_k).

The normalization $X_n := \frac{Y_n - EY_n}{n}$ satisfies the recursive equation

$$Y_n \stackrel{d}{=} \frac{I_n}{n} X_{I_n} + \frac{n - I_n + 1}{n} \overline{X}_{n-I_n+1} + C_n(I_n) \tag{5.3}$$

and leads to the limit equation

$$X \stackrel{d}{=} UX + (1-U)\overline{X} + C(U), \tag{5.4}$$

where the entropy function $C(x) = x \ln x + (1-x)\ln(1-x) + 1$ arises as the limit of C_n and $U \stackrel{d}{=} U[0,1]$ as the limit of $\frac{I_n}{n}$. The contraction method then gives general conditions which imply existence and uniqueness of solutions of (5.4) and convergence of X_n to this solution.

The solution of (5.4) resp. (5.1) can be described as fixpoint of the operator

$$T : M_s \to M_s, \quad T\mu = \mathcal{L}\Big(\sum_{i=1}^{n} (A_i X_i + b_i)\Big) \tag{5.5}$$

where (X_i) are iid and $X_i \stackrel{d}{=} \mu$. M_s is the class of distributions with finite s-th moments. A natural contraction condition is:

$$A_i, b_i \in L^s, \quad \sum_{i=1}^{K} E|A_i|^s < 1. \tag{5.6}$$

This has been shown to imply the fixpoint result for $0 \le s \le 1$ and for $s = 2$ (with the additional restriction of a fixed first moment). For $1 < s < 2$ one has only an inequality of the form

$$\ell_s^s(T\mu, T\nu) \le K_s^s \Big(\sum_{i=1}^{K} E|A_i|^s\Big) \ell_s^s(\mu, \nu) \tag{5.7}$$

with some constant $K_s > 1$. This yields only an existence result for the fixpoint equation (5.1) under restrictive conditions. To solve this problem a new modification of the minimal ℓ_s-metric was introduced in Rü (2006). Define

$$\ell_s^0(\mu, v) := \inf\{\|X - Y\|_s : X \stackrel{d}{=} \mu, Y \stackrel{d}{=} \nu, \text{ and } X \approx Y\} \tag{5.8}$$

Here $X \approx Y$ is defined by the conditions

$$E(X - Y) = 0, \quad E|X - Y|^s < \infty. \tag{5.9}$$

Note that condition (5.9) does not need that μ, $\nu \in M_s$.

With this new variant of the minimal ℓ_s-metric which is defined by a modification of the transportation problem one obtains the existence of fixpoints under the natural condition (5.6) on the coefficients (see Rü (2006)).

Theorem 5.1 *If A_i, $b_i \in L^s$, $1 \le s \le 2$, $E\sum_{i=1}^K |A_i|^s < 1$ and if for some $\mu_0 \in M$, $\ell_s^0(\mu_0, T\mu_0) < \infty$, then there exists a unique solution of* (5.1) *in*

$$M_s^0(\mu_0) = \{\mu \in M;\ \ \ell_s^0(\mu, \mu_0) < \infty\}. \tag{5.10}$$

In the proof of this theorem it is shown that $(M_s^0(\mu_0), \ell_s^0)$ is a complete metric space and with the help of weighted branching processes it is shown that some power T^{n_0} of T satisfies a contraction condition on $M_s^0(\mu_0)$ w.r.t. the new variant ℓ_s^0 of the minimal ℓ_s-metric.

An interesting corollary of this result is an equivalence principle for homogeneous and inhomogeneous stochastic equations.

Corollary 5.2 *Under the assumptions of Theorem 5.1 there is a one-to-one relationship between solutions of the homogeneous stochastic equation*

$$X \stackrel{d}{=} \sum_{i=1}^K A_i X_i \tag{5.11}$$

and the inhomogeneous stochastic equation

$$Y \stackrel{d}{=} \sum_{i=1}^K (A_i Y_i + b_i). \tag{5.12}$$

More exactly: For any solution X of (5.11) *there exists exactly one solution Y of* (5.12) *with distribution $\mu \in M_s^0(\mu_0)$, where $\mu_0 = \mathcal{L}(X)$ and conversely.*

Thus the modification of the minimal ℓ_s-metric allows to investigate solutions of fixpoint equations (5.1) without any moment conditions.

References

Ambrosio, L. (2003). Lecture notes on optimal transport problems. In P. C. et al. (Ed.), *Mathematical Aspects of Evolving Interfaces. Lectures given at the C.I.M.–C.I.M.E. joint Euro-summer school, Madeira, Funchal, Portugal, July 3–9, 2000*, Volume 1812 of *Lecture Notes in Mathematics*, pp. 1–52. Springer.

Ambrosio, L., N. Gigli, and G. Savaré (2005). *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics, ETH Zürich. Birkhäuser.

Ambrosio, L. and A. Pratelli (2003). Existence and stability results in the L^1 theory of optimal transportation. In *Optimal Transportation and Applications (Martina Franca, Italy, 2001)*, Volume 1813 of *Lecture Notes in Mathematics*, pp. 123–160. Springer.

Appel, P. (1887). Mémoire sur les déblai et les remblais des systeèmes continus ou discontinus. *Mémoires présentés par divers Savants à l'Académie des Sciences de l'Institut de France, Paris 28*, 1–208. Available at `http:gallica.bnf.fr.`

Benamou, J.-D. and Y. Brenier (1999). A numerical method for the optimal time-continuous mass transport problem and related problems. In *Monge Ampère Equation: Applications to Geometry and Optimization (Deerfield Beach, Florida, USA, 1997)*, Volume 226 of *Contemp. Math.*, pp. 1–11. American Mathematical Society.

Brenier, Y. (1987). Décomposition polaire et réarrangement monotone des champs de vecteurs. *Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique 305*(19), 805–808.

Brenier, Y. (1989). The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society 2*, 225–255.

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics 44*, 375–417.

Brenier, Y. (1999). Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Communications on Pure and Applied Mathematics 52*, 411–452.

Burton, G. R. and R. J. Douglas (1998). Rearrangements and polar factorisation of countably degenerate functions. *Proceedings of the Royal Society of Edinburgh. Section A. 128*, 671–681.

Burton, G. R. and R. J. Douglas (2003). Uniqueness of the polar factorisation and projection of a vector-valued mapping. *Annales de l'Institut Henri Poincaré (C) Non linear Analysis 20*(3), 405–418.

Caffarelli, L. A. (1992). The regularity of mappings with convex potentials. In *Partial Differential Equations and Related Subjects (Trento, Italy, 1990)*, Volume 269 of *Pitman Res. Notes Math. Ser.*, pp. 53–58. Longman Sci. Tech.

Caffarelli, L. A. (1996). Boundary regularity of maps with convex potential. *Annals of Mathematics (2) 144*, 453–496.

Caffarelli, L. A. (2000). Monotonicity properties of optimal transportation and the FKG and related inequalities. *Commun. Math. Phys. 214*, 547–563.

Cuesta-Albertos, J. A., C. Matrán, S. T. Rachev, and L. Rüschendorf (1996). Mass transportation problems in probability theory. *Mathematical Scientist 21*, 37–72.

Cuesta-Albertos, J. A., L. Rüschendorf, and A. Tuero-Diaz (1993). Optimal coupling of multivariate distributions and stochastic processes. *Journal of Multivariate Analysis 46*, 335–361.

Cuesta-Albertos, J. A. and A. Tuero-Diaz (1993). A characterization for the solution of the Monge–Kantorovich mass transference problem. *Statist. Probab. Letters 16*, 147–152.

Dall'Aglio, G. (1956). Sugli estremi dei momenti delle funzioni di ripartizione doppie. *Ann. Scuola Normale Superiore Di Pisa, Cl. Sci. 3*(1), 33–74.

de Acosta, A. (1982). Invariance principles in probability for triangle arrays of B-valued random vectors and some applications. *Annals of Probability 10*, 346–373.

Dietrich, H. (1988). Zur c-Konvexität und c-Subdifferenzierbarkeit von Funktionalen. *Optimization 19*, 355–371.

Dobrushin, L. (1970). Prescribing a system of random variables by conditional distributions. *Theory of Probability and Its Applications 15*, 458–486.

Dowson, D. C. and B. V. Landau (1982). The Fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis 12*, 450–455.

Dudley, R. M. (1976). *Probabilities and Metrics. Convergence of laws on metric spaces, with a view to statistical testing*, Volume 45 of *Lecture Notes in Statistics.* Matematisk Institut, Aarhus Universitet, Aarhus.

Elster, K.-H. and R. Nehse (1974). Zur Theorie der Polarfunktionale. *Math. Operationsforschung Statist. 5*, 3–21.

Fernique, X. (1981). Sur le théorème de Kantorovitch–Rubinstein dans les espaces polonais. In *Seminaire de Probabilites XV (Univ. Strasbourg, Strasbourg, 1979/1980)*, Volume 850 of *Lecture Notes in Mathematics*, pp. 6–10. Springer. In French.

Fortet, R. and E. Mourier (1953). Convergence de la répartition empirique vers la répartition théoretique. *Ann. Sci. Ecole Norm. Sup.* (70), 267–285.

Fréchet, M. (1940). *Les probabilités associées à un système d'événements compatibles et dépendants; I. Événements en nombre fini fixe*, Volume 859 of *Actual. sci. industr.*

Gaffke, N. and L. Rüschendorf (1981). On a class of extremal problems in statistics. *Math. Operationsforschung Statist. 12*, 123–135.

Gangbo, W. and R. J. McCann (1995). Optimal maps in Monge's mass transport problem. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique 321*(12), 1653–1658.

Gangbo, W. and R. J. McCann (1996). The geometry of optimal transformations. *Acta Mathematica 177*, 113–161.

Gangbo, W. and A. Święch (1998). Optimal maps for the multidimensional Monge-Kantorovich problem. *Communications on Pure and Applied Mathematics 51*, 23–45.

Gini, C. (1913/1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, A.A. 73*, 1203–1248.

Givens, C. R. and R. M. Shortt (1984). A class of Wasserstein metrics for probability distributions. *Mich. Math. J. 31*, 231–240.

Hoeffding, W. (1940). *Maßstabinvariante Korrelationstheorie.* Ph.D.Thesis, Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin 5, 181–233 (1940) u. Berlin: Diss.

Kantorovich, L. V. (1942). On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS, n. Ser. 37*, 199–201.

Kantorovich, L. V. (1948). On a problem of Monge. *Uspekhi Mat. Nauk 3*, 225–226. In Russian.

Kantorovich, L. V. and G. P. Akilov (1982). *Functional Analysis* (2nd ed.). Transl. from the Russian by H. L. Silcock, Oxford.

Kantorovich, L. V. and G. S. Rubinstein (1957). On a function space in certain extremal problems. *Dokl. Akad. Nauk USSR 115*(6), 1058–1061.

Kellerer, H. G. (1984a). Duality theorems and probability metrics. In *Proc. 7th Brasov Conf. (Bucuresti, Romania, 1982)*, pp. 211–220.

Kellerer, H. G. (1984b). Duality theorems for marginal problems. *Z. Wahrscheinlichkeitstheorie Verw. Geb. 67*, 399–432.

Knott, M. and C. S. Smith (1984). On the optimal mapping of distributions. *Journal of Optimization Theory and Applications 43*, 39–49.

Knott, M. and C. S. Smith (1987). Note on the optimal transportation of distributions. *Journal of Optimization Theory and Applications 52*, 323–329.

Knott, M. and C. S. Smith (1994). On a generalization of cyclic monotonicity and distances among random vectors. *Linear Algebra and its Applications 199*, 363–371.

Levin, V. L. (1975). On the problem of mass transfer. *Soviet Math. Doklady 16*, 1349–1353.

Levin, V. L. (1978). The Monge–Kantorovich problem on mass transfer. In *Methods of Functional Analysis in Mathematical Economics*, pp. 23–55. Moscow: Nauka.

Levin, V. L. and A. A. Milyutin (1979). The mass transfer problem with discontinuous cost function and a mass setting for the problem of duality of convex extremum problems. *Trans Russian Math. Surveys 34*, 1–78.

Makarov, G. D. (1981). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability and Its Applications 26*, 803–806.

Mallows, C. L. (1972). A note on asymptotic joint normality. *Annals of Mathematical Statistics 43*, 508–515.

McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal 80*, 309–323.

McCann, R. J. (1999). Exact solutions to the transportation problem on the line. *The Royal Society of London. Proceedings. Series A. Mathematical, Physical and Engineering Sciences 455*(1984), 1341–1380.

McCann, R. J. (2001). Polar factorization of maps on Riemannian manifolds. *Geometric and Functional Analysis 11*, 589–608.

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématiques et de Physique, pour la même Année, Tirés des Registres de cette Académie, Paris*, pp. 666–704. Mémoires de l'Académie Royale.

Neininger, R. and L. Rüschendorf (2004). General limit theorem for recursive algorithms and combinatorial structures. *Annals of Applied Probability 14*, 378–418.

Neveu, J. and R. Dudley (1980). On Kantorovich–Rubinstein theorems. (Transcript).

Olkin, I. and F. Pukelsheim (1982). The distance between two random vectors with given dispersion matrices. *Journal of Linear Algebra and Its Applications 48*, 257–263.

Olkin, I. and S. T. Rachev (1993). Maximum submatrix traces for positive definite matrices. *SIAM J. Matrix Anal. Appl. 14*, 390–397.

Otto, F. (2001). The geometry of dissipative evolution equations: The porous medium equation. *Commun. Partial Differential Equations 26*(1-2), 101–174.

Otto, F. and C. Villani (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis 173*, 361–400.

Rachev, S. T. (1985). The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory of Probability and Its Applications 29*, 647–676.

Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models.* Wiley.

Rachev, S. T. and L. Rüschendorf (1990). A transformation property of minimal metrics. *Theory of Probability and Its Applications 35*, 110–117.

Rachev, S. T. and L. Rüschendorf (1998a). *Mass Transportation Problems. Vol. I: Theorey,.* Probability and its Applications. Springer.

Rachev, S. T. and L. Rüschendorf (1998b). *Mass Transportation Problems. Vol. II: Applications.* Probability and its Applications. Springer.

Rachev, S. T. and R. M. Shortt (1990). Duality theorems for Kantorovich–Rubinstein and Wasserstein functionals. *Dissertationes Mathematicae 299*, 647–676.

Ramachandran, D. (1979a). *Perfect Measures. Part I: Basic Theory*, Volume 5. New Delhi: Macmillan.

Ramachandran, D. (1979b). *Perfect Measures. Part II: Special Topics*, Volume 7. New Delhi: Macmillan.

Ramachandran, D. and L. Rüschendorf (1995). A general duality theorem for marginal problems. *Probability Theory and Related Fields 101*, 311–319.

Ramachandran, D. and L. Rüschendorf (2000). On the validity of the Monge–Kantorovich duality theorem. *Theory of Probability and Its Applications 45*, 350–356.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press.

Rösler, U. and L. Rüschendorf (2001). The contraction method for recursive algorithms. *Algorithmica 29*, 3–33.

Rüschendorf, L. (1979). *Vergleich von Zufallsvariablen bzgl. integralinduzierter Halbordnungen.* Habilitationsschrift, Mathematisch-Naturwissenschaftliche Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen. 133 S. .

Rüschendorf, L. (1980). Inequalites for the expectation of Δ-monotone functions. *Z. Wahrscheinlichkeitstheorie Verw. Geb. 54*, 341–354.

Rüschendorf, L. (1981). Sharpness of Fréchet-bounds. *Z. Wahrscheinlichkeitstheorie Verw. Geb. 57*, 293–302.

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability 14*, 623–632.

Rüschendorf, L. (1983). Solution of a statistical optimization problem by rearrangement methods. *Metrika 30*, 55–62.

Rüschendorf, L. (1985). The Wasserstein distance and approximation theorems. *Z. Wahrscheinlichkeitstheorie Verw. Geb. 70*, 117–129.

Rüschendorf, L. (1991a). Bounds for distributions with multivariate marginals. In K. Mosler and M. Scarsini (Eds.), *Stochastic Order and Decision under Risk*, Volume 19, pp. 285–310. IMS Lecture Notes.

Rüschendorf, L. (1991b). Fréchet-bounds and their applications. In G. Dall'Aglio, S. Kotz, and G. Salinetti (Eds.), *Advances in Probability Measure with Given Marginals (Rome, Italy, 1990)*, Volume 67 of *Math. Appl.*, pp. 151–187. Kluwer.

Rüschendorf, L. (1995). Optimal solutions of multivariate coupling problems. *Appl. Mathematicae 23*, 325–338.

Rüschendorf, L. (1996a). Developments on Fréchet bounds. In *Proceedings of Distributions with Fixed Marginals and Related Topics*, Volume 28, pp. 273–296. IMS Lecture Notes Monograph Series.

Rüschendorf, L. (1996b). On c-optimal random variables. *Statistics Prob. Letters 27*, 267–270.

Rüschendorf, L. (2006). On stochastic recursive equations of sum and max-type. *Journal of Applied Probability 43*, 687–703.

Rüschendorf, L. and S. T. Rachev (1990). A characterization of random variables with minimum L^2-distance. *Journal of Multivariate Analysis 32*, 48–54.

Rüschendorf, L. and W. Thomsen (1993). Note on the Schrödinger equation and I-projections. *Statistics and Prob. Letters 17*, 369–375.

Rüschendorf, L. and W. Thomsen (1997). Closedness of sum spaces and the generalized Schrödinger problem. *Theory of Probability and Its Applications 42*, 483–494.

Rüschendorf, L. and L. Uckelmann (1997). On optimal multivariate couplings. In *Distribution with given marginals and moment problems (Prague, Czech Republic, 1996)*, pp. 261–273. Kluwer.

Rüschendorf, L. and L. Uckelmann (2000). Numerical and analytical results for the transportation problem of Monge-Kantorovich. *Metrika. International Journal for Theoretical and Applied Statistics 51*, 245–258.

Rüschendorf, L. and L. Uckelmann (2002). On the n-coupling problem. *Journal of Multivariate Analysis 81*, 242–258.

Salvemini, T. (1949). Nuovi procedimenti di calcolo degli indici di dissomiglianza e di connessione. *Statistica, Milano 9*, 3–26.

Schachermayer, W. and J. Teichmann (2006). Solution of a problem in Villani's book. Technical report.

Smith, C. and M. Knott (1992). On Hoeffding–Fréchet bounds and cyclic monotone relations. *Journal of Multivariate Analysis 40*, 328–334.

Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statistics 36*, 423–439.

Sturm, K.-T. (2006). On the geometry of metric measure spaces. I. *Acta Math. 196*, 65–131.

Szulga, A. (1978). On the Wasserstein metric. In *Transactions of the 8th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Volume B, Prague, pp. 267–273. Akademia Praha.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probability 8*, 814–827.

Uckelmann, L. (1997). Optimal couplings between one-dimensional distributions. In V. Beneš et al. (Eds.), *Distributions With Given Marginals and Moment Problems Proceedings of the 1996 conference, Prague, Czech Republic*, pp. 275–281. Kluwer Academic Publishers.

Vallander, S. S. (1973). Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability and Its Applications 18*, 784–786.

Vasershtein, L. N. (1969). Markov processes over denumerable products of spaces describing large system of automata. *Problemy Peredači Informacii 5*(3), 64–72.

Villani, C. (2003). *Topics in Optimal Transportation.* Number 58 in *Graduate Studies in Mathematics*. American Mathematical Society, Providence.

Villani, C. (2006). Optimal Transport, Old and New. Book manuscript.

Zolotarev, V. M. (1976). Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory of Probability and Its Applications 21*, 721–737.